

BACKGROUND TO THE STUDY OF PSYCHOLOGY

“Know thyself.” These two words were inscribed on the shrine of the Oracle of Apollo, at Delphi, Greece, in the sixth century B.C. Throughout recorded history, human beings have striven to understand the nature of being human, to fathom the mysteries of the human mind and human behavior. Today that endeavor is pursued as a science, the science of psychology. In this first, background unit, we examine some fundamental ideas that helped to bring about a science of psychology, and we preview some of the methods that help to make psychology a science.

part 1

Foundations for the Study of Psychology

THREE FOUNDATION IDEAS FOR PSYCHOLOGY: A HISTORICAL OVERVIEW

- The Idea of Physical Causation of Behavior
- The Idea That the Mind and Behavior Are Shaped by Experience
- The Idea That the Machinery of Behavior and Mind Evolved Through Natural Selection

THE SCOPE OF PSYCHOLOGY

- Varieties of Explanations in Psychology, and Their Application to Sexual Jealousy
- The Connections of Psychology to Other Scholarly Fields
- Psychology as a Profession

THOUGHTS ABOUT USING THIS BOOK AND ITS SPECIAL FEATURES

The human being, as far as any human being can tell, is the only creature that contemplates itself. We not only think, feel, dream, and act but also wonder how and why we do these things. Such contemplation has taken many forms, ranging from just plain wondering to folk tales and popular songs, to poetry and literature, to formal theologies and philosophies. A little more than a century ago, human self-contemplation took a scientific turn, and we call that science psychology.

Welcome! Welcome to *Psychology* and to psychology—that is, to this book and to the field of study it is about. I hope you will enjoy them both. The big question of psychology is one of the most fascinating that anyone can ask: What makes people feel, think, and behave the way they do? In this book you will read about many routes toward answering that big question, and you will discover many dozens of specific findings and ideas that help to answer it.

It is useful to begin with a formal definition of our subject: **Psychology** is the *science* of *behavior* and the *mind*. In this definition **behavior** refers to the observable actions of a person or an animal. **Mind** refers to an individual's sensations, perceptions, memories, thoughts, dreams, motives, emotional feelings, and other subjective experiences. It also refers to all of the unconscious knowledge and operating rules that are built into or stored in the brain, and that provide the foundation for organizing behavior and conscious experience. **Science** refers to all attempts to answer questions through the systematic collection and logical analysis of objectively observable data. Most of the data in psychology are based on observations of behavior, because behavior is directly observable and mind is not; but psychologists often use those data to make inferences about the mind.

In this opening chapter, I want to do three things, all aimed at helping to prepare you for the rest of the book. First, I want to present you with a little bit of the history and philosophy

chapter 1

1

How might the focus questions (such as this one) in the text's margins be used to guide initial reading and review?

that predate and underlie modern psychology. More specifically, I will say something about the historical origins of three ideas that are so basic to our science that I refer to them as “foundation ideas for psychology.” Second, I want to tell you something about the scope of modern psychology, especially about the various levels of analysis that psychologists use in learning about and explaining mental experiences and behavior. Third, I want to tell you something about the features of this book and how you might use them to maximize your enjoyment of it and your learning from it. I put that section last, because I thought you might learn more from it after you have read a bit into the book than you would if it came first. If you prefer to read that section first, please do. It starts on page 19.

<<< There is one feature of the book that I want you to notice right now, however. In the margins of the text, throughout the book, you will find numbered *focus questions*. The first such question appears in the margin next to the paragraph you are reading right now. These are the questions that I am trying to answer in the text, and they are also good test questions. An effective way to study this book is to read and think about each focus question, as you come to it, *before* you read the adjacent paragraphs of text, which are aimed at answering that question. This method of study will help you focus your attention on the text and understand and remember what you read. If you read with the active intention of answering the focus questions, your mind is less likely to drift, and you are more likely to understand and think about what you read than if you read passively just to “learn” or “absorb” the material.

Three Foundation Ideas for Psychology: A Historical Overview

The founding of psychology as a formal, recognized, scientific discipline is commonly dated to 1879, when Wilhelm Wundt, in Germany, opened the first university-based psychology laboratory. At about that same time, Wundt also authored the first textbook of psychology and began mentoring the first official graduate students of psychology. The first people to earn Ph.D. degrees in psychology were Wundt's students.

But the roots of psychology predate Wundt. They were developed by people who called themselves philosophers, physicists, physiologists, and naturalists. In this section we shall examine three fundamental ideas of psychology, all of which were conceived of and debated before the establishment of psychology as a recognized scientific discipline. Briefly, the ideas are these:

1. Behavior and mental experiences have physical causes, so they are amenable to scientific analysis.
2. The way a person behaves, thinks, and feels is modified, over time, by the person's experiences in his or her environment.
3. The body's machinery, which produces behavior and mental experiences, is a product of evolution by natural selection.

René Descartes Descartes's speculations, in the seventeenth century, about reflexes and the interaction of the body and soul in controlling voluntary actions were an important step toward a scientific analysis of human behavior.



The Idea of Physical Causation of Behavior

Before a science of psychology could emerge, people had to conceive of and accept the idea that questions about human behavior and the mind can, in principle, be answered scientifically. Seeds for this conception can be found in the writings of some of the ancient Greeks, who speculated about the senses, the human intellect, and the physical basis of the mind in ways that seem remarkably modern. But such ideas became dormant in the Middle Ages and did not begin to sprout again until the fifteenth century (the Renaissance) or to take firm hold until the eighteenth century (the Enlightenment).

Philosophy was tightly bound to religion through the seventeenth century. The church maintained that each human being consists of two distinct but intimately

conjoined entities, a material body and an immaterial soul—a view referred to today as **dualism**. The body is part of the natural world and can be studied scientifically, just as inanimate matter can be studied. The soul, in contrast, is a supernatural entity that operates according to its own free will, not natural law, and therefore cannot be studied scientifically. This was the accepted religious doctrine, which—at least in most of Europe—could not be challenged publicly without risk of a charge of heresy and consequent execution. Yet the doctrine left some room for play, and one who played dangerously near the limits was the great French mathematician, physiologist, and philosopher René Descartes (1596–1650).

Descartes's Version of Dualism: Focus on the Body

Prior to Descartes, most dualists assigned all the interesting qualities of the human being to the soul. The soul was deemed responsible for the body's heat, for its ability to move, for life itself. In *Treatise of Man* (1637/1972), and even more explicitly in *The Passions of the Soul* (1649/1985), Descartes challenged this view. He had performed dissections of animals and of human cadavers, was familiar with research on the flow of blood, and began to regard the body as an intricate, complex machine that generates its own heat and is capable of moving even without the influence of the soul. Although little was known about the nervous system in his time, Descartes's conception of the mechanical control of movement resembles our modern understanding of *reflexes*, which are involuntary responses to stimuli (see Figure 1.1).

Descartes believed that even quite complex behaviors can occur through purely mechanical means, without involvement of the soul. Consistent with church doctrine, he contended that non-human animals do not have souls, and he pointed out a logical implication of this belief: Any activity performed by humans that is qualitatively no different from the behavior of a non-human animal can, in theory, occur without the soul. If my dog (who can do some wondrous things) is just a machine, then a good deal of what I do might occur purely mechanically as well.

In Descartes's view, the one essential ability that I have but my dog does not is *thought*, which Descartes defined as conscious deliberation and judgment. Whereas previous philosophers ascribed many functions to the soul, Descartes ascribed just one—thought. But even in his discussion of thought, Descartes tended to focus on the body's machinery. To be useful, thought must be responsive to the sensory input channeled into the body through the eyes, ears, and other sense organs, and it must be capable of directing the body's movements by acting on the muscles.

How can the thinking soul interact with the physical machine—the sense organs, muscles, and other parts of the body? Descartes suggested that the soul, though not physical, acts on the body at a particular physical location. Its place of action is a small organ (now known as the pineal body) buried between the two hemispheres (halves) of the brain (see Figure 1.2). Threadlike structures, which we now call nerves, bring sensory information by physical means into the brain, where the soul receives the information and, by non-physical means, thinks about it. On the basis of those thoughts, the soul then wills movements to occur and executes its will by triggering physical actions in nerves that, in turn, act on muscles.

Descartes's dualism, with its heavy emphasis on the body, certainly helped open the door for a science of psychology. It is a popular theory among non-scientists even today, because it acknowledges the roles of sense organs, nerves, and muscles in behavior without violating people's religious beliefs or intuitive feelings that conscious thought occurs on a non-physical plane. But the theory has

What was Descartes's version of dualism? How did it help pave the way for a science of psychology?



FIGURE 1.1 | Descartes's depiction of a reflex Descartes believed that reflexes occur through purely mechanical means. In describing this figure, Descartes (1637/1972) suggested that the fire causes movement in the nearby particles of skin, pulling on a "thread" (that runs "C" to "C" along the back) going to the brain, which, in turn, causes a pore to open in the brain, allowing fluid to flow through a "small conduit" to the muscles that withdraw the foot. What Descartes called a "thread" and a "small conduit" are today called nerves, and we now know that nerves operate through electrical means, not through physical pulling or the shunting of fluids.

Why was Descartes's theory, despite its intuitive appeal, unsuitable for a complete psychology?

Descartes, R. (1972). *Treatise of Man*.
Cambridge, MA: Harvard University Press.

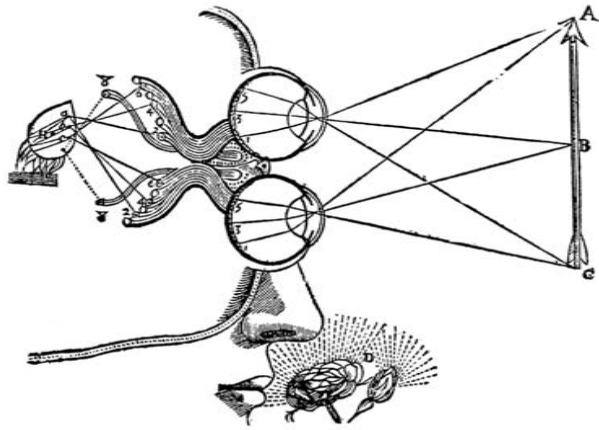


FIGURE 1.2 | Descartes's depiction of how the soul receives information through the eyes Descartes believed that the human soul is housed in the pineal gland, depicted here as the tear-shaped structure in the center of the head. In describing this figure, Descartes (1637/1972) suggested that light from the arrow enters the eyes and opens pores in structures that we now know as the optic nerves. Fluid flows from the eyes through the opened pores, causing movement in the pineal gland, which, in Descartes's words, "renders the idea" of the arrow to the soul.

4

How did Hobbes's materialism help lay the groundwork for a science of psychology?

to repeat his heresy, lived to the ripe age of 91.

<<< In *Leviathan*, and in a shorter work called *Human Nature*, Hobbes argued that spirit, or soul, is a meaningless concept and that nothing exists but matter and energy, a philosophy now known as **materialism**. In Hobbes's view, all human behavior, including the seemingly voluntary choices we make, can in theory be understood in terms of physical processes in the body, especially the brain. Conscious thought, he maintained, is purely a product of the brain's machinery and therefore subject to natural law. With this philosophy there is no theoretical limit to what psychologists might study scientifically. Most of Hobbes's work was directed toward the implications of materialism for politics and government, but his ideas helped inspire, in England, a school of thought about the mind known as *empiricism*, to which we shall soon turn.

Nineteenth-Century Physiology: Learning About the Machine

The idea that the body, including the brain, is a machine, amenable to scientific study, helped to promote the science of physiology—the study of the body's machinery. By the beginning of the nineteenth century, considerable progress had been made in this endeavor, and during that century discoveries were made about the nervous system that contributed significantly to the origins of scientific psychology.

Increased Understanding of Reflexes One especially important development for the later emergence of psychology was an increased understanding of reflexes. The basic arrangement of the nervous system—consisting of a central nervous system (brain and spinal cord) and peripheral nerves that connect the central nervous

Smithsonian Institute



A seventeenth-century mechanical man Mechanical clocks represented the pinnacle of technological achievement of the seventeenth century, comparable to computers today. For amusement, clocklike mechanisms were used to operate robotic, humanoid figures, as illustrated here. Such mechanical men helped to inspire, in Descartes and Hobbes, the idea that actual human beings might also operate by mechanical means, not requiring a nonmaterial spirit to move them.

system to sense organs and muscles—was well understood by the beginning of the nineteenth century. In 1822 in France, François Magendie demonstrated that nerves entering the spinal cord contain two separate pathways: one for carrying messages into the central nervous system from the skin's sensory receptors and one for carrying messages out to operate muscles. Through experiments with animals, scientists began to learn about the neural connections that underlie simple reflexes, such as the automatic withdrawal response to a pin prick, and found brain areas that, when active, could either enhance or inhibit such reflexes.

Some of these physiologists began to suggest that all human behavior occurs through reflexes, that even so-called voluntary actions are actually complex reflexes involving higher parts of the brain. One of the most eloquent proponents of this view, known as *reflexology*, was the Russian physiologist I. M. Sechenov. In his monograph *Reflexes of the Brain*, Sechenov (1863/1935) argued that every human action, “[b]e it a child laughing at the sight of toys, or . . . Newton enunciating universal laws and writing them on paper,” can in theory be understood as a reflex. All human actions, he claimed, are initiated by stimuli in the environment. The stimuli act on a person's sensory receptors, setting in motion a chain of events in the nervous system that culminates in the muscle movements that constitute the action. Sechenov's work inspired another Russian physiologist, Ivan Pavlov (1849–1936), whose work on reflexes played a critical role in the development, in North America, of a school of thought in psychology called *behaviorism* (discussed in Chapter 4).

The Concept of Localization of Function in the Brain Another important advance in nineteenth-century physiology was the concept of *localization of function* in the brain, the idea that specific parts of the brain serve specific functions in the production of mental experience and behavior. In Germany, Johannes Müller (1838/1965) proposed that the different qualities of sensory experience come about because the nerves from different sense organs excite different parts of the brain. Thus we experience vision when one part of the brain is active, hearing when another part is active, and so on. In France, Pierre Flourens (1824/1965) performed experiments with animals showing that damage to different parts of the brain produces different kinds of deficits in animals' ability to move. And Paul Broca (1861/1965), also in France, published evidence that people who suffer injury to a very specific area of the brain's left hemisphere lose the ability to speak but do not lose other mental abilities. All such evidence about the relationships between mind and brain helped lay the groundwork for a scientific psychology, because it gave substance to the idea of a material basis for mental processes.

How did the nineteenth-century understanding of the nervous system inspire a theory of behavior called reflexology?

How did discoveries of localization of function in the brain help establish the idea that the mind can be studied scientifically?



Early evidence for localization of function Shown here is the preserved brain of Paul Broca's patient known as Tan, who lost his ability to speak after suffering brain damage. The damage is in the left frontal lobe, in an area now called Broca's area (discussed in Chapter 5).

The Idea That the Mind and Behavior Are Shaped by Experience

Besides helping to inspire research in physiology, the materialist philosophy of seventeenth-century England led quite directly to a school of thought about the mind known as *British empiricism*, carried on by such British philosophers as John Locke (1632–1704), David Hartley (1705–1759), James Mill (1773–1836), and John Stuart Mill (1806–1873). **Empiricism**, in this context, refers to the idea that human knowledge and thought derive ultimately from sensory experience (vision, hearing, touch, and so forth). If we are machines, we are machines that learn. Our senses provide the input that allows us to acquire knowledge of the world around us, and this knowledge allows us to think about that world and behave adaptively within it. The essence of empiricist philosophy is poetically expressed in the following often-quoted passage from Locke's *An Essay Concerning Human Understanding* (1690/1975, p. 104):

Let us suppose the mind to be, as we say, white paper, void of all characters, without any ideas; how comes it to be furnished? Whence comes it by that vast store, which the busy and boundless fancy of man has painted on it, with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, from experience. In that, all our knowledge is founded; and from that it ultimately derives itself.

The Empiricist Concept of Association by Contiguity

7 How did the British empiricists explain the origin of complex ideas and thoughts? What role did the law of association by contiguity play in their philosophy?

<<< In keeping with materialist philosophy, Locke and the other British empiricists argued that thoughts are not products of free will, but reflections of one's experiences in the physical and social environment. All the contents of the mind derive from the environment and bear direct relationship to that environment. According to the empiricists, the fundamental units of the mind are elementary ideas, which derive directly from sensory experiences, and these become linked together, in lawful ways, to form complex ideas and thoughts.

The most basic operating principle of the mind's machinery, according to the empiricists, is the *law of association by contiguity*, an idea originally proposed by Aristotle in the fourth century B.C. *Contiguity* refers to closeness in space or time, and the law of association by contiguity can be stated as follows: If a person experiences two environmental events (stimuli, or sensations) at the same time or one right after the other (contiguously), those two events will become associated (bound together) in the person's mind, such that the thought of one event will, in the future, tend to elicit the thought of the other.

As a simple illustration, consider a child's experiences when seeing and biting into an apple. The child receives, from the apple, a set of sensations, which produce in her mind such elementary ideas as red color, spherical shape, and sweet and tart taste. The child may also, at the same time, hear the sound *apple* emanating from the vocal cords of a nearby adult. Because all these sensations are experienced together, they become associated in the child's mind. Together, they form the complex idea "apple." Because of association by contiguity, the thought of any of the sensory qualities of the apple will tend to call forth the thought of all the apple's other sensory qualities. Thus when the child hears *apple*, she will think of the red color, the spherical shape, and the sweet, tart taste. Or, when the child sees an apple, she will think of the sound *apple* and imagine the taste.

The empiricists contended that even their own most complex philosophical ponderings could, in theory, be understood as amalgams of elementary ideas that became linked together in their minds as a result of contiguities in their experiences. John Stuart Mill (1843/1875) referred to this sort of analysis of the mind as *mental chemistry*. Complex ideas and thoughts are formed from combinations of elementary ideas, much as chemical compounds are formed from combinations of chemical elements.

8 What influence has empiricist philosophy had on psychology?

<<< As you will discover in Chapters 4 and 9 of this textbook, the law of association by contiguity is still regarded as a fundamental principle of learning and memory. More broadly, most of psychology—throughout its history—has been devoted to

© Lake County Museum / Corbis



A complex idea To the empiricist philosophers, even as simple a concept as that of "apple" is a complex idea, consisting of a set of elementary sensations—of shape, color, and taste—that become associated in the person's mind through experiences with apples.

the study of the effects of experience on one's thoughts, feelings, and behavior. As examples, social psychologists study the ways in which people are influenced by their social environment; cultural psychologists study differences among people that arise from growing up in different cultures; developmental psychologists are interested in the experiences that lead children to acquire language and to think in ever more complex ways; clinical psychologists are interested in the experiences that can lead to mental disorders or can help one overcome such disorders; and cognitive psychologists strive to understand the mental machinery that makes learning and memory possible. The impact of empiricist philosophy on psychology has been enormous.

The Nativist Response to Empiricism

For every philosophy that contains part of the truth, there is an opposite philosophy that contains another part of it. The opposite of empiricism is **nativism**, the view that some knowledge and rules of operation are native to the human mind—that is, are inborn and do not have to be acquired from experience.

Take a sheet of white paper and present it with all the learning experiences that a normal human child might encounter (a suggestion made by Ornstein, 1991). The paper will learn nothing. Talk to it, sing to it, give it apples and oranges, take it for trips in the country, hug it and kiss it; it will learn nothing about language, music, fruit, nature, or love. To learn anything, any entity must contain some initial machinery, already built into it. At a minimum, that machinery must include an ability to sense some aspects of the environment, some means of interpreting and recording those sensations, some rules for storing and combining those sensory records, and some rules for recalling them when needed. The mind, contrary to Locke's poetic assertion, must come with some initial furnishings in order for it to be furnished further through experience.

While empiricist philosophy flourished in England, nativist philosophy took root in Germany, led by such thinkers as Gottfried Wilhelm von Leibniz (1646–1716) and Immanuel Kant (1724–1804). In his *Critique of Pure Reason* (1781/1908), Kant distinguished between *a priori* knowledge, which is built into the human brain and does not have to be learned, and *a posteriori* knowledge, which one gains from experience in the environment. Without the first, a person could not acquire the second. As an illustration, Kant referred to a child's learning of language. The specific words and grammar that the child acquires are *a posteriori* knowledge, but the child's ability to learn a language at all depends on *a priori* knowledge. The latter includes built-in rules about what to attend to and how to store and organize the verbal sounds that are heard in ways that allow the child eventually to make sense of them. Kant also argued that to make any sense of the physical world, the child must already have, built into its mind, certain fundamental physical concepts, such as the concepts of space and time. Without such concepts, a child would have no capacity for seeing an apple as spherical or for detecting the temporal contiguity of two events.

A fun robot that can learn This computer-driven robot, designed by researchers at the University of California, San Diego, can keep track of some of its previous interactions with children and incorporate them into its future responses to the children. Such machines can learn only the kinds of information that they are programmed to learn. Similarly, according to nativist philosophers, human learning is limited by the information and operating rules that are genetically programmed into the human brain.



Alan Decker / University of California, San Diego

Why is the ability to learn dependent on inborn knowledge? In Kant's nativist philosophy, what is the distinction between *a priori* knowledge and *a posteriori* knowledge?

The Idea That the Machinery of Behavior and Mind Evolved Through Natural Selection

Kant understood that the human mind has some innate furnishings, but he had no scientific explanation of how those furnishings could have been built or why they function as they do. That understanding came, at last, in 1859, when the English naturalist Charles Darwin (1809–1882) published *The Origin of Species*, a book that was destined to revolutionize biology, mark a new age in philosophy, and provide, along with the developments in physiology, a biological grounding for psychology.

Natural Selection and the Analysis of the Functions of Behavior

10

How did Darwin's theory of natural selection offer a scientific foundation for explaining behavior by describing its functions? How did it provide a basis for understanding the origin of *a priori* knowledge?

<<< Darwin's fundamental idea (explained much more fully in Chapter 3) was that living things evolve gradually, over generations, by a process of natural selection. Those individuals whose inherited characteristics are well adapted to their environment are more likely to survive and reproduce than are other individuals. At each generation, random changes in the hereditary material produce variations in offspring, and those variations that improve the chance of survival and reproduction are passed from generation to generation in increasing numbers.

Because of natural selection, species of plants and animals change gradually over time in ways that allow them to meet the changing demands of their environments. Because of evolution, the innate characteristics of any given species of plant or animal can be examined for the *functions* they serve in allowing individuals to survive and reproduce. To understand, for example, why finches have stout beaks and warblers have slender beaks, one must know what foods the birds eat and how they use their beaks to obtain those foods. The same principle that applies to anatomy applies to behavior. Through natural selection, living things have acquired instinctive tendencies to behave in ways that promote their survival and reproduction. A key word here is *function*. While physiologists were examining the neural mechanisms of behavior, and empiricist philosophers were analyzing lawful relationships between behavior and the environment, Darwin was studying the functions of behavior—the ways in which an organism's behavior helps it to survive and reproduce.

Application of Darwin's Ideas to Psychology

In *The Origin of Species*, Darwin discussed only plants and non-human animals, but in later writings he made it clear that he viewed humans as no exception. We humans also evolved through natural selection, and our anatomy and behavior can be analyzed in the same ways as can those of other living things.

In a book entitled *The Expression of the Emotions in Man and Animals*, Darwin (1872/1965) illustrated how evolutionary thinking can contribute to a scientific understanding of human behavior. He argued that the basic forms of human emotional expressions (such as laughing and crying) are inherited, as are those of other animals, and may have evolved because the ability to communicate one's emotions or intentions to others of one's kind improves one's chances of survival. Darwin's work provided psychology with a scientific way of thinking about all the inborn universal tendencies that constitute human nature. The innate mechanisms underlying human emotions, drives, perceptual abilities, learning abilities, and capacities for reason came about gradually because they promoted the survival and reproduction of our ancestors. One approach to understanding such characteristics is to analyze their evolutionary functions—their ways of promoting survival and reproduction.

If Kant had been privy to Darwin's insight, he would have said that the innate furnishings of the mind, which make it possible for children to learn language, to learn about the physical world, to experience human drives and emotions, and, more generally, to behave as human beings and not as blank sheets of paper, came about through the process of natural selection, which gradually built all these capacities into the brain's machinery. Darwin, perhaps more than anyone else,



Granger Collection

Charles Darwin Darwin's principle of evolution by natural selection helped provide a scientific footing for psychology. The principle links humans to the rest of the biological world and explains the origin of brain mechanisms that promote the individual's survival and reproduction.

helped convince the intellectual world that we humans, despite our pretensions, are part of the natural world and can be understood through the methods of science. In this way he helped make the world ripe for psychology.

We have now reached the end of the first section of the chapter, and, before moving on, I want to point out a second feature of the book that is designed to help you study it. At the end of each major section of each chapter is a *section review* that summarizes the section's main ideas. The chart is organized hierarchically: the main idea or topic of the section is at the top, the sub-ideas or subtopics are at the next level down, and specific facts and lines of evidence pertaining to each sub-idea or subtopic fill the lower parts. A good way to review, before going on, is to read each item in the chart and think about it. Start at the top and think about the main idea or purpose of the whole section. How would you explain it to another person? Then go down each column, one by one; think about how the idea or topic heading the column pertains to the larger idea or topic above it and how it is supported or elaborated upon by the more specific statements below it. If you are unclear about the meaning of any item in the chart, or can't elaborate on that item in a meaningful way, you may want to read the relevant part of the chapter again before moving on. Another way to review is to look back at the focus questions in the margins and make sure you can answer each of them.

>>> **11**
How can the section review at the end of each major section of each chapter be used to guide one's thought and review before going on to the next section?

SECTION REVIEW

Psychology—the science of behavior and the mind—rests on prior intellectual developments.		
Physical Causation of Behavior • Descartes's dualism placed more emphasis on the role of the body than had previous versions of dualism. Hobbes's materialism held that behavior is completely a product of the body and thus physically caused. • To the degree that behavior and the mind have a physical basis, they are open to study just like the rest of the natural world. • Nineteenth-century physiological studies of reflexes and of localization of function in the brain demonstrated the applicability of science to mental processes and behavior.	The Role of Experience • The British empiricists claimed that all thought and knowledge are rooted in sensory experience. • They used the concept of association by contiguity to explain how sensory experience could ultimately lead to complex thought. • In contrast to empiricism, nativism asserts that some knowledge is innate and that such knowledge enables learning from experience.	The Evolutionary Basis of Mind and Behavior • Darwin proposed that natural selection leads to the evolution of behavioral tendencies (as well as anatomical changes) that promote survival and reproduction. • Darwin's thinking led to a focus on the functions of behavior. • Natural selection also offered a scientific basis for considering nativist views.

The Scope of Psychology

Psychology is a vast and diverse field of research. Every question about behavior and mental experience that is potentially answerable by scientific means is within its scope. One way to become oriented to this grand science is to preview the various kinds of explanatory concepts that psychologists use.

Varieties of Explanations in Psychology, and Their Application to Sexual Jealousy

Psychologists strive to *explain* mental experiences and behavior. To explain is to identify causes. What causes us to do what we do, feel what we feel, perceive what we perceive, or believe what we believe? What causes us to eat in some conditions and not in others; to cooperate sometimes and to cheat at other times; to feel angry, frightened, happy, or guilty; to dream; to hate or love; to see red as different

The Kobal Collection



Sexual jealousy in humans Like any common human behavioral predisposition, sexual jealousy can be studied at the neural, genetic, evolutionary, learning, cognitive, social, cultural, and developmental levels of analysis. Here the theme of sexual jealousy is played out by Rita Hayworth, Tyrone Power, and Anthony Quinn in a scene from the 1941 film *Blood and Sand*.

from blue; to remember or forget; to suddenly see solutions to problems that we couldn't see before; to learn our native language so easily when we are very young; to become depressed or anxious? This is a sample of the kinds of questions that psychologists try to answer and that are addressed in this book.

The causes of mental experiences and behavior are complex and can be analyzed at various levels. The phrase **level of analysis**, as used in psychology and other sciences, refers to the level, or type, of causal process that is studied. More specifically, in psychology, a person's behavior or mental experience can be examined at the *neural* level (brain as cause), *genetic* level (genes as cause), *evolutionary* level (natural selection as cause), *learning* level (the individual's prior experiences with the environment as cause), *cognitive* level (the individuals' knowledge or beliefs as cause), *social* level (the influence of other people as cause), *cultural* level (the culture in which the person develops as cause), and *developmental* level (age-related changes as cause). It is

convenient to group these eight levels of analysis into two categories. The first category is biological, and consists of neural, genetic, and evolutionary explanations. The second category is experiential, and consists of learning, cognitive, social, cultural, and developmental explanations. You will find many examples of each of these eight levels of analysis in this book. Now, as an overview, I'll describe each of them very briefly.

Any given type of behavior or mental experience can, in principle, be analyzed at any of the eight levels. To illustrate that point, I will suggest, in the following descriptions, ways by which each level of analysis can be applied to an understanding of sexual jealousy. One reason for choosing sexual jealousy as the example is that it has not, in fact, been studied very much at most of the levels, so I can describe hypothetical ways of studying it without getting into too much factual detail. For our purposes, *sexual jealousy* can be defined as the set of emotions and behaviors that result when a person believes that his or her relationship with a sexual partner, or a potential sexual partner, is threatened by the partner's involvement with another person.

12

How do neural, genetic, and evolutionary explanations differ from one another? How might each be applied toward an understanding of jealousy?

<<< Explanations That Focus on Biological Processes

Neural Explanations All mental experiences and behavioral acts are products of the nervous system. One logical route to explanation in psychology, therefore, is to try to understand how the nervous system produces the specific type of experience or behavior being studied. Researchers who specialize in this level of explanation are referred to as *behavioral neuroscientists*.

Some behavioral neuroscientists study individual neurons (nerve cells) or small groups of neurons to determine how their characteristics contribute to particular psychological capacities, such as learning. Others map out and study larger brain regions and pathways that are directly involved in some particular category of behavior or experience, such as speaking grammatically, seeing the shapes of objects, or experiencing an emotion such as fear. Behavioral neuroscientists also study the ways that hormones and drugs act on the brain to alter behavior and experience, either in humans or in non-human animals.

To date, no neural studies of jealousy have been conducted with human subjects, but at least one such study has been conducted with macaque monkeys (Rilling & others, 2004). Researchers induced a jealous state in male monkeys by exposing each male to the sight of a female, with which he had previously mated, being courted by another male. During this experience, they measured the animal's brain activity using a technique called positron emission tomography (PET, described in



Hoa Qui / Picture Quest

Viewing the active brain In recent years the field of behavioral neuroscience has advanced greatly, due in part to new techniques for assessing the amount of activity that occurs in specific brain locations as a person performs mental tasks. These neuroimaging techniques are discussed in Chapter 5.

Chapter 5). The result was a preliminary mapping of specific brain areas that become especially active during the experience of sexual jealousy in male macaques.

A next step, not yet taken, might be for neuroscientists to try to increase or decrease jealous behavior in monkeys by artificially activating or inactivating those same areas of the brain. They might also examine people who have, through strokes or other accidents, suffered damage to those brain areas to determine if they have any deficits in the experience of jealousy. These are the kinds of techniques regularly used by behavioral neuroscientists.

Genetic Explanations Genes are the units of heredity that provide the codes for building the entire body, including the brain. Differences among individuals in the genes they inherit can cause differences in the brain and, therefore, differences in mental experiences and behavior. Researchers who attempt to explain psychological differences among individuals in terms of differences in their genes are called *behavioral geneticists*.

Some behavioral geneticists study non-human animals. They may, for example, deliberately modify animals' genes to observe the effects on behavior. Others study people. To estimate the extent to which variation among people in some particular trait is the result of variation in genes, researchers might assess the degree to which genetic relatedness between people correlates with their degree of similarity in that trait. A finding that close genetic relatives are more similar in the trait than are more distant relatives would suggest that genes contribute to variation in the trait. Behavioral geneticists might also try to identify specific genes that contribute to a particular trait by analyzing the DNA (genetic material) of people who differ in that trait.

People differ in the degree to which they are prone to sexual jealousy. Some are easily made jealous; others are not. To measure the degree to which such differences are the result of genetic differences, researchers might assess sexual jealousy in twins. If identical twins, who share all their genes with each other, are much more similar in jealousy than are same-sex non-identical twins, who are no more closely related than other siblings, that would indicate that much of the variation among people in sexual jealousy is caused by variation in genes. A next step might be to find out just which genes are involved in these differences and how they act on the brain to influence jealousy. So far, no such studies have been done of sexual jealousy, but you will read later of such studies of intelligence (in Chapter 10), personality traits such as extraversion (in Chapter 15), and predispositions to mental disorders such as schizophrenia and depression (in Chapter 16).

Duck jealousy Ducks, such as these blue-winged teals, form monogamous pair bonds, at least for the duration of the breeding season. When an intruder encroaches, the member of the pair that is the same sex as the intruder drives the intruder away. Such jealous-like behavior helps to keep the mated pair intact.



© Cliff Beittel

Evolutionary Explanations All the basic biological machinery underlying behavior and mind, coded by genes, is a product of evolution by natural selection. One way to explain universal human characteristics, therefore, is to explain how or why that characteristic came about in evolution. Researchers who specialize in this level of analysis are called *evolutionary psychologists*.

Some evolutionary psychologists are interested in the actual routes by which particular behavioral capacities or tendencies evolved. For instance, researchers studying the evolution of smiling have gained clues about how smiling originated in our ancestors by examining smile-like behaviors in other primates, including chimpanzees (discussed in Chapter 3). Most evolutionary psychologists are interested in identifying the *evolutionary functions*—that is, the survival or reproductive benefits—of the types of behaviors and mental experiences that they study. In later chapters you will read about evolutionary, functional explanations of many human behavioral tendencies, drives, and emotions.

Evolutionary psychologists have examined the forms and consequences of human jealousy in some detail in order to identify its possible benefits for reproduction (Buss, 2000). Toward the same end, researchers have also examined jealous-like behavior in many other animal species. Such research supports the view that jealousy functions to promote long-term mating bonds. In all animals that form such bonds, behavioral mechanisms (discussed in Chapter 3) have evolved that motivate the mated male or female to actively drive off, or in other ways discourage, competitors for his or her mate.

Explanations That Focus on Environmental Experiences, Knowledge, and Development

13

How do learning and cognitive explanations differ from each other? How might each be applied toward an understanding of jealousy?

Learning Explanations Essentially all forms of human behavior and mental experience are modifiable by learning; that is, they can be influenced by prior experiences. Prior experiences can affect our emotions, drives, perceptions, thoughts, skills, and habits. Most psychologists are in one way or another interested in the role that experience plays in shaping the types of behavior that they study. Those who explain behavior most directly in terms of past experiences with the environment refer to themselves as *learning psychologists*. These specialists might, for example, attempt to explain compulsive gambling in terms of the pattern of rewards that the person has experienced in the past while gambling, or might attempt to explain particular fears in terms of previous experiences with the feared objects or situations. They might also conduct research, with animals as well as people, to understand the most efficient ways to acquire new skills (discussed in Chapter 4).

Differences among individuals in jealousy derive partly from genetic differences, but they also derive partly from differences in past experiences. Jealous reactions that prove to be effective in obtaining rewards—such as those that succeed in repelling competitors or attracting renewed affection from the beloved—may increase in frequency with experience, and ineffective reactions may decrease. People and animals may also learn, through experience, what sorts of cues are potential signs of infidelity in their mates, and those cues may come to trigger jealous reactions. The intensity of sexual jealousy, the specific manner in which it is expressed, and the environmental cues that trigger it can all be influenced, in many ways, by learning. Learning psychologists might study all of those effects.

Cognitive Explanations The term *cognition* refers to information in the mind—that is, to information that is somehow stored and activated by the workings of the brain. Such information includes thoughts, beliefs, and all forms of memories. Some information is innate to the human mind, as nativist philosophers pointed out, and other information is acquired through learning, as empiricist philosophers pointed out. Some information is conscious, in the sense that the person is aware of it and can describe it, and other information is unconscious but can still influence one's conscious experiences and behavior. One way to explain any behavioral action or mental experience is to relate it to the cognitions (items of mental information) that underlie that action or experience. Researchers who specialize in this level of analysis are called *cognitive psychologists*.

You can think of mental information as analogous to the operating rules (software) and data that are stored in a computer, which influence the way that the computer responds to particular forms of input. Cognitive psychologists are interested in specifying, as clearly as possible, the types of mental information that underlie and make possible the behaviors that they study. For instance, a cognitive psychologist who is interested in reasoning might attempt to understand the rules by which people manipulate information in their minds in order to solve particular classes of problems (discussed in Chapter 10). A cognitive psychologist who is interested in racial prejudice might attempt to specify the particular beliefs—including unconscious as well as conscious beliefs—that promote prejudiced behavior (discussed in Chapter 13). Cognitive psychologists are also interested in the basic processes by which learned information is stored and organized in the mind, which means that they are particularly interested in memory (discussed in Chapter 9).

In general, cognitive psychology differs from the psychology of learning in its focus on the mind. Learning psychologists generally attempt to relate learning experiences directly to behavioral changes and are relatively unconcerned with the mental processes that mediate such relationships. To a learning psychologist: *experience in the environment leads to change in behavior*. To a cognitive psychologist: *experience in the environment leads to change in knowledge or beliefs, and that change leads to change in behavior*.

A cognitive psychologist interested in jealousy would define jealousy first and foremost as a set of beliefs—beliefs about the behavior of one's beloved and some third party, about the vulnerability of one's own relationship with the beloved, and about the appropriateness or inappropriateness of possible ways to react. One way to study jealousy, from a cognitive perspective, is to ask people to recall episodes of it from their own lives and to describe the thoughts that went through their minds, the emotions they felt, and the actions they took. Such work reveals that a wide variety of thoughts can enter one's mind in the jealous state, which can lead to actions ranging from romantic expressions of love to murderous violence (Guerrero & others, 2005). Psychotherapists who use cognitive methods to treat cases of pathological jealousy try to help their clients change their thought patterns, so they will no longer misperceive every instance of attention that the beloved pays to someone else as a threat to their relationship, and so they will focus on constructive rather than destructive ways of reacting to actual threats (Bishay & others, 1996).



Walter Dawn / Photo Researchers, Inc.

Rat learning to press a lever To identify basic principles of learning, some learning psychologists study the processes by which animals learn simple responses for rewards. This thirsty rat receives a drop of water to drink each time it presses the lever.

14

How do social and cultural explanations differ from each other? How might each of these be applied toward understanding jealousy?

<<< Social Explanations We humans are, by nature, social animals. We need to cooperate and get along with others of our species in order to survive and reproduce. For this reason, our behavior is strongly influenced by our perceptions of others. We use others as models of how to behave, and we often strive, consciously or unconsciously, to behave in ways that will lead others to accept us and approve of us. One way to explain mental experiences and behavior, therefore, is to identify how it is influenced by other people or by one's beliefs about other people (discussed in Chapters 13 and 14). Researchers who specialize in this level of analysis are called *social psychologists*.

According to an often-quoted definition (originally from Allport, 1968), "social psychology is the attempt to understand and explain how the thought, feeling, and behavior of individuals are influenced by the actual, imagined, or implied presence of others." Social psychologists often explain behavior in terms of conformity to social norms, or obedience to authority, or living up to others' expectations. A popular term for all such influences is "social pressure."

Social-psychological explanations are often phrased in terms of people's conscious or unconscious beliefs about the potential social consequences of acting in a particular way. This means that many social-psychological explanations are also cognitive explanations. Indeed, many modern social psychologists refer to their specialty as *social cognition*. A social psychologist interested in physical fitness, for example, might attempt to explain how people's willingness to exercise is influenced by their beliefs about others' fitness efforts and their beliefs about how others will react to them if they do or do not exercise.

A social psychologist interested in jealousy might focus on the norms and beliefs concerning romance, mating, and jealousy that surround and influence the jealous person. How do others react in similar situations? Are the beloved's flirtations with a third person within or outside the realm of what is considered acceptable by other dating or married couples? Would violent revenge be approved of or disapproved of by others who are important to the jealous person? Implicitly or explicitly, the answers to such questions influence the way the jealous person feels and behaves. An understanding of such influences constitutes a social-psychological explanation of the person's feelings and behavior.

Cultural Explanations We can predict some aspects of a person's behavior by knowing what culture that person grew up in. Cultures vary in language or dialect, in the values and attitudes they foster, and in the kinds of behaviors and emotions they encourage or discourage. Researchers have found consistent cultural differences even in the ways that people perceive and remember aspects of their physical environment (discussed in Chapter 10). Researchers who explain mental experiences and behavior in terms of the culture in which the person developed are called *cultural psychologists*.

Cultural and social psychology are very closely related, but differ in emphasis. While social psychologists emphasize the immediate social influences that act on individuals, cultural psychologists strive to characterize entire cultures in terms of the typical ways that people within them feel, think, and act. While social psychologists use concepts such as conformity and obedience to explain an individual's behavior, cultural psychologists more often refer to the unique history, economy, and religious or philosophical traditions of a culture to explain the values, norms, and habits of its people. For example, a cultural psychologist might contend that the frontier history of North America, in which individuals and families often had to struggle on their



"Mom! Everybody at school says we're just a bunch of crazy Victorians."

own, with little established social support, helps explain why North Americans value independence and individuality so strongly.

Concerning jealousy, a cultural psychologist would point to significant cultural differences in romantic and sexual mores. Some cultures, for example, are more tolerant of extramarital affairs than are others, and this difference affects the degree and quality of jealousy that is experienced. In some cultures a strong double standard exists, by which women are condemned far more harshly than are men for sexual infidelity, and in those cultures violent revenge on the part of a jealous man may be socially sanctioned (Bhugra, 1993; Delgado & others, 1997). In other cultures, the same violence would dishonor the perpetrator and land him in jail. A full cultural analysis would include an account of the cultures' histories, which led to differences in the ways that infidelity is understood and treated.

Developmental Explanations We can predict some aspects of a person's behavior by knowing his or her age. Four-year-olds behave differently from two-year-olds, and middle-aged adults differently from adolescents. *Developmental psychologists* document and describe the typical age differences that occur in the ways that people feel, think, and act. For any given type of behavior or mental capacity, they may describe the sequence of changes that occur from infancy to adulthood. For example, developmental psychologists who study language have described a sequence of stages in speech production that goes from cooing to babbling, then to first recognizable words, to frequent one-word utterances, to two-word utterances, and so on, with each stage beginning, on average, at a certain age. At a superficial level, then, age can itself be an explanation: "She talks in such-and-such a way because she is 3 years old, and that is how most 3-year-olds talk."

Looking deeper, developmental psychologists are also interested in the processes that produce the age-related changes that they document. Those processes include physical maturation of the body (including the brain), instinctive behavioral tendencies that are genetically timed to emerge at particular ages, the accumulated effects of many learning experiences, and new pressures and opportunities provided by the social environment or the cultural milieu as one gets older. At this deeper level, then, developmental psychology is an approach that brings together the other levels of analysis. Neural, genetic, evolutionary, learning, cognitive, social, and cultural explanations might all be brought to bear on the task of explaining behavioral changes that occur with age. Developmental psychologists are particularly interested in understanding how experiences at any given stage of development can influence behavior at later stages.

A developmental analysis of jealousy might begin with a description of age-related changes in jealousy that correspond with age-related changes in social relationships. Infants become jealous when their mother or other primary caregiver devotes extended attention to someone else (Hart & Carrington, 2002). Children of middle-school age, especially girls, often become jealous when their same-sex "best friend" becomes best friends with someone else (Parker & others, 2005). These early forms of jealousy are similar in form and function to sexual jealousy, which typically emerges along with the first serious romantic attachment, in adolescence or young adulthood. Researchers have found evidence of continuity between early attachments to parents and friends and later attachments to romantic partners (discussed in Chapter 12). People who, in childhood, develop secure relationships with their parents and friends also tend, later on, to develop secure relationships with romantic partners, relatively untroubled by jealousy (Fraley, 2002; Guerrero, 1998).

A Comment on Psychological Specialties

Because of psychology's vast scope, research psychologists generally identify their work as belonging to specific subfields, or specialties, of research. To some degree, as I have implied in the foregoing discussion, different psychological research specialties correspond to different levels of analysis. This is most true of the eight specialties to which I have already referred: *behavioral neuroscience*, *behavioral genetics*,

>>> 15
What constitutes a developmental explanation? How can such an explanation be applied toward understanding jealousy?

16

What are some research specialties in psychology that are not defined primarily by the level of analysis employed?

evolutionary psychology, psychology of learning, cognitive psychology, social psychology, cultural psychology, and developmental psychology.

<<< Other specialties, however, are defined more in terms of the topics of study than in terms of the level of analysis. For example, *sensory psychology* is the study of basic abilities to see, hear, touch, taste, and smell the environment; and *perceptual psychology* is the study of how people and animals make sense of, or interpret, the input they receive through their senses. Similarly, some psychologists identify their specialty as the *psychology of motivation* or the *psychology of emotion*. These specialists might use any or all of psychology's modes of explanation to understand particular phenomena related to the topics they study.

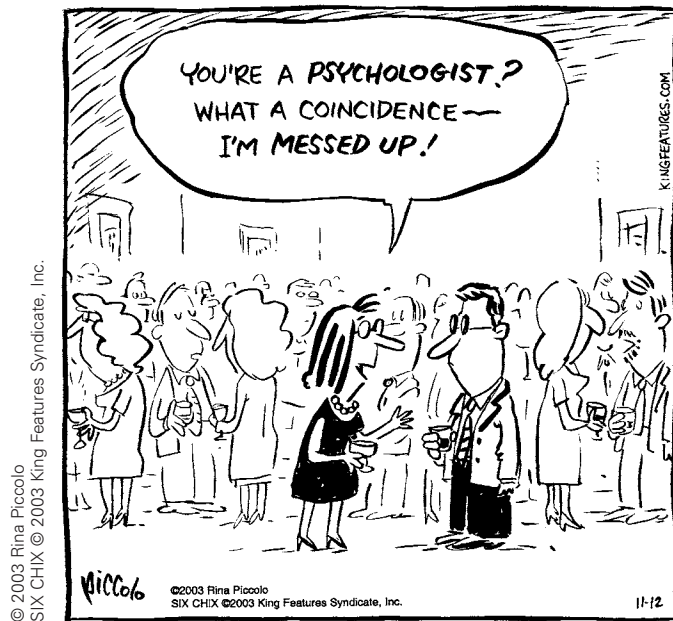
Two large specialties, which are closely related to each other, are devoted to the task of understanding individual differences among people. One of these is *personality psychology* (discussed in Chapter 15), which is concerned with normal differences in people's general ways of thinking, feeling, and behaving—referred to as personality traits. The other is *abnormal psychology* (discussed in Chapter 16), which is concerned with variations in psychological traits that are sufficiently extreme and disruptive to people's lives as to be classified as mental disorders. Personality and abnormal psychologists use various levels of analysis. Differences

in the nervous system, differences in genes, different learning experiences, different beliefs or expectations gained from learning, and different social pressures or cultural milieus may all contribute to an understanding of differences in personality and in susceptibility to particular mental disorders.

Closely related to abnormal psychology is *clinical psychology* (discussed in Chapter 17), the specialty that is concerned with helping people who have mental disorders or less serious psychological problems. Most clinical psychologists are practitioners rather than researchers. They offer psychotherapy or drug treatments, or both, to help people cope with or overcome their disorders or problems. Clinical psychologists who conduct research are usually interested in identifying or developing better treatment methods.

In general, research specialties in psychology are not rigidly defined. They are simply convenient labels aimed at classifying, roughly, the different levels of analysis and topics of study that characterize the work of different research psychologists. Regardless of what they call themselves, good researchers often use several different levels of analysis in their research and may study a variety of topics that in some way relate to one another.

My main reason for listing and briefly describing some of the specialties here has been to give you an overview of the broad scope of psychological science.



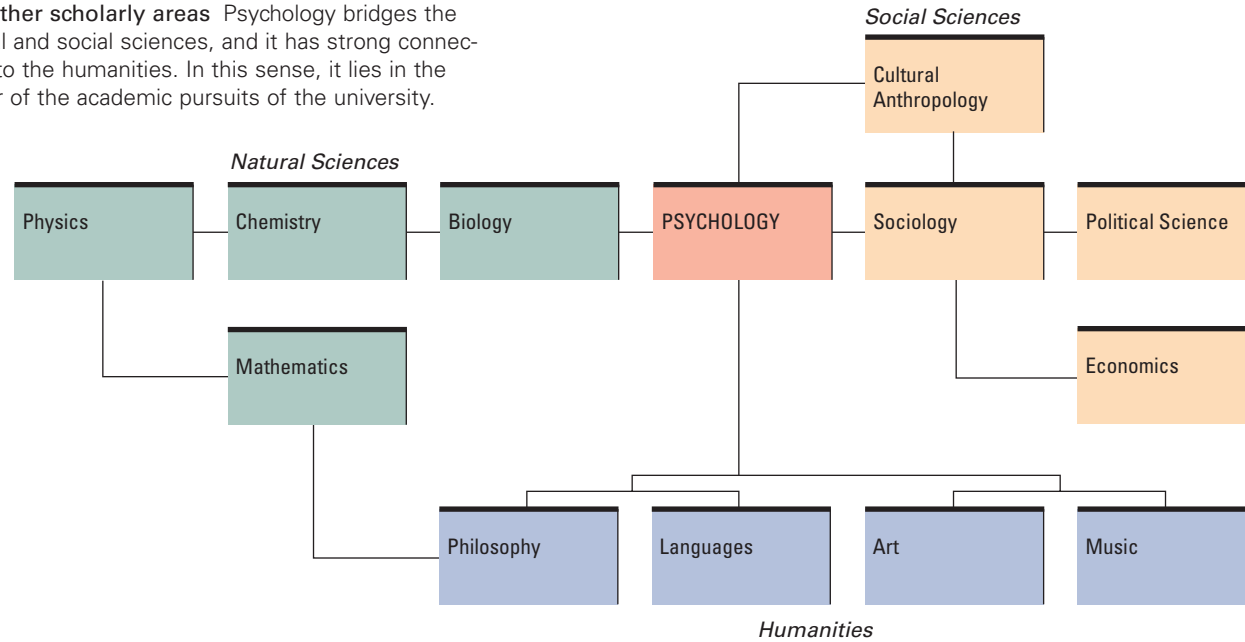
The Connections of Psychology to Other Scholarly Fields

17

How does psychology link the three main divisions of academic studies?

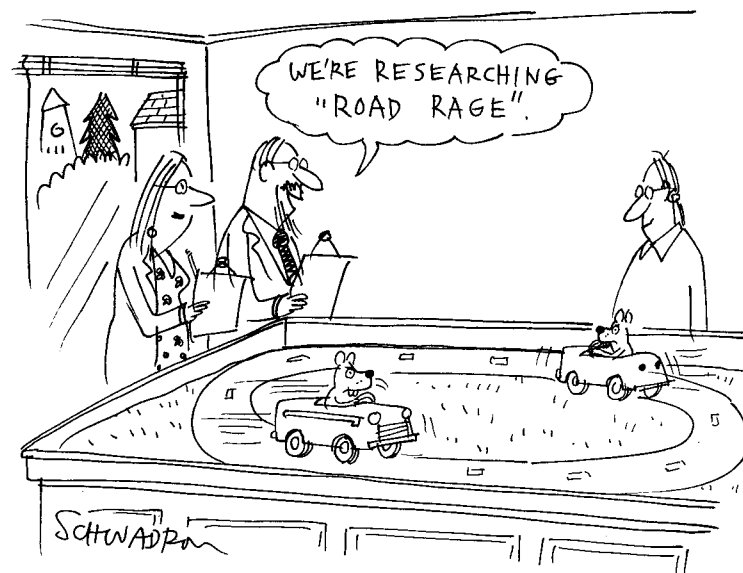
<<< Another way to characterize psychology is to picture its place in the spectrum of disciplines that form the departments of a typical college of arts and sciences. Figure 1.3 illustrates a scheme that I call (with tongue only partly in cheek) the *psychocentric theory of the university*. The disciplines are divided roughly into three broad areas. One division is the *natural sciences*, including physics, chemistry, and biology, shown on the left side of the figure. The second division is the *social sciences*, including sociology, anthropology, political science, and economics, shown on the right side of the figure. The third division is the *humanities*—including languages, philosophy, art, and music—shown in the lower part of the figure. The humanities represent things that *humans* do. Humans, unlike other animals, talk to one another, develop philosophies, and create art and music.

FIGURE 1.3 | Connections between psychology and other scholarly areas Psychology bridges the natural and social sciences, and it has strong connections to the humanities. In this sense, it lies in the center of the academic pursuits of the university.



Where does psychology fit into this scheme? Directly in the center, tied to all three of the broad divisions. On the natural science end, it is tied most directly to biology by way of behavioral neuroscience, behavioral genetics, and evolutionary psychology. On the social science end, it is tied most directly to sociology and anthropology by way of social and cultural psychology. In addition to bridging the natural and social sciences, psychology ties the whole spectrum of sciences to the humanities, through its interest in how people produce and understand languages, philosophies, art, and music.

If you were to look at the range of research activities of psychologists, you would find that some behave like natural scientists. They have labs, wear white coats, perform experiments, and spend a good deal of time in university departments of biology, chemistry, and physics exchanging ideas and equipment. Others, you would find, behave more like social scientists. They use surveys, historical documents, and interviews as sources of data, and they find intellectual companionship in the various social science departments of the universities where they work. The distinction between these two groups, however, is often fuzzy. Social psychologists often have labs and fancy data-collecting equipment, and biologically oriented psychologists can be found whose data come mainly through their own two eyes, out in the field.



Psychology as a Profession

Psychology is not only an academic discipline but also a profession. The profession includes both academic psychologists, who are engaged in research and teaching, and practicing psychologists, who apply psychological knowledge and ideas in clinics, businesses, and other settings. The majority of professional psychologists in the United States hold doctoral degrees in psychology, and most of the rest hold master's degrees (Landrum & others, 2000). The main settings in which they work and the kinds of services they perform in each are:

- **Academic departments in universities and colleges** Academic psychologists are employed to conduct basic research, such as that which fills this book, and to teach psychology courses.
- **Clinical settings** Clinical and counseling psychologists work with clients who have psychological problems or disorders, in independent practice or in hospitals, mental health centers, clinics, and counseling or guidance centers.
- **Elementary and secondary schools** School psychologists administer psychological tests, supervise programs for children who have special needs, and may help teachers develop more effective classroom techniques.
- **Business and government** Psychologists are hired by businesses and government agencies for such varied purposes as conducting research, screening candidates for employment, helping to design more pleasant and efficient work environments, and counseling personnel who have work-related problems.

The decision to major in psychology as an undergraduate does not necessarily imply a choice of psychology as a career. Most students who major in psychology do so primarily because they find the subject fun to learn and think about. Most go on to careers in other fields—such as social work, law, education, and business—where they are quite likely to find their psychology background helpful (Carroll & others, 1992; Landrum & others, 2000). If you are considering psychology as a career, you might want to look at one or more of the books concerning careers in psychology that I have listed in the Further Reading section at the end of this chapter.

ARLO & JANIS® by Jimmy Johnson



SECTION REVIEW

Psychology is a broad and diverse field of research, and is a profession.

Levels of Causal Analysis and Topics of Study in Psychology

- Three types of biological causal explanations are used in psychology—neural, genetic, and evolutionary explanations.
- Five other types of causal explanations in psychology are learning, cognitive, social, cultural, and developmental explanations.
- As demonstrated with jealousy, each level of analysis can be applied to any given type of behavior or mental experience.
- Some subfields in psychology are defined primarily by the level of analysis; others are defined more by the topics studied.

A Discipline Among Disciplines

- Broadly speaking, scholarly disciplines can be classified as belonging to natural sciences, social sciences, or humanities.
- Psychology has strong ties to and commonalities with each class of disciplines.

The Profession of Psychology

- The profession includes academic psychologists, who teach and do research, and practicing psychologists, who apply psychological knowledge and principles to real-world issues.
- Psychologists work in various settings—including universities, clinical settings, and businesses—and typically hold advanced degrees.

Thoughts About Using This Book and Its Special Features

Most of you reading this book are doing so because it was assigned to you as part of a college course. You are reading it not because you chose it, but because someone else chose it for you. This creates a special difficulty.

When you read nonfiction outside of a course, you usually do so because you have some question or questions in mind that you want to answer. You are curious about some issue, and you choose a book that seems to address it. In that case you read actively, constantly thinking about what you read to see if it helps answer your questions, and you may well skip parts that don't seem relevant to the issue that motivates you. But when a book is assigned to you for a course, the situation is different. You don't necessarily have particular questions in mind that you want to answer; you don't usually have the option of skipping parts that fail to interest you; and your job is the rather vaguely defined one of "learning the material." All this leads, often, to a passive and rather ineffective mode of reading, aimed more at memorizing than at thinking and understanding.

Our minds are not designed for memorizing what we don't understand or what we have not thought about actively. Our mental machinery evolved for the purpose of making sense of things, and we don't remember much of what doesn't make sense. So, when we read for the passive purpose of "learning" or "absorbing" the material, our minds often wander. We often find that we have read long passages—or, rather, that our eyes have moved across the passages and the words have been detected—without our having any idea what we just read. Such conditions are also excellent for inducing sleep.

My sympathies are with you. I really want you to enjoy this book. I want it to keep you awake. I want you to question it, argue with it, and get excited about some of the ideas in it. Toward that end, I have done my best to present psychology as a set of ideas to think about, not as a set of facts to memorize. I have tried to give you enough information about each idea, enough of the evidence and logic supporting it, to enable you to have something to think about and argue with. Most of all, I do not want you to read this book as Truth with a capital T. Psychology is a science, and the essence of science is this: We do not accept anything on authority. It doesn't matter *who* says that something is or isn't true; what matters is the evidence and logic behind the statement. Each page of this book is offered for your consideration, not for your unquestioned acceptance.

Using the Focus Questions to Guide Your Study

In the introduction to this chapter, I pointed out the numbered *focus questions* that appear in the book's margins. I suggested there a way to use these questions to guide both your initial reading and your review of the text. Here I'll elaborate on their use.

Each focus question is the main question that I am trying to answer in the portion of text that lies adjacent to and immediately below the question. You can make your reading of the text more interesting and active if you read and think about each focus question as you come to it, before reading the paragraphs aimed at answering it. One way to think about the question is to formulate a preliminary, possible answer based on what you already know or believe. You might also put the question into your own words, to make it *your* question, rather than just mine. This will prepare you to read the relevant portion of text with a clear purpose in mind, that of finding out how I answer the question and how my answer compares to your preliminary thoughts about it.

As an illustration, consider Focus Question 2, on page 3 of this chapter. This question consists of two parts: *What was Descartes's version of dualism? How did it help pave the way for a science of psychology?* When you first came to this question, you already had some good grounds for forming a preliminary answer. You had just read a definition of dualism, in the previous paragraph, in which that term

>>> **18**
Why is it often more difficult to read a textbook for a course than to read nonfiction that you have chosen on your own?

>>> **19**
How can students use the focus questions in this textbook to make their reading more thought-provoking and effective?

appeared in bold italics. You had read that dualism distinguishes between the body, which is physical and can be studied scientifically, and the soul, which is supernatural and cannot be studied scientifically. You may have also noticed that the section heading just above the focus question reads, *Descartes's Version of Dualism: Focus on the Body*, and that the larger section heading above that (on page 2) reads, *The Idea of Physical Causation of Behavior*. So, in thinking about Focus Question 2, you might have said something like the following to yourself: “Okay, maybe Descartes's version of dualism placed greater emphasis on the physical body and less emphasis on the soul than did previous versions. Now, I wonder if that guess is correct. If it is correct, I wonder just how Descartes developed and supported this view. What attributes did he ascribe to the body that had previously been ascribed to the soul, and why?” Having said all this to yourself, you would be ready to read the adjacent portion of text with great understanding.

After reading the portion of text that is relevant to answering a given focus question, it is a good idea to stop and think about that question again and how you would answer it. You might jot down, next to the question, the gist of the answer that you would now give. If you aren't sure how to answer it, you might want to read that portion again. If you still aren't sure after that, you might want to mark that question as one to ask of your course instructor or study companions. Perhaps I didn't answer the question sufficiently clearly in the text, and perhaps a discussion with others will throw some light on it.

In later chapters you will discover that many focus questions ask about the evidence for or against some idea. Be sure to think especially carefully about the answers you read to those questions and ask yourself whether or not the evidence seems convincing.

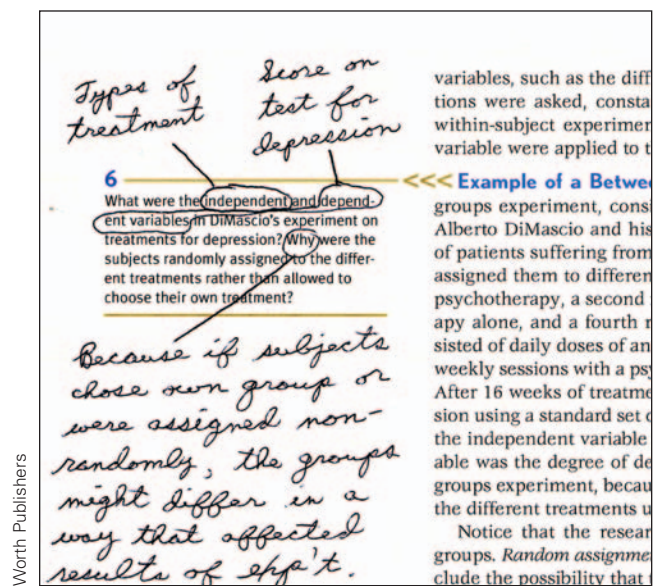
Admittedly, this approach to study will slow down your initial reading of each chapter. At first, stopping to think about each focus question may seem awkward and annoying. Your natural tendency may be to read straight through the text and ignore the focus questions, or to glance at them without thinking about them as you read. But most of the students I have taught, using

previous editions of this book, have told me on course surveys that the focus-question approach begins to seem natural with practice and that it improves their comprehension, enjoyment, and test performance. In the long run, for most students, it saves study time. Having understood and thought about the material the first time through, later study and review become relatively easy. Some students have even told me that they transfer this study skill to their reading of textbooks in other courses. In those books they do not find focus questions already written for them, but they use section headings and opening sentences to generate their own focus questions as they read. For more information about this study method, you might turn to Chapter 9, where textbook reading is discussed in the context of a more general discussion of ways to improve memory.

I should add, however, that a few students—roughly 10 percent of those I have taught—find that they do not need the focus questions to read very actively, thoughtfully, and effectively. These are students who naturally form questions as they read. They don't have the problem of a drifting mind that most of us have when reading assigned material. If you are one of those, then you may happily choose to ignore the focus questions on your first reading and use them just for review after you have finished each major section.

Using the Headings and Section Reviews to See the Hierarchical Structure of Ideas

Textbooks are always hierarchically organized. Each main heading refers to some major idea or argument, each subheading refers to a more limited idea that is part of the larger one, and each sub-subheading refers to a still more limited idea or ob-



Example of a student's notes (see page 30) Don't be afraid to write your own notes in the margin of this textbook. Your note taking will help you think about what you are reading and will be useful for review.

servation that is part of the subheading's idea. In this book I have tried to write all of the headings in such a way as to give you a pretty good idea what each section, subsection, and sub-subsection is about. By turning pages and reading all the headings within a main section before you read that section, you can preview the material and give yourself a head start toward thinking about it. You will see the basic structure of the argument that you are about to read, which will help you make sense of it as you read.

At the end of each main section of this book you will find a *section review*, a chart that reflects the hierarchical organization of the ideas and observations described in that section. I already described (on p. 9) how to use these charts for review. Reviewing in this way allows you to reflect back on each observation and idea as it relates to the larger idea to which it pertains. It helps you to see the individual elements of the section not as separate nuggets to memorize but as integral parts of a larger argument, and that helps you to make sense of them and remember them. A common mistake in studying is to write out lists of key terms and definitions and then study only that list. That method causes the student to lose track of the larger arguments, which tie the various elements together and make them interesting and memorable. The section review charts offer you an alternative way to view the key ideas—a way that maintains their relationship to one another.

For example, the section review on page 9 of this chapter depicts the organization of my argument that three foundation ideas preceded scientific psychology historically and provided a conceptual base on which psychology could build. That chart should help you think about the individual concepts, such as *empiricism* and *nativism*, in relation to the argument and should discourage you from memorizing definitions out of context. Your main goal is not to memorize definitions of these terms but to think about how these philosophical ideas helped to provide a foundation for the emergence of a science of psychology.

Using the Book's Other Features

Numbered Figures As you read each chapter, pay attention to the *numbered figures*. Whenever the text says, “see Figure Such-and-Such,” take a few moments to study the figure and read the caption. Many of the figures are graphs of data that back up an idea described in the text. Others are photographs or drawings that are designed to help you understand a point that would be difficult to grasp from words alone.

Bold Italicized Terms and Glossary Another feature that runs through the text is the use of ***bold italics*** to highlight technical terms that are likely to be new to many students. I suggest that you *not* devote much effort, on your first reading, to learning term definitions. Rather, read with the aim of understanding and thinking about the main ideas and the lines of evidence for or against them. In that process you will learn most of the terms, in the context of the ideas, without explicitly trying to learn them. The bold italics will be more useful in your later review. While you are reviewing the focus questions, look also for each of the terms in bold italics and check your knowledge of its meaning. These terms are also defined in the *Glossary* at the back of the book. If an important term has been defined in an earlier chapter, it is sometimes, but not always, defined again when it reappears. If it is not defined, you can use the Glossary to find both the term's definition and the number of the page on which it was first used.

Concluding Thoughts Section Each chapter ends with a section called *Concluding Thoughts*. This is not a summary; rather, it provides some ideas and integrating themes that emerge from reflecting back on the chapter as a whole. It also suggests, in some cases, some alternative ways for you to think about and review the whole chapter. This section is in a different format from the rest of the chapter, and there are no focus questions related to it; but that doesn't mean it is less important than the rest of the chapter. Some instructors, myself included, consider any new ideas or themes that emerge in the Concluding Thoughts section to be fair game for the test. In this first chapter, for example, you will find a discussion

20

How can students use the section and subsection headings, and the section review charts, to preview and review each major idea or argument? Why is the hierarchical organization of these study tools useful?

21

What purposes are served, for students, by the numbered figures, the use of bold italics, the Glossary, the Concluding Thoughts sections, the reference citations, the Further Reading sections, and the Subject Index in this book?

of the concept of *mind* in the Concluding Thoughts section that may be especially useful. Don't skip it.

Reference Citations, Further Reading, and Subject Index A feature that this book shares with other books and articles in psychology is the use of *reference citations*. This first chapter has relatively few of them, but in most chapters you will find them on nearly every page. Each citation consists of the name of one or more researchers followed by a year. Sometimes both the name (or names) and the year are in parentheses, such as (*Jones & Smith, 2001*); at other times, when the name or names are part of the sentence, only the year is in parentheses, such as *According to Alice Jones (2004) . . .* In either case, the year refers to the year of publication of an article or book, by the person or persons named, which describes more fully the idea or the research study being mentioned or discussed. The full reference to that article or book can be found in the *References* section at the back of the textbook, which is arranged alphabetically by authors' last names. At first you may find these citations disruptive to the flow of your reading, but you will soon learn to read right through them. Their purpose is to give credit to the people whose work or ideas are being described and to give you the opportunity to look up, and read more about, any ideas or research findings that intrigue you. To entice you to read further, you will also find, at the very end of each chapter, in a section called *Further Reading*, brief reviews of several interesting books that pertain to topics that were discussed in that chapter.

Many students save their introductory psychology textbook and use it, after the course is over, as a reference book to review topics that are relevant to other courses that they take. For them, the *Subject Index* at the back of the book, which lists topics alphabetically and indicates all the pages where each topic is discussed, is very useful. They also find the reference citations and Further Readings to be useful as sources for term papers, not just in psychology courses, but also in other social science courses, education courses, business courses, nursing courses, and some courses in such diverse subjects as biology, philosophy, and English. Those who decide to apply to graduate programs in psychology often use the book to help them prepare for the Graduate Record Examination in Psychology.

SECTION REVIEW

Using this book's special features can markedly increase your learning.		
Focus Questions • The focus questions in the margins are designed to help you read in an active, thoughtful way, which will help you understand and remember what you read. • The focus questions are also very useful in reviewing and testing your knowledge after you have read a section or chapter.	Headings and Section Reviews • The book's hierarchically arranged headings can help you quickly preview a section and prepare an organized mental structure for thinking about it. • Section reviews can help you to check your understanding of content, visualize relationships among the ideas and facts discussed, and consolidate your learning.	Other Features • Exploring the numbered figures, checking your understanding of bold italicized terms, using the Glossary as needed, and considering the integrative discussions in Concluding Thoughts can also benefit your study. • Reference citations, the Further Reading sections, and the Subject Index can help you access material of interest to you within and beyond the book.

Concluding Thoughts

In some cases, the Concluding Thoughts section at the end of a chapter is intended to help you organize your review of the chapter. In other cases, it makes explicit some idea that was implicit in the chapter or that emerges in reflecting back on the chapter as a whole. Here are two concluding thoughts, of the latter type, for the chapter you have just read.

1. Psychology as the study of normal behavior In the world outside of colleges and universities, psychology is often associated with the study of mental problems or disorders and how to help people overcome them. When people "go to see a psychologist," they are seeking help. But, as you have gleaned from this chapter, psychology as a research field is primarily aimed at

understanding normal human ways of feeling, thinking, and behaving. From an intellectual point of view, the problem of how any entity normally works—whether it is an automobile, a computer, or a human being—is much more interesting than the problem of how it breaks down. Philosophers such as Descartes, Hobbes, and the British empiricists were fascinated by the workings of the normal human mind, and that is what fascinates most academic psychologists today. The normal mental experiences that we take for granted in our everyday lives—such as our ability to see the color red, or to remember anything, or to learn our native language—become intriguing mysteries when we stop to think about them. As you go through this book, I hope you will allow yourself to become fascinated by all the normal things that you and other human beings do. If your ultimate interest is clinical psychology, keep in mind that some knowledge of normal functioning is essential background for figuring out where a breakdown has occurred and how normality might be restored. After all, automobile mechanics and computer repair people need to understand the normal operations of the machines they fix.

2. The concept of the mind as a product of the brain After reading this chapter, you may be confused about the meaning of the term *mind*. Join the club. Even today, *mind* ranks as perhaps the most debated of all philosophical concepts. Psychologists use the term in two quite different ways. In one use, it refers to people's *conscious experiences*—to the sensations, perceptions, memories, thoughts, desires, and emotions that run through our conscious awareness such that we can describe them to ourselves and to others. This is the usage that generates the most debate in philosophy, because it is hard to imagine just what “consciousness” is or how it emerges from physical matter. This difficulty leads some philosophers even today to be dualists. In psychology we usually sidestep that philosophical mystery by

defining consciousness as whatever it is that a person can tell us about his or her experiences. If a person can tell us about a memory, or about something that he or she is looking at, then we say that the person is conscious of that memory or that visual perception.

In its other use, *mind* refers to all of the *knowledge and operating rules* that are somehow built into or stored in the brain and that provide the foundation for organizing behavior and conscious experiences. By this definition, mind is analogous to the knowledge and rules that are programmed into a computer to organize its ability to calculate and perform all the operations we expect of it. In this usage, mind is not equated with consciousness. People are not aware of most of the inner knowledge and processes that organize their feelings, thoughts, and behavior, any more than a computer is aware of its software programs.

I have used the term *mind* in both of these ways in various parts of this book, but it should be clear from the context which meaning I am using at any given time. Some psychologists have argued—and some still argue—that we should avoid the concept of mind entirely. They point out that by either definition, *mind* refers to something that we cannot see directly. What we observe is people's behavior, including their verbal reports of their thoughts and feelings. With modern techniques of monitoring the brain, we can also observe physical changes in the brain, but, again, that is not the same as observing the mind. According to some psychologists, we should define our science as the study of behavior, or as the study of the brain and behavior, and leave the unobservable mind out of it. Most psychologists, however, find the concept of mind to be very useful. We can *infer* characteristics of the mind by observing behavior, and then we can use those inferences to make predictions about further behavior. Gravity can't be seen directly, either—it is inferred by observing the behavior of physical entities; yet physicists find the concept of gravity to be very useful.

Further Reading

DAVID HOTHERSALL (2004). *History of psychology* (4th ed.). New York: McGraw-Hill. This lively historical account begins with the psychological theories of the ancient Greeks and ends with theories that were dominant in the mid-twentieth century. The book takes primarily a biographical approach to history; it describes the ideas of the notable figures in psychology's history in the context of information about their lives.

JEROEN JANSZ & PETER VAN DRUNEN (Eds.) (2004). *A social history of psychology*. Malden, MA: Blackwell.

While most histories of psychology focus on the ideas and discoveries of philosophers and scientists, this history focuses on larger developments in society as a whole that influenced and motivated psychological thought and research. The book includes chapters on child rearing and education, madness and mental

health, work and organizations, culture and ethnicity, and delinquency and law, as well as three more general chapters that describe relationships between psychology and social developments.

ERIC LANDRUM, STEPHEN DAVIS, & TERESA LANDRUM (2000). *The psychology major: Career options and strategies for success*. Upper Saddle River, NJ: Prentice-Hall.

If you are considering psychology as a major, this 160-page paperback can help you decide. If you have already chosen to major in psychology, it can help guide you through your years of study. It is packed with practical information and suggestions—about jobs and careers, how to conduct library research and write papers, how to find research and internship opportunities, and how to prepare for and apply for graduate study.

MITCHELL PRINSTEIN & MARCUS PATTERSON (Eds.) (2003). *The portable mentor: Expert*

guide to a successful career in psychology. New York: Plenum.

This book was written primarily for graduate students in psychology, but would also be a useful source for undergraduates who are trying to decide whether or not to aim toward a career in psychology. Various chapters deal with the kinds of training needed for particular kinds of psychological careers, employment opportunities, and predicted future trends in psychological employment.

JULIE DeGALAN & STEPHEN LAMPERT (2001). *Great jobs for psychology majors* (2nd ed.). Lincolnwood, IL: NTC Contemporary Publishing.

This book contains useful information about what kinds of jobs are available for psychology majors (including jobs that do not require a postgraduate degree), how to find such jobs, and how to apply for them.

FPO

Methods of Psychology

LESSONS FROM CLEVER HANS

TYPES OF RESEARCH STRATEGIES

- Research Designs
- Research Settings
- Data-Collection Methods

STATISTICAL METHODS IN PSYCHOLOGY

- Descriptive Statistics
- Inferential Statistics

MINIMIZING BIAS IN PSYCHOLOGICAL RESEARCH

- Avoiding Biased Samples
- Avoiding Measurement Bias
- Avoiding Biases from Observers' and
Subjects' Expectancies

ETHICAL ISSUES IN PSYCHOLOGICAL RESEARCH

In Chapter 1, *psychology* was defined as the *science* of behavior and the mind. But what does it mean to say that psychology is a science? Science is the attempt to answer questions through the systematic collection and analysis of objective, publicly observable data (data that all observers can agree on). In psychology, the data are usually measures or descriptions of some form of behavior produced by humans or animals. Special problems exist in choosing what data to collect, collecting the data, and drawing conclusions from them.

This chapter is about scientific methods as applied to psychology. You will read sections on the research strategies psychologists use to answer questions, the statistical procedures they use to analyze data, the safeguards they employ to avoid biased results, and the protections they provide to human and animal research subjects. But first, to ease ourselves into the topic, here is a story about a horse—and a psychologist.

Lessons from Clever Hans

This story, a true one, took place in Germany near the beginning of the twentieth century. The horse was Clever Hans, famous throughout Europe for his ability to answer questions, and the psychologist was Oskar Pfungst. In a preface to the original account (Pfungst, 1911/1965), James Angell wrote, "Were it offered as fiction, it would take high rank as a work of imagination. Being in reality a sober fact, it verges on the miraculous." I tell the story here because of the lessons it teaches about scientific attitude and methods.

The Mystery

Hans's owner, a Mr. von Osten, was an eccentric retired school-teacher and devoted horseman who had long believed that horses would prove to be as intelligent as people if only they were given a proper education. To test his theory, von Osten spent 4 years tutoring Hans in the manner employed in the most reputable German schools for children. Using flash cards, counting frames, and the like, he set about teaching his horse reading, arithmetic, history, and other scholarly disciplines. He always began with simple problems and worked toward more complex ones, and he rewarded Hans frequently with praise as well as carrots. Recognizing that horses lack the vocal apparatus needed for speech, von Osten taught Hans to spell out words using a code in which the letters of the alphabet were translated into hoof taps and to answer yes-no questions by tossing his head up and down for "yes" and back and forth for

Pfungst, O. (1965). *Clever horse: The horse of Mr. Von Osten*. New York: Holt, Rinehart & Winston



Clever Hans at a mathematics lesson
Mr. von Osten believed that his horse was intellectually gifted, and so did many other people until the psychologist Oskar Pfungst performed some simple experiments.

that von Osten's claims were true. The evidence that most convinced them was Hans's ability to answer questions even when von Osten was not present, a finding that seemed to rule out the possibility that the horse depended on secret signals from his master. Moreover, several circus trainers, who specialized in training animals to give the appearance of answering questions, studied Hans and could find no evidence of trickery.

The Solution

1 How did Clever Hans give the appearance of answering questions, and how did Oskar Pfungst unveil Hans's methods?

<<< Hans's downfall finally came when the psychologist Oskar Pfungst performed a few simple experiments. Pfungst (1911/1965) theorized that Hans answered questions not through understanding them and knowing the answers but through responding to visual signals inadvertently produced by the questioner or other observers. Consistent with this theory, Pfungst found that the animal failed to answer questions when he was fitted with blinders so that he could not see anyone, and that even without blinders he could not answer questions unless at least one person in his sight knew the answer. With further study, Pfungst discovered just what the signals were.

Immediately after asking a question that demanded a hoof-tap answer, the questioner and other observers would naturally move their heads down just a bit to observe the horse's hoof. This, it turned out, was the signal for Hans to start tapping. To determine whether Hans would be correct or not, the questioner and other observers would then count the taps and, unintentionally, make another response as

soon as the correct number had been reached. This response varied from person to person, but a common component was a slight upward movement of either the whole head or some facial feature, such as the eyebrows. This, it turned out, was the signal for Hans to stop tapping. Hans's yes-no head-shake responses were also controlled by visual signals. Questioners and observers would unconsciously produce slight up-and-down head movements when they expected the horse to answer yes and slight back-and-forth head movements when they expected him to answer no, and Hans would shake his head accordingly.

Cues from the audience When members of the audience knew the answer to a question asked of Clever Hans, they inadvertently signaled the horse as to when to start and stop tapping, or which way to shake his head, through their own head movements.

The Granger Collection



All the signals that controlled Hans's responses were so subtle that even the most astute observers had failed to notice them until Pfungst pointed them out. And Pfungst himself reported that the signals occurred so naturally that, even after he had learned what they were, he had to make a conscious effort to prevent himself from sending them after asking a question. For 4 years, von Osten had believed that he was communicating scholarly information to Hans, when all he had really done was teach the horse to make a few simple responses to a few simple, though minute, gestures.

Facts, Theories, and Hypotheses

The story of Clever Hans illustrates the roles of facts, theories, and hypotheses in scientific research. A **fact**, also referred to as an *observation*, is an objective statement, usually based on direct observation, that reasonable observers agree is true. In psychology, facts are usually particular behaviors, or reliable patterns of behaviors, of persons or animals. When Hans was tested in the manner typically employed by von Osten, the horse's hoof taps or head shakes gave the appearance that he was answering questions correctly. That is a fact, which no one involved in the adventure disputed.

A **theory** is an idea, or a conceptual model, that is designed to explain existing facts and make predictions about new facts that might be discovered. Any prediction about new facts that is made from a theory is called a **hypothesis**. Nobody knows what facts (or perhaps delusions) led von Osten to develop his theory that horses have humanlike intelligence. However, once he conceived his theory, he used it to hypothesize that his horse, Hans, could learn to give correct answers to verbally stated problems and questions. The psychologist, Pfungst, had a quite different theory of equine intelligence: Horses don't think like humans and can't understand human language. In keeping with this theory, and to explain the fact that Hans seemed to answer questions correctly, Pfungst developed the more specific theory that the horse responded to visual cues produced by people who were present and knew the answers. This theory led Pfungst to hypothesize that Hans would not answer questions correctly if fitted with blinders or if asked questions to which nobody present knew the answer.

Facts lead to theories, which lead to hypotheses, which are tested with experiments or other research studies, which lead to new facts, which sometimes lead to new theories, which That is the cycle of science.

The Lessons

In addition to illustrating the roles of fact, theory, and hypothesis, the story contains three more specific lessons about scientific research:

1. *The value of skepticism.* People are fascinated by extraordinary claims and often act as though they want to believe them. This is as true today as it was in the time of Clever Hans. We have no trouble at all finding otherwise rational people who believe in astrology, psychokinesis, water divining, telepathy, or other occult phenomena, despite the fact that all these have consistently failed when subjected to controlled tests (Hines, 2003). Von Osten clearly wanted to believe that his horse could do amazing things, and to a lesser degree, the same may have been true of the scholars who had studied the horse before Pfungst. Pfungst learned the truth partly because he was highly skeptical of such claims. Instead of setting out to prove them correct, he set out to prove them wrong. His skepticism led him to look more carefully; to notice what others had missed; to think of an alternative, more mundane explanation; and to pit the mundane explanation against the paranormal one in controlled tests.
Skepticism should be applied not only to extraordinary theories that come from outside of science but also to the usually more sober theories produced by scientists themselves. The ideal scientist always tries to disprove theories, even those that are his or her own. The theories that scientists accept as correct, or

How are facts, theories, and hypotheses related to one another in scientific research?

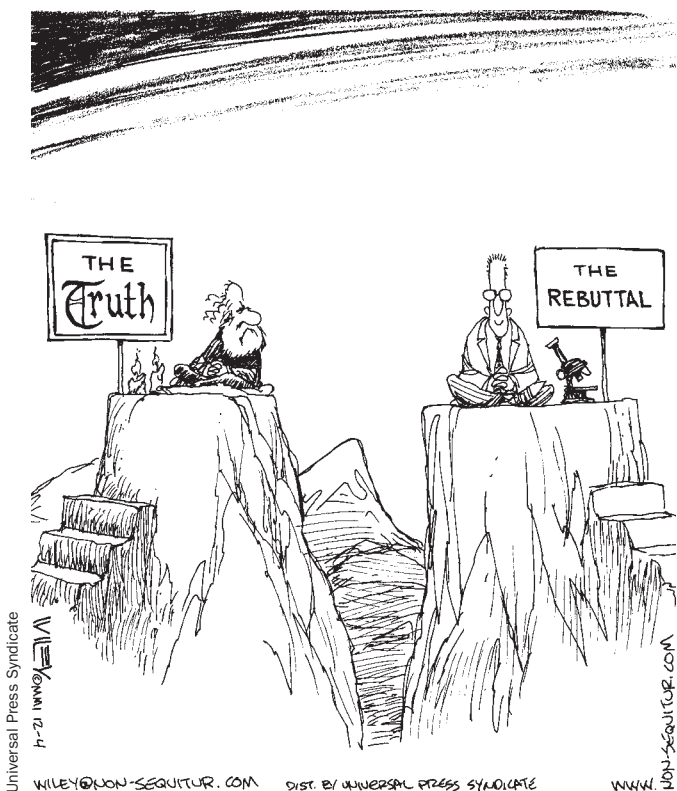
How does the Clever Hans story illustrate (1) the value of skepticism, (2) the value of controlled experimentation, and (3) the need for researchers to avoid communicating their expectations to subjects?

most likely to be correct, are those that could potentially be disproved but have survived all attempts so far to disprove them.

2. *The value of careful observations under controlled conditions.* Pfungst solved the mystery of Clever Hans by identifying the conditions under which the horse could and could not respond correctly to questions. He tested Hans repeatedly, with and without blinders, and recorded the percentage of correct answers in each condition. The results were consistent with the theory that the animal relied on visual signals.

Pfungst then pursued the theory further by carefully observing Hans's examiners to see what cues they might be sending. And when he had an idea what they might be, he performed further experiments, in which he deliberately produced or withheld the signals and recorded their effects on Hans's hoof-tapping and head-shaking responses. Careful observation under controlled conditions is a hallmark of the scientific method.

3. *The problem of observer-expectancy effects.* In studies of humans and other sentient animals the observers (the people conducting the research) may unintentionally communicate to subjects (the individuals being studied) their expectations about how the subjects “should” behave, and the subjects, intentionally or not, may respond by doing just what the researchers expect. The same is true in any situation in which one person administers a test to another. Have you ever taken an oral quiz and found that you could tell whether you were on the right track by noting the facial expression of your examiner? By tentatively trying various tracks, you may have finally hit on just the answer that your examiner wanted. Clever Hans's entire ability depended on his picking up such cues. We will discuss the general problem of expectancy effects later in this chapter (in the section on avoiding biases).



SECTION REVIEW

The case of the horse named Clever Hans illustrates a number of issues fundamental to scientific research.

Facts, Theories, and Hypotheses - Objective observations about behavior (facts) lead psychologists to create conceptual models or explanations (theories), which make specific, testable predictions (hypotheses). - Pfungst drew testable hypotheses from his theory that Hans was guided by visual cues from onlookers.	The Importance of Skepticism - Skeptics seek to disprove claims. That is the logical foundation of scientific testing. - A scientific theory becomes more believable as repeated, genuine attempts to disprove it fail. - Pfungst's skepticism caused him to test rather than simply accept claims about Hans's abilities.	Observation and Control - To test hypotheses, scientists control the conditions in which they make observations, so as to rule out alternative explanations. - Pfungst measured Hans's performance in conditions arranged specifically to test his hypothesis—with and without blinders, for example.	Observer-Expectancy Effects - Science is carried out by people who come to their research with certain expectations. - In psychology, the subjects—the people and animals under study—may perceive the observer's expectations and behave accordingly. - Cues from observers led Hans to give responses that many misinterpreted as signs of vast knowledge.
---	--	---	--

Types of Research Strategies

In their quest to understand the mind and behavior of humans and other animals, psychologists employ a variety of research strategies. Pfungst's strategy was to observe Clever Hans's behavior in controlled experiments. But not all research studies done by psychologists are experiments. A useful way to categorize the various research strategies used by psychologists is to think of them as varying along three dimensions (Hendricks & others, 1990). These dimensions are:

1. The *research design*, of which there are three basic types—experiments, correlational studies, and descriptive studies.
2. The *setting* in which the study is conducted, of which there are two basic types—field and laboratory.
3. The *data-collection method*, of which there are again two basic types—self-report and observation.

Each of these dimensions can vary independently from the others, resulting in 12 different kinds of studies, each defined by its combination of design, setting, and data-collection method. In what follows, we will first discuss the three types of research designs and then, more briefly, the other two dimensions.

Research Designs

Experiments

An experiment is the most direct and conclusive approach to testing a hypothesis about a cause–effect relation between two variables. A *variable* is simply anything that can vary. It might be a condition of the environment, such as temperature or amount of noise; or it might be a measure of behavior, such as score on a test. In describing an experiment, the variable that is hypothesized to cause some effect on another variable is called the **independent variable**, and the variable that is hypothesized to be affected is called the **dependent variable**. The aim of any experiment is to learn how the dependent variable is affected by (*depends on*) the independent variable. In psychology, dependent variables are usually measures of behavior, and independent variables are factors that are hypothesized to influence those measures.

More specifically, an **experiment** can be defined as a procedure in which a researcher systematically manipulates (varies) one or more independent variables and looks for changes in one or more dependent variables, *while keeping all other variables constant*. If all other variables are kept constant, and only the independent variable is changed, then the experimenter can reasonably conclude that any change observed in the dependent variable is *caused by* the change in the independent variable, and not by some other factor that happened to vary.

In any research study, the people or animals that are studied are referred to as the *subjects* of the study. In some experiments, called *within-subject* experiments, the different conditions of the independent variable are applied to each subject. In other experiments, called *between-groups experiments*, the different conditions are applied across the different groups of subjects.

Example of a Within-Subject Experiment In Pfungst's experiments with Clever Hans, there was just one subject, Hans. In each experiment, Hans was tested repeatedly, under varying conditions of the independent variable. In one experiment, to determine whether or not visual cues were critical to Hans's ability to respond correctly to questions, Pfungst tested the horse sometimes with blinders and sometimes without. In that experiment the independent variable was the presence or absence of blinders, and the dependent variable was the percentage of questions the horse answered correctly. The experiment could be described as a study of the effect of blinders (independent variable) on Hans's percentage of

>>> 4
How can an experiment prove the existence of a cause–effect relation between two variables?

>>> 5
What were the independent and dependent variables in Pfungst's experiment with Clever Hans?

correct responses to questions (dependent variable). Pfungst took care to keep other variables, such as the difficulty of the questions and the setting in which the questions were asked, constant across the two test conditions. This experiment is a within-subject experiment, because the different conditions of the independent variable were applied to the same subject.

6

What were the independent and dependent variables in DiMascio's experiment on treatments for depression? Why were the subjects randomly assigned to the different treatments rather than allowed to choose their own treatment?

<<< Example of a Between-Groups Experiment As an example of a between-groups experiment, consider an experiment in clinical psychology conducted by Alberto DiMascio and his colleagues (1979). These researchers identified a group of patients suffering from major depression (defined in Chapter 16) and randomly assigned them to different treatments. One group received both drug therapy and psychotherapy, a second received drug therapy alone, a third received psychotherapy alone, and a fourth received no scheduled treatment. The drug therapy consisted of daily doses of an antidepressant drug, and the psychotherapy consisted of weekly sessions with a psychiatrist that focused on the person's social relationships. After 16 weeks of treatment, the researchers rated each patient's degree of depression using a standard set of questions about mood and behavior. In this experiment, the independent variable was the kind of treatment given, and the dependent variable was the degree of depression after 16 weeks of treatment. This is a between-groups experiment, because the manipulations of the independent variable (that is, the different treatments used) were applied to different groups of subjects.

Notice that the researchers randomly assigned the subjects to the treatment groups. *Random assignment* is regularly used in between-group experiments to preclude the possibility that the groups differ, at the outset of the experiment, in some

systematic way that could bias the results. If DiMascio and his colleagues had allowed the subjects to choose their own treatment group, those who were most likely to improve even without treatment—maybe because they were more motivated to improve—might have disproportionately chosen one treatment condition over the others. In that case we would have no way to know whether the greater improvement of one group compared with the others derived from the treatment or from preexisting differences in the subjects. With random assignment, any differences among the groups that do not stem from the differing treatments must be the result of chance, and, as you will see later, researchers have statistical tools for taking chance into account in analyzing their data.

The results of this experiment are shown in Figure 2.1. Following a common convention in graphing experimental results, which is used throughout this book, the figure depicts variation in the independent variable along the horizontal axis and variation in the dependent variable along the

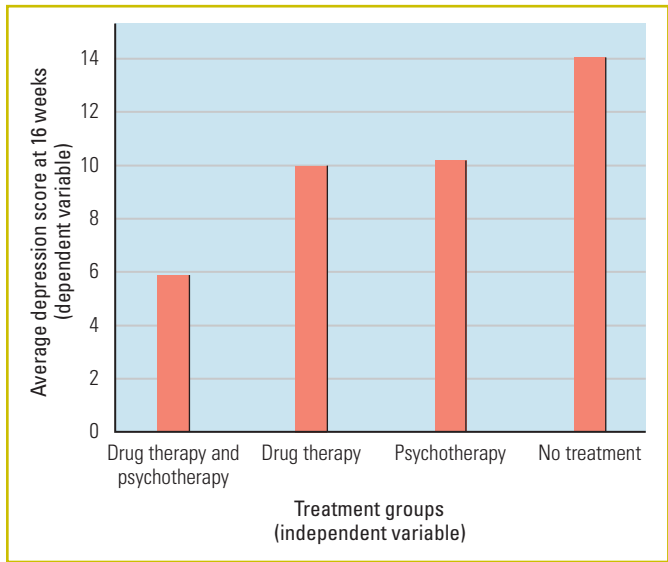


FIGURE 2.1 Effect of treatment condition on depression Subjects who received both drugs and psychotherapy were the least depressed at the end of the 16-week treatment period (according to the results of a standard interview procedure scored on a 17-point scale). In contrast, subjects who received no treatment were the most depressed. (From DiMascio & others, 1979.)

vertical axis. As you can see in the figure, those in the drug-plus-psychotherapy group were the least depressed after the 16-week period, and those in the no-treatment group were the most depressed. The results support the hypothesis that both drug therapy and psychotherapy help relieve depression, and that the two treatments together have a greater effect than either alone.

Correlational Studies

For many of the most interesting questions that psychologists ask, it is not possible to assign subjects to particular experimental conditions and control their experiences. Suppose, for example, that you are interested in the relation between the disciplinary styles of parents and the psychological development of their children. Perhaps you entertain the theory that strict punishment is harmful, that it promotes aggressiveness or other unwanted characteristics. To test that theory with an experiment, you would have to manipulate the discipline variable and then meas-

ure some aspect of the children's behavior. You might consider randomly assigning some families to a strict punishment condition and others to other conditions. The parents would then have to raise their children in the manners you prescribe. But you know that you cannot control families that way; it's not practical, not legal, and not ethical. So, instead, you conduct a correlational study.

A **correlational study** can be defined as a study in which the researcher does not manipulate any variable but observes or measures two or more variables to find relationships between them. Correlational studies can identify lawful relationships between variables, and thereby help us make predictions about one variable based on knowledge of another, but they do not tell us in any direct way whether change in one variable is the cause of change in another.

Example of a Correlational Study A classic example of a correlational study in psychology is Diana Baumrind's (1971) study of the relationship between parents' disciplinary styles and children's development. Instead of manipulating disciplinary style, she measured it. Through questionnaires and home observations, Baumrind classified disciplinary styles into three categories: *authoritarian* (high exertion of parental power), *authoritative* (a kinder and more democratic style, but with the parents still clearly in charge), and *permissive* (parental laxity in the face of their children's disruptive behaviors). She also rated the children on various aspects of behavior, such as cooperation and friendliness, through observations in their nursery schools. The main finding (discussed more fully in Chapter 12) was that children of authoritative parents scored better on the measures of behavior than did children of authoritarian or permissive parents.

Why Cause and Effect Cannot Be Determined from a Correlational Study

It is tempting to treat Baumrind's study as though it were an experiment and interpret the results in cause-effect terms. More specifically, it is tempting to think of the parents' discipline style as the independent variable and the children's behavior as the dependent variable and to conclude that differences in the former caused the differences in the latter. But because the study was not an experiment, we cannot justifiably come to that conclusion. The researcher did not control either variable, so we cannot be sure what was cause and what was effect. Maybe the differences in the parents' styles did cause the differences in the children's behavior, but other interpretations are possible. Here are some of these possibilities:

- Differences in children's behavior may cause differences in parents' discipline style, rather than the other way around. Some children may be better behaved than others for reasons quite separate from parental style, and parents with well-behaved children may simply glide into an authoritative mode of parenting, while those with more difficult children fall into one of the other two approaches as a way of coping.
- The causal relationship may go in both directions, with parents and children influencing each other's behavior. For example, children's disruptive behavior may promote authoritarian parenting, which may promote even more disruptive behavior.
- A third variable, not measured in Baumrind's study, may influence both parental style and children's behavior in such a way as to cause the observed correlation. For example, anything that makes families feel good about themselves (such as having good neighbors, good health, and an adequate income) might promote an authoritative style in parents and, quite independently, also lead children to behave well. Or maybe the causal variable has to do with the fact that children are genetically similar to their parents and therefore have similar personalities: The same genes that predispose parents to behave in a kind but firm manner may predispose children to behave well, and the same genes that predispose parents to be either highly punitive or neglectful may predispose children to misbehave.

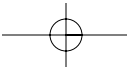
What are the differences between a correlational study and an experiment, in procedure and in types of conclusions that can be drawn?

How does an analysis of Baumrind's classic study of parental discipline styles illustrate the difficulty of trying to infer cause and effect from a correlation?

What causes what? Although many correlational studies have found a relationship between the viewing of televised violence and the displaying of aggressive behavior, such studies cannot tell us whether television inspires the aggressive behavior or whether aggressive individuals are more likely than others to watch violent television programs.



Peter Byron / Monkmeyer Press



In some correlational studies, one causal hypothesis may be deemed to be more plausible than others, but that is a judgment based on logical thought about possible causal mechanisms or on evidence from other sources, not from the correlation itself. For example, if I found a correlation between brain damage to a certain part of the brain, due to accidents, and the onset of a certain type of mental disorder immediately following the accident, I would probably be correct in inferring that the brain damage caused the mental disorder rather than the other way around. That possibility seems far more plausible than any other possible causal relationship between the two variables.

In Baumrind's study, one variable (parents' discipline style) was used to place subjects into separate groups, and the other (children's behavior) was compared across those groups. Many correlational studies are analyzed in that way, but in many others both variables are measured numerically and neither is used to assign subjects to groups. For example, a researcher might be interested in the correlation between the height of tenth grade boys (measured in centimeters or inches) and their popularity (measured by counting the number of classmates who list the boy as a friend). In such cases, the data are assessed by a statistic called the *correlation coefficient*, which will be discussed later (in the section on statistical methods).

Descriptive Studies

9

How do descriptive studies differ, in method and purpose, from experiments and from correlational studies?

<<< Sometimes the aim of research is to describe the behavior of an individual or set of individuals without systematically investigating relationships between specific variables. A study of this sort is called a **descriptive study**. Descriptive studies may or may not make use of numbers. As an example of one involving numbers, researchers might survey the members of a given community to determine the percentage who suffer from various mental disorders. This is a descriptive study if its aim is simply to describe the prevalence of each disorder without correlating the disorders to other characteristics of the community's members. As an example of a descriptive study not involving numbers, an animal behaviorist might observe the courtship behaviors of mallard ducks to describe in detail the sequence of movements that are involved. Some descriptive studies are narrow in focus, concentrating on one specific aspect of behavior, and others are very broad, aiming to learn as much as possible about the habits of a particular group of people or species of animal. One of the most extensive and heroic descriptive studies ever conducted is Jane Goodall's research on wild chimpanzees in Africa. She observed the apes' behavior over a period of 30 years and provided a wealth of information about every aspect of their lives.

Research Settings

The second dimension of research strategy is the research setting, which can be either the laboratory or the field. A **laboratory study** is any research study in which the subjects are brought to a specially designated area that has been set up to facilitate the researcher's collection of data or control over environmental conditions. A **field study** is any research study conducted in a setting other than a laboratory. Field studies in psychology may be conducted in people's homes, at their workplaces, at shopping malls, or in any place that is part of the subjects' typical environment.

10

What are the relative advantages and disadvantages of laboratory studies and field studies?

<<< Laboratory and field settings offer opposite sets of advantages and disadvantages. The laboratory allows the researcher to collect data under more uniform, controlled conditions than are possible in the field. However, the strangeness or artificiality of the laboratory may induce behaviors that obscure those that the researcher wants to study. A laboratory study of parent-child interactions, for example, may produce results that reflect not so much the subjects' normal ways of interacting as their reactions to a strange environment in which they know that they are being observed. To counteract such problems, some researchers combine laboratory and field studies. When, as often happens, the same conclusions emerge



FIGURE 2.2 | Sign used in a field experiment in Petrified Forest National Park. Park visitors often see signs describing and depicting undesired behaviors. Psychologist Robert Cialdini (2003) wondered if such signs are effective in reducing the behaviors that they are designed to reduce. As a test, he varied, on different days, the signs that were placed along trails in Petrified Forest National Park, and he measured, by observation, the amount of theft of petrified wood that occurred. When this sign was used, visitors actually took *more* pieces of petrified wood than when there was no sign. An explanation of why this might have happened is offered in Chapter 14.

from tightly controlled laboratory studies and less controlled but more natural field studies, researchers can be reasonably confident that the conclusions are meaningful (Anderson & others, 1999).

As you might expect, experiments are most often conducted in the laboratory because greater control of variables is possible in that setting, and correlational and descriptive studies are more often conducted in the field. But these relationships between research design and setting are by no means inevitable. Experiments are sometimes performed in the field, and correlational and descriptive studies are sometimes carried out in the laboratory. For an example of a field experiment, read the caption to Figure 2.2.

Data-Collection Methods

The third dimension of research is the data-collection method, of which there are two broad categories: self-report methods and observational methods.

Self-report methods are procedures in which people are asked to rate or describe their own behavior or mental state in some way. This might be done through a *questionnaire*, in which people produce self-descriptions either by checking off items on a printed list or by writing answers to essay questions. Or it might be done through an *interview*, in which people describe themselves orally in a dialogue with the interviewer. An interview may be tightly structured, with the interviewer asking questions according to a completely planned sequence, or it may be more loosely structured, with the interviewer following up on some of the subject's responses with additional questions. Researchers have developed numerical methods for scoring some structured interviews. For an example, look back at the method used to rate depression in the experiment depicted in Figure 2.1.

Observational methods include all procedures by which researchers observe and record the behavior of interest, rather than rely on subjects' self-reports. In one subcategory, **naturalistic observation**, the researcher avoids interfering with the subject's behavior. For example, an animal behaviorist might unobtrusively watch a duck's courtship behavior, or a developmental psychologist might observe children through a one-way window in a laboratory playroom. In the other subcategory, **tests**, the researcher deliberately presents stimuli or problems to which the subject responds. Examples include reaction-time tests, in which a person is asked to respond in a particular way as quickly as possible when a specific stimulus is presented; and problem-solving tests, such as maze problems for animals or paper-and-pencil logic problems for people.

None of these data-collection methods is in any absolute sense superior to another. Each has its purposes, advantages, and limitations. Questionnaires and interviews can provide information that researchers

How do self-report methods, naturalistic observations, and tests differ from one another? What are some advantages and disadvantages of each?



Dave Coverly / © 2004 Creators Syndicate, Inc.

Observing without interfering An obstacle to naturalistic observation is that the researchers may inadvertently, by their mere presence, influence the behavior they are observing. To minimize the degree to which her presence would affect the behavior of the Efe people she was studying, Gilda Morelli (at left) spent many months living among the Efe in Africa. People go about their daily activities more naturally in the presence of someone who is familiar and has proved trustworthy than they do in the presence of a stranger.



Courtesy of David S. Wilke

would not be able to obtain by watching subjects directly. However, the validity of such data is limited by the subjects' ability to observe and remember their own behaviors or moods and by their willingness to report those observations frankly, without distorting them to look good or please the researcher. Naturalistic observations allow researchers to learn firsthand about their subjects' natural behaviors, but the practicality of such methods is limited by the great amount of time they take, the difficulty of observing ongoing behavior without disrupting it, and the difficulty of coding results in a form that can be used for statistical analysis. Tests are convenient and easily scored but are by nature artificial, and the relationships between test results and everyday behaviors are not always clear. What is the relationship between a rat's ability to run a maze and the kinds of behaviors that rats engage in normally? What is the relationship between a person's score on an IQ test and his or her ability to solve the problems of daily living? These are the kinds of questions that psychologists must address whenever they wish to draw conclusions from test results about behaviors outside of the test environment.

SECTION REVIEW

Research strategies used by psychologists vary in their design, setting, and data-collection method.

Research Designs	Research Settings	Data-Collection Methods
● In an experiment (such as the experiment on treatments for depression), the researcher can test hypotheses about causation by manipulating the independent variable(s) and looking for corresponding differences in the dependent variable(s), while keeping all other variables constant. ● In a correlational study, a researcher measures two or more variables to see if there are systematic relationships among them. Such studies do not tell us about causation. ● Descriptive studies are designed only to characterize and record what is observed, not to test hypotheses about relationships among variables.	● Laboratory settings allow researchers the greatest control over variables, but they may interfere with the behavior being studied, by virtue of being unfamiliar or artificial. ● Field studies, done in “real-life” settings, have the opposite advantages and disadvantages, offering less control but perhaps more natural behavior.	● Self-report methods require the people being studied to rate or describe themselves, usually in questionnaires or interviews. ● Observational methods require the researcher to observe and record the subjects’ behavior, through naturalistic observation or some form of test. ● Each data-collection method has advantages and disadvantages.

Statistical Methods in Psychology

In order to make sense of the data collected in a research study, one must have some way of summarizing and analyzing the data to see if there are patterns that are not likely to be the result simply of chance. The statistical procedures used for these purposes can be divided into two categories: (1) *descriptive statistics*, which are used to summarize sets of data, and (2) *inferential statistics*, which help researchers decide how confident they can be in judging that the results observed are not due to chance. We shall look briefly here at some commonly used descriptive statistics and then at the rationale behind inferential statistics. A more detailed discussion of some of these procedures, with examples, can be found in the Statistical Appendix at the back of this book.

Descriptive Statistics

Describing a Set of Scores

Descriptive statistics include all numerical methods for summarizing a set of data. If we classified a group of individuals according to whether or not they showed a particular behavior, we might summarize the data by calculating the percentage who showed it. If our original data were a set of numerical measurements (such as depression ratings), we might summarize them by calculating either the mean or the median. The *mean* is simply the arithmetic average, determined by adding the scores and dividing the sum by the number of scores. The *median* is the center score, determined by ranking the scores from highest to lowest and finding the score that has the same number of scores above it as below it. (The Statistical Appendix explains when the mean or the median is the more appropriate statistic.)

For certain kinds of comparisons, researchers need to describe not only the central tendency (the mean or median) but also the variability of a set of numbers. *Variability* refers to the degree to which the numbers in the set differ from one another or from their mean. In Table 2.1 you can see two sets of numbers that have identical means but different variabilities. In set A the scores cluster close to the mean (low variability), and in set B they differ widely from the mean (high variability). A common measure of variability is the *standard deviation*. As illustrated in Table 2.1, the further most individual scores are from the mean, the greater is the standard deviation. This is calculated by a formula described in the Statistical Appendix.

>>> 12
How do the mean, median, and standard deviation help describe a set of numbers?

TABLE 2.1 |
Two sets of data, with the same mean but different amounts of variability

Set A	Set B
7	2
7	4
8	8
11	9
12	14
12	16
13	17
Median = 11	Median = 9
Total = 70	Total = 70
Mean = 70/7 = 10	Mean = 70/7 = 10
Standard deviation = 2.39	Standard deviation = 5.42

Describing a Correlation

Correlational studies, you recall, assess two or more variables to determine whether or not a relationship exists between them. When both variables are measured numerically, the strength and direction of the relationship can be assessed by a statistic called the *correlation coefficient*.

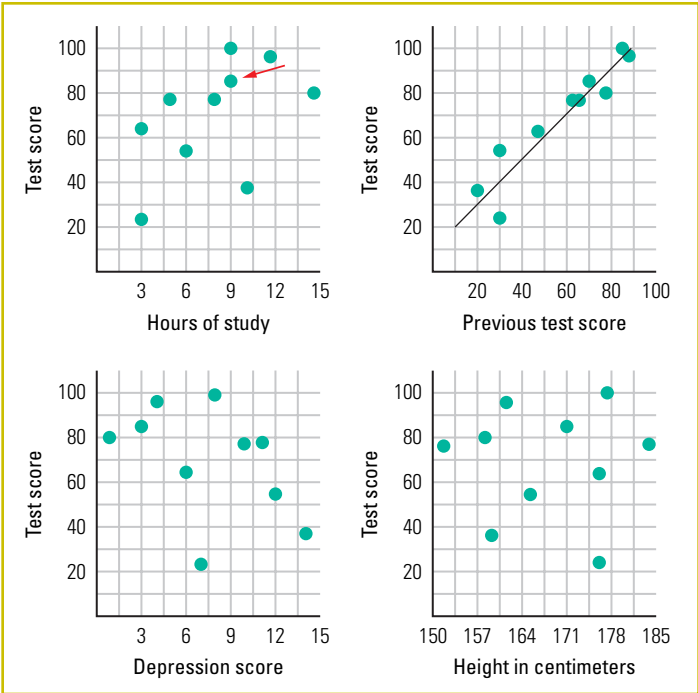
Correlation coefficients are calculated by a formula (described in the Statistical Appendix) that produces a result ranging from +1.00 to -1.00. The sign (+ or -) indicates the direction of the correlation (positive or negative). In a *positive correlation*, an increase in one variable coincides with a tendency for the other variable to increase; in a *negative correlation*, an increase in one variable coincides with a tendency for the other variable to decrease. The absolute value of the correlation coefficient (the value with sign removed) indicates the *strength* of the correlation. To the degree that a correlation is strong (close to +1.00 or -1.00), you can predict the value of one variable by knowing the other.

As an example, consider a hypothetical research study conducted with 10 students in a college course, aimed at assessing the correlation between the students' most recent test score and each of four other variables: (1) the hours they spent studying for the test; (2) the score they got on the previous test; (3) their level of

>>> 13
How does a correlation coefficient describe the direction and strength of a correlation? How can correlations be depicted in scatter plots?

Student ranked by test score	Test score	Hrs of study	Prvs. Test Score	Depres- sion	Height (cm)
1	100	9	85	8	177
2	97	11	88	4	160
3	85	9	70	3	171
4	80	14	79	1	158
5	78	8	62	10	184
6	78	5	65	11	152
7	62	3	44	6	176
8	54	6	30	12	165
9	38	10	20	14	159
10	22	3	30	7	176
Correlation with test score		+ .51	+ .93	- .43	- .04

FIGURE 2.3 | Examples of correlations, using made-up data The table at the top shows the most recent test score and four other measures for each of 10 students in a college course. At the bottom of each of the four right-hand columns is the correlation coefficient relating the data of that column to the test score. (The study is fictitious and the data were made up for illustrative purposes.) The four scatter plots depict, separately, the correlation between the test score and each of the four other measures. By comparing the plots, you can see the difference between weaker and stronger correlations and between positive and negative correlations.



psychological depression, measured a day before the test; and (4) their height in centimeters. Suppose the data collected in the study are those depicted in the chart at the top of Figure 2.3. Each row in the chart shows the data for a different student, and the students are rank ordered in accordance with their scores on the test.

To visualize the relationship between the test score and any of the other four variables, the researcher might produce a *scatter plot*, in which each student's test score and that student's value for one of the other measurements are designated by a single point on the graph. The scatter plots relating test score to each of the other variables are shown in Figure 2.3:

- Plot A illustrates the relation between test score and hours of study. Notice that each point represents both the test score and the hours of study for a single student. Thus the point indicated by the red arrow denotes a student whose test score is 85 and who spent 9 hours studying. By looking at the whole constellation of points, you can see that, in general, higher test scores correspond with more hours spent studying. This is what makes the correlation positive. But the correlation is far from perfect. It can be described as a *moderate positive correlation*. (The calculated correlation coefficient is +.51.)
- Plot B, which relates this test score to the score on the previous test, illustrates a *strong positive correlation*. Notice that in this plot the points fall very close to an upwardly slanted line. The closer the points are to forming a straight line, the stronger is the correlation between the two variables. In this study, score on the previous test is an excellent predictor of score on the new test. (The correlation coefficient is +.93.)
- Plot C, which shows the relation between test score and depression, illustrates a *moderate negative correlation*. Test score tends to decrease as depression increases. (The correlation coefficient here is -.43.)



Leo Cullum © 1994 from The New Yorker Collection. All Rights Reserved.

- Plot D, which shows the relation between test score and height, illustrates uncorrelated data—a correlation coefficient close to or equal to 0. Knowing a person's height provides no help in predicting the person's test score. (The correlation coefficient here is $-.04$.)

Inferential Statistics

Any set of data collected in a research study contains some degree of variability that can be attributed to chance. In the experiment comparing treatments for depression summarized back in Figure 2.1, the average depression scores obtained for the four groups reflect not just the effects of treatment but also random effects caused by uncontrollable variables. For example, more patients who were predisposed to improve could by chance have been assigned to one treatment group rather than to another. Or measurement error stemming from imperfections in the rating procedure could have contributed to differences in the depression scores. If the experiment were repeated several times, the results would be somewhat different each time because of such uncontrollable random variables. Given that results can vary due to chance, how confident can a researcher be in inferring a general conclusion from the study's data? Inferential statistics are ways of answering that question using the laws of probability.

>>> 14
Why is it necessary to perform inferential statistics before drawing conclusions from the data in a research study?

Statistical Significance

When two groups of subjects in an experiment have different mean scores, the difference might be meaningful or might be just the result of chance. Similarly, a non-zero correlation coefficient in a correlational study might indicate a meaningful relationship between two variables, or it might be just the result of chance. Inferential statistical methods, applied to either an experiment or a correlational study, are procedures for calculating the probability that the observed results could derive from chance alone.

Using such methods, researchers calculate a statistic referred to as p , or the *level of significance*. When two means are being compared, p is the probability that a difference as great as or greater than that observed would occur by chance if, in the larger population, there were no difference between the two means. ("Larger population" here means the entire set of scores that would be obtained if the experiment were repeated an infinite number of times with all possible subjects.) In the case of a correlational study, p is the probability that a correlation coefficient as large as or larger than that observed (in absolute value) would occur by chance if, in the larger population, the two variables were uncorrelated. By convention, results are usually labeled as **statistically significant** if the value of p is less than .05 (5 percent). To say that results are statistically significant is to say that the probability is small (generally less than 5 percent) that they could be caused by chance alone. All of the results of experiments and correlational studies discussed in this textbook are statistically significant at the .05 level or better.

>>> 15
What does it mean to say that a result from a research study is statistically significant at the 5 percent level?

The Components of a Test of Statistical Significance

There is no need to present here the precise formulas used to calculate p values for various kinds of research studies, but it is worthwhile to think a bit about the elements that go into such calculations. They are:

1. The *size of the observed effect*. Other things being equal, a large effect is more significant than a small one. For example, the larger the difference found between the mean scores for one group compared to another in an experiment, or the larger the absolute value of the correlation coefficient in a correlational study, the more likely it is that the effect is statistically significant. A large effect is less likely to be caused just by chance than is a small one.
2. The *number of individual subjects or observations* in the study. Other things being equal, the more subjects or observations included in a research study,

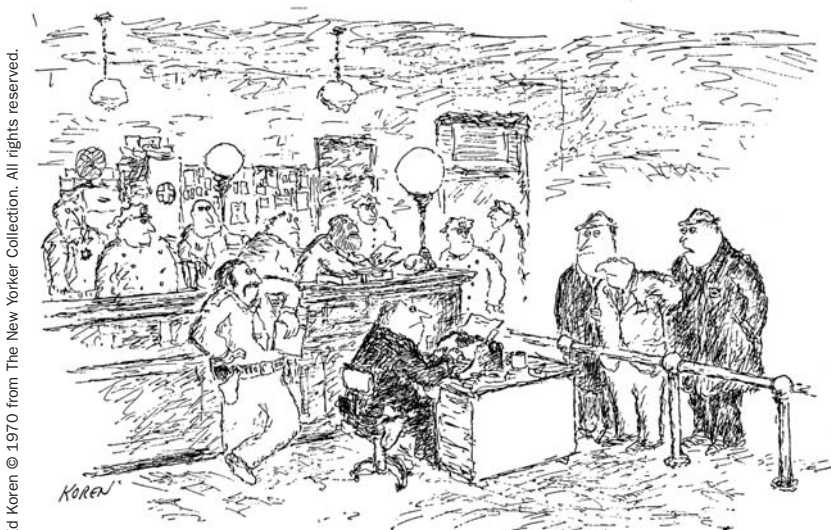
>>> 16
How is statistical significance affected by the size of the effect, the number of subjects or observations, and the variability of the scores within each group?

the more significant are the results. Large samples of data are less distorted by chance than are small samples. The larger the sample, the more accurately an observed mean, or an observed correlation coefficient, reflects the true mean, or correlation coefficient, of the population from which it was drawn. If the number of subjects or observations is huge, then even very small effects will be statistically significant.

3. The *variability of the data within each group*. This element applies to cases in which group means are compared to one another and an index of variability, such as the standard deviation, can be calculated for each group. Variability can be thought of as an index of the degree to which uncontrolled, chance factors influence the scores in a set of data. For example, in the experiment assessing treatments for depression, greater variability in the depression scores within each treatment group would indicate greater randomness attributable to chance. Other things being equal, the *less* the variability within each group, the more significant are the results. If all of the scores within each group are close to the group mean, then even a small difference between the means of different groups may be significant.

In short, a large observed effect, a large number of observations, and a small degree of variability in scores within groups all reduce the likelihood that the effect is due to chance and increase the statistical significance of the results.

Statistical significance tells us that a result probably did not come about by chance, but it does not, by itself, tell us that the result has practical value. Don't confuse statistical significance with practical significance. If I were to test a new weight-loss drug in an experiment that compared 1,000 people taking the drug with a similar number not taking it, I might find a high degree of statistical significance even if the drug produced an average weight loss of only a few ounces (or only a few grams). In that case, most people would agree that, despite the high statistical significance, the drug has no practical significance in a weight-loss program.



Ed Koren © 1970 from The New Yorker Collection. All rights reserved.

"He's charged with expressing contempt for data-processing."

SECTION REVIEW

Researchers use statistics to analyze and interpret the results of their studies.

Descriptive Statistics

- Descriptive statistics help to summarize sets of data.
- The central tendency of a set of data can be represented with the mean (the arithmetic average) or the median (the middle score).
- The standard deviation is a measure of variability, the extent to which scores in a set of data differ from one another or from the mean.
- Correlation coefficients represent the strength and direction of a relationship between two numerical variables.

Inferential Statistics

- Inferential statistics help us assess the likelihood that relationships observed are real and repeatable or due merely to chance.
- Statistically significant results are those in which the observed relationships are very *unlikely* to be merely the result of chance, so we expect them to be repeatable.
- Researchers calculate a statistic called *p*, which must generally be .05 or lower (indicating a 5% or lower probability that the results are due to chance) before the results are considered to be statistically significant.
- The calculation of a *p* value takes into account the size of the observed effect, the number of subjects or observations, and the variability of data within each group.

Minimizing Bias in Psychological Research

Good scientists strive to minimize both error and bias, especially the latter, in their research. **Error**, as a technical term, refers to random variability in results. Some degree of error is inevitable in psychological research, as a researcher can never precisely control all the extraneous variables that can influence a measure of behavior. The occurrence of error does not imply that the researcher has made a mistake. Individual differences among the research subjects and imperfections in the measure of behavior, for example, contribute inevitably to error. Because error is random, its consequences tend to disappear when averages are calculated, especially when the data set is large. Moreover, researchers can measure error precisely, by calculating the standard deviation, and can take it into account in their inferential statistics. Therefore, error is not a devastating problem in research. But bias is a devastating problem. **Bias**, as a technical term, refers to nonrandom (directed) effects caused by some factor or factors extraneous to the research hypothesis.

The difference between error and bias can be visualized by thinking of the difference between the sets of bullet holes produced by two men engaged in target practice. One is a novice. He hasn't learned to hold the gun steady, so it wavers randomly as he shoots. The bullets rarely hit the bull's-eye but scatter randomly around it (target A of Figure 2.4). His average shot, calculated as the average geometric location of his entire set of shots, is on or near the bull's-eye, even if few or none of his individual shots are on or near it. Those bullet holes exemplify error. The other man is an experienced marksman, but the sights on his rifle are out of alignment. Because of that defect, all his bullets strike the target some distance to the right of the bull's-eye (target B). Those misses exemplify bias, and they are not correctable by averaging. No matter how many times the man shoots or how carefully he aims, the average location of the whole set of bullet holes will be off the bull's-eye. Of course, error and bias can occur together, as would happen if a novice shooter were given a defective rifle (target C). In that case the holes would be widely scattered around a center that was some distance away from the bull's-eye.

Bias is a very serious problem in research because statistical techniques cannot identify it or correct for it. Whereas error only reduces the chance that researchers will find statistically significant results (by increasing the variability of the data), bias can lead researchers to the false conclusion that their hypothesis has been supported when, in fact, some factor irrelevant to the hypothesis has caused the observed results.

Avoiding Biased Samples

One source of bias in research has to do with the way in which the individuals to be studied are selected or assigned to groups. If the members of a particular group are initially different, in some systematic way, from those of another group, or are different from the larger population that the researcher is interested in, then that

>>> **17**
What is the difference between error and bias, and why is bias the more serious problem?

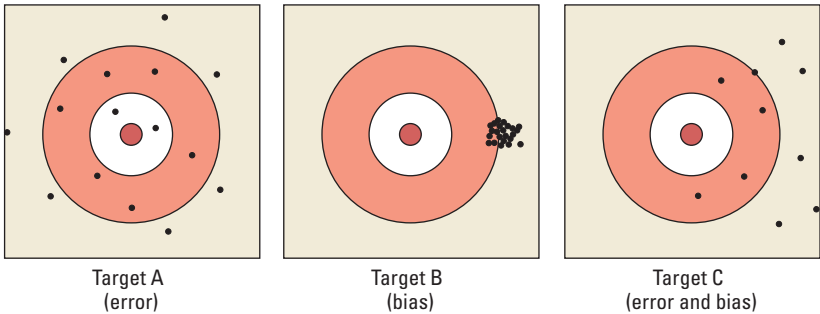
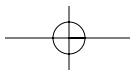


FIGURE 2.4 | Error, bias, and both
The difference between error and bias in research is like the difference between the sets of bullet holes produced by a novice shooter (target A) and by a skilled marksman whose rifle sights are misaligned (target B).



18

How can a nonrepresentative selection of research participants introduce bias into (a) an experiment and (b) a descriptive study?

group is a **biased sample**. Conducting research with a biased sample is like shooting a rifle whose sights are misaligned. No matter how large the sample, the results will be off target.

<<< Suppose that in the experiment on depression (Figure 2.1), the researchers had not randomly assigned the subjects to the different treatments but had allowed them to choose their own. In that case, biased samples could have resulted, because people's choices might have been based on preexisting differences among them. For example, those who felt most motivated to overcome their depression might have been particularly likely to choose the psychotherapy condition. Thus any greater improvement by those in psychotherapy compared with the others might have resulted not from the psychotherapy but from the preexisting difference in motivation. When subjects are assigned randomly to groups, their individual differences are merely a source of error (and can be taken into account by the statistics); but when subjects are not assigned randomly, their differences can be a source of bias as well as error.

A sample is biased when those in it are not representative of the larger population that the researchers are trying to describe. A classic example of the effect of a biased sample in descriptive research is the *Literary Digest's* poll of U.S. voters in 1936, which led the *Digest* to announce that Alf Landon would beat Franklin D. Roosevelt in the presidential election that year by a margin of 2 to 1 (Huff, 1954). It turned out that the publication's conclusion could not have been more mistaken—Roosevelt won by a landslide. The *Digest* had conducted its poll by telephoning its subscribers. In 1936, in the midst of the Great Depression, people who could afford magazine subscriptions and telephones may indeed have favored Landon, but the great majority of voters, as the election showed, did not.

Avoiding Measurement Bias

19

What is the difference between the reliability and the validity of a measurement procedure? How can lack of validity contribute to bias?

<<< Psychological researchers must give careful thought to their measures of behavior. A good measure is reliable and valid. Reliability has to do with measurement error, not bias. A measure is **reliable** to the degree that it yields similar results each time it is used with a particular subject under a particular set of conditions. Measuring height with a cloth measuring tape is not as reliable as measuring it with a metal measuring tape because the elasticity of the cloth causes the results to vary from one time to the next. A psychological test is not reliable if the scores vary widely because of uncontrolled variations in how it is administered or how subjects respond to it. Because it is a source of error, low reliability decreases the chance of finding statistical significance in a research study.

Validity is an even more critical issue than reliability, because lack of validity can be a source of bias. A measurement procedure is **valid** if it measures or predicts what it is intended to measure or predict. A procedure may be reliable and yet not be valid. For example, assessing intelligence in adults by measuring thumb length is highly reliable (you would get nearly the same score for a given person each time) but almost certainly not valid (thumb length is almost assuredly unrelated to adult intelligence). This invalid measure exemplifies bias because it would produce false conclusions, such as the conclusion that tall people (who would have long thumbs) are more intelligent than short people. If common sense tells us that a measurement procedure assesses the intended characteristic, we say the procedure has *face validity*. A test of ability to solve logical problems has face validity as a measure of intelligence, but thumb length does not.

20

How can the validity of a measurement procedure be assessed?

<<< A more certain way to gauge the validity of a measurement procedure is to correlate its scores with another, more direct index of the characteristic that we wish to measure or predict. In that case, the more direct index is called the *criterion*, and the validity is called *criterion validity*. Suppose, for example, that I defined intelligence as the quality of mind that allows a person to achieve greatness in any of various realms, including business, diplomacy, science, literature, and art. With this definition, I might use the actual achievement of such greatness as my criterion

for intelligence. I might identify a group of people who have achieved such greatness and a group who, despite similar environmental opportunities, have not, and assess the degree to which they differ on various potential measures of intelligence. The more the two groups differ on any of the measures, the greater the correlation between that measure and my criterion for intelligence, and the more valid the test. As you can see from this example, the assessment of validity requires a clear definition of the characteristic to be measured or predicted. If your definition of intelligence differs from mine, you will choose a different criterion from mine for assessing the validity of possible intelligence tests.

As another illustration of how an invalid measure can bias a research study, think once more about the study of depression (Figure 2.1). Suppose that the interview procedure used in the study was not a valid measure of depression but actually measured subjects' desire to appear depressed or not. In that case, the results would still be statistically significant, but the researchers' conclusion (that treatment affects depression) would be mistaken. A more accurate conclusion would be that some treatments lead to a stronger desire to appear not depressed than do others.

Avoiding Biases from Observers' and Subjects' Expectancies

Being human, researchers inevitably have wishes and expectations that can affect how they behave and what they observe when recording data. The resulting biases are called **observer-expectancy effects**. A researcher who desires or expects a subject to respond in a particular way may unintentionally communicate that expectation and thereby influence the subject's behavior. As you recall, Pfungst discovered that this sort of effect provided the entire basis for Clever Hans's apparent ability to answer questions. That episode occurred nearly a century ago, but the power of observer expectancies to delude us is as strong today as ever. A dramatic recent example concerns a technique designed to enable people with autism to communicate.

The Facilitated-Communication Debacle

Autism is a congenital (present at birth) disorder of development, characterized principally by a deficit in the ability to form emotional bonds and to communicate with other people. Some people with autism fail almost completely to develop either spoken or written language. Some years ago, in the prestigious *Harvard Educational Review*, Douglas Biklen (1990) described an apparently remarkable discovery, made originally by Rosemary Crossley in Australia. The discovery was that people with severe autism, who had previously shown almost no language ability, could type meaningful statements with one finger on a keyboard. They could answer questions intelligently, describe their feelings, display humor, and write emotionally moving poetry by typing. To do this, however, a "facilitator" had to help by holding the typing hand and finger of the autistic person. According to Crossley and Biklen, the hand-holding was needed to calm the person, to keep the typing hand steady, and to prevent repeated typing of the same letter. (People with autism tend to repeat their actions.)

The community concerned with autism—including special educators and parents of autistic children—responded with great excitement to this apparent discovery. It is hard, emotionally, to care for and work with people who don't communicate their thoughts and feelings; it takes enormous dedication. You can imagine the thrill that parents and teachers felt when their autistic children typed, for the first time, something like "I love you." The motivation to believe in this new method was enormous. Workshops were held to teach people to be facilitators, and thousands of teachers and parents learned the technique. By 1993, over \$100 million a year was being spent by the U.S. educational system on equipment and personnel for facilitated communication (Levine & others, 1994).

>>> **21**
How can the supposed phenomenon of facilitated communication by people with autism be explained as an observer-expectancy effect?

Who is typing the message? In the technique of facilitated communication, a facilitator holds the hand of the disabled person in order to “help” that person type a message on a keyboard. Experiments have shown that when this technique is used with autistic individuals, it is the facilitator, not the autistic person, who composes the message and controls the typing, even though the facilitator is not conscious of doing so.



© Bob Mahoney / The Image Works

Yet, from the beginning, there were skeptics. The credulity of some observers was strained by the sudden appearance of literary skills and erudition in people who had never previously shown evidence that they could read. As an alternative theory, the skeptics proposed that the messages were not communications from the autistic persons but unconscious creations of the facilitators (Dillon, 1993; Levine & others, 1994). The skeptics suggested that hand movements, made unconsciously by the facilitator, guided the autistic person's finger to the keys. Consistent with this view, some skeptics noticed that the autistic persons often did not even look at the keyboard as they ostensibly did their typing but that the facilitators always looked. The issue soon became important for moral and legal reasons as well as educational and scientific ones. Some autistic children, working with facilitators, typed out messages that accused parents or other caregivers of sexually abusing them (Bligh & Kupperman, 1993; Heckler, 1994). Could facilitated messages be used by child-welfare authorities as a basis for taking a child from a parent?

Partly in response to ensuing court cases, many experiments were performed in the 1990s to test whether facilitated messages are creations of the autistic person or of the facilitator. In a typical experiment, pairs consisting of a facilitator and an autistic person, chosen because of their experience working together and their putative skill at the technique, were tested under two conditions. In one condition, the item of information that the autistic person was asked to communicate was also shown to the facilitator, and in the other condition it was not. For example, the autistic person might be shown a picture of a common object and asked to type the name of that object, under conditions in which the facilitator either saw the object or did not. The inevitable result was that many correct responses were typed in the first condition (in which the facilitator was shown what the autistic person was shown), but not in the second condition. When the facilitator did not know what the autistic person was shown, the number of correct responses was no more than what would be produced by random guessing (Jacobson & others, 1995; Mostert, 2001). Subsequent research revealed that people newly trained in facilitated communication immediately and unconsciously begin to control the other person's hand movements (Wegner & others, 2003). It does not feel to them that they are creating the messages and controlling the typing, even though they are.

The original observers of facilitated communication—who were also the original facilitators—were deluded by a powerful effect of their own expectations. They be-

lieved that an autistic child could type out meaningful messages, and, at any given instance during the facilitation, they had some idea in mind (perhaps unconsciously) of what that message was and what letter should come next. Their expectation that a particular letter would be typed led them to feel that the autistic child's finger was moving toward that letter on the keyboard, and they experienced themselves as merely “facilitating” that movement, when actually they were creating it.

Avoiding Observer-Expectancy Effects in Typical Experiments

In the experiments testing the validity of facilitated communication, the influence of observer expectancy was the focus of study and was assessed by varying the observers' (in this case, the facilitators') knowledge. In a more typical psychological experiment, the objective is not to study observer-expectancy effects but to eliminate them in order to observe other effects without bias. A researcher who expects subjects in one condition to behave differently from those in another condition may send different unintentional signals in the two conditions, and thereby elicit the expected difference in behavior. In addition to influencing subjects' behavior, expectancy may affect the observer's perception or judgment of that behavior. For example, an observer who expects to find that people smile more in one condition than in another may interpret an ambiguous facial expression as a smile in the one condition and as something else in the other.

The best way to prevent observer-expectancy effects is to keep the observer **blind**—that is, uninformed—about those aspects of the study's design that could lead to expectations about the outcome. Thus in a between-groups experiment, a blind observer would not be told which subjects received which treatment. Not knowing who is in which group, the blind observer has no basis for expecting a particular behavior from a particular subject. In the study of treatments for depression (illustrated in Figure 2.1), the clinicians who evaluated the patients' depression at the end of the treatment period were blind to treatment condition. To keep them blind, patients were instructed not to say anything about their treatment during the evaluation interview.

Avoiding Subject-Expectancy Effects

Subjects also have expectations. If different treatments in an experiment induce different expectations in subjects, then the expectations, rather than anything else about the treatments, may account for observed differences in how the subjects respond. Effects of this sort are called **subject-expectancy effects**. For example, people who take a drug may subsequently behave or feel a certain way simply because they believe that the drug causes such a behavior or feeling. Similarly, subjects who receive psychotherapy may improve simply because they believe that psychotherapy will help them.

Ideally, to prevent bias from subject expectancy, subjects should be kept blind about the treatment they are receiving. Any experiment in which both the observer and the subjects are kept blind in this way is called a **double-blind experiment**. In double-blind drug studies, for example, some subjects receive the drug while other subjects receive a **placebo**, an inactive substance that looks like the drug, and neither subjects nor observers are told who received the drug and who did not. Consequently, any observed difference between those who got the drug and those who did not must be due to the drug's chemical qualities, not to the subjects' or observers' expectancies.

Subjects cannot always be kept blind concerning their treatment. For instance, it is impossible to administer psychotherapy to people without their knowing it. As a partial control in some psychotherapy experiments, subjects in the non-psychotherapy group are given a fake form of therapy designed to induce subject expectancies equivalent to those induced by actual psychotherapy. In the depression experiment described earlier, subjects were not blind about their treatment. Those in the non-drug groups did not receive a placebo, and those in the non-

>>> 22
What are two ways by which an observer's expectations can bias results in a typical experiment? How does blind observation prevent such bias?

>>> 23
How can subjects' expectancies bias the results of an experiment? How does a double-blind procedure control both subjects' and observers' expectancies?

psychotherapy groups did not receive fake psychotherapy. The results depicted in Figure 2.1 might possibly, therefore, be *placebo effects*, that is, subject-expectancy effects caused by subjects' beliefs that the treatment would help them. Other experiments (discussed in Chapter 17) have shown that placebos can indeed reduce depression (Kirsch & others, 2002).

SECTION REVIEW

Bias—nonrandom effects caused by extraneous factors—must be avoided.		
Biased Samples • Unless subjects in a between-groups experiment are assigned to groups randomly, an observed difference in the dependent variable may be caused by systematic <i>a priori</i> differences between the groups rather than by the independent variable. • If the subjects in any type of study are not representative of the group to which the researcher wants to generalize, false conclusions may be drawn.	Measurement Error and Bias. • A good measure is reliable—able to yield similar results with repeated use on the same subjects in the same conditions. • An unreliable measure leads to error—random variability that makes it harder to establish statistical significance. • A good measure is also valid—able to measure what it is intended to measure. Invalid measures are sources of bias. • A measure that seems valid on the basis of common senses has face validity. A measure that actually correlates with another, more direct, measure of the variable has criterion validity.	Expectancy Effects • A researcher's expectations about a study's results can influence those results. This is observer-expectancy bias. • Subjects' expectations as to how they should respond can also influence results. This is subject-expectancy bias. • Such expectancy effects can occur without intention or even awareness. • In observer-blind studies, the observer is deliberately kept ignorant of information that could create expectancies. In double-blind studies, observers <i>and</i> subjects are kept ignorant of such information.

Ethical Issues in Psychological Research

Psychologists must consider ethical as well as scientific issues in designing their studies. DiMascio and his colleagues could, from a scientific vantage point, have improved their study of treatments for depression by using a placebo in the non-drug conditions and a fake form of psychotherapy in the non-psychotherapy conditions to reduce differences among the groups in subject expectancies. But the researchers felt that their subjects should know which form of treatment they were getting so that they could make an informed decision about whether or not to participate and could understand any side effects that might arise as treatment progressed.

Research with Humans

24 What are the ethical concerns pertaining to privacy, discomfort, deception, and animal welfare in psychological research? How do researchers strive to minimize problems related to these concerns?

<<< In research with humans, ethical considerations revolve around three interrelated issues:

1. *The person's right to privacy.* Common safeguards to protect privacy include obtaining informed consent from subjects before they take part, informing subjects that they do not have to share any information about themselves that they do not wish to share, and keeping reports and records in ways that ensure anonymity.
2. *The possibility of discomfort or harm.* If a planned research study involves some risk of discomfort or harm to subjects, researchers are obliged to determine whether the same question can be answered in a study that involves less risk. If the answer is no, a determination must be made that the risk is minimal and is outweighed by the human benefits of the study. In addition, in any research

study, human subjects must be advised that they are free to quit at any time. (In reality, the great majority of psychological studies involve completely harmless procedures, such as reading rapidly flashed letters, memorizing lists of terms, or carrying on a discussion with other research participants.)

3. *The use of deception.* This is the most controversial ethical issue in human psychological research. In a small minority of experiments, the independent variable involves a lie. Subjects may be falsely told or led to believe that something is happening, or is going to happen, so that the researcher can study the effects of that belief. Some psychologists are opposed to all use of deception. They argue that deception (a) is intrinsically unethical and (b) undermines the possibility of obtaining truly informed consent (Carroll & others, 1985). Others, however, justify deception on the grounds that some psychological processes cannot be studied effectively without it. These psychologists contend, further, that research deception usually takes the form of benign “white lies,” which are cleared up when the researcher informs the subjects of the true nature of the study after the session has ended. They also point out that informed consent can still be obtained by telling subjects about any realistic dangers they will face and by telling them that some details of the study must be withheld until the data have been collected.

Research with Animals

The use of nonhuman animals in research presents another area of ethical controversy. Most people agree that some procedures that would be unethical to use with humans can be used ethically with other animal species. One can breed animals in controlled ways, raise them in controlled environments, and surgically intervene in their physiology for research purposes. Basic biological mechanisms underlying animal behavior are similar to those underlying human behavior, so such research contributes to our understanding of humans as well as the species studied. Still, research on nonhuman animals can sometimes cause them to suffer, and any researcher who employs animals has an ethical obligation to balance the animals' suffering against the potential benefits of the research. Animals must be well cared for and not subjected to unnecessary deprivation or pain.

Some people question whether subjecting animals to pain or deprivation for research purposes is ever justifiable. But others, pointing to the enormous gains in knowledge and the reduction in human (and animal) suffering that have come from such research, have turned the ethical question around. In the words of one (N. E. Miller, 1986): “Is it morally justifiable to prolong human (and animal) suffering in order to reduce suffering by experimental animals?”

Formal Principles and Safeguards for Ethical Research

The American Psychological Association (2002) has established a set of ethical principles for psychological research, which researchers must follow if they are to publish their results in the research journals of that association. Moreover, in the United States, Canada, and many other countries, publicly funded research institutions are required by law to establish ethics review panels, whose task is to evaluate all proposed research studies that have any potential for ethical controversy. Such panels often turn down research proposals that once were regarded as quite acceptable. A few studies that are now considered classics and are cited in most general psychology textbooks, including this one, would not be approved today. As you read about some of the studies in the chapters that follow, questions of ethics may well occur to you from time to time. Such questions are always legitimate, as are those about the scientific merit of a study's design and the interpretation of its results. Psychology needs and usually welcomes people who raise those questions.



Joel Gordon

A rat with an electrode in its brain
Experiments in behavioral neuroscience frequently involve operations on animal brains. For scientific as well as ethical reasons, conscientious researchers are scrupulous about minimizing discomfort to the animals. Discomfort can produce behaviors that interfere with those that the researcher wishes to study.

SECTION REVIEW

Psychologists must deal with ethical concerns in conducting research.

Human Subjects

- A human subject’s right to privacy must be protected.
- The risk of discomfort or harm to human subjects must be minimal.
- Deceiving subjects about some aspect of a study is both common and controversial.
- Routine measures to protect subjects include obtaining informed consent, letting subjects know they can quit at any time, ensuring anonymity in results, and debriefing subjects about deception after the study ends.

Animal Subjects

- Many procedures that would be unethical with humans—such as controlled breeding and surgical interventions—are performed with animals.
- The benefits of the knowledge gained from such research are the primary ethical justification for them, since common biological mechanisms often enable us to apply findings from animal studies to humans.
- Animals used in research must be well cared for, must not suffer unnecessary deprivation or pain, and must have their suffering balanced against the potential value of the knowledge gained.

Concluding Thoughts

As you review and think about the concepts in this chapter, you might think also about the following two questions.

1. How does science compare with everyday observation and thought? No sharp dividing line exists between science and the kinds of observation and thought that all of us use every day to learn about the world around us. In our everyday learning, we begin with the data of our senses and use those data to draw tentative conclusions (make inferences) about specific aspects of our world. For example, we might one day observe someone from town X acting politely and someone from town Y acting rudely and infer from those observations that people from X are more polite than people from Y. Most of us make such inferences all the time, often on scarcely more evidence than that. Science is simply the attempt to improve on our natural ways of learning by systematizing the data-collection procedures, controlling conditions to be more certain about which variables are having which effects, striving to eliminate sources of bias, deliberately thinking of alternative explanations, and using statistical procedures to assess the degree of confidence we should have in our tentative conclusions.

As you review each of the main concepts discussed in the sections on research strategies, statistical methods, and sources of bias in this chapter, you might think about how that concept applies—somewhat less formally—to the distinctions between good and poor observation and thought in everyday life. We are observing and thinking poorly when we draw firm conclusions from too little evidence, or neglect to think about alternative explanations, or fail to see what is really there because of our biased expectations.

2. What is a science of psychology for? I remember, as a college freshman on my first visit home, expressing pride about an A that I had received in calculus. My mother, hearing me brag and having a knack for deflating undue pride and putting things into perspective, asked a simple question: What is calculus for? I was floored. I could rattle off terms and equations

about calculus, and I could solve the problems as they were given to me in the class, but I had no understanding at all of what calculus was for. Perhaps that is why, by a few months after the class had ended, I had completely forgotten the terms, the equations, and the way to solve them. So what is a science of psychology for?

Some people think of psychology purely in applied terms, as a means of solving human problems. These people are likely to appreciate the study on treatments for depression (illustrated in Figure 2.1) but are less likely to appreciate research motivated by simple curiosity rather than by pragmatic concerns. The issue of pragmatism occurs in other sciences as well. What good does it do us to know what the other side of the moon looks like?

For the most part, people go into psychological research, or any other research field, because they are curious or because they are thrilled by the prospect of being the first to uncover some mystery of nature, large or small. So psychology, like any other science, has two purposes: to solve practical problems and to satisfy the human quest for knowledge. It is hard to separate the two, however, because research done solely to satisfy curiosity very often reveals solutions to practical problems that at first seemed unrelated to the research. As you read the remaining chapters of this book, I hope you will allow yourself to become engaged by the questions for their own sake, regardless of whether you think they have practical applications. Each chapter contains mysteries—some solved, some not.

Furthur Reading

RANDOLPH A. SMITH & STEPHEN F. DAVIS (2001). *The psychologist as detective: An introduction to conducting research in psychology* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

By presenting the research psychologist as a detective, and by giving many examples of the solving of psychological mysteries, this book enlivens the often dry stuff of research methodology. It deals with all aspects of the research process: forming hypotheses or questions, using the library, designing research studies, considering ethical issues, analyzing and graphing data, and writing reports.

KEITH E. STANOVICH (2003). *How to think straight about psychology* (7th ed.). Boston: Allyn & Bacon.

This fun-to-read book deals with popular misconceptions about psychological phenomena, the faulty uses of evidence and logic that lead to such misconceptions, and the processes through which enduring psychological knowledge has been developed. The author's goal, well

achieved, is to help readers acquire the tools needed for critical thinking in psychology.

DARRELL HUFF (1954). *How to lie with statistics*. New York: Norton.

This witty paperback, filled with anecdotes and cartoons, has a serious message. It tells you how not to lie with statistics and how to spot lies when they occur. It is guaranteed to make you a more critical consumer of statistical information. Don't be put off by the publication date; the book is reprinted regularly and is widely available.

TERENCE HINES (2003). *Pseudoscience and the paranormal* (2nd ed.). Amherst, NY: Prometheus Books.

This is a book about the "evidence," faulty reasoning, and (in some cases) trickery that lead many people to believe in phenomena that violate our normal expectations of how the world works. Among the beliefs discussed are those concerning psychic readings, prophetic dreams, astrology, biorhythm theory, ESP, psy-

chokinesis, UFOs, alien abductions, faith healing, psychic surgery, certain forms of pseudo-psychotherapy, facilitated communication, and water dowsing. Central to the book is the idea of an unfalsifiable theory, a theory or belief stated in such a way that it cannot be proved wrong. With such a theory any evidence that would seem to contradict it is discounted.

ROBERT J. STERNBERG (2003). *The psychologist's companion: A guide to scientific writing for students and researchers* (4th ed.). Cambridge, England: Cambridge University Press.

Psychologists, like scholars in every other field, must communicate their ideas and findings in writing. This book, by one of psychology's most prolific writers, is a practical guide to the writing of psychological papers of all types, ranging from library papers and lab reports by first-year college students to papers for publication by full-fledged researchers.

FPO

THE ADAPTIVENESS OF BEHAVIOR

We are the products of our genes and our environments. Our genes have been shaped by millions of years of evolution, adapting us to the general conditions of human life on earth. Through this process, we have acquired, among other things, an immense capacity to learn. In this unit, Chapter 3 examines the role of genes and evolution in the production of the underlying mechanisms of behavior, and Chapter 4 deals with basic processes of learning, through which an individual's behavior is constantly modified to meet the unique conditions of that person's life.

part 2

Genetic and Evolutionary Foundations of Behavior

chapter 3

REVIEW OF BASIC GENETIC MECHANISMS

- How Genes Affect Behavior
- How Genes Are Passed Along in Sexual Reproduction

INHERITANCE OF BEHAVIORAL TRAITS

- Examples of Single-Gene (Mendelian) Behavioral Traits
- Polygenic Characteristics and Selective Breeding

EVOLUTION BY NATURAL SELECTION

NATURAL SELECTION AS A FOUNDATION FOR FUNCTIONALISM

- Ultimate and Proximate Explanations of Behavior
- Limitations on Functionalist Thinking

NATURAL SELECTION AS A FOUNDATION FOR UNDERSTANDING SPECIES-TYPICAL BEHAVIORS

- Species-Typical Behaviors in Humans
- The Value of Cross-Species Comparisons of Species-Typical Behaviors

EVOLUTIONARY ANALYSES OF MATING PATTERNS

EVOLUTIONARY ANALYSES OF HURTING AND HELPING

- Sex Differences in Aggression
- Patterns of Helping
- Final Words of Caution: Two Fallacies to Avoid

Have you ever stood before a chimpanzee enclosure at a zoo and watched for a while? If you haven't, I urge you to seize the next opportunity to do so. It is impossible, I think, to look for long without sensing strongly the animal's kinship to us. Its facial expressions, its curiosity, even its sense of humor, are so like ours that we intuitively see it as a hairy, long-armed cousin. Indeed, the chimpanzee is our cousin. It—along with the bonobo, a chimp-like ape discussed later in this chapter—is one of our two closest animal relatives. Geneticists have lined up the DNA of chimpanzees against that of humans and found that they match at 98.8 percent of the nearly three billion individual base units that make up the DNA molecules for each species (Cryanoski, 2002). In genetic material, you and I are just 1.2 percent different from a chimpanzee. Language and culture, and the knowledge these have given us, have in some ways separated us markedly from our cousins. But in our genes—and in our basic drives, emotions, perceptual processes, and ways of learning, which are partly products of those genes—we are kin not just to chimpanzees, but in varying degrees to all of the mammals, and in lesser degrees to other animals as well.

Nearly 150 years ago, in *On the Origin of Species*, Charles Darwin (1859/1963) presented a theory of evolution that explains both the similarities and the differences among the animal species. According to Darwin, all species are to varying degrees similar to one another because of common ancestry, and all species are to some degree unique because natural selection has adapted each species to the unique aspects of the environment in which it lives and reproduces. Darwin presented massive amounts of evidence for his theory, and essentially everything that scientists have learned since, about our own and other species, is consistent with it.

This chapter is primarily about the application of evolutionary theory to the behavior of humans and other animals. It is also the first of a two-chapter sequence on the *adaptiveness of*

behavior. Adaptation refers to modification to meet changed life circumstances. Evolution is the long-term adaptive process, spanning generations, that equips each species for life in its ever-changing natural habitat. The next chapter is on learning, a set of shorter-term adaptive processes that occur within the life span of each individual. The mechanisms that permit learning to occur are themselves products of evolution.

Darwin developed his theory of evolution without any knowledge of genes, but the theory is best understood today in the light of such knowledge. Therefore, the chapter begins with a discussion of basic genetic mechanisms and their implications for the inheritance of behavioral characteristics. With that as background, the rest of the chapter is concerned with the evolution of behavior and with ways in which we can learn about our own behavior by comparing it to that of our animal relatives.

Review of Basic Genetic Mechanisms

You have almost certainly studied the mechanisms of gene action and reproduction in a biology course at one time or another, so I will not go into detail on these processes here. But a brief review of them, focused on their implications for psychology, may be useful.

How Genes Affect Behavior

1

How can genes affect behavioral traits through their role in protein synthesis?

<<< Sometimes, as a sort of shorthand (which I will use occasionally in this book), researchers speak of genes “for” particular behavioral traits. For example, they might speak of genes *for* singing ability, *for* aggression, or *for* cooperation. But genes never produce or control behavior directly. All the effects that genes have on behavior occur through their role in building and modifying the physical structures of the body. Those structures, interacting with the environment, produce behavior. Thus, a gene might influence musical ability by promoting the development of a brain system that analyzes sounds or by promoting certain physical aspects of the vocal cords. Similarly, a gene might affect aggressiveness by fostering the growth of brain systems that organize aggressive behavior in response to irritating stimuli. In a sense, all genes that contribute to the body’s development are “for” behavior, since all parts of the body are involved in behavior. Especially relevant for behavior, however, are genes that contribute to the development of sensory systems, motor systems (muscles and other organs involved in movement), and, most especially, the nervous system (which includes the brain).

Genes Provide the Codes for Proteins

Genes affect the body’s development through, and only through, their influence on the production of *protein molecules*. We are what we are, biologically speaking, because of our proteins. A class of proteins called *structural proteins* forms the structure of every cell of the body. Another, much larger class called *enzymes* controls the rate of every chemical reaction in every cell.

How do genes organize the production of proteins? Physically, genes are components of extremely long molecules of a substance called DNA (deoxyribonucleic acid). These molecules exist in the egg and sperm cells that join to form a new individual, and they replicate themselves during each cell division in the course of the body’s growth and development. A replica of your whole unique set of DNA molecules exists in the nucleus of each of your body’s cells, where it serves as a template (that is, a mold or pattern) for producing another molecular substance called RNA (ribonucleic acid), which in turn serves as a template for producing protein molecules. Each protein molecule consists of a long chain of smaller molecules called amino acids. The 20 distinct amino acids can be arranged in countless sequences to form different protein molecules. The job of each gene is to provide

the code that dictates the particular sequence of amino acids for a single type of protein. From a molecular vantage point, a *gene* can be defined as the segment of a DNA molecule that contains the code for manufacturing a particular type of protein molecule. With this definition, geneticists have determined that humans, chimpanzees, and mice all have nearly the same number of genes—about 30,000 of them (Tecott, 2003).

Genes Work Only Through Interaction with the Environment

At every level, from biochemical to behavioral, the effects of genes are entwined with the effects of the environment. *Environment*, as used in this context, refers to every aspect of an individual and his or her surroundings except the genes themselves. It includes the nourishing womb and maternal bloodstream before birth; the internal chemical environment of the developing individual; and all the events, objects, and other individuals encountered after birth. Foods—a part of the environment—supply genes with amino acids, which are needed to manufacture proteins. Environmental effects also turn genes “on” and “off,” resulting in bodily changes that alter the individual’s behavioral capacity. Such changes can occur in adulthood as well as earlier in development. For example, physical exercise modifies the chemical environment of muscle cells in a way that activates genes that promote further growth of the muscle. One’s body and behavioral capacities result from a continuous, complex interplay between genes and environment (see Figure 3.1). In no sense is one more basic than the other.

Researchers have begun recently to learn about specific mechanisms through which experiences can activate genes and thereby alter the individual’s brain and behavior. A well-studied example has to do with parental behavior in mice and rats. Adult mice and rats that have not given birth will normally avoid newborns of their species that are placed in their cage. However, if exposed to newborns continuously for several hours or more, they gradually begin to care for them. This change in behavior involves the environmental induction of gene activity (Brown & others, 1996; Stack & Numan, 2000). The sight, sound, or smell of newborns activates a particular gene. The activated gene produces a protein molecule that stimulates activity in a specific cluster of brain cells that are known to be crucial for the motivation and organization of such behaviors as retrieving young to a nest and hovering over them. The result is that a mouse or rat that previously did not take care of young is transformed into a mouse or rat that does.

There is good reason to believe that all sorts of prolonged behavioral effects that derive from experience, including those that we call “learning,” involve the activation of genes (Johnston & Edwards, 2002). The experience activates genes, which produce proteins, which in turn alter the function of some of the neural circuits in the brain and thereby change the manner in which the individual behaves. More about this is discussed in Chapter 5.

Distinction Between Genotype and Phenotype

Geneticists use the term *genotype* to refer to the set of genes that the individual inherits and the term *phenotype* to refer to the observable properties of the body and behavioral traits. The same genes can have different effects, depending on the environment and the mix of other genes. Two individuals with the same genotype can be quite different in phenotype as a result of differences in their environments. Genetically identical rats will differ phenotypically in their behavior toward infant rats if one has been previously exposed to infant rats and the other has not. Genetically identical human twins will differ in size if they have been exposed differently to growth-promoting factors in their environments (see Figure 3.2) and will differ in behavior if they have been subjected to different learning experiences.

What does it mean to say that genes can influence behavioral traits only through interaction with the environment? How, in general, are genes involved in long-term behavioral changes that derive from experience?

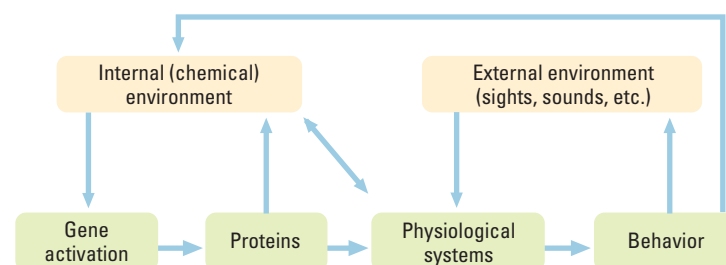


FIGURE 3.1 Route through which genes affect behavior Genes build proteins, which form or alter the body’s physiological systems (including brain systems), which, in turn, produce behavior. Each step in this process involves interaction with the environment. Aspects of the internal environment control gene activation, and aspects of both the internal and the external environments act on physiological systems to control behavior. Behavior, in turn, can affect gene activation through direct and indirect effects on the internal environment.

How can the same genotype produce various phenotypes?



Van Bucher / Photo Researchers

FIGURE 3.2 | Identical twins These 13-year-old girls have the same genotype, but they obviously differ in at least one aspect of their phenotype. It is uncertain what caused this difference. It may have derived from their occupying different positions in the womb such that one received more early nutrition than the other, which activated genes promoting more growth.

4 How does meiosis produce egg or sperm cells that are all genetically different from one another?

How Genes Are Passed Along in Sexual Reproduction

Genes not only provide the codes for building proteins; they also serve as the biological units of heredity. They are replicated and passed along from parents to offspring, and they are the cause of offspring's resemblance to parents.

To understand how genes are passed along in sexual reproduction, it is useful to know something about their arrangement within cells. The genetic material (strands of DNA) exists in each cell in structures called **chromosomes**, which are usually dispersed throughout the cell nucleus and are not visible. Just prior to cell division, however, the chromosomes condense into compact forms that can be stained, viewed through a microscope, and photographed. The normal human cell has 23 pairs of chromosomes. Twenty-two of these are true pairs in both the male and the female, in the sense that each chromosome looks like its mate. The remaining pair is made up of the sex chromosomes. In the normal human male cell, that pair consists of a large chromosome labeled X and a small chromosome labeled Y (see Figure 3.3). Genetically, the only difference between the sexes is that females have two X chromosomes (XX) rather than the XY of the male.

The Production of Genetically Diverse Egg and Sperm Cells

When cells divide to produce new cells *other than* egg or sperm cells, they do so by a process called **mitosis**. In mitosis, each chromosome precisely replicates itself and then the cell divides, with one copy of each chromosome moving into each of the two cell nuclei thus formed. Because of the faithful copying of genetic material in mitosis, all your body's cells, except your egg or sperm cells, are genetically identical to one another. The differences among different cells in your body—such as muscle cells and skin cells—arise from the differential activation of their genes, not from different gene content.

When cells divide to produce egg or sperm cells, they do so by a different process, called **meiosis**, which results in cells that are not genetically alike. During meiosis, each chromosome replicates itself once, but then the cell divides twice. Before the first cell division, the chromosomes of each pair line up next to one another and exchange genetic material in a random manner. Although the chromosomes in each pair look the same, they do not contain precisely the same genes. The result of this random exchange of genetic material and of the subsequent cell

divisions is that each egg or sperm cell produced is genetically different from any other egg or sperm cell and contains only half of the full number of chromosomes (one member of each of the 23 pairs).

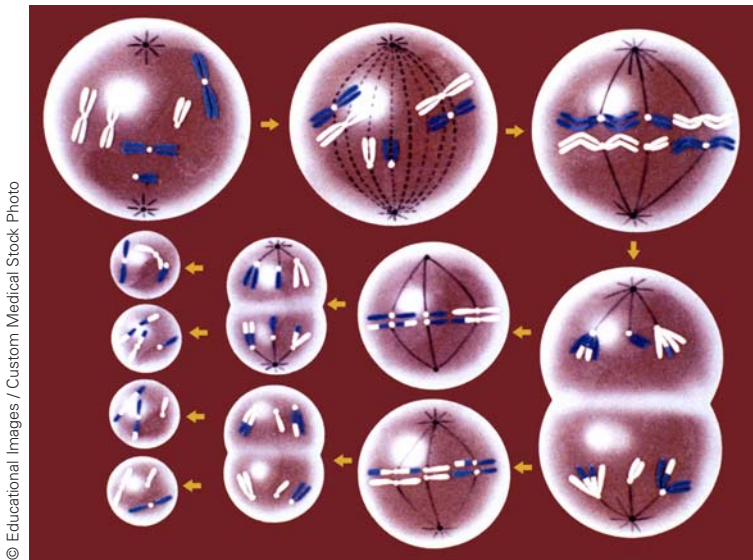
The Genetic Diversity of Offspring

It may seem ironic that the very cells you use for “reproduction” are the only cells in your body that cannot, in theory, reproduce you. They are the only cells in your body that do not have all your genes. In sexual reproduction you are, of course, not really reproducing yourself. Rather, you are creating a genetically unique individual who has half of your genes and half of your partner's. When a sperm and an egg unite, the result is a single new cell, the **zygote**, which contains the full complement of 23 paired chromosomes. One member of each of these pairs comes from



Pueschel, S. M., & Goldstein, A. (1983). Genetic counseling. In J. L. Maton & J. A. Mulick (Eds.), *Handbook of Mental Retardation*. London: Pergamon Press PLC. Reprinted with permission.

FIGURE 3.3 | Chromosomes of a normal human male cell The 22 numbered pairs of chromosomes are the same in a normal female cell as they are in a normal male cell. The remaining two, labeled X and Y, are the sex chromosomes. The normal human female cell (not shown) has a second X chromosome instead of a Y.



Schematic illustration of meiosis in sperm production
This illustration is for a creature that has only three pairs of chromosomes rather than the 23 pairs that humans have. At the beginning of this sequence, each chromosome has already replicated itself and remains attached to its replica. The pairs of replicated chromosomes (one blue and one white in the diagram) then line up next to one another and exchange genetic material. The cell then divides twice, resulting in four sperm cells, each with just one member of each pair of chromosomes. Notice that each sperm cell is genetically different from the others, having a different mix of the original (blue and white) material from the parental pairs of chromosomes. Meiosis in egg production is similar to that in sperm production, but only one of the two cells produced at each division survives.

each parent. The zygote then grows, through mitosis, eventually to become a new person. Because each sperm or egg is different from any other sperm or egg (even from the same parent), each zygote is different from any other.

The value of sex, as opposed to simple cloning (the asexual production of genetically identical offspring), apparently lies in the production of genetically diverse offspring. In a world where the environment keeps changing, genes have a better chance of surviving if they are rearranged at each generation in many different ways, to produce different kinds of bodies, than if they are all put into the same kind of body. This is an almost literal example of the old saying, “Don’t put all your eggs in the same basket.” By producing diverse offspring, parents are reducing the chance that all of their offspring will die as a result of some unforeseen change in the environment. In Chapter 15 you will see how this idea—that there is value in genetic diversity—may apply in the realm of human personality.

The only people who are genetically identical to each other are **identical twins**. They are formed when two bundles of cells separate from each other during the early mitotic divisions following the formation of a zygote. Because they originate from one zygote, identical twins are also known as *monozygotic twins*. **Fraternal twins**, or *dizygotic twins*, originate from two zygotes, formed from two different egg and sperm cells. Fraternal twins are no more or less similar to each other genetically than are any two non-twin siblings. In later chapters (especially Chapter 10), you will see how psychologists make use of twins in research aimed at understanding how much of the variability among people, in certain psychological traits, results from differences in their genes and how much results from differences in their environments.

Consequences of the Fact That Genes Come in Pairs

You have seen that genes exist on long DNA strands in chromosomes, rather like beads on a string, and that chromosomes come in pairs. The two genes that occupy the same **locus** (location) on a pair of chromosomes are sometimes identical to each other and sometimes not. When they are identical, the individual is said to be **homozygous** [home-oh-zai-gus] at that locus, and when they are not identical, the individual is said to be **heterozygous** [het-er-oh-zai-gus] at that locus (see Figure 3.4). Different genes that can occupy the same locus, and thus can potentially pair with each other, are referred to as **alleles**.

>>> **5**
What is the advantage of producing genetically diverse offspring?

Genetic Diversity As exemplified by these brothers, siblings who are not identical twins can differ considerably in personality.



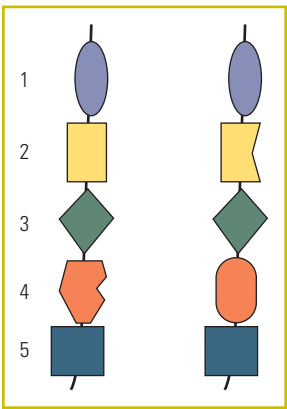
© Sonda Dawes / The Image Works

6

What is the difference between a dominant and a recessive gene (or allele)?

FIGURE 3.4 | Schematic illustration of gene loci on a pair of chromosomes Successive genes are depicted here as beads on a string. This pair of chromosomes is homozygous at loci 1, 3, and 5 (the paired genes there are identical to each other) and heterozygous at loci 2 and 4 (the paired genes there are not identical to each other). Nonidentical genes that can occupy the same locus on a pair of chromosomes are referred to as alleles of each other. Thus the two genes at locus 2 are alleles, as are the two at locus 4.

<<< For example, a gene for brown eyes and a gene for blue eyes in humans are alleles because they can occupy the same locus. If you are homozygous for brown eyes, you have two copies of a gene that manufactures an enzyme that makes your eyes brown. What if you were heterozygous for eye color, with one copy of the allele for brown eyes and one copy for blue eyes? In this case, you would have brown



eyes, no different from the eye color you would have if you were homozygous for brown eyes. This effect is described by saying that the allele for brown eyes is *dominant* and the one for blue eyes is *recessive*. A **dominant** gene (or allele) is one that will produce its observable effects in either the homozygous or the heterozygous condition, and a **recessive** gene (or allele) is one that will produce its effects in the homozygous condition only. But not all pairs of alleles manifest dominance or recessiveness. Some pairs blend their effects. For example, if you cross red four-o'clocks (a kind of flower) with white four-o'clocks, the offspring will be pink, because neither the red nor the white allele is dominant over the other.

Mendelian Pattern of Heredity

The idea that the units of heredity come in pairs and that one member of a pair can be dominant over the other was developed in the mid-nineteenth century by an Austrian monk named Gregor Mendel. In a typical experiment, Mendel would start with two purebred strains of peas that differed in one or more easily observed traits. He could cross-pollinate them to observe the traits of the offspring, called the F_1 (first filial) generation. Then he would pollinate the F_1 peas with pollen from other F_1 peas to produce the F_2 (second filial) generation. In one experiment, for example, Mendel cross-pollinated a strain of peas that regularly produced round seeds with a strain that regularly produced wrinkled seeds. His famous findings were that (a) all of the F_1 generation had round seeds and (b) three-fourths of the F_2 generation had round seeds and one-fourth had wrinkled seeds.

7

Why do three-fourths of the offspring of two heterozygous parents show the dominant trait and one-fourth show the recessive trait?

<<< Mendel's findings make perfect sense if we assume that seed texture is controlled by a single pair of genes, with the allele for round dominant over that for wrinkled. To illustrate this, let us use the capital letter R to stand for the dominant, round-producing allele and the small letter r for the recessive, wrinkle-producing allele. The purebred round strain is homozygous for the "round" allele (RR), and the purebred wrinkled strain is homozygous for the "wrinkled" allele (rr). (Purebred strains are homozygous for all traits.) Because one allele must come from each parent, the only possible result for the F_1 generation, produced by crossing the two purebred strains, is the heterozygous condition (Rr). This explains why all the F_1 peas in Mendel's experiment were round. At the next step, when Rr peas receive pollen from other Rr peas to produce the F_2 generation, four equally likely combinations can occur: (1) an R from each parent (RR), (2) an R from the female parent and an r from the male (Rr), (3) an r from the female parent and an R from the male (rR), and (4) an r from each parent (rr). (See Figure 3.5.) Since only one of these possible outcomes (rr) is wrinkled, the expectation is that one-fourth of the F_2 generation will be wrinkled and the other three-fourths, round. This is just what Mendel found.

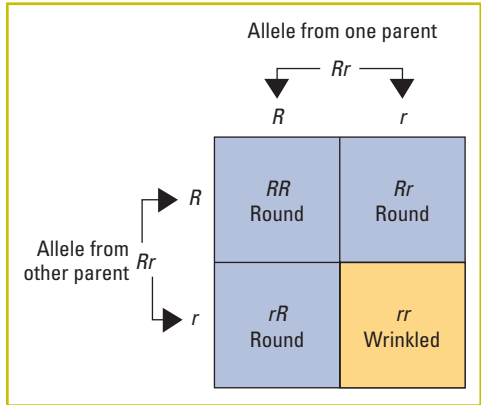


FIGURE 3.5 | Explanation of Mendel's 3:1 ratio When a pea plant that is heterozygous for round versus wrinkled seeds is pollinated by another pea plant that is similarly heterozygous, four possible gene combinations occur in the offspring. Here R stands for the dominant, round-producing allele, and r for the recessive, wrinkle-producing allele. The phenotype of three of the offspring will be round and that of one, wrinkled. This 3:1 ratio was Mendel's famous finding.

Whenever a trait is inherited in a pattern like that observed by Mendel, we can assume that the trait results from variation at a single gene locus.

SECTION REVIEW

Genes affect behavior by affecting the bodily structures involved in behavior.		
Nature of Genetic Influence - Genes affect bodily structures, and thereby affect behavior, by influencing protein synthesis. - Genes act in concert with the environment, not in isolation. For example, environmental cues can activate genes that make rats or mice nurturant to newborns.	Meiosis and Sexual Reproduction - Meiosis results in egg and sperm cells that contain just one member of each pair of genes. - Meiosis involves random assortment of paired genes. - Genetic diversity produced by sexual reproduction promotes the genes' survival.	Gene Pairing - Paired genes, which occupy the same locus (location) on a pair of chromosomes, may be identical (homozygous) or different (heterozygous). Gene variations that can occupy the same locus are called alleles. - Mendel's discovery of consistent ratios of traits in offspring of cross-pollinated strains of peas led to the gene concept and to the concepts of dominance and recessiveness.

Inheritance of Behavioral Traits

Variation in genes contributes to the variation in behavior that one sees in any animal species. Some behavioral characteristics are inherited in accordance with the same pattern that Mendel observed, indicative of control by a single pair of genes. Most behavioral characteristics, however, depend on many genes. In this section we shall look first at two examples of single-gene traits and then at traits that are affected by many genes.

Examples of Single-Gene (Mendelian) Behavioral Traits

Mendelian Inheritance of Fearfulness in Dogs

One of the first demonstrations of single-gene control of a behavioral trait in dogs occurred more than 40 years ago. In pioneering research on the role of genes in behavior, John Paul Scott and John Fuller (1965) studied the behavior of basenji [buh-sen-jee] hounds, cocker spaniels, and their mixed-breed offspring. Basenjis are timid dogs, showing fear of people until they have received much gentle handling. Cocker spaniels, in contrast, show little fear under normal rearing conditions. In a standard test with 5-week-old puppies, Scott and Fuller found that all the basenji puppies yelped and/or ran away when approached by a strange person, whereas only a few of the cocker puppies showed these reactions. When cockers and basenjis were crossbred (see Figure 3.6), the offspring (F_1 hybrids) were like basenjis in this test: All showed signs of fear when approached. Since this was as true of hybrids raised by cocker mothers as of those raised by basenji mothers, Scott and Fuller concluded that the effect stemmed from the hybrid dogs' genes and not from the way their mothers treated them.

The fact that the F_1 hybrids were as fearful as the purebred basenjis suggested to Scott and Fuller that the difference in fearfulness between the two purebred strains might be controlled by a single gene locus, with the allele promoting fear dominant over that promoting confidence. If this were so, then mating F_1 hybrids with each other should produce a group of offspring (F_2 generation) in which three-fourths would show basenji-like fear and one-fourth would show cocker-like confidence—

>>> 8 How did Scott and Fuller show that the difference in fearfulness between cocker spaniels and basenji hounds is controlled by a single gene locus, with the “fear” allele dominant over the “non-fear” allele?

Scott, J. P. & Fuller, J. L. (1965). *Genetic and social behavior of the dog*. Chicago: University of Chicago Press.

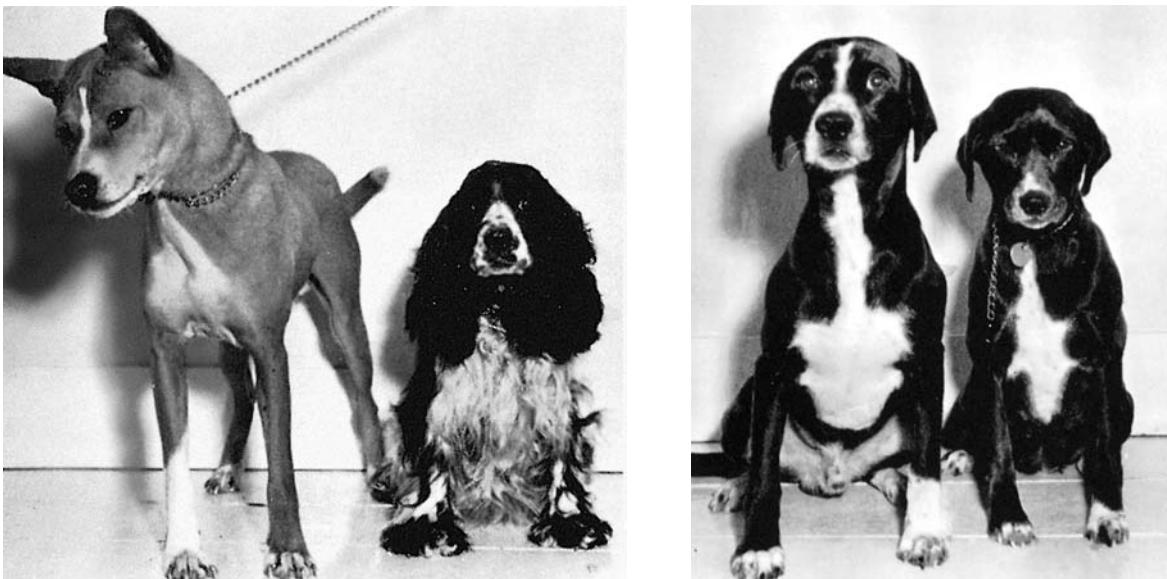


FIGURE 3.6 Dogs used in Scott and Fuller's research. At left are a male basenji and a female cocker spaniel; at right are two F_1 (first-generation) hybrids resulting from a basenji-cocker cross.

the same ratios that Mendel had found with seed texture in peas. Scott and Fuller did this experiment and, indeed, found ratios very close to those predicted. As additional evidence, they also mated F_1 hybrids with purebred cockers. About half the offspring of those *backcrosses* were basenji-like in fear, and the other half were cocker-like in confidence—just as can be expected if the “fear” allele is dominant over the “non-fear” allele (see Figure 3.7).

9

Why would it be a mistake to conclude, from Scott's and Fuller's work, that fear in dogs is caused just by one gene or that it is caused just by genes and not by the environment?

Be careful not to misinterpret this finding. It concerns a difference between two breeds of dogs in certain behavioral tests. It would not be reasonable to conclude that fear in all its various forms is controlled by a single gene. Many different genes must contribute to building the complex neural structure needed to experience fear and express it in behavior. Scott's and Fuller's work demonstrates only that the difference between cocker spaniels and basenji hounds in a particular test of fear is controlled by a single gene. Recognize also that their studies do not diminish the role of environmental influences. Scott and Fuller could detect the effect of a specific gene pair because they raised all the dogs in similar environments. In other research, Scott (1963) showed that any puppy isolated from people for the first 4 months of life will be fearful of humans. Had Scott and Fuller isolated the cockers from all human contact and given the basenjis lots of kind handling before the behavioral test, they might well have found the cockers to be more fearful than the basenjis, despite the genetic predispositions toward the opposite behavior.

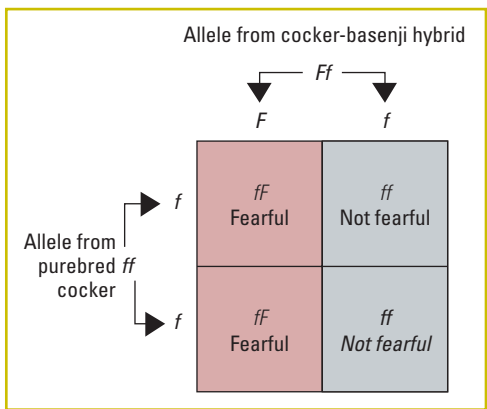


FIGURE 3.7 Explanation of Scott and Fuller's results of mating basenji-cocker hybrids with purebred cockers. The finding that half the offspring were fearful and half were not makes sense if fearfulness results from a dominant allele (F) and lack of fearfulness results from a recessive allele (f). Because half the offspring receive F from their hybrid parent and all receive f from the purebred parent, half the offspring will be Ff (phenotypically fearful) and the other half, ff (not fearful).

Mendelian Inheritance of a Specific Language Disorder

Most of the behaviorally relevant traits in humans that derive from alteration at a single gene locus are brain disorders, caused by relatively rare, mutant, malfunctioning genes that are passed from generation to generation.

A particularly interesting example is a specific language impairment that has been studied extensively in three generations of one family, known as the KE family. The disorder, which is very rare outside of the KE family, is characterized primarily by difficulty in articulating words, distinguishing speech sounds from other sounds, and learning grammatical rules (Gopnik, 1999; Marcus & Fisher, 2003). With much effort, people with this disorder learn to speak and understand language well enough to get along, but language never becomes natural to them. For example, they never develop the intuitive understanding of grammatical categories and rules that even normal 3-year-olds manifest. Brain scans of people with this disorder have revealed underdevelopment in several areas of the brain that are known to be critical for producing and understanding language (Liégeois & others, 2003).

The pattern of heredity of this disorder in the KE family is depicted in Figure 3.8. As you can see by examining the results for the third generation, when neither parent had the specific language impairment (SLI), no child had it, and when one parent had it, about half of the offspring had it—just what you would expect if the impairment results from a single dominant gene. The logic behind this expectation is identical to that shown in Figure 3.7 for fearfulness in dogs. If one parent has one copy of the abnormal gene and the other parent lacks the abnormal gene, each offspring has a 50 percent chance of inheriting a copy of that gene. Because the abnormal gene is dominant, each person inheriting it is linguistically impaired.

Recently, researchers have isolated the gene that causes this language impairment (Lai & others, 2001; Fisher, 2003). It is located on Chromosome 7, and in its normal (unimpaired) form it codes for a type of protein known as a *transcription factor*. Transcription factors are proteins that interact with the regulatory regions of other genes and in that way control the rate at which those genes produce their protein molecules. Apparently, the normal version of this gene is responsible for activating a set of other genes that are involved in the development of various areas

How did the pattern of inheritance of a disorder in language ability, in a particular family, show that the disorder is caused by a single dominant gene? Also, what other general ideas about genetic influences are illustrated by this example?

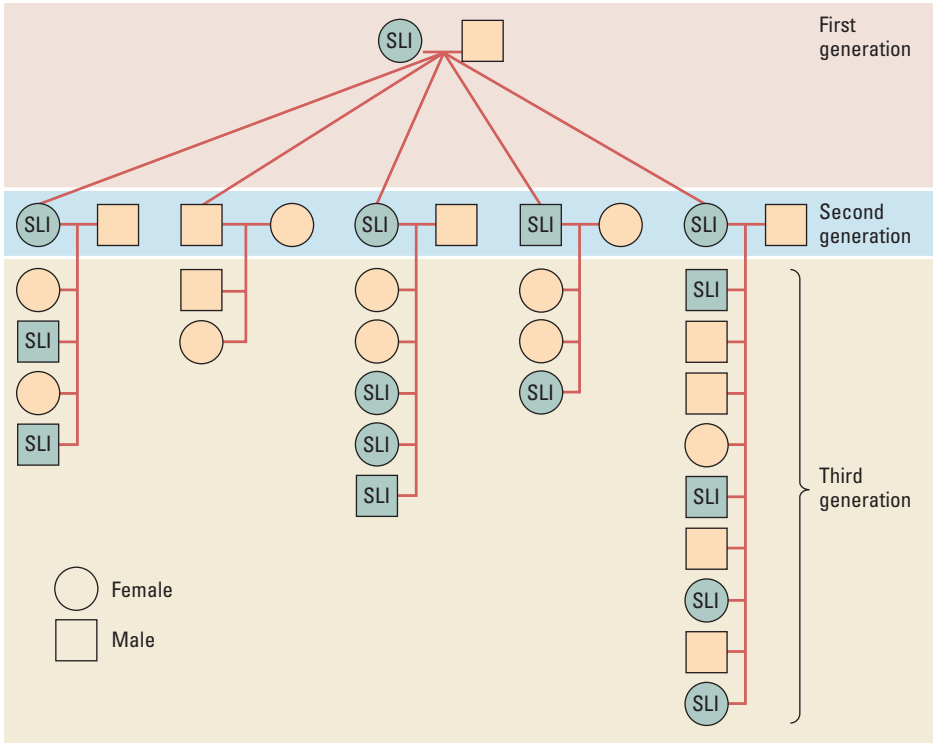


FIGURE 3.8 | Inheritance of a specific language impairment This diagram shows the members of three generations of a family in which a specific language impairment (SLI) recurred. Circles depict females; squares, males; horizontal lines, marriage bonds; and slanted or vertical lines, lines of descent. "SLI" indicates presence of the disorder. Notice that approximately half (10 of 21) of the third-generation children who had one parent with the disorder also had the disorder. This pattern is consistent with the theory that the disorder is inherited through a single dominant gene. (Based on Gopnik & Crago, 1991.)

of the brain that are crucial for human language. The mutated (SLI) version of the gene fails to activate the other genes. Two copies of the normal gene are required to activate the other genes sufficiently for normal development of the language areas of the brain. Thus the disorder appears in people who have one normal gene and one SLI gene.

The normal human version of this gene appears to be unique to humans. The chimpanzee version of the same gene is slightly different from that of the human version and produces a slightly different protein molecule (Marcus & Fisher, 2003). Apparently, modification of this particular gene was one of the evolutionary steps that enabled language to occur in humans and helped to distinguish us from chimpanzees and other apes.

I chose this particular example of a single-gene trait not just because it illustrates Mendelian inheritance in human behavior but also because it illustrates four other general ideas about genes:

- Genes can influence behavior by influencing the development of particular areas of the brain.
- A single gene can have multiple effects (in this case the gene affects various aspects of linguistic ability by altering the brain in various locations).
- Some genes exert their effects by activating sets of other genes, thereby controlling the production of several or many different protein molecules.
- The evolution of human beings (and of other species) involves alterations in anatomy and behavior that derive from alterations in genes.

Polygenic Characteristics and Selective Breeding

11

How does the distribution of scores for a polygenic trait differ from that usually obtained for a single-gene trait?

<<< Characteristics that derive from variation at a single gene locus are typically *categorical* in nature. That is, they are characteristics that sharply differentiate one group (category) from another. Peas are either round or wrinkled; mixed-breed basenji-cockers differ so sharply from one another in fearfulness that they can be categorized into two distinct groups; members of the KE family either have or don't have the specific language disorder (none of them “sort of have it”).

But most measurable anatomical and behavioral differences among individuals of any species are in degree, not type. They are *continuous* rather than categorical. That is, the measures taken from individuals do not fall into two or more distinct groups but can lie anywhere within the observed range of scores. Most often, the set of scores obtained on such measures approximate a **normal distribution**, meaning that most scores fall near the middle of the range and the frequency tapers off toward the two extremes (see Figure 3.9). Measures of aggressiveness in mice, of maze learning in rats, and of conscientiousness in people are just a few of the behavioral measures that are consistently found to fit a normal distribution.

Characteristics that vary in a continuous way are generally affected by many genes and are therefore called **polygenic characteristics** (the prefix *poly-* means “many”). Of course, these traits are also influenced by variation in the environment, so the variability observed in a graph such as Figure 3.9 results from a combination

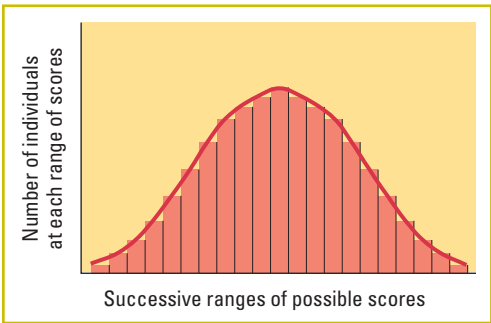


FIGURE 3.9 | Normal distribution
When many individuals are tested for a polygenic characteristic, the majority usually fall in the middle of the range of scores and the frequency tapers off toward zero at the extremes. Mathematically, this defines a normal curve. (For a more complete description, see the Statistical Appendix at the end of the book.)

of genetic differences at many gene loci and environmental differences. In animals it is possible to study the role of genes in polygenic traits through the procedure of selective breeding.

Selective Breeding for Behavioral Characteristics in Animals

To the degree that individuals within a species differ in any measurable characteristic because of differences in their genes, that characteristic can be modified over successive generations through *selective breeding*. This procedure involves the mating of individuals that lie toward the desired extreme on the measure in question. For single-gene characteristics the effects of selective breeding are immediate, but for polygenic characteristics the effects are gradual and cumulative over generations.

The basic procedure of selective breeding is by no means new. For thousands of years before a formal science of genetics existed, plant and animal breeders used selective breeding to produce new and better strains of every sort of domesticated species. Grains were bred for plumper seeds; cows, for docility and greater milk production; horses, along separate lines for working and racing; canaries, for their song; and dogs, along dozens of different lines for such varied purposes as following a trail, herding sheep (running around them instead of at them), and being gentle playmates for children. The procedure in every case was essentially the same: The members of each generation that best approximated the desired type were mated to produce the next generation, resulting in a continuous genetic molding toward the varieties we see today.

Under controlled laboratory conditions, researchers have used selective breeding to produce many behaviorally specialized strains of animals. Fruit flies have been bred to move instinctively either toward or away from a source of light, mice to be either more or less inclined to fight, and rats to either prefer or not prefer alcohol over water (Crabbe & others, 1999a; Wimer & Wimer, 1985). That selective breeding can influence essentially any behavioral trait should come as no surprise. It follows logically from the fact that all behaviors depend on particular sensory, motor, and neural structures, all of which are built from proteins whose production depends on genes.

Selective Breeding for Maze Learning: Tryon's Classic Research

The first long-term, systematic study of selective breeding in psychology was begun in the 1920s by Robert Tryon (1942), partly in reaction to the belief then held by some psychologists that individual differences in behavior are due entirely to environmental, not at all to genetic, differences. Tryon wanted to demonstrate that a type of behavior frequently studied by psychologists could be strongly influenced by variation in genes.

Tryon began by testing a genetically diverse group of rats for their ability to learn a particular maze. Then he mated the males and females that had made the fewest errors in the maze to begin what he called the "maze bright" strain and those that had made the most errors to begin the "maze dull" strain. When the offspring of succeeding generations reached adulthood, he tested them in the same maze and mated the best-performing members of the bright strain, and the worst-performing members of the dull strain, to continue the two lines.

Some of his results are shown in Figure 3.10. As you can see, with each generation the two strains became increasingly distinct, until by the seventh there was almost no overlap between them. Almost all seventh-generation bright rats made fewer errors in the maze than even the best dull rats. To control for the possibility that the offspring were somehow learning to be bright or dull from their mothers, Tryon cross-fostered the rats so that some of the offspring from each strain were raised by mothers in the other strain. He found that rats in the bright strain were equally good in the maze, and those in the dull strain equally poor, regardless of which mothers raised them.

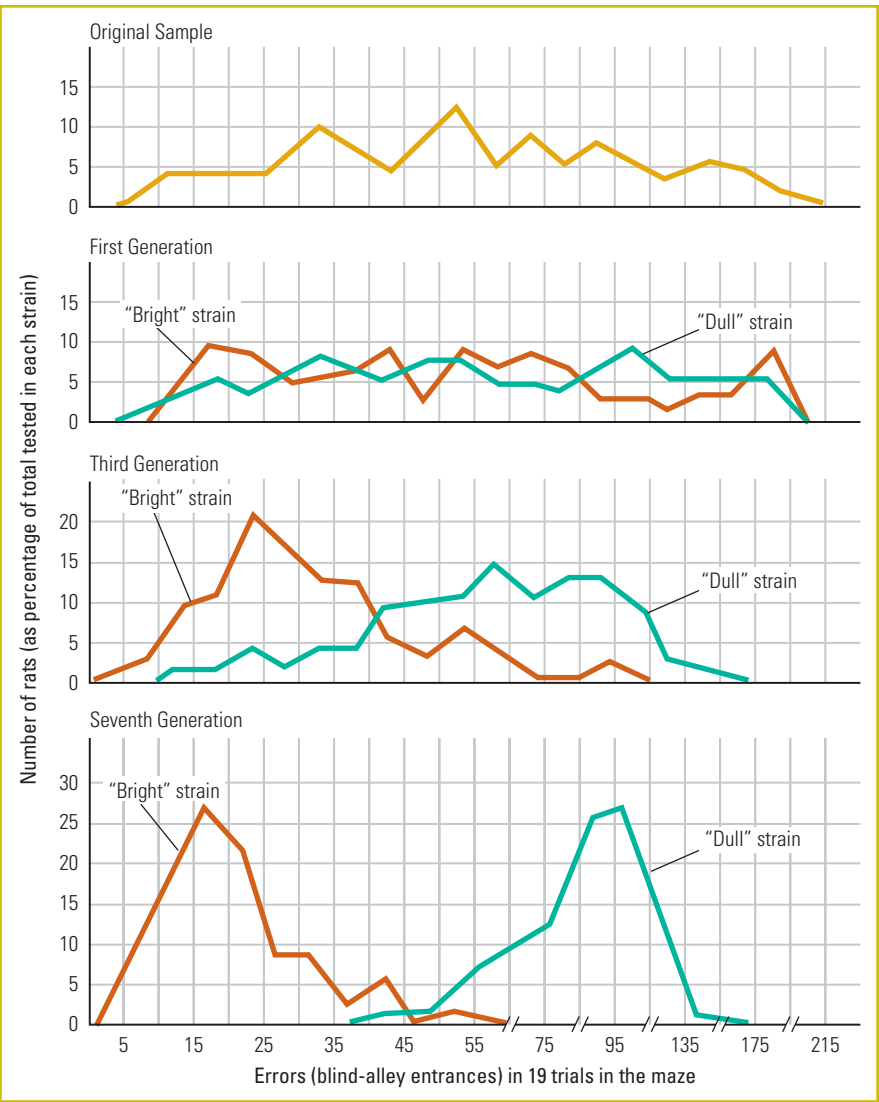
>>> 12
How are the characteristics of animals shaped through selective breeding?



A fox bred for tameness Since 1959, researchers in Russia have been selectively breeding silver foxes for tameness. At each generation, only those foxes that show the least fear and aggression and the most affection to humans have been bred. The result, after 30 to 35 generations, is a breed of foxes that are as friendly to humans as are dogs (Trut, 1999).

>>> 13
How did Tryon produce "maze bright" and "maze dull" strains of rats? How did he show that the difference was the result of genes, not rearing?

FIGURE 3.10 Selective breeding for “maze brightness” and “maze dullness” in rats. The top graph shows, for the original parent stock, the distribution of rats according to the number of errors they made in the maze. Subsequent graphs show this distribution for the “bright” and “dull” rats. With successive generations, an increasing percentage in the “bright” strain made few errors and an increasing percentage in the “dull” strain made many errors. (From Tryon, 1942.)



14

Why is the strain difference produced by Tryon not properly characterized in terms of “brightness” or “dullness”?

Once a strain has been bred to show some behavioral characteristic, the question arises as to what other behavioral or physiological changes accompany it. Tryon referred to his two strains of rats as “bright” and “dull,” but all he had measured was their performance in a particular type of maze. Performance in the maze no doubt depended on many sensory, motor, motivational, and learning processes, and specific changes in any of them could in theory have mediated the effects that Tryon observed. In theory, Tryon’s “dull” rats could simply have been those that had less acute vision, or were less interested in the variety of food used as a reward, or were more interested in exploring the maze’s blind alleys.

In later studies, another researcher found that Tryon’s “dull” rats were as good as the “bright” ones, and sometimes even better, at other learning tasks (Searle, 1949). We do not know what underlying abilities or dispositions changed in Tryon’s two strains of rats to produce their difference in maze performance, but the change was apparently not one of general learning ability. This problem still occurs in modern behavioral genetics research, in which new strains of animals (usually mice) are created by adding, deleting, or modifying known genes using sophisticated genetic-engineering methods. The behavioral differences between two strains found in one laboratory often fail to occur in another laboratory, apparently because of subtle differences in the way the animals are housed or tested (Cabib & others, 2000; Crabbe & others, 1999b).

Polygenic Behavioral Characteristics in Humans

Most of the measures of human traits that interest psychologists—such as scores on personality tests—are continuous and normally distributed, and are affected by many genes as well as by environmental variables. Some psychologists are interested in the degree to which the differences in such scores, for a particular group of people, are the result of differences in their genes or differences in their environmental experiences. Of course, psychologists can't perform selective breeding studies with humans, but they have developed other methods to estimate that degree. Those methods involve comparing the average difference in test scores for people who are closely related to one another with that for people who are less closely related, using people who are chosen in such a way that the environments of the closely related people are no more similar to one another than are those of the less closely related people. Comparisons of identical twins with fraternal twins, and of biologically related siblings with adoptive siblings, have proven particularly useful. In Chapters 10, 15, and 16 you will read about such methods as they apply, respectively, to intelligence tests, personality tests, and predisposition to various mental disorders.

SECTION REVIEW

Hereditary effects on behavioral traits can involve just one gene, but usually involve many.

Single-Gene Traits

- Single-gene traits (controlled by one pair of genes) are categorical (all or none) in nature.
- Mendelian patterns of inheritance indicate single-gene control.
- Examples are breed differences in fearfulness in dogs and a language disorder in the KE family. General lessons concerning inheritance are illustrated with each example.

Polygenic Traits

- Polygenic traits (influenced by many gene pairs) are continuous (present in varying degrees) and often fit a normal distribution.
- Through selective breeding, a trait can be strengthened or weakened gradually over generations.
- An example is Tryon's breeding of rats for maze ability, which illustrates certain general lessons.

Evolution by Natural Selection

Darwin's Insight: Selective Breeding Occurs in Nature

The idea that plants and animals can be modified by selective breeding was understood long before Darwin was born. In the first chapter of *On the Origin of Species*, Darwin (1859/1963) referred to such human-controlled selective breeding as **artificial selection**, and he reminded readers of the enormously diverse varieties of plants and animals that had been produced through that procedure. He then pointed out—and this was his true, original insight—that breeding in nature is also selective and can also produce changes in living things over generations.

Selective breeding in nature, which Darwin labeled **natural selection**, is dictated not by the needs and whims of humans but by the obstacles to reproduction that are imposed by the natural environment. Those obstacles include predators, limited food supplies, extremes of temperature, difficulty in finding and attracting mates for sexual reproduction—anything that can cut life short or otherwise prevent an organism from producing offspring. Animals and plants that have characteristics that help them overcome such obstacles are, by definition, more likely to have offspring than those that don't have these characteristics.

Darwin's essential point was this: Individuals of a species vary in the number of offspring they produce. Some produce none, because they die early or fail to mate,

>>> 15
What insight led Darwin to his theory of evolution? How is natural selection similar to and different from artificial selection?

NON SEQUITUR by Wiley

HOW NATURAL SELECTION
REALLY WORKS...

WILEY © 1997 WASHINGTON POST WRITERS GROUP

WELL, I'M NOT MOVING UNTIL SOMEONE CAN
EXPLAIN WHY WE SHOULD RUN WHEN WE HAVE
HIM OUT-NUMBERED...Non Sequitur by Wiley
© 1997 Washington Post Writers Group

and others produce many. Any inherited trait that increases the number of offspring that an individual produces is automatically “selected for,” as the trait is passed on to those offspring. Conversely, any inherited trait that decreases the number of one’s offspring is automatically “selected against,” appearing in fewer members of the next generation. Thus, as long as inheritable differences exist among individuals in an interbreeding population, and as long as some of those differences affect survival and reproduction, evolution will occur.

Genetic Diversity Provides the Fodder for Natural Selection

16

How are genes involved in evolution?
What are the sources of genetic diversity
on which natural selection acts?

<<< Darwin knew nothing of genes. He realized that something, passed along through eggs and sperm, must provide the hereditary foundation for evolution, but he did not know what. Mendel’s work, which was the first step toward our modern knowledge of genes, was unknown to most scientists until about 1900. Today we know that genes are the units of heredity and that evolution entails changes, from generation to generation, in the frequencies of particular genes in an interbreeding population. Genes that improve an individual’s ability to survive and reproduce in the existing environment increase from generation to generation, and genes that impede this ability decrease from generation to generation.

The genetic variability on which natural selection acts has two main sources: (1) the reshuffling of genes that occurs in sexual reproduction (already discussed) and (2) mutations.

Mutations are errors that occasionally and unpredictably occur during DNA replication, causing the “replica” to be not quite identical to the original. In the long run of evolution, mutation is the ultimate source of all genetic variation. As would be expected of any random change in a complex, organized structure, new mutations are more often harmful than helpful, and natural selection usually weeds them out. But occasionally a mutation is useful, producing a protein that affects the organism’s development in a way that increases its ability to reproduce. Because of its effect on reproduction, the gene arising from such a mutation increases in frequency from generation to generation. At the level of the gene, that is evolution.

Prior to the modern understanding of genes, many people believed that changes in an individual that stem from practice or experience could be inherited and therefore provide a basis for evolution. For example, some argued that early giraffes, by frequently stretching their necks to reach leaves in trees, slightly elongated their necks in the course of their lives and that this change was passed on to their offspring, resulting, over many generations, in the long-necked giraffes we see today. That idea, referred to as the *inheritance of acquired characteristics*, is most often attributed to Jean-Baptiste de Lamarck (1744–1829), although many other evolutionists, both before and after Lamarck, held the same view (Futuyma, 1997). Even Darwin did not reject that idea but added to it the concepts of random variation and natural selection.

Now, however, we know that evolution is based entirely on genetic changes and that no amount of practice or experience can change one’s genes in a way that affects the next generation. Random change in genes followed by natural selection, not directed change stemming from individual experience, provides the basis for evolution.

Environmental Change Provides the Force for Natural Selection

Evolution is spurred by changes in the environment. If the environment were completely stable, organisms would adapt as fully as possible and change little or not at all thereafter. But the environment keeps changing. Climates change, sources of food change, predators change, and so on. When the conditions of life change, what was previously a useful characteristic may become harmful, and vice versa.

Darwin believed that evolution was a slow and steady process. But today we know that it can occur rapidly, slowly, or almost not at all, depending on the rate and nature of environmental change and on the degree to which genetic variability already exists in a population (Gould & Eldredge, 1993). Environmental change spurs evolution *not* by causing the appropriate mutations to occur but by promoting natural selection. Some mutations that previously would not have been advantageous, and would have been weeded out by natural selection, are advantageous in the new environment, so they are passed along in increasing numbers from generation to generation. Evolution sometimes occurs so quickly that people can see it happen. In fact, scientists since Darwin's time have reported more than a hundred cases of observed evolution (Endler, 1986; Weiner, 1994).

One of the most well-documented examples of observed evolution comes from the work of Peter and Rosemary Grant, who for 30 years studied a species of finch that inhabits one of the Galápagos Islands, about 600 miles off the coast of Ecuador (Grant, 1991; Weiner, 1994). The Grants found that the members of this species differ somewhat in body size, that the variation is inheritable, and that environmental changes sometimes result in selection for large size and other times for small size. During years of drought, when most of the finches died because the plants that produce the seeds they eat failed to grow, the birds that survived and produced offspring tended to be the larger ones, apparently because their beaks were thick and powerful enough to crack open the large, hard-shelled seeds that were last to be consumed (see Figure 3.11). During rainy years, when an abundance of seeds caused a population explosion among the finches, the smaller birds did better. Apparently, in those conditions, their young could grow to full size and reproduce more quickly than the young of the larger finches.

The evolution of simple or small changes, such as in skin pigmentation or in body size, can occur in a few generations when selection conditions are strong, but more complex changes require much more time. The difference between, say, a chimpanzee brain and a human brain could not have come about in a few generations, as it must have involved many mutations, each of which would have promoted a slight selective advantage to the chimpanzee (in its environment) or to the human (in our environment). When evolutionists talk about “rapid” evolution of complex changes, they are usually talking about periods measured in hundreds of thousands of years (Gould & Eldredge, 1993).

Evolution Has No Foresight

People sometimes mistakenly think of evolution as a mystical force working toward a predetermined end. One manifestation of this belief is the idea that evolution could produce changes for some future purpose, even though they are useless or harmful at the time that the change occurs. But evolution has no foresight. The finches studied by the Grants could not have evolved thicker beaks in anticipation of drought. Only genetic changes that increase survival and reproduction in the immediate environment can proliferate through natural selection.

Another manifestation of the belief in foresight is the idea that present-day organisms can be ranked according to the distance they have moved along a set evolutionary route, toward some planned end (Gee, 2002). For example, some may think of humans as the “most evolved” creatures, with chimpanzees next and amoebas way down on the list. But evolution has no set route or planned end. Humans, chimps, and amoebas have taken their different forms and behavioral

>>>

17

How does change in the environment affect the direction and speed of evolution? How did a study of finches illustrate the role of environmental change in evolution?

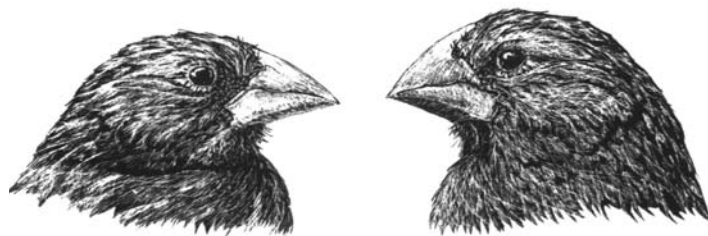


FIGURE 3.11 | Rapid evolution During years of drought, natural selection quickly produces the larger body size and the thicker beak, shown at right, in the finches studied by Peter and Rosemary Grant.

>>>

18

What are three mistaken beliefs about evolution, all related to the misconception that foresight is involved?

characteristics because of chance events that led them to occupy different niches in the environment, where the selection criteria differed. The present-day amoeba is not an early step toward humans but a creature that is at least as well adapted to its environment as we are to ours. The amoeba has no more chance of evolving to become like us than we have of evolving to become like it.

A third manifestation of the belief in foresight is the idea that natural selection is a moral force, that its operation and its products are in some sense right or good. In everyday talk, people sometimes imply that whatever is natural (including natural selection) is good and that evil stems from society or human contrivances that go beyond nature. But nature is neither good nor bad, moral nor immoral. To say that natural selection led to such and such a characteristic does not lend any moral virtue to that characteristic. As you will see, fighting is as much a product of evolution as is cooperation, but that is no reason to consider them morally equivalent.

SECTION REVIEW

Natural selection is the mechanism that produces evolutionary change.		
How Natural Selection Works • To the degree that a trait enhances survival and reproduction, genes producing that trait are passed on to offspring. The result is that such genes become more frequent over generations. • Mutations and reshuffling of genes in sexual reproduction provide genetic diversity on which natural selection operates.	Role of Environmental Change • The rate and nature of environmental change affect the rate and course of evolution. An example is the effect of drought on the evolution of beak thickness in finches. • Complex changes, requiring many mutations, require a long time to evolve.	Evolution Lacks Foresight • Natural selection can only lead to changes that are immediately adaptive; it cannot anticipate future needs. • There is no preset pathway for evolution. • Natural selection is not a moral force.

Natural Selection as a Foundation for Functionalism

19 How does an understanding of evolution provide a basis for functionalism in psychology?

<<< The mechanisms underlying behavior are products of natural selection, and, like all products of natural selection, they came about because they promoted survival and reproduction. Just as Tryon, through artificial selection, bred rats to be better at learning a particular maze, natural selection automatically breeds animals to be better at doing what they must to survive and reproduce in their natural environments. This idea provides a foundation for the psychological approach known as *functionalism*—the attempt to explain behavior in terms of what it accomplishes for the behaving individual.

When applied to understanding the behavioral traits of a species, the functionalist approach in psychology is essentially the same as the functionalist approach in anatomy: Why do giraffes have long necks? Why do humans lack fur? Why do male songbirds sing in the spring? Why do humans have such an irrepressible ability to learn language? The anatomist trying to answer the first two questions, and the behavioral researcher or psychologist trying to answer the latter two, would look for ways by which each trait helped ancestral members of the species to survive and reproduce.

Ultimate and Proximate Explanations of Behavior

Biologists and psychologists who think in evolutionary terms find it useful to distinguish between two kinds of explanations of behavior—ultimate and proximate:

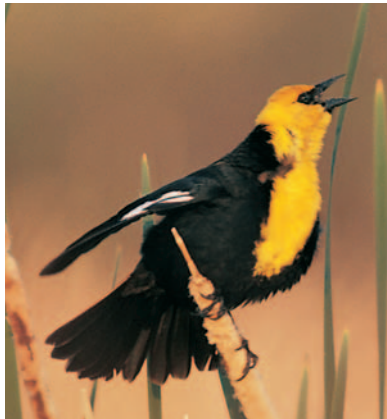
- **Ultimate explanations** are functional explanations at the evolutionary level. In other words, they are statements of the role that the behavior plays in the animal's survival and reproduction. Viewed from the vantage point of the gene, they are statements of how the behavior helps the individual's genes make it into the next generation.
- **Proximate explanations** are explanations that deal not with function but with mechanism; they are statements of the immediate conditions, both inside and outside the animal, that bring on the behavior.

>>> **20**
How are ultimate explanations of behavior different from, but complementary to, proximate explanations?

Illustration of the Complementarity of Ultimate and Proximate Explanations

As an illustration of these two modes of explanation, consider how they might be applied to the question of why male songbirds (of many species) sing in the spring. An *ultimate explanation* goes something like this (Koodsma & Byers, 1991): Over the course of evolution, songbirds have adapted to a mating system that takes place in the spring. The male's song serves to attract a female with which to mate and to warn other males to stay away from the singer's territory in order to avoid a fight. In the evolution of these birds, males whose genes promoted such singing produced more offspring (more copies of their genes) than did males whose genes failed to promote such singing.

A *proximate explanation*, in contrast, might go as follows (Ball & Hulse, 1998): The increased period of daylight in the spring triggers, through the birds' visual system, a physiological mechanism that leads to the increased production of the sex hormone testosterone, which in turn acts on certain areas of the brain (which we might call the "song areas"), promoting the drive to sing. Notice the complementary nature of these explanations. The ultimate explanation states the survival or reproductive value of the behavior, and the proximate explanation states the stimuli and physiological mechanisms through which the behavior occurs.



Robert Ballou / Animals, Animals

A yellow-throated western blackbird at home This male's singing warns other males of the species to stay away.

The Search for Ultimate Explanations in Human Psychology

All of the complex biological mechanisms underlying human behavior and experience—including the basic mechanisms of perception, learning, memory, thought, motivation, and emotion—are products of evolution by natural selection. They all came about because each small step in their evolution tended to promote the survival and reproduction of our ancestors. Thus for any basic psychological characteristic that is part of human nature—for any basic drive or emotion, for example—it is legitimate to ask: How did this characteristic improve the survival and reproductive chances of our ancestors? How did it help our ancestors get their genes into the next generation?

The ultimate explanations of some human traits (especially those that we share with all other mammals) are relatively obvious. We have strong drives to breathe air, drink fluids, and consume foods because our bodies need these things to remain alive. We have strong drives to copulate, under certain conditions, because that is the means by which our genes get from one generation to the next. Individuals who lacked such drives are ancestors to nobody today; their genes died out. The ultimate explanations of some other human traits, however, are not so obvious. It is not obvious, for example, why humans everywhere tend to sleep about eight hours each night (discussed in Chapter 6), or why humans everywhere under certain conditions experience the painful emotion of guilt (discussed in Chapter 14). In various places in this book, including in the last sections of this chapter, you will encounter examples of ultimate explanations that are not intuitively obvious but are supported by research evidence. As you will see, evidence for or against any particular ultimate explanation can come from detailed analysis of the behavior or trait in question, from cross-species comparisons, and sometimes from studies showing what happens when the behavior or trait is missing.

Limitations on Functionalist Thinking

21

What are four reasons for the existence of traits or behaviors that do not serve survival and reproductive functions?

<<< Before we go deeper into discussions of ultimate functions, it is useful to know something about the limitations of functionalist thinking. It is not the case that every detail of every trait serves a useful function, and some traits that were once functional may not be so today. Here are four reasons why a particular trait or behavior may not be functional.

Eibl-Eibesfeldt, I. (1975). *Human ethology*. Hawthorne, NY: Holt, Rinehart & Winston.

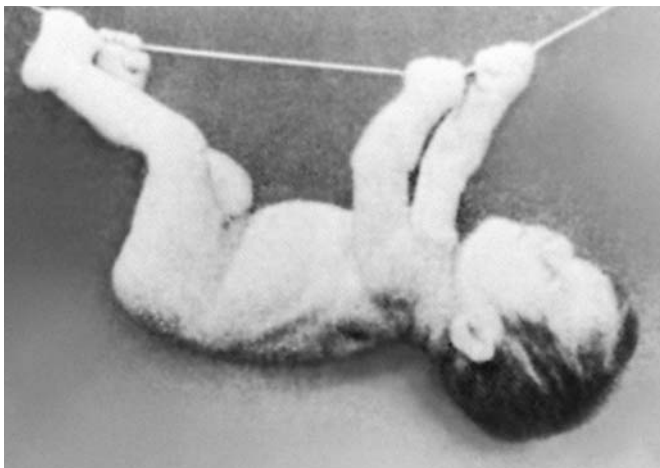


FIGURE 3.12 | Premature infant clinging with hands and toes This ability may be a vestigial carryover from an earlier evolutionary time, when the infants of our ancestors clung to their parents' fur.

Some Traits Are Vestigial

Some traits that evolved to serve the needs of our ancestors are no longer functional today, yet they remain. These remnants of our past are called **vestigial characteristics**. As an example, consider the *grasp reflex* by which newborn infants close their fingers tightly around objects in their hands. This reflex may well be useful today in the development of the infant's ability to hold and manipulate objects, but that does not explain why prematurely born infants grasp so strongly that they can support their own weight, why they grasp with their toes as well as their hands (see Figure 3.12), and why the best stimulus for eliciting the response is a clump of hair (Eibl-Eibesfeldt, 1975). These aspects of the reflex make more sense when we observe them in other primates. To survive, infant monkeys and apes cling tightly with hands

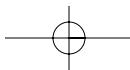
and feet to their mother's fur while she swings in trees or goes about her other daily business. In the course of our evolution from apelike ancestors, we lost our fur, so our infants can no longer cling to us in this way, but the reflex remains.

The concept of vestigial traits becomes especially relevant to psychologists when applied to our inherited drives, or motives. Because of culture, our habitats and lifestyles have changed dramatically in just a few centuries, a speck on the evolutionary time scale. Essentially all of our evolution as a species occurred in conditions that were quite different from those present today, and some of our inherited tendencies may be harmful, rather than helpful, in the habitat that we occupy today. An example is our great appetite for sugar. In the world of our ancestors, sugar was a rare and valuable commodity. It existed primarily in fruits and provided energy needed for life in the wild, as we can see by looking at the diets of present-day monkeys and apes. But today sugar is readily available in most areas of the world, and life (for many of us) is less physically strenuous. Yet our preference for sugar persists as strong as ever, despite such negative consequences as tooth decay and obesity.

Some Traits Are Side Effects of Natural Selection for Other Traits

Useless changes can come about in evolution as by-products of natural selection for other, useful changes. A simple example, in humans, is the belly button (suggested by Buss & others, 1998). To the best of anybody's knowledge, the belly button serves no function related to survival or reproduction. It is simply a remnant left from the umbilical cord. The umbilical cord, of course, does serve a survival and reproductive function, that of conveying nutrients from the mother to the developing fetus. An anatomist from Mars who observed belly buttons on adult earthlings, but who never observed a fetus or the birth process, would be at a loss to explain why such a structure would have evolved.

It seems quite possible that some human psychological capacities, even some that are so general that we would consider them to be part of human nature, came about as side effects of the evolution of other capacities. It is reasonable to ask, for example, whether the universal human proclivities for art and music are direct effects of natural selection or side effects. Perhaps these proclivities served to attract



mates during much of our evolutionary history (as they seem to today), and were therefore selected for directly, much as song was selected for in birds. It is also possible, however, that they emerged simply as by-products of selection for other proclivities, such as those for planning, constructing tools, and communicating through language. A third possibility, combining the first two, is that proclivities for art and music may have initially emerged as by-products and then been selected for because of their usefulness for attracting mates or other helpers. At present, we do not have evidence to support strongly any of these theories over the others.

Some Traits Result Simply from Chance

Some inheritable characteristics that result from just one or two mutations are inconsequential for survival and reproduction. Different races of people have somewhat differently shaped noses. *Maybe* that variation is caused by natural selection. Perhaps one shape worked best in one climate and another worked best in another climate, so natural selection molded the noses differently. But we can't automatically assume that. The different shapes might be a result of mutations that didn't matter and therefore were never weeded out by natural selection. Maybe the small group of people who migrated to a specific part of the world, and who were the ancestors of a particular racial group, just happened to carry along genes for a nose shape that was different from the average for the group they left. Such variation, due to chance alone without selection, is referred to as *genetic drift*.

Many years ago, researchers discovered that the incidence of schizophrenia (a serious mental disorder, discussed in Chapter 16) is three times greater among people living in northern Sweden, above the Arctic Circle, than among people in most other parts of the world (Huxley & others, 1964). There are at least three possible explanations of this observation: (a) Environmental conditions, such as the harsh climate or the isolation it produces, might tend to bring on schizophrenia in people who are prone to it. (b) Natural selection might have increased the frequency of schizophrenia-promoting genes among these people, perhaps because such genes help protect people from harmful effects of physical stressors such as cold climate. (This was the hypothesis suggested by Huxley and his colleagues.) (c) The Arctic population may have been founded by a small group of Swedish migrants who, just by chance, had a higher proportion of schizophrenia-promoting genes than the population at large. This last possibility (also mentioned by Huxley and his colleagues) would be an example of genetic drift. To this day, scientists do not know which of these theories is correct.

Evolved Mechanisms Cannot Deal Effectively with Every Situation

Our basic drives, emotions, and other behavioral tendencies came about in evolution because, on balance, they promoted survival and reproduction more often than they interfered with survival and reproduction. That does not mean, however, that every instance of activation of such a drive, emotion, or tendency serves survival or reproductive ends. In Chapter 14 I will present evidence that the emotion of guilt serves the ultimate function of helping us to preserve our relationships with people whose help we need for survival and reproduction. When we hurt someone on whom we depend, we feel guilty, which motivates us to make amends and patch up the relationship. That does not mean, however, that every manifestation of guilt in every person serves that function. Sometimes guilt can be crippling; sometimes our capacity for guilt is exploited by others for their ends at the expense of ours. The best that natural selection could do was to develop a guilt mechanism that is triggered by certain general conditions. It could not build a mechanism capable of distinguishing every possible condition from every other and triggering guilt only when it is useful, never when it is harmful. The same is true for all of our other evolved emotions and drives (and that is why we have psychotherapists).

SECTION REVIEW

The concept of natural selection provides a secure footing for functionalism.

The Functionalist Approach

- Functionalism is an approach to psychology that focuses on the usefulness of a particular behavior to the individual engaging in it.
- Ultimate explanations are functional explanations, stating the role that specific behaviors play in survival and reproduction.
- Proximate explanations are complementary to ultimate explanations; they are concerned with mechanisms that bring about behavior.

Limitations of Functionalism

- Some traits are vestigial; they once served a function but no longer do.
- Some traits are side effects of other traits that arose through natural selection.
- Some traits that require few mutations are products just of chance, not natural selection.
- Even evolved mechanisms, such as that for guilt, are not useful in every situation in which they are active.

Natural Selection as a Foundation for Understanding Species-Typical Behaviors

Suppose you saw an animal that looked exactly like a dog, but it went “meow,” climbed trees, and ignored the mail carrier. Would you call it a dog or a cat? Clearly, we identify animals as much by their behavior as by their anatomy. Every species of animal has certain characteristic ways of behaving. These are commonly called *instincts*, but a more technical term for them is **species-typical behaviors**. Meowing, tree climbing, and acting aloof are species-typical behaviors of cats. Dam building is species-typical of beavers. Smiling, talking, and two-legged walking are species-typical of humans.

Species-Typical Behaviors in Humans

Species-typical behaviors are products of evolution, but that does not mean that they are necessarily rigid in form or uninfluenced by learning. To understand more fully the concept of species-typical behaviors, let us examine some examples in humans.

Human Emotional Expressions as Examples of Species-Typical Behaviors

22

What evidence supports the idea that many human emotional expressions are examples of species-typical behaviors?

<<< Darwin noted that humans, like other animals, automatically communicate moods and behavioral intentions to one another through body postures, movements, and facial expressions. In his book *The Expression of the Emotions in Man and Animals*, Darwin (1872/1965) argued that specific facial expressions accompany specific emotional states in humans and that these expressions are universal, occurring in people throughout the world and even in people who were born blind and thus could not have learned them through observation.

In an extension of Darwin’s pioneering work, Paul Ekman and Wallace Friesen (1975, 1982) developed an atlas that describes and depicts the exact facial-muscle movements that make up each of six basic emotional expressions in people: surprise, fear, disgust, anger, happiness, and sadness (see Figure 3.13). They then showed photographs of each expression to individuals in many different cultures—including members of a preliterate tribe in the highlands of New Guinea who had little previous contact with other cultures—and found that people in every culture described each depicted emotion in a way that was consistent with descriptions in the United States (Ekman, 1973; Ekman & others, 1987). In a reversal of this proce-

Ekman, P. & Friesen, W. (1975).
Unmasking the face. Englewood Cliffs,
NJ: Prentice Hall.

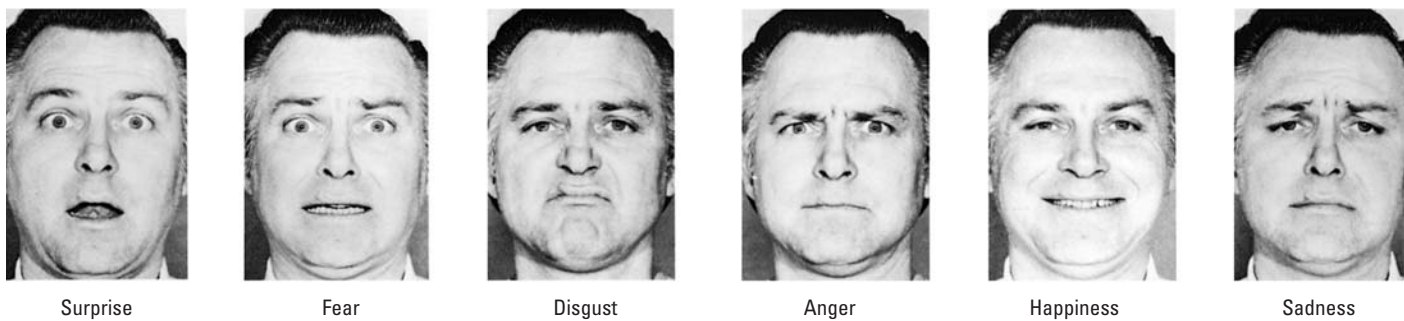


FIGURE 3.13 | Six basic human emotional expressions
These expressions, taken from Ekman and Friesen's atlas of emotional expressions, were produced by a model who was asked to move specific facial muscles in specific ways. As you study each figure, try to describe the positions of the facial features for each

expression. For example, surprise can be described as follows: The brows are pulled upward, producing horizontal wrinkles across the forehead; the eyes are opened wide, revealing white above the iris; and the lower jaw is dropped, with no tension around the mouth.

ture, they also photographed members of the New Guinea tribe who had been asked to act out various emotions and showed the photographs to college students in the United States. The college students were quite accurate in labeling the emotions portrayed by the New Guineans.

In a further extension of Darwin's work, Irenäus Eibl-Eibesfeldt documented the cross-cultural universality of many nonverbal signals, including one that he labeled the *eyebrow flash*, a momentary raising of the eyebrows lasting about one-sixth of a second, usually accompanied by a smile and an upward nod of the head (see Figure 3.14). He observed this expression in every culture he studied—including those in New Guinea, Samoa, and various parts of Africa, Asia, South America, and Europe—and concluded that it is a universal sign of greeting among friends (Eibl-Eibesfeldt, 1989). Raised eyebrows are also a component of the emotional expression of surprise (look at Figure 3.13 again), so the eyebrow flash with its accompanying smile might be interpreted as a nonverbal way of saying, "What a happy surprise to see you."

Eibl-Eibesfeldt (1975) also filmed children who were born blind, or both blind and deaf, and found that they manifest emotions in the same basic ways as sighted children do (see Figure 3.15). Such observations provide the most direct evidence that at least some human expressions do not have to be learned through observing them in others or hearing descriptions of them.

Taking all the evidence together, there can be little doubt that we are biologically predisposed to express certain emotions in certain species-typical ways. It is also clear, however, that we can control and modify our emotional expressions and

>>> 23
How do human emotional expressions illustrate the point that species-typical behaviors can be modified by learning?

Eibl-Eibesfeldt, I. (1989). *Human ethology*.
Hawthorne, NY: Walter de Gruyter, Inc.



FIGURE 3.14 | The eyebrow flash This universal signal of greeting is shown in adjacent frames from films of (a) a French woman and (b) a Yanomami man (of the Brazil-Venezuela border).

Eibl-Eibesfeldt, I. (1975). *Human ethology*. Hawthorne, NY: Holt, Rinehart & Winston.



FIGURE 3.15 Some emotional expressions need not be learned through observation. This young girl, manifesting joy, has been blind and deaf since birth.

24

How is the point that species-typical behaviors may depend on learning illustrated by the examples of two-legged walking and language in humans, and singing in white-crowned sparrows?

The Role of Learning in the Development of Species-Typical Behaviors

To say that a behavior is instinctive, or species-typical, is not to say that it is unaffected by learning. As I just pointed out, our basic emotional expressions are species-typical, but cultural differences among them are learned. The role of learning is even more obvious in two of our most characteristic species-specific behaviors—our manner of walking and our use of language.

<<< A scientist from Mars would almost certainly point to two-legged walking and use of a grammar-based language as being among the defining behavioral characteristics of the human species. These characterize humans everywhere and clearly depend on inherited predispositions, yet their development also clearly depends on learning. During the peak months of learning to walk (generally during the second year of life), infants spend an average of about six hours per day practicing balancing and walking and, on a typical day, take about 9,000 walking steps and travel the length of 29 football fields (Adolph & others, 2003). By the time they are proficient walkers, they have spent thousands of hours practicing, on their own initiative. During those same months, infants also, on their own initiative, intensely practice talking. With language, infants do not just learn the motor coordination needed to produce the sounds, but they also, and more amazingly, learn the basic vocabulary and grammar of the language that they hear around them (to be discussed in Chapter 11). Two-legged walking and talking are species-typical behaviors in humans, but a human raised in an environment where either of these capacities was impossible to practice would not develop that capacity. Such an in-human environment would not produce a normal human being.

Learning plays crucial roles in the development of species-specific behaviors in other animals as well. For example, white-crowned sparrows develop the ability to sing their species-typical song only if they are permitted to hear it during the first summer after hatching (Marler, 1970). Indeed, populations of the species living in different areas have somewhat different dialects, and a white-crowned sparrow learns to sing the dialect of the adult that it hears (usually its father). Yet the range of possible songs that the birds can learn is limited by their biology. No matter what environmental experiences it has, a white-crowned sparrow cannot learn to sing like a canary or like any species other than a white-crowned sparrow. In Chapter 4 you will encounter other examples of species-typical behaviors that require specific learning experiences to develop normally.

25

How is the concept of biological preparedness related to that of species-typical behavior? How is biological preparedness illustrated with the examples of human walking and talking?

Biological Preparedness as the Basis for Species-Typical Behaviors

The difference between behaviors that we call instinctive, or species typical, and those that we do not so label has to do with the degree of *biological preparedness*.

<<< Evolution has equipped each species with anatomical structures, in the brain and



© Jennie Woodcock; Reflections Photolibrary / Corbis



© Geri Engberg / The Image Works

Biological preparedness Infants are born with a “stepping reflex,” which shows that they have, inborn, the neural mechanisms for two-legged walking. Perfecting that behavior, however, requires extensive practice, during the toddling stage, when the legs have become strong enough to support the child’s weight.

elsewhere, that operate to ensure that normal individuals of the species, raised in a normal environment for that species, will be physically able to perform the species-typical behaviors and will be motivated to learn what they must for adequate performance.

Human beings clearly come into the world biologically prepared to learn to walk on two legs. Evolution has provided us with anatomical features—such as strong hindlimbs with feet, weaker forelimbs without feet, an upwardly tilted pelvis, and a short, stiff neck—that combine to make it more convenient for us to walk upright than on all fours. Moreover, we are born with neural systems in the brain and spinal cord that enable us to move our legs correctly for coordinated two-legged walking and with neural structures that motivate us to practice this behavior at the appropriate stage in our development. Consider the difference between two-legged walking in humans and in dogs. Dogs are capable of learning to walk on two legs, and much is made of that fact by circus trainers, but they are never very good at it. They do not have the appropriate muscular and skeletal systems to coordinate the behavior properly, and they have no natural impulse to walk in this manner. A dog, unlike a human child, will practice two-legged walking only if it receives immediate rewards, such as food, for doing so. Thus two-legged walking is not a species-typical behavior in dogs.

The same is true for talking. Humans are born with anatomical structures, including a tongue and larynx, that can produce a wide range of sounds and with a brain that has special neural centers for understanding and producing language (discussed in Chapter 5). Infants begin talking at a certain stage even if they receive little encouragement from those around them. Chimpanzees can be taught to simulate some aspects of human language, just as dogs can be taught to walk on their hind legs, but they require lots of encouragement and are never very good at it (discussed in Chapter 11).

The Relative Nature of the Concept of Species-Typical Behavior

Having characterized the concept of species-typical behavior in terms of biological preparedness, I must now add that the concept is relative rather than absolute. No behavior stems just from biological preparation; some sort of experience with the environment is always involved. Conversely, any behavior that an animal can produce—no matter how artificial it may seem or how much training is required—must make use of the animal’s inherited biological capacities. The concept of

>>> **26**
Why is the concept of species-typical behavior relative rather than absolute?

species-typical behavior is useful as long as we accept it as relative and do not argue about whether one or another behavior really should or should not be called species-typical. *Big* and *little* are useful words in our vocabulary, but there is no point in arguing about whether a breadbox is properly called one or the other. Two-legged walking is more species-typical for humans than for dogs, as a breadbox is bigger than a matchbox.

The question to ask when we study a particular behavior is not, Is this a species-typical behavior? Rather, the meaningful questions are these:

- What are the environmental conditions needed for the full development of this behavior?
- What internal mechanisms are involved in producing this behavior?
- What consequences does this behavior have in the individual's daily life?
- In the course of evolution, why would the genes that make this behavior possible have been favored by natural selection?

These questions can be asked of any behavior, regardless of whether or not it is thought of as species-typical.

The Value of Cross-Species Comparisons of Species-Typical Behaviors

In psychology as well as biology, scientists have learned a lot about human beings by comparing humans to other animals. The basic rationales for learning about any one species by comparing it with other species are found in the principle of evolution by natural selection.

Two Forms of Cross-Species Comparison: Homologies and Analogies

27

What is the difference between a homology and an analogy, and how can researchers tell whether a similarity between two species in some trait is one or the other?

<<< An understanding of evolution makes it clear that two conceptually different classes of similarities exist across species: homologies and analogies.

A **homology** is any similarity that exists because of the different species' common ancestry. All animals originated from a common ancestor, so it is not surprising that some homologies—such as those in the basic structure of DNA molecules and of certain enzymes—can be found between any two species. But the more closely related two species are, the more homologies they will show.

An **analogy**, in contrast, is any similarity that stems not from common ancestry but from *convergent evolution*. Convergent evolution occurs when different species, because of some similarity in their habitats or lifestyles, independently evolve a common characteristic.

As an illustration, consider some comparisons among species that can fly. Flying has arisen separately in three taxonomic groups: birds, some insects (such as butterflies), and some mammals (bats). Similarities across these three groups in

FIGURE 3.16 | Analogous wings
Similarities in the wings and flying behavior of birds, bats, and butterflies are considered to be analogies, not homologies, because they arose independently in evolution.



Joe McDonald / Earth Scenes / Animals, Animals



M. Tuttle / Photo Researchers



James Carmich / Bruce Coleman

their flying motions, and in the anatomical structures that permit flight, are examples of analogies because they do not result from common ancestry (see Figure 3.16). However, similarities in flight and wings among species within any of these groups, such as between crows and sparrows, are likely to be homologies. The last common ancestor between a crow and a sparrow was itself a bird with wings, but the last common ancestor between a crow and a butterfly, or between a crow and a bat, did not have wings.

Aside from evidence based on knowledge about the degree of relatedness of the compared species, analogies and homologies can often be distinguished by the nature of the observed similarity (Lorenz, 1974). Analogies entail similarity in function and gross form, but not in detail and underlying mechanism. Thus the wings of birds, bats, and butterflies are similar at the functional and gross anatomical level in that they provide broad, flappable surfaces that enable flight; but they are very different from one another in the details of their construction and in the neural and muscular mechanisms that control them. The difference in detail is great because they evolved independently, through different sets of genetic mutations. In contrast, because homologies arise from shared genes, they entail similarities in their underlying construction and physiological mechanisms, even when, because of divergent evolution, large differences have emerged in gross form or function (for example, see Figure 3.17).

The Value for Psychology of Studying Homologies

Homologies are useful for research on the physiological mechanisms of behavior (that is, research on how the brain and other biological structures operate to produce the behavior being studied). Because convergent evolution can produce similar behaviors that operate through different mechanisms, researchers who seek to understand the physiological mechanism of some behavior in humans through experiments on other species must study species in which the relevant behavior is homologous, not analogous, to that in humans. You will find many examples of such research in Chapters 4 to 8 of this book. Many basic mechanisms of learning, motivation (such as hunger), and sensation (such as vision) are homologous across all or at least most species of mammals, and we have learned much about these by studying them in mice, rats, cats, and other laboratory mammals.

Homologies are also useful for inferring the pathways along which species-typical behaviors evolved. By comparing the different forms of a particular species-typical behavior in closely related species, it is often possible to reconstruct how the more complex of these forms evolved through a series of steps from the simpler form. Darwin (1859/1963) himself used this method to figure out the evolutionary steps through which honeybees acquired their marvelous ability to construct complex hives consisting of wax combs of closely fitting hexagonal cells in which to store honey and larvae (see Figure 3.18).

Homologies as Clues to the Evolutionary Origins of Two Human Smiles

In research that is more directly relevant to psychology, Darwin also used homologies to understand the origins of species-typical emotional expressions in humans. At the London Zoo, he watched monkeys and apes and noted that a number of their expressions seemed to be homologous to human expressions, including the smile (Darwin, 1872/1965). Research following up on Darwin's work has suggested that people may produce two kinds of smiles, which may have separate evolutionary origins.

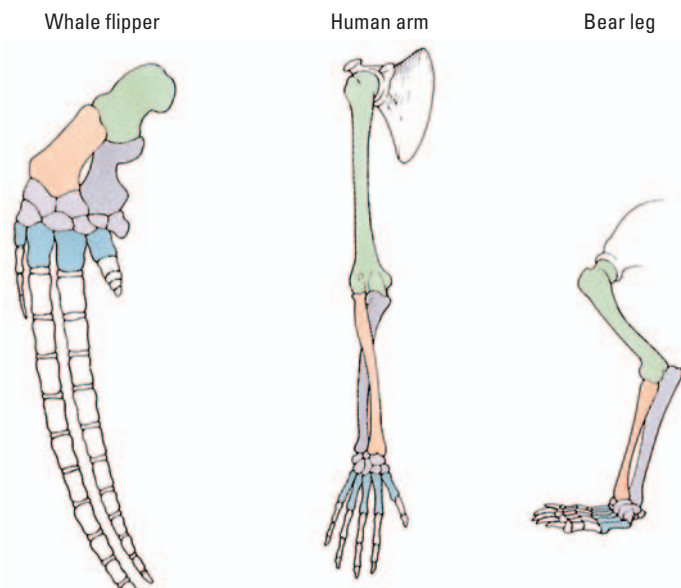


FIGURE 3.17 | Homologous forelimbs Similarities in the forelimbs of different species of mammals are considered to be homologies because they arose from common ancestry. Although the limbs of the whale, human, and bear differ in function and gross structure, they are similar in certain structural details, as is characteristic of homologies. Behaviors, too, can be homologous, and a key to their homology is similarity in mechanism and detail, even when function differs. (Adapted from Lorenz, 1974.)

28

How are homologies used for learning about (a) the physiological mechanisms and (b) the evolutionary pathways of species-typical traits?



M. A. Chappell / Animals, Animals

(a)



D. M. Shale / Earth Scenes / Animals, Animals

(b)

FIGURE 3.18 | Cells built by honeybees and bumblebees
Honeybees build hives with hexagonally shaped cells (a), which are the optimal shape for storing large amounts of honey and larvae using the least amount of precious wax. In order to understand how such behavior might have come about through small steps in natural selection, Darwin studied the homologous but simpler storage structures built by related bee species. The simplest, produced by bumble bees (b), consists simply of a cluster of spherical cells, which the bees easily build by sweeping their

abdomens compass-like to carve out the spheres. Other species, more closely related to honeybees, build intermediate structures, with spherical cells that intersect and are patched up with flat wax walls at the places of intersection. From such observations, Darwin suggested that ancestors of modern honeybees built their hives in a way similar to that of modern bumblebees but, through evolution, began placing their cells ever closer together and more regularly spaced and patching up the intersections, resulting eventually in the hexagonal cells that honeybees build today.

People smile in two quite different contexts: (1) when genuinely happy and (2) when wishing to show another person that they are favorably disposed toward that person. The latter situation need not entail happiness at all; in fact, people are especially likely to smile at others in potentially tense circumstances, apparently as a means of reducing the tension (Goldenthal & others, 1981). Darwin (1872/1965) pointed out that these two smiles are anatomically distinct. The *happy smile* involves not just the turning up of the corners of the lips but also the pulling in of the skin near the outside corners of the eyes. This creates the creases called crow's feet, which radiate from the eyes and seem to make them sparkle. The other smile, in contrast, typically involves the lips alone, without the eyes. This distinction has been confirmed in many studies with both adults and young children (Ekman, 1992; Sarra & Otta, 2001). In one study, for example, 10-month-old infants smiled with eyes and mouth when approached by their mother (presumably a happy situation) but smiled with mouth alone when approached by a stranger (a possibly tense situation) (Fox & Davidson, 1988).

Ekman (1992) considers the mouth-alone smile to be a derivative of the happy smile. He emphasizes its use in masking one's true feelings and calls it a "false smile." An alternative possibility, supported by research with monkeys and apes, is that the mouth-alone smile is a unique expression—let's call it the *greeting smile*—that arrived through a different evolutionary route from the happy smile (Redican, 1982; van Hooff, 1976).

29

How do studies of homologies between humans and other primates support the view that the human greeting smile and the human happy smile have separate evolutionary origins?

<<< Possible Origin of the Greeting Smile Nonhuman primates manifest two distinct smile-like displays. The one that seems most clearly to be homologous to the human greeting smile is the *silent bared-teeth display* (see Figure 3.19a). If you have ever watched a cage of macaque monkeys at a zoo, you have almost certainly observed this display, a grimace usually shown by the more submissive of two monkeys as it glances nervously toward the more dominant. A direct stare in macaques (and other primates) is an aggressive signal, which precedes attack and can precipitate an attack by the other, and the silent bared-teeth display seems to have evolved as a means for a more submissive monkey to look at a more dominant one without provoking a fight. If it could be translated into words, it might be rendered

FIGURE 3.19 Possible homologues to two types of human smiles The silent bared-teeth display (a) is believed to be homologous to the human greeting smile, and the relaxed open-mouth display (b) is believed to be homologous to the human laugh and happy smile. The animals in both photos are chimpanzees.

as, “I’m looking at you but I’m not going to attack, so please don’t attack me.” In some monkey species, it is also used to promote affiliation after an aggressive encounter (Preuschoft & van Hooff, 1997).

J. A. van Hooff (1972, 1976) found that among chimpanzees this display takes on a new function, more similar to that of the human smile of greeting. *Both* the more submissive and the more dominant of two chimpanzees show the display upon meeting, and it usually precedes friendly interaction between them. From such observations, van Hooff proposed that the silent bared-teeth display originated in monkeys as a submissive gesture, but in the evolutionary line leading to chimpanzees and humans it evolved further into a general form of greeting. As used by the more submissive individual, it may retain its original meaning, “Please don’t attack me,” but as used by the more dominant, it may mean, “Rest assured, I won’t attack,” and as used by both it may mean, “Let’s be friends.”

Possible Origin of Laughter and the Happy Smile The other primate smile-like expression is the *relaxed open-mouth display*, or *play face* (see Figure 3.19b), which occurs mostly in young primates during playful fighting and chasing and may be homologous to both laughter and the happy smile in humans. In chimpanzees, it is often accompanied by a vocalized *ahh ahh ahh*, which sounds like a throaty human laugh. Van Hooff believes that this display originated as a means for young primates to signal to each other that their aggressive-like behavior is not to be taken seriously; nobody will really get hurt. Interestingly, in human children, laughter occurs during playful fighting and chasing more reliably than during any other form of play (Blurton-Jones, 1967), and even among us “sophisticated” adults, pie throwing, chase scenes, mock insults, and other forms of fake aggression are among the most reliable ways to elicit laughter. Thus our laughter not only is similar in form to the relaxed open-mouth display of other primates but, at least in some cases, seems to serve a similar function as well.

The smile of laughter is similar in form to the non-laughing happy smile, so the latter, too, may have its roots in the relaxed open-mouth display (Redican, 1982). There is, it seems to me, some poetry in the thought that the smile of happiness may have originated from a signal indicating that, although the world can be frightening and full of conflict, the aggression going on now is just in fun and we are safe.

The Value, for Psychology, of Studying Analogies

You have just seen examples of how homologies can be used to make inferences about the evolutionary origins of species-typical behaviors. Analogies, in contrast, are not useful for tracing evolutionary origins, but are useful for making inferences about the ultimate functions of species-typical behaviors. If different species have independently evolved a particular behavioral trait, then comparing the species may reveal commonalities of habitat and lifestyle that are clues to the ultimate function of that trait. You will see examples of this use of analogies in the remaining sections of this chapter, as applied to patterns of mating, patterns of aggression, and patterns of helping.



Tim Davis / Animals, Animals

(a)



Norman Tornalin / Bruce Coleman

(b)

>>> **30**
How can analogies be used to make inferences about the ultimate functions of species-typical traits?

SECTION REVIEW

Species-typical behaviors have come to exist through natural selection.

Species-Typical Behaviors

- Commonly called instincts, species-typical behaviors are ways of behaving that characterize a species—such as cats meowing and humans walking upright.
- They may be influenced by learning or even require learning, as exemplified by cultural differences in the eyebrow flash, human language learning, and white-crowned sparrows’ song development.
- They depend upon biological preparedness—i.e., upon having anatomical structures that permit and motivate the behavior).

Homologies and Analogies

- Homologies are similarities due to common ancestry. They are useful for studying underlying mechanisms and for tracing the evolutionary course of species-typical behaviors, exemplified by research on the greeting smile and happy smile in humans.
- Analogies are similarities due to convergent evolution (independent evolution of similar traits). They are useful for inferring ultimate functions.

Evolutionary Analysis of Mating Patterns

From an evolutionary perspective, no class of behavior is more important than mating. Mating is the means by which all sexually reproducing animals get their genes into the next generation. Mating is also interesting because it is the most basic form of social behavior. Sex is the foundation of society. Were it not necessary for female and male to come together to reproduce, members of a species could, in theory, go through life completely oblivious to one another.

Countless varieties of male–female arrangements for sexual reproduction have evolved in different species of animals. One way to classify them is according to the number of partners a male or female typically mates with over a given period of time, such as a breeding season. Four broad classes are generally recognized: **polygyny** [pah-**lij**-inee], in which one male mates with more than one female; **polyandry** [pah-lee-**an**-dree], in which one female mates with more than one male; **monogamy**, in which one male mates with one female; and **polygynandry** [pah-lee-**j**in-an-dree], in which members of a group consisting of more than one male and more than one female mate with one another (Rees & Harvey, 1991). (These terms are easy to remember if you know that *poly-* means “many”; *mono-*, “one”; *-gyn*, “female”; and *-andr*, “male.” Thus, for example, *polygynandry* literally means “many females and males.”) As illustrated in Figure 3.19, a feature of both polygyny and polyandry is that some individuals are necessarily deprived of a mating opportunity—a state of affairs associated with considerable conflict.

A Theory Relating Mating Patterns to Parental Investment

31 What is Trivers’s theory of parental investment?

<<< In a now-classic article, Robert Trivers (1972) outlined a theory relating courtship and mating patterns to sex differences in amount of **parental investment**. Parental investment can be defined roughly as the time, energy, and risk to survival that are involved in producing, feeding, and otherwise caring for each offspring. More precisely, Trivers defined it as the loss, to the adult, of future reproductive capacity that results from the production and nurturance of any given offspring. Every offspring in a sexually reproducing species has two parents, one of each sex, but the amount of parental investment from the two is usually not equal. The essence of Trivers’s theory is this: *In general, for species in which parental investment is unequal, the more parentally invested sex will (a) be more vigorously competed for than the other and (b) be more discriminating than the other when choosing mates.*

To illustrate and elaborate on this theory—and to see how it is supported by cross-species comparisons focusing on analogies—let us apply it to each of the four general classes of mating patterns.

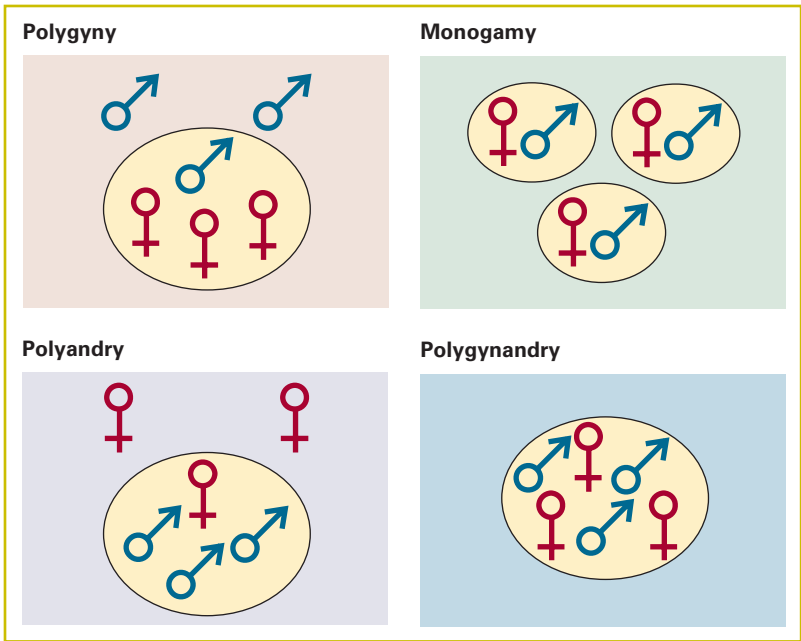


FIGURE 3.20 Four mating systems In a polygynous system (common in mammals), the unmated males are a threat to the mated male, and in a polyandrous system (present in some birds and fishes), the unmated females are a threat to the mated female. Threat is reduced by monogamy and polygynandry, because with those systems, most individuals find mates.

Polygyny Is Related to High Female and Low Male Parental Investment

Most species of mammals are polygynous, and Trivers's theory helps explain why. Mammalian reproductive physiology is such that the female necessarily invests a great deal in the offspring she bears. The young must first develop within her body and then must obtain nourishment from her in the form of milk. Because of the female's high investment, the number of offspring she can produce in a breeding season or a lifetime is limited. A female whose gestation and lactation periods are such that she can bear at most four young a year can produce no more than that regardless of the number of different males with which she mates.

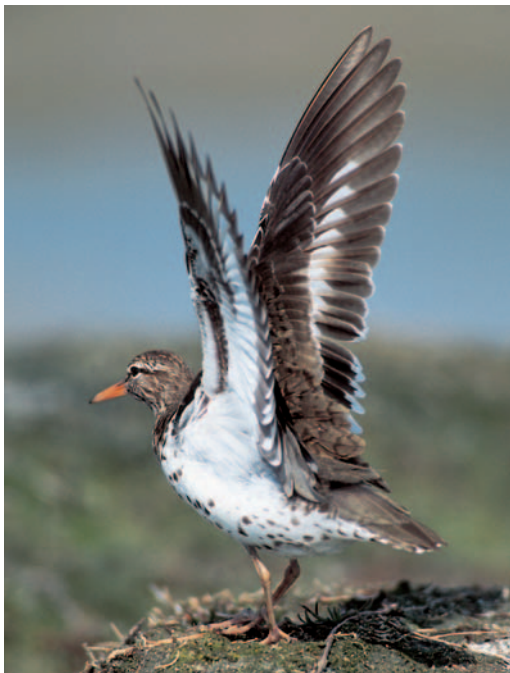
Things are different for the male. His involvement with offspring is, at minimum, simply the production of sperm cells and the act of copulation. These require little time and energy, so his maximum reproductive potential is limited not by parental investment but by the number of fertile females with which he mates.

>>> 32 Based on Trivers's theory of parental investment, why does high investment by the female lead to (a) polygyny, (b) large size of males, and (c) high selectivity in the female's choice of mate?



Who's bigger and stronger? These male elephant seals are sizing each other up for possible battle over mating rights to the many females in the background. Because the larger combatant usually wins, male elephant seals have through natural selection become huge compared with females.

Natural History, August 1996, p. 57. Photo by Tom Vezo.



An aggressive female The spotted sandpiper is a polyandrous species. The female mates with several males and defends her territory from invading females. This female is stretching her wings in a threat display.

A male that mates with 20 females, each of which can bear four young, can in theory produce 80 offspring a year. When the evolutionary advantage in mating with multiple partners is greater for males than for females, a pattern evolves in which males compete with one another to mate with as many females as they can.

Among mammals, competition among males for females often involves one-on-one battles, which the larger and stronger combatant most often wins. This leads to a selective advantage for increased size and strength in males, up to some maximum beyond which the size advantage in obtaining mates is outweighed by disadvantages, such as difficulty in finding sufficient food to support the large size. In general, the more polygynous a species, the greater is the average size difference between males and females. An extreme example is the elephant seal. Males of this species fight one another, sometimes to the death, for mating rights to groups averaging about 50 females, and the males outweigh females several-fold (Hoelzel & others, 1999). In the evolution of elephant seals, those males whose genes made them large, strong, and ferocious enough to defeat other males sent many copies of their genes on to the next generation, while their weaker or less aggressive opponents sent few or none.

For the same reason that the female mammal usually has less evolutionary incentive than the male to mate with many individuals, she has more incentive to be discriminating in her choice of mate (Trivers, 1972). Because she invests so much, risking her life and decreasing her future reproductive potential whenever she becomes pregnant, her genetic interests lie in producing offspring that will have the highest possible chance themselves to survive and reproduce. To the degree that the male affects the young, either through his genes or through other resources he provides, females would be expected to select males whose contribution will be most beneficial. In elephant seals, it is presumably to the female's evolutionary advantage to mate with the winner of battles. The male victor's genes increase the chance that the female's sons will win battles in the future and produce many young themselves.

Polyandry Is Related to High Male and Low Female Parental Investment

33

What conditions promote the evolution of polyandry? How do sex differences within polyandrous species support Trivers's theory?

<<< Polyandry is not the primary mating pattern for any species of mammal, but it is for some species of fishes and birds (Clutton-Brock & Vincent, 1991; Erckmann, 1983). Polyandry is more likely to evolve in egg-laying species than in mammals, because a smaller proportion of an egg layer's reproductive cycle is tied to the female's body. Once the eggs are laid, they can be cared for by either parent, and, depending on other conditions, evolution can lead to greater male than female parental investment.

Consistent with Trivers's theory, females of polyandrous species are the more active and aggressive courters, and they have evolved to be larger, stronger, and in some cases more brightly colored than the males (Berglund & Rosenqvist, 2001). An example is the spotted sandpiper, a common freshwater shorebird. A female spotted sandpiper can lay up to three clutches of eggs in rapid succession, each cared for by a different male that has mated with her (Oring, 1995). At the beginning of the breeding season, the females—which outweigh the males by about 20 percent and have somewhat more conspicuous spots—stake out territories where they actively court males and drive out other females.

Monogamy Is Related to Equivalent Male and Female Parental Investment

34

What conditions promote the evolution of monogamy? Why are sex differences in size and strength generally lacking in monogamous species?

<<< According to Trivers's theory, when the two sexes make approximately equal investments in their young, their degree of competition for mates will also be approximately equal, and monogamy will prevail. Equal parental investment is most likely to come about when conditions make it impossible for a single adult to raise

the young but quite possible for two to raise them. Under these circumstances, if either parent leaves, the young fail to survive, so natural selection favors genes that lead parents to stay together and care for the young together. Because neither sex is much more likely than the other to fight over mates, there is little or no natural selection for sex differences in size and strength, and, in general, males and females of monogamous species are nearly identical in these characteristics.

Consistent with the view that monogamy arises from the need for more than one adult to care for offspring, over 90 percent of bird species are predominantly monogamous (Cézilly & Zayan, 2000; Lack, 1968). Among most species of birds, unlike most mammals, a single parent would usually not be able to raise the young. Birds must incubate and protect their eggs until they hatch and then must guard the hatchlings and fetch food for them until they can fly. One parent alone cannot simultaneously guard the nest and leave it to get food, but two together can. Among mammals, monogamy has arisen in some species that are like birds in the sense that their young must be given food other than milk, of a type that the male can provide. The best-known examples are certain carnivores, including foxes and coyotes (Malcolm, 1985). Young carnivores must be fed meat until they have acquired the necessary strength, agility, and skills to hunt on their own, and two parents are much better than one at accomplishing this task. Monogamy also occurs in several species of rodents, where the male may play a crucial role in protecting the young from predators while the mother forages (Sommer, 2000).

With modern DNA techniques to determine paternity, researchers have learned that social monogamy (the faithful pairing of female and male for raising young) does not necessarily imply sexual monogamy (fidelity in copulation between that female and male). Researchers commonly find that between 5 and 35 percent of offspring in socially monogamous birds are sired by a neighboring male rather than by the male at the nest (Birkhead & Moller, 1992); for one species, the superb fairy wren, that average is 75 percent (Mulder, 1994). Why does such extra-mate copulation occur? From the female's evolutionary perspective, copulation with a male that is genetically superior to her own mate (as manifested in song and feathers) results in genetically superior young, and copulation with any additional male increases the chance that all her eggs will be fertilized by viable sperm (Zeh & Zeh, 2001). For the male, evolutionary advantage rests in driving neighboring males away from his own mate whenever possible and in copulating with neighboring females whenever possible. Genes that build brain mechanisms that promote such behaviors are passed along to more offspring than are genes that do not.

Polygynandry Is Related to Investment in the Group

Among the clearest examples of polygynandrous species are chimpanzees and bonobos, which happen to be our two closest animal relatives (see Figure 3.21). Bonobos are similar in appearance to chimpanzees but are rarer and have only recently been studied in the wild. The basic social structure of both species is the colony, which consists usually of two to three dozen adults of both sexes and their offspring. When the female is ovulating, she develops on her rump a prominent pink swelling, which she actively displays to advertise her condition. During the time of this swelling, which lasts about a week in chimps and 3 weeks in bonobos, she is likely to mate with most of the adult males of the colony, though she may actively choose to mate with some more often than with others (Goodall, 1986; Kano, 1992; Stumpf & Boesch, 2005).

Polygynandry has apparently evolved in these ape species because it permits a group of adult males and females to live together in relative harmony, without too much fighting over who mates with whom. A related advantage, from the female's perspective, is paternity confusion (Hrdy, 1981, 2000). Among many species of



Natural History, November 1994, p. 60. Photo by C. Allan Morgan.

A not-so-faithful couple The superb fairy wren is socially but not sexually monogamous. The male (at the left) and the female stay together at the nest and raise the young together, but DNA testing has shown that about 75 percent of the offspring, on average, are sired by neighboring males.

>>> 35 For what evolutionary reasons might monogamously mated females and males sometimes copulate with partners other than their mates?

>>> 36 What appear to be the evolutionary advantages of polygynandry for chimpanzees and bonobos, and in what ways is polygynandry more fully developed for the latter than the former?

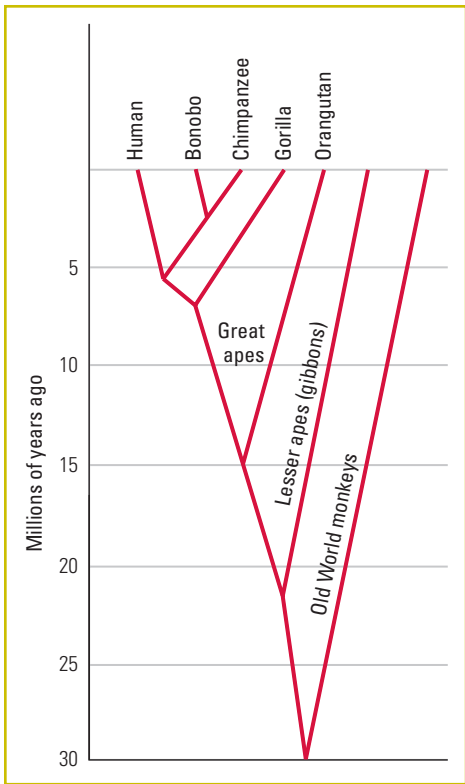


FIGURE 3.21 Relationship of humans to apes and Old World monkeys The ancestral line leading to humans split off from that leading to Old World monkeys 30 million years ago, and it split off from that leading to bonobos and chimpanzees 5.6 million years ago. (Data from Corballis, 1999.)



Frans Lanting / Minden Pictures

Bonobo sex Bonobos seem to live by the motto, “Make love, not war.” Research suggests that they are the most sexually active and the most peaceful of all primates. Here a male has mounted a female in a face-to-face position—a position long thought to be unique to humans. But bonobo sex occurs in all possible partner combinations (homosexual as well as heterosexual) and essentially all imaginable positions.

primates, males kill young that are not their own, and such behavior has been observed in chimpanzees when a female migrates into a colony bringing with her an infant that was sired elsewhere (Wrangham, 1993). Because almost any chimp or bonobo male in the colony could be the father of any infant born within the colony, each male's evolutionary interest lies not in attacking the young but in helping to protect and care for the group as a whole.

Polygynandry seems to be more fully developed in bonobos than in chimps. Male chimps do fight with one another over females that are at peak fertility, and sometimes a single male manages to monopolize the sexual activity of a female throughout her ovulatory cycle (Goodall, 1986; Wrangham, 1993). In contrast, bonobo males rarely compete physically to copulate. In fact, for bonobos sex appears to be more a reducer of aggression than a cause of it (Parish & de Waal, 2000; Wrangham, 1993). Unlike any other apes, female bonobos copulate at all times of their reproductive cycle, not just near the time of ovulation. In addition to their frequent heterosexual activity, bonobos of the same sex often rub their genitals together, and genital stimulation of all types occurs most often following conflict and in situations that could potentially elicit conflict, such as when a favorite food is discovered (Hohmann & Fruth, 2000; Parish, 1996). Field studies suggest that bonobos are the most peaceful of primates and that their frequent polygynandrous sexual activity helps keep them that way (Kano, 1992).

What About Human Mating Patterns?

37

What evidence suggests that humans evolved as a partly monogamous, partly polygynous species? How is this consistent with Trivers's parental investment theory?

<<< When we apply the same criteria that are used to classify the mating systems of other species, we find that humans fall on the boundary between monogamy and polygyny (Dewsbury, 1988). In no culture are human beings as sexually promiscuous as are our closest ape relatives, the chimpanzees and bonobos. In every culture people form long-term mating bonds, which are usually legitimized through some sort of culturally recognized marriage contract. Anthropologists have found that the great majority of non-Western cultures, where Western influence has not made

polygyny illegal, practice a mixture of monogamy and polygyny (Marlowe, 2000; Murdock, 1981). In such cultures, a few men, who have sufficient wealth or status, have two or more wives, while the great majority of men have one wife and a few have none. Thus, even in cultures that permit and idealize polygyny, most marriages are monogamous.

Human children, more so than the young of any other primates, require an extended period of care before they can play full adult roles in activities such as food gathering. Cross-cultural research shows that in every culture mothers provide most of the direct physical care of children, but fathers contribute in various ways. In many cultures—especially in hunter-gatherer cultures—fathers share to some degree in the physical care of their offspring (Marlowe, 2000), and in nearly all cultures fathers provide indirect care in the form of food and other material provisions. In fact, in 77 percent of the cultures for which data are available, fathers provide more of the provisions for young than do mothers (Marlowe, 2000). Taking both direct and indirect care into account, we are a species in which fathers typically lag somewhat behind mothers, but not greatly behind them, in degree of parental investment. This, in line with Trivers's parental investment theory, is consistent with our being a primarily monogamous but moderately polygynous species. The moderate size difference between men and women is also consistent with this conclusion (Dewsbury, 1988). The average size difference between males and females in humans is nowhere near that observed in highly polygynous species, such as elephant seals and gorillas, but is greater than that observed in monogamous species.

Our biological equipment that predisposes us for mating bonds includes brain mechanisms that promote the twin emotions of romantic love and sexual jealousy. These emotions are found in people of every culture that has been studied (Buss, 2000; Fisher, 1992). People everywhere develop strong emotional ties to those toward whom they are sexually drawn. This is experienced as a need to be constantly near the other person. People also, everywhere, feel intensely jealous when “their” mates appear to be sexually drawn to others. While love tends to create mating bonds, jealousy tends to preserve such bonds by motivating each member of a mated pair to act in ways designed to prevent the other from having an affair with someone else.

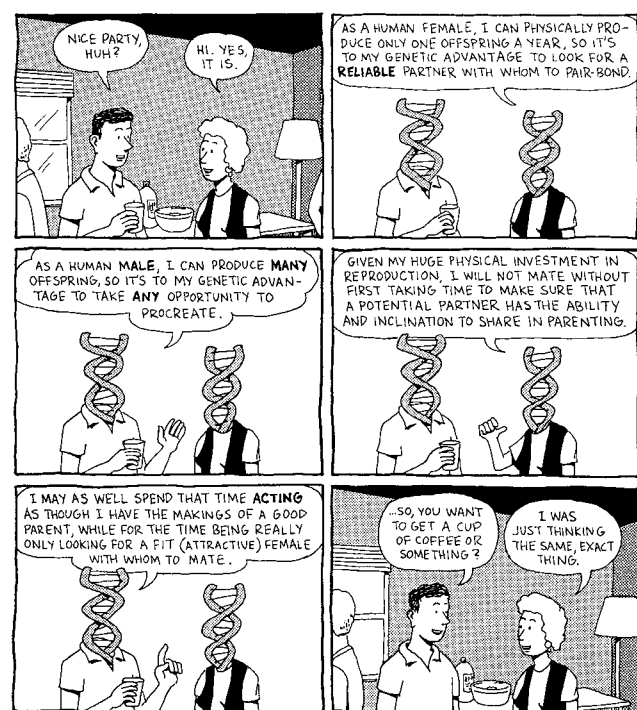
Other animals that form long-term mating bonds show evidence of emotions that are functionally similar to human love and jealousy. In this sense, we are more like monogamous birds than we are like our closest ape relatives. The similarities between humans and birds in sexual love and jealousy are clearly analogies, not homologies (Lorenz, 1974). These evolved separately in humans and birds as means to create and preserve mating bonds long enough to promote biparental care of offspring. Chimpanzees and bonobos (especially the latter) can engage in open, promiscuous sex with little emotional consequence, because they have not evolved strong emotions of sexual love and jealousy, but humans and monogamous birds cannot. The difference ultimately has to do with species differences in the need for care from both parents.

At the same time that love and jealousy tend to promote bonding, lust (another product of evolution) tends to motivate both men and women to engage surreptitiously in sex outside of such bonds. In this sense we are like those socially monogamous birds that are sexually unfaithful. A man who can inseminate women beyond his wife may send more copies of his genes into the next generation than a completely faithful man. A woman who has sex with men other than her husband may benefit, evolutionarily, by (a) increasing her chances of conception (hedging

A
“TOM THE
DANCING
BUT”
COMIC STRIP

BY
RUBEN
BOLLING

THE EVOLUTIONARY PSYCHOLOGY OF LOVE



©1994 Ruben Bolling. Distributed by Quaternary Features.

From an evolutionary perspective, what are the functions of romantic love and sexual jealousy, and how is this supported by cross-species comparisons? How is sexual unfaithfulness explained?

against the possibility that her husband's sperm are not viable or are genetically incompatible with her eggs); (b) increasing the evolutionary fitness of her offspring (by mating with a man who has evolutionarily superior genes compared to those of her husband); and/or (c) gaining provisions from more than one man (Hrdy, 2000). And so the human soap opera continues, much like that of the superb fairy wren, though not to such an extreme. Studies involving DNA testing, in cultures ranging from hunter-gatherer groups to modern Western societies, suggest that somewhere between 2 and 10 percent of children in socially monogamous families are sired by someone other than the mother's husband (Marlowe, 2000).

SECTION REVIEW

An evolutionary perspective offers functionalist explanations of mating patterns.

Mating Patterns Related to Parental Investment

- Trivers theorized that sex differences in parental investment (time, energy, risk involved in bearing and raising young) explain mating patterns and sex differences in size, aggressiveness, competition for mates, and selectivity in choosing mates.
- Consistent with Trivers's theory, polygyny is associated with high female and low male parental investment; polyandry is associated with the opposite; and monogamy is associated with approximately equal investment by the two sexes.
- Polygynandry, common to chimps and bonobos, seems to be associated with high investment in the group.

Human Mating Patterns

- Parental investment is somewhat lower for fathers than for mothers, consistent with the human mix of monogamy and polygyny.
- Romantic love and jealousy help promote and preserve bonding of mates, permitting two-parent care of offspring.
- Sex outside of mating bonds may yield evolutionary benefits.

Evolutionary Analyses of Hurting and Helping

Human beings, like other animals, are motivated both to hurt and to help one another in the struggle to survive and reproduce. From an evolutionary perspective, other members of one's species are competitors for food, mates, safe places to live, and other limited resources. Ultimately, such competition is the foundation of aggression. Yet, at the same time, others of one's kind are potential helpmates. Many life-promoting tasks can be accomplished better by two or more individuals working together than by one struggling alone. The human drama, like that of other social species, involves the balancing of competitiveness with the need for others' help. Let us look first at the grim topic of aggression and then end, more pleasantly, with cooperation.

Sex Differences in Aggression

Aggression, as the term is used here, refers to fighting and threats of fighting among members of the same species. Brain mechanisms that motivate and organize such behavior have evolved because they help animals acquire and retain resources needed to survive and reproduce. As you saw in the previous section, much animal aggression centers on mating. Polygynous males and polyandrous females fight for mates; monogamous males fight to prevent other males from copulating with their mates; and monogamous females fight to keep other females from leading their mates away (Slagsvold & others, 1999; Tobias & Seddon, 2000). Aggression can also serve to protect a feeding ground for oneself and one's offspring, to drive away individuals that may be a threat to one's young, and to elevate one's status within a social group. Much could be said from an evolutionary perspective about all aspects of aggression, but here I will focus just on sex differences in how aggression is manifested.

Why Male Primates Are Generally More Violent Than Female Primates

Among most species of mammals, and especially among primates, males are much more violently aggressive than are females. Female primates are not unaggressive, but their aggression is typically aimed directly toward obtaining resources and defending their young. When they have achieved their ends, they stop fighting. Male primates, in contrast, seem at times to go out of their way to pick fights, and they are far more likely to maim or kill their opponents than are females.

Most of the violence perpetrated by male primates has to do directly or indirectly with sex. Male monkeys and apes of many species have been observed to kill infants fathered by others, apparently as a means to get the females to stop lactating so they will ovulate again and become sexually active. Males also fight with one another, sometimes brutally, to raise their ranks in the dominance hierarchy of the colony. High rank generally increases both their attractiveness to females and their ability to intimidate sexual rivals (Cowlshaw & Dunbar, 1991). Males are also often violent toward females; they use violence to force copulation or to prevent the female from copulating with other males. All of these behaviors have been observed in chimpanzees and many other primate species (Goodall, 1986; Smuts, 1992; Wittig & Boesch, 2003).

Evolution, remember, is not a moral force. It promotes those behaviors that tend to get one's genes passed on to the next generation. Female primates, because of their higher parental investment, don't need to fight to get the opposite sex interested in them. Moreover, aggression may have a higher cost for females than for males. The female at battle risks not just her life but also that of any fetus she is gestating or young she is nursing—the repositories of her genes. The male at battle risks just himself, and, in the cold calculus of evolution, his life isn't worth anything unless he can get a female to mate with him. Genes that promote mating, by whatever means, proliferate, and genes that fail to promote it vanish.

Humans are no exception to the usual primate rule. Cross-cultural studies show that everywhere men are more violent, more likely to maim or kill, than are women. In fact, in a survey of cross-cultural data on this issue, Martin Daly and Margo Wilson (1988) were unable to find any society in which the number of women who killed other women was even one-tenth as great as the number of men who killed other men. On average, in the data they examined, male-male killings outnumbered female-female killings by more than 30 to 1. One might construe a scenario through which such a difference in violence would be purely a product of learning in every culture, but the hypothesis that the difference resides at least partly in inherited sex differences seems more plausible.

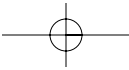
According to Daly's and Wilson's analyses, the apparent motives underlying male violence and homicide are very much in accord with predictions from evolutionary theory. Among the leading motives for murder among men in every culture is sexual jealousy. In some cultures, men are *expected* to kill other men who have sex with their wives (Symons, 1979), and in others, such murders are common even though they are illegal (Daly & Wilson, 1988). Men also fight over issues of status (Daly & Wilson, 1990). One man insults another and then the two fight it out—with fists, knives, or guns. And men, like many male monkeys and apes, often use violence to control females. In the United States and Canada between 25 and 30 percent of women are battered by their current or former mate at some time in their lives

>>> **39**
How is male violence toward infants, toward other males, and toward females explained from an evolutionary perspective?

Tough young males Male mammals of many species compete with one another for dominance. Much of their competition, however, involves threat and bluff rather than bloodshed, as illustrated by these two young mountain gorillas.



Michael Nichols / National Geographic Image Collection



(Randall & Haskell, 1995; Williams & Hawkins, 1989). Analyses of domestic violence cases indicate that much of it has to do with the man's belief (often unfounded) that his partner has been or might become sexually unfaithful (Peters & others, 2002; Wilson & Daly, 1993).

How Female Bonobos Dominate Males

40

How have bonobo females countered the mammalian tendency for males to control females through violence?

<<< Bonobos are an exception to the primate rule that males are the more violent sex. Bonobos, as you recall, are as closely related to us as are chimpanzees, and bonobos and chimpanzees are more closely related to each other than either species is to us (look back at Figure 3.21). Yet, in terms of their patterns of aggression, bonobos are very different from chimpanzees. They are much less aggressive than chimpanzees overall, and, to the degree that they are aggressive, females dominate males. It took researchers a long time to come to this conclusion, because it was so unexpected, but by now the researchers who are most familiar with bonobo behavior agree that female bonobos are aggressively dominant over males, despite the fact that the males are physically larger and stronger than the females (Kano, 1998; Parish & de Waal, 2000).

The females dominate because they form strong alliances with one another and help one another whenever one of them has a dispute with a male (Parish & de Waal, 2000). A single female bonobo generally cannot defeat a male in a fight, but a group of two or three can easily do so. When a valued but limited source of food is found by a colony of bonobos, the females eat first. A male who attempts to violate that rule is beaten off by the females as a group. If a male attempts to mount a female who does not want to be mounted, she screams and other females come to her rescue. A male who offends a female in any way—even one who offends by refusing a female's invitation to mount her—risks serious attack and possible injury at the hands and teeth of the female's friends (Parish & de Waal, 2000). Even the most dominant male is subject to such attacks.

In bonobos, a strong tendency for females to form close alliances to control males seems to derive largely from the genetic makeup of the species. It is seen among all bonobo groups that have been studied, both in the wild and in captivity. You might wonder why the males don't form alliances for counterattack; nobody knows the answer. Human females, by contrast, don't universally form alliances to control males, yet such alliances are within human capacity. In a cross-cultural analysis, Barbara Smuts (1992) showed that men's violence toward women is lowest in those cultures and communities in which women have strong alliances with one another and men have relatively weak alliances with one another. In those communities, women hear about and come to the aid of a woman who is being abused by a man, and, through humiliation or other means (usually not physical violence), teach the offending man a lesson. Other cross-cultural analyses reveal that women have relatively more control over economic resources in societies where alliances among women are strong than in societies where such alliances are weak (Yanca & Low, 2004).

Patterns of Helping

From an evolutionary perspective, **helping** can be defined as any behavior that increases the survival chance or reproductive capacity of another individual. Given this definition, it is useful to distinguish between two categories of helping: cooperation and altruism.

Cooperation occurs when an individual helps another while helping itself. This sort of helping happens all the time in the animal world and is easy to understand from an evolutionary perspective. It occurs when a mated pair of foxes work together to raise their young, a pack of wolves work together to kill an antelope, or a group of chimpanzees work together to chase off a predator or a rival group of chimpanzees. Most of the advantages of social living lie in cooperation. By working with others for common ends, each individual has a better chance of survival and

reproduction than it would have alone. Whatever costs accrue are more than repaid by the benefits. Human beings everywhere live in social groups and derive the benefits of cooperation. Those who live as our ancestors did cooperate in hunting and gathering food, caring for children, building dwellings, defending against predators and human invaders, and, most human of all, in exchanging, through language, information that bears on all aspects of the struggle for survival.

Altruism, in contrast, occurs when an individual helps another while *decreasing* its own survival chance or reproductive capacity. This is less common than cooperation, but many animals do behave in ways that at least appear to be altruistic. For example, some animals, including female ground squirrels, emit a loud, distinctive call when they spot an approaching predator. The cry warns others of the predator's approach and, at the same time, tends to attract the predator's attention to the caller (Sherman, 1977). (See Figure 3.22.) The selfish response would be to remain quiet and hidden, or to move away quietly, rather than risk being detected by warning others. How can such behavior be explained from an evolutionary perspective? As Trivers (1971) pointed out long ago, any evolutionary account of apparent altruism must operate by showing that from a broader perspective, focusing on the propagation of one's genes, the behavior is not truly altruistic. Evolutionists have developed two broad theories to account for ostensible altruism in animals: the kin selection theory and the reciprocity theory.

The Kin Selection Theory of Altruism

The **kin selection theory** holds that behavior that seems to be altruistic came about through natural selection because it preferentially helps close relatives, who are genetically most similar to the helper (Hamilton, 1964). What actually survives over evolutionary time, of course, is not the individual but the individual's genes. Any gene that promotes the production and preservation of copies of itself can be a fit gene, from the vantage point of natural selection, even if it reduces the survival chances of a particular carrier of the gene.

Imagine a ground squirrel with a rare gene that promotes the behavior of calling out when a predator is near. The mathematics of inheritance are such that, on average, one-half of the offspring or siblings of the individual with this gene would be expected to have the same gene, as would one-fourth of the nieces or nephews and one-eighth of the cousins. Thus if the altruist incurred a small risk (Δ) to its own life while increasing an offspring's or a sibling's chances of survival by more than 2Δ , a niece's or nephew's by more than 4Δ , or a cousin's by more than 8Δ , the gene would increase in the population from one generation to the next.

Many research studies have shown that animals do help kin more than non-kin. For example, ground squirrels living with kin are more likely to emit alarm calls than are those living with non-kin (Sherman, 1977). Chimpanzees and other primates are more likely to help kin than non-kin in all sorts of ways, including sharing food, providing assistance in fights, and helping take care of young (Goodall, 1986; Nishida, 1990; Silk, 2002). Consistent with the mathematics of genetic relatedness, Macaque monkeys have been observed to help their brothers and sisters more readily than their cousins and their cousins more readily than more distant relatives (Silk, 2002). For these and many other examples, helpers can apparently distinguish kin from non-kin, and this ability allows them to direct help selectively to kin (Pfennig & Sherman, 1995; Silk, 2002). In theory, however, altruistic behavior can evolve through kin selection even without such discrimination. A tendency to help any member of one's species, indiscriminately, can evolve if the animal's usual living arrangements are such that, by chance alone, a sufficiently high percentage of help is directed toward kin.

Among humans, the selective aiding of kin is called *nepotism*, and cross-cultural research shows that such behavior is common everywhere (Essock-Vitale & McGuire, 1980). If a mother dies or for other reasons is unable to care for a child, the child's grandmother, aunt, or other close relative is by far the most likely adopter (Kurland, 1979). Close kin are also most likely to share dwellings or land,

>>> **41**
How do the kin selection and reciprocity theories take the altruism out of "altruism"? What observations show that both theories apply to humans as well as to other animals?



L. D. Klein / Photo Researchers

FIGURE 3.22 | An alarm-calling ground squirrel When they spot a predator, female ground squirrels often emit an alarm call, especially if they are living in a group of close kin. Males are less likely to live near close kin and do not show this response.

hunt together, or form other collaborative arrangements. On the other side of the same coin, studies in Western culture indicate that genetic kin living in the same household are less often violent toward one another than are non-kin living in the same household (Daly & Wilson, 1988), and studies in other cultures have shown that villages in which most people are closely related have less internal friction than those in which people are less closely related (Chagnon, 1979). People report feeling emotionally closer to their kin than to their non-kin friends, even if they live farther away from the former than from the latter and see them less often (Neyer & Lang, 2003).

When leaders call for patriotic sacrifice or universal cooperation, they commonly employ kinship terms (Johnson, 1987). At times of war, political leaders ask citizens to fight for the “motherland” or “fatherland”; at other times, religious leaders and humanists strive to promote world peace by speaking of our “brothers and sisters” everywhere. The terms appeal to our instincts to be kind to relatives. Our imagination and intelligence allow us, at least sometimes, to extend our concept of kinship to all humanity.

The Reciprocity Theory of Apparent Altruism

The *reciprocity theory* provides an account of how acts of apparent altruism can arise even among non-kin. According to this theory, behaviors that seem to be altruistic are actually forms of long-term cooperation (Trivers, 1971). Computer simulations of evolution have shown that a genetically induced tendency to help non-kin can evolve if it is tempered by (a) an ability to remember which individuals have reciprocated such help in the past and (b) a tendency to refrain from helping again those who failed to reciprocate previous help (discussed in Chapter 14).

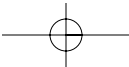
Under these conditions, helping another is selfish because it increases the chance of receiving help from that other in the future. Behavior fitting this pattern is found in various niches of the animal world. As one example, vampire bats frequently share food with unrelated members of their species that have shared food with them in the past (Wilkinson, 1988). As another example, bonobo females that establish friendship coalitions and are so very effective in helping one another are often unrelated to one another, having immigrated from different natal colonies (Kano, 1992; Parish & de Waal, 2000). The help each gives the others, in such acts as chasing off offending males, is reciprocated at another time.

The greatest reciprocal helpers of all, by far, are human beings. People in every culture feel a strong drive to return help that is given to them (Gouldner, 1960; Hill, 2002). Humans, more than any other species, can keep track of help given, remember it over a long period of time, and think of a wide variety of ways of reciprocating. Moreover, to ensure reciprocity, people everywhere have a highly developed sense of fairness and behave in ways that punish those who fail to fulfill their parts in reciprocal relationships (Fehr & Fischbacher, 2003). Certain human emotions seem to be well designed, by natural selection, to promote reciprocity. We feel gratitude (a sense of wanting to give) toward those who help us, pride when we return such help, guilt when we fail to return help, and anger when someone fails repeatedly to return help that we have given (as discussed more fully in Chapter 14). Humans also help others, including others who may never be able to reciprocate, in order to develop a good reputation in the community at large, and those with a good reputation are valued and helped by the community (Fehr & Fischbacher, 2003).



Gunter Ziesler / Peter Arnold, Inc.

Helpful little demons Vampire bats are gregarious mammals that demonstrate reciprocal altruism. After gorging itself on a blood meal, a bat will share some of what it has ingested with another bat, usually one that has fed it in the past.



Final Words of Caution: Two Fallacies to Avoid

Before closing this chapter it would be worthwhile to reiterate and expand on two points that I have already made: (1) Natural selection is not a moral force, and (2) our genes do not control our behavior in ways that are independent of the environment. These statements contradict two fallacious ideas that often creep into the evolutionary thinking of people who don't fully understand natural selection and the nature of genetic influences.

The Naturalistic Fallacy

The *naturalistic fallacy* is the equation of “natural” with “moral” or “right.” If male mammals in nature dominate females through force, then aggressive dominance of women by men is right. If natural selection promotes self-interested struggle among individuals, then selfishness is right. Such equations are logically indefensible, because nature itself is neither moral nor immoral except as judged so by us. Morality is a product of the human mind. We have acquired, in our evolution, the capacity to think in moral terms, and we can use that ability to develop moral philosophies that go in many possible directions, including those that favor real altruism and constrain individual self-interest for the good of the larger community.

The term *naturalistic fallacy* was coined by the British philosopher G. E. Moore (1903) as part of an argument against the views of another British philosopher, Herbert Spencer (1879). A contemporary of Darwin, Spencer considered himself a strong believer in Darwin's theory, and his goal was to apply it to the spheres of social philosophy and ethics. Unlike Darwin, Spencer seemed to imply in his writings that natural selection is guided by a moral force. He distinguished between “more evolved” and “less evolved” behaviors in nature and suggested that “more evolved” meant more moral. Although Spencer discussed cooperation as highly evolved and virtuous, his philosophy leaned more toward the virtues of individualism and competition. Spencer's writings were especially popular in the United States, where they were championed by such industrialists as John D. Rockefeller and Andrew Carnegie (Rachels, 1990). It was Spencer, not Darwin, who popularized the phrase “survival of the fittest,” and some of the so-called *social Darwinists*, who were inspired by Spencer, used that phrase to justify even the most ruthless extremes of capitalism. In their view, the fittest were those who rose to the top in unchecked capitalism, and the unfit were those who fell into poverty or starvation.

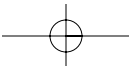
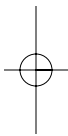
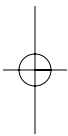
Darwin himself was not seduced by the naturalistic fallacy. He was repulsed by much of what he saw in nature and marveled at the human ability to rise, sometimes, above it. He conscientiously avoided phrases such as “more evolved” that would imply that evolution is a moral force, and he felt frustrated by his inability to stop others from assuming that it is. In a letter to a friend, shortly after publication of *On the Origin of Species*, he wrote wryly, “I have noted in a Manchester newspaper a rather good squib, showing that I have proved ‘might is right’ and therefore Napoleon is right and every cheating tradesman is also right” (Rachels, 1990).

The Deterministic Fallacy

The second general error, called the *deterministic fallacy*, is the assumption that genetic influences on our behavior take the form of genetic control of our behavior, which we can do nothing about (short of modifying our genes). The mistake here is assuming or implying that genes influence behavior directly, rather than through the indirect means of working with the environment to build or modify biological structures that then, in interplay with the environment, produce behavior. Some popular books on human evolution have exhibited the deterministic fallacy by implying that one or another form of behavior—such as fighting for territories—is unavoidable because it is controlled by our genes. That implication is unreasonable even when applied to nonhuman animals. Territorial birds, for example, defend territories only when the environmental conditions are ripe for them to do so. We humans can control our environment and thereby control ourselves. We can

>>> **42**
Why is the equation of “natural” with “right” considered a fallacy? How did that fallacy figure into the philosophy of Herbert Spencer?

>>> **43**
Why is it a mistake to believe that characteristics that are influenced by genes cannot be changed except by modifying genes?



either enhance or reduce the environmental ingredients needed for a particular behavioral tendency to develop and manifest itself.

We also can and regularly do, through conscious self-control and well-learned social habits, behave in ways that are contrary to biases built into our biology. One might even argue that our capacity for self-control is the essence of our humanity. In our evolution, we acquired that ability in a greater dose than seems apparent in any other species, perhaps because of its role in permitting us to live in complex social groups. Our ability for self-control, itself, is part of our biological heritage, and it liberates us to some degree—but by no means completely—from that heritage.

Our great capacity for knowledge, including knowledge of our own biological nature, can also be a liberating force. Most evolutionary psychologists contend that an understanding of human nature, far from implying fatalism, can be a step toward human betterment. For example, in her evolutionary analysis of men's violence against women, Smuts (1992) wrote:

Although an evolutionary analysis assumes that male aggression against women reflects selection pressures operating during our species' evolutionary history, it in no way implies that male domination of women is genetically determined, or that frequent male aggression toward women is an immutable feature of human nature. In some societies male aggressive coercion of women is very rare, and even in societies with frequent male aggression toward women, some men do not show these behaviors. Thus, the challenge is to identify the situational factors that predispose members of a particular society toward or away from the use of sexual aggression. I argue that an evolutionary framework can be very useful in this regard.

As you review all of the human species-typical behaviors and tendencies discussed in this chapter, you might consider how each depends on environmental conditions and is modifiable by changes in those conditions.

SECTION REVIEW

An evolutionary perspective offers functionalist explanations of aggression and helping.		
Male Violence - Male primates, including men, are generally more violent than are females of their species. - Most aggression and violence in male primates relate directly or indirectly to sex. Genes that promote violence are passed to offspring to the degree that they increase mating. - Among bonobos, female alliances counter male violence.	Helping - Helping (promoting another's survival or reproduction) takes two forms—cooperation and altruism. - Cooperation (helping others while also helping oneself, as in the case of wolves hunting together) is easy to understand evolutionarily. - Apparent acts of altruism (helping others at a net cost to oneself) make evolutionary sense if explained by the kin selection or reciprocity theories.	Common Fallacies - The naturalistic fallacy is the equation of what is natural with what is right. It was indulged in by indulged in by Social Darwinists. - The deterministic fallacy is the belief that genes control behavior in ways that cannot be altered by environmental experiences or conscious decisions.

Concluding Thoughts

1. The indirect nature of genetic influences on behavior Genes are simply DNA molecules that provide the code for building the body's proteins. Variation in genes across species provides the basis for species-typical behaviors, and variation in genes among members of a species is one source of individual differences in behavior within a species. But genes never produce behaviors directly. Genes always work in conjunction with the environment, and so their effects depend on environmental conditions. Neither genes nor environment “deter-

mines” our behavior. Our behavior results from an interplay between the environment in which we live and our bodies' biological mechanisms, which themselves were built through an interplay between genes and environment.

2. The unconscious nature of ultimate functions Sigmund Freud (discussed in Chapters 15 and 17) is famous for his claim that we are often unconscious of the real reasons for our actions. On that point, at least, modern evolutionary psycholo-

gists and Freud agree. Our species-typical drives and behavioral tendencies evolved to promote functions related to survival and reproduction, but we rarely think of those functions, and we are often completely ignorant of them. Toddlers toddle and play with words because it is “fun” to do so, without any thought about the value of such play in learning to walk and talk. We all smile, instinctively or because it seems like the right thing to do, when we are happy or when we meet someone, without thinking about the functions that smiling might serve. When we fall in love, we are far more likely to attribute that feeling to the sweet, charming, and irresistible nature of the beloved person than to anything having to do with the value of bonding for producing and raising children. When we feel jealous because of attention another is paying to our beloved, we think angry thoughts about betrayal and unfaithfulness, not about the role of jealousy in preserving monogamy. The reasons we give ourselves for what we do are an aspect of the proximate causation of our behavior. We are often no more aware of the ultimate functions of our actions than the cabbage fly is of why it is irresistibly drawn to the cabbage plant as the only proper place to lay its eggs.

3. Evolution as an integrative theme in psychology The evolutionary perspective provides the broadest view that we can take in psychology. It is concerned with the origins and ultimate functions of all aspects of human nature (and the nature of other animals). It is a perspective that can be applied to the whole vast range of topics with which psychology is concerned. All of the complex biological mechanisms that underlie our psychological nature came about because they helped our ancestors to survive and reproduce. We can expect, therefore, that all of our basic motivational and emotional mechanisms are biased toward generating behaviors that promote survival and reproduction; and we can expect that our sensory, perceptual, memory, and reasoning mechanisms are biased toward picking up and using information essential to those purposes. We are not general learning or thinking machines that indiscriminately analyze all information available; we are biological survival machines designed to use information selectively to achieve our ends. As you go through the rest of this book, crossing the whole range of psychology, you will see this idea applied in every chapter.

Furthur Reading

CHARLES DARWIN (1859; reprinted 1963). *The origin of species*. New York: Washington Square Press.

Darwin was an engaging writer as well as a brilliant thinker. Why not read at least part of this book, which revolutionized the intellectual world? The most relevant chapter for psychologists is Chapter 8, entitled “Instinct,” which includes Darwin’s research on hive building in bees and many other insights about the behavior of wild and domesticated animals.

MATT RIDLEY (2003). *Nature via nurture*. New York: HarperCollins.

The recent explosion of research on human genetics—including the full sequencing of the human genome—is greatly increasing our knowledge of the genetic differences and similarities between humans and other animals and of the pathways by which genes influence behavior. Ridley summarizes, in highly readable and well-documented form, some of the most dramatic recent discoveries. As the title implies, the more

we know about genes the more we realize that nature (genetic influence) and nurture (environmental influence) are inextricably entwined.

BOBBI S. LOW (2000). *Why sex matters: A Darwinian look at human behavior*. Princeton, NJ: Princeton University Press.

Low is an anthropologist who has conducted wide-ranging research into sex differences in animals and people, the latter across cultures and historical time. This scholarly but highly readable book is about sex, the differences between the sexes, and all that relates to these—beauty, bonding, reproduction, parenting, division of labor, cooperation, competition, wealth, power, and politics. It’s an excellent introduction to evolutionary psychology from an anthropological perspective.

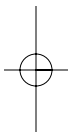
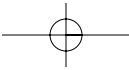
JANE GOODALL (1988). *In the shadow of man* (rev. ed.). Boston: Houghton Mifflin. Goodall’s study of wild chimpanzees, which began in 1960, ranks as one of the most courageous and scientifically valu-

able studies of animal behavior ever undertaken. This book, first published in 1971, is an exciting account of her early struggle to locate and study the animals and of her early findings about their behavior. For a more complete and scientific account of her research, I recommend Goodall’s 1986 book, *The Chimpanzees of Gombe*.

FRANS DE WAAL, WITH PHOTOGRAPHS BY FRANS LANTING (1997). *Bonobo: The forgotten ape*. Berkeley: University of California Press.

Bonobos—the apes that are tied with chimpanzees as our closest animal relatives—were rediscovered in the 1980s, after decades of scientific neglect. Among the scientists who have studied them intensively is Frans de Waal, and here he describes their ecology and habits in a narrative that is accompanied by dozens of full-color, full-page photographs of these endangered apes, famous for their make-love-not-war style of life. Chapter 4 is X-rated.

FPO



Basic Processes of Learning

CLASSICAL CONDITIONING I: FUNDAMENTALS

CLASSICAL CONDITIONING II: BEYOND THE FUNDAMENTALS

- Relevance of Pavlov's Work to the Emergence of Behaviorism
- What Is Learned in Classical Conditioning?
- Conditioned Fear, Hunger, and Sexual Arousal
- Conditioned Drug Reactions

OPERANT CONDITIONING I: FUNDAMENTALS

- From the Law of Effect to Operant Conditioning: From Thorndike to Skinner
- Principles of Reinforcement

OPERANT CONDITIONING II: WHAT IS LEARNED?

- Learning When a Response Will Be Rewarded
- Learning to Expect a Certain Reward

ACTIVITIES DESIGNED FOR LEARNING: PLAY, EXPLORATION, AND OBSERVATION

- Play as a Vehicle for Learning How to Do Things
- Exploration as Vehicle for Learning About the Environment
- Learning by Watching Others

SPECIALIZED LEARNING ABILITIES: FILLING THE BLANKS IN SPECIES- TYPICAL BEHAVIOR PATTERNS

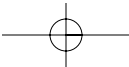
- Special Abilities for Learning What to Eat
- Other Examples of Special Learning Abilities

To survive, animals must adapt to their environments. Evolution by natural selection, discussed in Chapter 3, is the long-term adaptive process that equips each species for life within a certain range of environmental conditions. But the environment is never constant; it changes from place to place and from time to time, even during short periods of an individual's life. To be efficient in finding foods, finding mates, avoiding predators, and carrying out the other necessities of survival and reproduction, animals must adapt to the ever-changing conditions of the specific environments in which they live. In other words, they must learn.

The term *learning* is used in various ways by different psychologists to refer to a wide variety of phenomena. For our purposes, however, we can define it broadly as *any process through which experience at one time can alter an individual's behavior at a future time*. *Experience* in this definition refers to any effects of the environment that are mediated by the individual's sensory systems (vision, hearing, touch, and so on). *Behavior at a future time* refers to any subsequent behavior that is not part of the individual's immediate response to the sensory stimulation during the learning experience. If I make a clicking sound just before flashing a bright light into your eyes, your immediate response to the click or to the light (such as blinking) does not exemplify learning. Your increased tendency to blink to the click alone, the next time I present that sound, however, does exemplify learning.

Most of psychology is in one way or another concerned with the effects of experience on subsequent behavior. Social psychologists try to explain people's beliefs and social behaviors in terms of their past experiences. Clinical psychologists try to explain people's emotional problems in terms of their past experiences. Cognitive psychologists try to understand the basic processes of perception, memory, and thought that are involved in people's ability to learn. Thus most of the chapters in this book, or in any other introductory psychology text, are in one way or another about learning.

In this chapter we are still in the part of the book that links human psychology to the psychology of animals in general. From an evolutionary perspective, learning is a quite ancient set of abilities. All animals that have any kind of nervous system have acquired, through natural selection, some abilities to learn. We humans are in some ways unique but in many ways similar to other species in our basic mechanisms of learning. You will read more about our unique learning abilities—such as our ability to learn a grammar-based language—later in this book. Our present focus is on mechanisms of learning that characterize mammals (and other vertebrates) in general, including humans. We start with a very general variety of learning referred to as classical conditioning.



Classical Conditioning I: Fundamentals

1

What is a reflex, and how can it change through habituation?

<<< Classical conditioning is a learning process that has to do with the formation of new reflexes. A **reflex** is a simple, relatively automatic, stimulus-response sequence mediated by the nervous system. If your knee is tapped with a rubber mallet, your leg will jerk. If a bright light is flashed in your eyes, you will blink. If lemon juice is squirted into your mouth, you will salivate. If a loud alarm suddenly clangs, your muscles will tighten. In each of these examples, a particular well-defined event in the environment, a **stimulus**, results in a particular well-defined bit of behavior, a **response**. The tap on the knee, the flash of light, the squirt of lemon juice, and the sudden alarm are stimuli (note that *stimuli* is the plural form of *stimulus*). The leg jerk, the eye blink, the salivation, and the muscle tightening are responses.

To be considered a reflex, the response to a stimulus must be mediated by the nervous system. Messages carried by nerves from the eyes, ears, or other sensory organs enter the spinal cord or brain and act there to produce messages in nerves running outward to muscles and glands. If something hits you and you fall down as a result of the direct force of the impact, that is not a reflex. If something hits you and your muscles respond in a way that tends to keep you from falling down, that is a reflex. Because reflexes are mediated by the nervous system, they can be modified by experience.

One simple effect of experience on reflexes is **habituation**, defined as a decline in the magnitude of a reflexive response when the stimulus is repeated several times in succession. Not all reflexes undergo habituation. One that does is the startle response to a loud sound. You might jump the first time the sound occurs, but each time the sound is repeated you respond less, and soon show no visible response at all. Habituation is one of the simplest forms of learning. It does not produce a new stimulus-response sequence but only weakens an already existing one. Classical conditioning, in contrast, is a form of reflex learning that does produce a new stimulus-response sequence. It was first described and most extensively studied by a Russian physiologist, Ivan Pavlov.

Pavlov's Initial Discovery of Classical Conditioning

Ivan Petrovich Pavlov (1849–1936) was the personification of the dedicated scientist. By the time of his most famous research on classical conditioning, he was in his fifties and had already earned a Nobel prize for studies of the reflexes involved in digestion. His research so engulfed his life that he is said to have hardly noticed such events as the Bolshevik Revolution of 1917, which radically transformed his country. One former co-worker (Gantt, in 1975, as quoted by Hothersall, 1995) recalled Pavlov's angry scolding of an assistant who arrived 10 minutes late to start an experiment: "But Professor," exclaimed the assistant, "there's a revolution going on, with shooting in the streets." To which Pavlov replied, "What the ____ difference does that make when you've work to do in the laboratory? Next time there's a revolution, get up earlier!"

2

How did Pavlov discover the conditioned reflex?

<<< Pavlov's initial discovery of what we now call classical conditioning emerged from his earlier studies of digestive reflexes in dogs. Using permanently implanted tubes to collect salivary and stomach juices from dogs (see Figure 4.1), he and his team of researchers found, for example, that a dog salivates differently when different kinds of food are placed in its mouth. Juicy meat triggers a very thick saliva; dry bread, a wetter saliva; and acidic fluids, a wetter one yet. In a fine-grained analysis, these represent three different reflexes, with three different stimuli eliciting measurably different salivary secretions.

In the course of these studies, Pavlov encountered a problem. Dogs that had been given food on previous occasions in Pavlov's experiments would salivate before they received any food. Apparently, signals that regularly preceded food, such as the sight of the food or the sound associated with its delivery, alerted the dogs to

the upcoming stimulation and caused them to salivate. At first Pavlov was content to treat this simply as a source of experimental error. He called it “psychic secretion,” implying that it was outside the physiologist’s realm of study, and he attempted to eliminate it by developing ways to introduce the food into the dog’s mouth without any warning. But then it occurred to Pavlov that this might well be a phenomenon that could be studied physiologically. Rather than call it psychic secretion, perhaps he could consider it a reflex and analyze it objectively, just as he had analyzed the reflexive salivary response to food in the mouth. This insight led Pavlov (1927/1960) to his first experiments on conditioned reflexes.

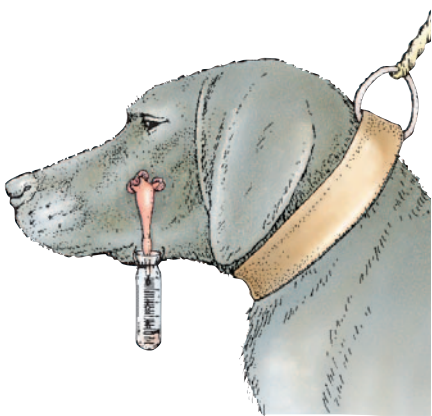


FIGURE 4.1 Pavlov’s method for measuring salivation One of the dog’s salivary ducts is surgically connected to a glass tube. In his early experiments, Pavlov learned that dogs produce different salivary secretions in response to different kinds of food. Later he found that the dogs could be conditioned to produce these secretions in response to stimuli that reliably precede the presentation of food. (Adapted from Yerkes & Morgulis, 1909.)

The Procedure and Generality of Classical Conditioning

To study such reflexes, Pavlov deliberately controlled the signals that preceded food. In one experiment he sounded a bell just before placing food in the dog’s mouth. After several such pairings of a bell with food, the dog would salivate in response to the bell sound alone; no food was necessary. Pavlov referred to this new reflex as a **conditioned reflex** because it depended on the unique conditions present in the dog’s previous experience—the pairing of the bell sound with the food-in-mouth stimulus. He referred to the stimulus in a conditioned reflex (the bell sound, in this case) as a **conditioned stimulus** and to the learned response to it (salivation) as a **conditioned response**. Likewise, he referred to the original, unlearned reflex as an **unconditioned reflex** and to its stimulus (food placed in the mouth) and response (salivation) as an **unconditioned stimulus** and **unconditioned response**. For a diagram of Pavlov’s basic procedure, called **classical conditioning**, see Figure 4.2.

Pavlov (1927/1960) was impressed in these studies by the similarity between the dog’s salivary response to a conditioned stimulus and its response to the unconditioned stimulus. A sound that had been paired with meat elicited a thick saliva, similar to that elicited by meat; and a sound that had been paired with bread elicited a thinner, wetter saliva, similar to that elicited by bread. In other experiments, Pavlov and his colleagues varied the stimulus used as the conditioned stimulus. They concluded that essentially any environmental event that the animal could detect could become a conditioned stimulus for salivation. Sounds produced by bells, buzzers, or tuning forks were highly effective and used most often because

After his initial discovery, how did Pavlov systematize the process of conditioning, and what names did he give to the relevant stimuli and responses?

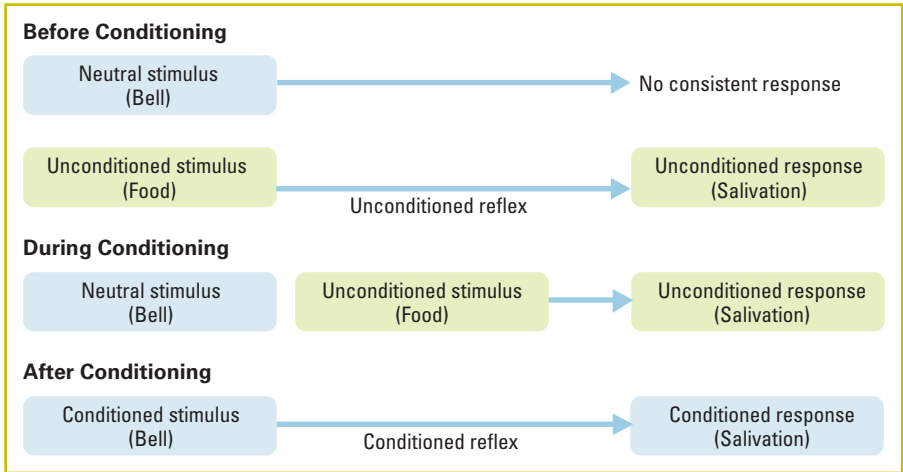
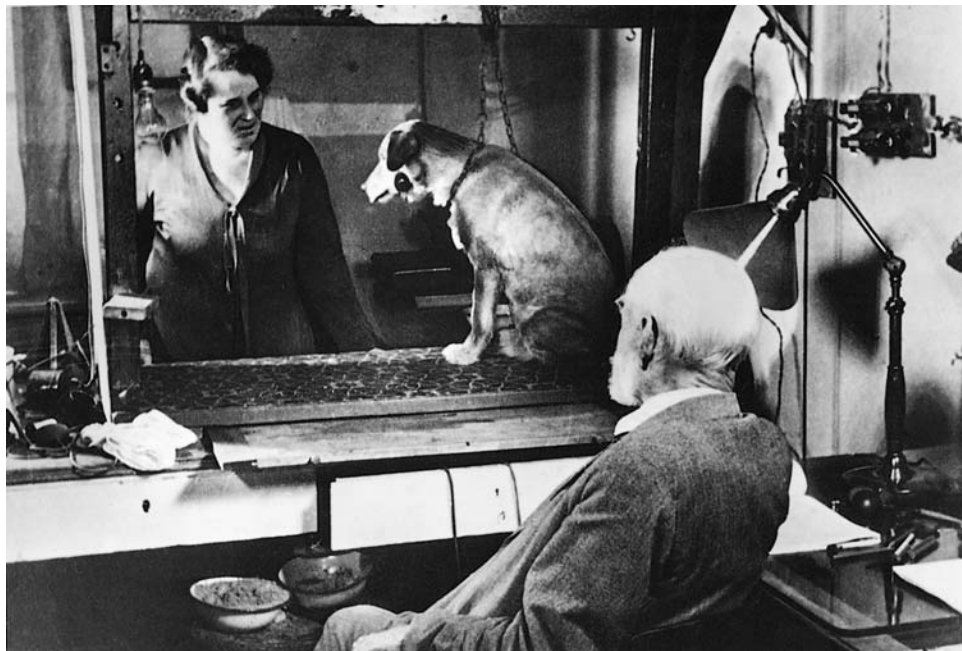


FIGURE 4.2 Classical-conditioning procedure A neutral stimulus initially does not elicit a response. After it is paired for several trials with an unconditioned stimulus, however, it becomes a conditioned stimulus and does elicit a response.

Pavlov conducting an experiment
To conduct his pioneering research on conditioning, Pavlov developed a means to control the presentation of stimuli, such as the sound of a bell and the presentation of food, and to measure the dog's reflexive salivary response.



Sovfoto

they were the easiest to control. But Pavlov's group also produced conditioned responses to visual stimuli, such as a black square; to olfactory stimuli, such as the odor of camphor; and to tactile (touch) stimuli, such as pressure applied to a spot on the animal's skin. In each case, the stimulus initially did not elicit the salivary response, but it did so after having been paired with food a number of times.

Of course, classical conditioning is not limited to salivary responses. Researchers have demonstrated this in hundreds of laboratory experiments, and you have experienced countless examples of such conditioning in your everyday life. The sound of a dentist's drill may elicit a conditioned cringing response because of its previous pairing with pain. The mere smell of coffee may help wake you up because of its previous pairing with coffee's stimulating effect. The sight of the toilet when you enter a bathroom to comb your hair may elicit a previously unfelt urge to urinate due to previous pairings of that sight with that urge. If you once had an automobile accident at a curve in the road on a wet day, each new encounter with such a curve on such a day may elicit a conditioned tensing of muscles. If you go through a day recording instances of conditioned responses, you will find that the list quickly becomes quite long.

Pavlov and scientists following him performed hundreds of experiments on classical conditioning and identified many phenomena related to it, including the following.

Extinction of Conditioned Reflexes, and Recovery from Extinction

4

How can a conditioned reflex be extinguished? What evidence led Pavlov to conclude that extinction does not return the animal to its original, untrained state?

<<< One question that interested Pavlov had to do with the permanence of a conditioned reflex. Once a dog has learned to salivate to a bell, does this reflex continue to occur if the bell is sounded for many trials without the unconditioned food-in-mouth stimulus? Pavlov's group found that without food, the bell elicited less and less salivation on each trial and eventually none at all, a phenomenon that they labeled **extinction**. But they also found that extinction does not return the animal fully to the unconditioned state. The mere passage of time following extinction can partially renew the conditioned reflex, a phenomenon now known as **spontaneous recovery** (see Figure 4.3). Moreover, a single pairing of the conditioned stimulus with the unconditioned stimulus can fully renew the conditioned reflex, which can be extinguished again only by another series of trials in which the conditioned stimulus is presented without the unconditioned stimulus. On the basis of such

findings, Pavlov (1927/1960) concluded that the conditioned reflex is not truly lost during extinction, but is somehow inhibited, and that it can be disinhibited by such means as the passage of time or the recurrence of the unconditioned stimulus. This conclusion has been validated in many experiments since Pavlov's time (Bouton, 1994).

Recently, in research with conditioned eye-blink reflexes in rabbits, researchers have shown that conditioning and extinction involve different sets of neurons (nerve cells) in the brain (Medina & others, 2002). Neurons involved in the conditioning of this response excite neurons that promote eye blinking. Neurons involved in extinction inhibit those same eye-blink neurons, thereby counteracting but not abolishing the conditioned neural response, just as Pavlov would have predicted.

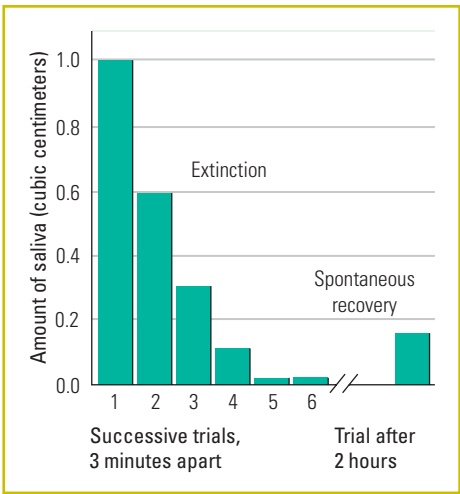


FIGURE 4.3 Extinction and spontaneous recovery of conditioned salivation The conditioned stimulus in this case was the sight of meat powder, presented repeatedly out of the animal's reach at 3-minute intervals. Extinction was complete by the fifth and sixth presentations, but when 2 hours were allowed to elapse before the seventh presentation, the reflex was partially renewed. (Data from Pavlov, 1927/1960.)

Generalization and Discrimination in Classical Conditioning

Pavlov and his colleagues found that after conditioning, animals would show the conditioned response not just to the original conditioned stimulus, but also to new stimuli that resembled that stimulus. This phenomenon is called **generalization**. They found, further, that the magnitude of the response to the new stimulus depended on the degree of similarity between the new stimulus and the original conditioned stimulus. Thus a dog conditioned to salivate to a 1,000-hertz (cycles-per-second) tone also salivated to tones of other frequencies, but the further the tone was in frequency from the original conditioned stimulus, the less the dog salivated to it (Pavlov, 1927/1960).

Generalization between two stimuli can be abolished if the response to one is reinforced while the response to the other is extinguished, a procedure called **discrimination training**. As an example of this procedure, Pavlov's group used a dog whose conditioning to the sight of a black square had generalized to a gray square. After a series of trials in which presentations of the gray square were never followed by food and presentations of the black square were always followed by food, the dog stopped salivating to the gray square but continued to salivate to the black one. The researchers continued this procedure with ever-darker shades of gray, until they eventually conditioned the dog to discriminate a black square from a gray one that was so nearly black that a human observer had difficulty telling them apart (Pavlov, 1927/1960).

Classical conditioning coupled with discrimination training provides an excellent tool for studying an animal's sensory capacities. A dog cannot tell you what it can or cannot hear, but you can find out by doing a conditioning experiment. If an animal can be conditioned to respond to a stimulus, we know that it can sense the stimulus; and if an animal can be trained to respond to one stimulus and not to another, we know that it can sense the difference between the two. Pavlov's team conditioned dogs to salivate to tones so high-pitched as to be inaudible to humans and to discriminate between pitches less than one-eighth of a note apart (Pavlov, 1927/1960).

Generalization as an Index of Subjective Similarity

In experiments with human subjects, researchers following Pavlov discovered that generalization occurs not just when two stimuli are physically similar to one another but also when they are similar in their subjective meaning to the person.

In one experiment, for example, Gregory Razran (1939) used college students as subjects, printed words as conditioned stimuli, and a squirt of lemon juice into the

5 How can generalization in classical conditioning be abolished through discrimination training? How can discrimination training be used to assess an animal's sensory capacities?

6 How have researchers shown that the meaning of a stimulus, not just its physical characteristics, can provide a basis for generalization in classical conditioning?

mouth as the unconditioned stimulus. By pairing each word with lemon juice, Razran conditioned students to salivate to the words *style*, *urn*, *freeze*, and *surf*. He then tested the students to see if the conditioned response would generalize to other words that had never been paired with lemon juice. Of most interest, he found that the students salivated more to the words *fashion*, *vase*, *chill*, and *wave* than to the words *stile*, *earn*, *frieze*, and *serf*. That is, the conditioned response generalized more to printed words that resembled the original conditioned stimuli in meaning than to those that resembled the originals in physical appearance or sound. Thus the true conditioned stimuli, in this case, were not the physical sights or sounds of the words but the subjects' interpretations of them.

An even more dramatic demonstration of stimulus generalization based on word meaning comes from the work of a Soviet psychologist (Volkova, 1953). A Russian schoolboy was conditioned to salivate to the Russian word for *good* and, by discrimination training, not to salivate to the word for *bad*. When the discrimination was well established, the boy was found to salivate copiously to such statements as *The Soviet army was victorious* and not at all to statements such as *The pupil was rude to the teacher*. In other words, those statements that the boy interpreted as good elicited salivation, and those that he interpreted as bad did not. A child who had a different set of conceptions of what is good or bad might have shown the opposite pattern of salivation. Thus, although classical conditioning is a very primitive phenomenon, it can be applied even toward understanding such uniquely human characteristics as a person's attitudes about particular, verbally described events.

SECTION REVIEW

Classical conditioning is the process involved in learning a new reflex.		
Conditioning Procedure • In classical conditioning one begins with an unconditioned reflex (e.g., food → salivation). • During conditioning, a conditioned stimulus (e.g., bell sound) is presented just before the unconditioned stimulus (e.g., food). • After sufficient pairings, the conditioned stimulus alone elicits a conditioned response. This is the conditioned reflex (e.g., bell sound → salivation).	Extinction • If a conditioned stimulus is repeatedly presented without the unconditioned stimulus, the conditioned response stops occurring. • The conditioned reflex is not unlearned in extinction; it is merely suppressed, as shown by spontaneous recovery, etc. • Conditioning and extinction apparently involve different sets of neurons, one promoting the conditioned response, the other inhibiting it.	Generalization and Discrimination • Generalization occurs when a stimulus similar to the conditioned stimulus also elicits the conditioned response. • Repeated presentation of the conditioned stimulus followed by the unconditioned stimulus, and of the similar stimulus followed by nothing, reduces generalization—that is, it promotes discrimination. • Discrimination training is useful in studying sensory abilities in animals. • Word meaning can provide a basis for generalization in humans.

Classical Conditioning II:
Beyond the Fundamentals

Having surveyed some of the basics of classical conditioning, let's now examine the process a little more deeply. What is actually learned in classical conditioning? What are some examples of conditioning that show its relevance to practical issues in psychology? Before considering these questions, it would be useful for you to know a little bit about an early school of thought in psychology referred to as behaviorism.

Relevance of Pavlov’s Work to the Emergence of Behaviorism

Early in the twentieth century, psychologists in North America were struggling to develop scientific methods appropriate for psychology. Some of these psychologists, who referred to their school of thought as **behaviorism**, argued that their science should avoid terms that refer to mental entities, such as thoughts, emotions, motives, and so on, because such entities cannot be directly observed. They believed that psychology should focus on the relationship between observable events in the environment (stimuli) and observable behavioral reactions to those events (responses).

The principal founder of behaviorism, John B. Watson (1913), put it this way: “In a system of psychology completely worked out, given the response the stimuli can be predicted, and given the stimuli the response can be predicted.” In Watson’s view, all of behavior is, in essence, reflex-like in nature. Neither Watson nor subsequent behaviorists denied the existence of mental processes, but they believed that these are too obscure to be studied scientifically, and they argued that behavior could be understood and described without reference to mental processes.

In addition to developing objective, stimulus-response descriptions of behavior, the early behaviorists established learning as their main explanatory concept. They maintained that a person’s behavior at any given time is determined by that person’s past experiences with the environment. Watson and the other early behaviorists were much inspired by Pavlov’s work on conditioning. In contrast to the empiricist philosophers (discussed in Chapter 1), who talked about learning in terms of unseen associations occurring in the mind, Pavlov seemed to provide an objective, stimulus-response way of studying and understanding learning. If all behavior is essentially reflexive, and if most behavior is learned, and if conditioning is the process by which reflexes are learned, then conditioning would appear to be psychology’s most basic explanatory concept. In one of his early books, Watson (1924) attempted to describe even complex examples of human learning in terms of what we now call classical conditioning.

7
What were the characteristics of early, North American behaviorism? Why were Pavlov’s findings on conditioning particularly appealing to behaviorists?

What Is Learned in Classical Conditioning?

We now return to the question raised earlier: What, really, is learned in classical conditioning? Watson’s (1924) and other early behaviorists’ answer to that question was simply that a new stimulus-response connection is learned. From their perspective, Pavlov’s conditioned dog salivated to the bell because of a direct, learned connection between the bell and salivation. This *stimulus-response* (S-R) theory of classical conditioning is diagrammed in the top part of Figure 4.4. Pavlov (1927/1960) himself, however, had a different theory.

Classical Conditioning Interpreted as Stimulus-Stimulus Association

Pavlov believed that the animal does not learn a direct stimulus-response connection but, rather, learns a connection between two stimuli, the conditioned stimulus and the unconditioned stimulus. Because the bell and food have been paired in past experience, a physiological bond is formed between their representations in the brain, such that the sound of the bell now activates the part of the brain that was formerly activated by food, and that, in turn, elicits salivation. Using mental rather than physiological terms, we could say that the dog salivates to the bell because the bell sound elicits in the dog a mental representation of food (bell → mental representation of food → salivation). This *stimulus-stimulus* (S-S) theory of classical conditioning is illustrated in the bottom part of Figure 4.4.

The S-S theory did not appeal to Watson and his followers because it posited the existence of an unobserved event in the animal’s mind, the mental representation

8
How did Pavlov’s S-S theory of classical conditioning differ from Watson’s S-R theory? How does an experiment involving habituation of the unconditioned stimulus support the S-S theory?

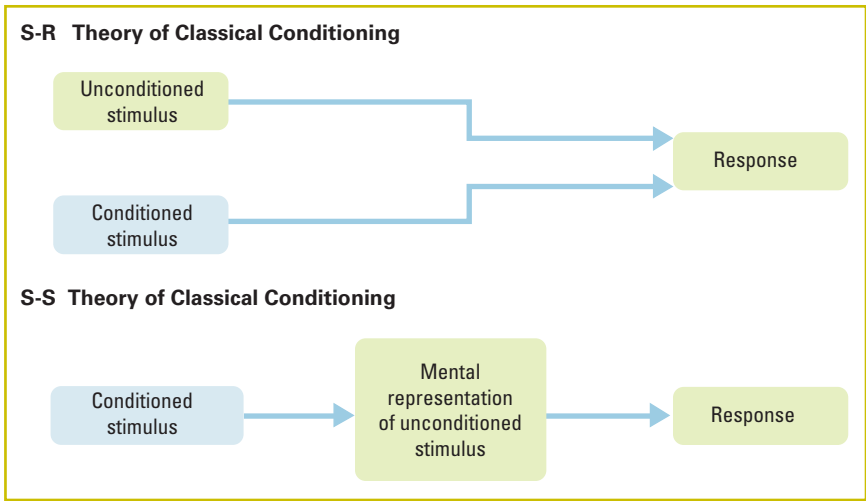


FIGURE 4.4 | Comparison of S-R and S-S theories of classical conditioning According to the S-R theory, conditioning produces a direct bond between the conditioned stimulus and the response. According to the S-S theory, conditioning produces a bond between the conditioned stimulus and a mental representation of the unconditioned stimulus, which, in turn, produces the response. Support for the S-S theory comes from experiments showing that weakening the unconditioned reflex (through habituation), after conditioning, also weakens the conditioned reflex.

of the original unconditioned stimulus. However, contrary to the early behaviorists' claims, reference to unseen events can be very useful in science if they lead to clear-cut predictions about events that can be seen. Gravity cannot be seen, yet theories of gravity can be developed and tested with experiments on falling objects. The S-S and S-R theories make different predictions about how animals will behave under certain conditions, so it is possible to test the theories with experiments. Many such experiments have by now been conducted, and the weight of the evidence favors the S-S theory for most examples of conditioning in mammals and birds (Anderson, 2000).

As an illustration, consider an experiment conducted by Robert Rescorla (1973) using rats as subjects, a loud sound as the unconditioned stimulus, and a signal light as the conditioned stimulus. The loud sound elicited freezing (a fear response in which the rat stands motionless) as an unconditioned response. By pairing the signal light with the loud sound, Rescorla conditioned rats to freeze when the signal light came on. Now, the question was: Did the rats freeze in response to the signal light because of a direct, learned connection between the light and freezing, in accordance with the S-R theory (light → freezing)? Or did they freeze because of a learned connection between the light and the loud sound, in accordance with the S-S theory (light → mental representation of loud sound → freezing)?

To answer the question, Rescorla habituated the response to the loud sound in half of the conditioned rats. That is, he presented the loud sound many times without the signal light until the rats no longer froze in response to it. Then he again tested the rats with the signal light. Would the rats that no longer froze in response to the loud sound continue to freeze in response to the light? The S-R and S-S theories make different predictions. According to the S-R theory, the habituated rats should continue to freeze in response to the light because conditioning would have produced a direct connection between the light and freezing. But according to the S-S theory, the habituated rats should not freeze in response to the light because conditioning would have produced a connection between the light and a representation of the loud sound, which itself no longer elicits freezing. Rescorla's results supported the S-S theory. Habituation to the sound greatly reduced the degree to which the rats froze in response to the light.

Classical Conditioning Interpreted as Learned Expectancy

The S-S theory of classical conditioning, by its nature, is more cognitive than the S-R theory. As explained in Chapter 1, the term *cognition* refers to knowledge, or information, held in the mind. Cognitive theories, contrary to the dictates of behaviorism, are theories in which hypothesized mental entities are used to explain and predict behavior. The S-S theory is cognitive because it holds that the observed stimulus-response relation is mediated by an inner, mental representation of the original unconditioned stimulus. Cognitive theorists have argued that this mental representation may be best understood as an *expectation* of the unconditioned stimulus. According to this view, Pavlov's dog learned to expect food when it heard the bell.

9 How does the cognitive construct of expectancy help explain the ways in which conditioned responses differ from unconditioned responses?

<<< The *expectancy theory* helps make sense of the observation that a conditioned response is often quite different from the unconditioned response. Consider again a dog being conditioned to a bell that precedes food. In response to food, the dog not only salivates but also chews (if it is solid food) and swallows. Salivation becomes conditioned to the bell, but chewing and swallowing usually do not. Moreover, the bell comes to elicit not only salivation but also responses that do not usually occur

in response to the food-in-mouth stimulus—such as tail wagging, food begging, and looking in the direction of the usual source of food (Jenkins & others, 1978). According to the expectancy theory, all these responses, including salivation, occur not because they were previously elicited by the unconditioned stimulus but because they are the dog's responses to the *expectation* of food:

bell → expectation of food → tail wagging, food begging, salivation, etc.

Rescorla (1988) has summed up his cognitive view of classical conditioning as follows: “[Classical] conditioning is not a stupid process by which the organism willy-nilly forms associations between any two stimuli that happen to co-occur. Rather, the organism is best seen as an information seeker using logical and perceptual relations among events, along with its own preconceptions, to form a sophisticated representation of its world.” By this, Rescorla does not mean that animals spend time consciously deliberating about these relationships. Rather, he means that animals have built-in neural mechanisms that automatically make the appropriate calculations.

Conditioning Depends on the Predictive Value of the Conditioned Stimulus

Support for Rescorla's view comes from research showing that classical conditioning occurs only, or at least mainly, when the new stimulus provides information that the animal can use to predict the arrival of the unconditioned stimulus. Here are three classes of such findings:

1. *The conditioned stimulus must precede the unconditioned stimulus.* Classical conditioning is most effective if the onset of the conditioned stimulus slightly precedes the unconditioned stimulus; it commonly doesn't occur at all if the conditioned stimulus comes either simultaneously with or just after the unconditioned stimulus (Lieberman, 2000). (See Figure 4.5a and b on the next page.) This observation makes sense if the animal is seeking predictive information; a stimulus that does not precede the unconditioned stimulus is useless as a predictor, and thus is ignored by the animal.
2. *The conditioned stimulus must signal heightened probability of occurrence of the unconditioned stimulus.* Conditioning depends not just on the number of pairings of the potential conditioned stimulus with the unconditioned stimulus, but also—negatively—on the number of times the potential conditioned stimulus occurs without being followed immediately by the unconditioned stimulus (Rescorla, 1988; Rescorla & Wagner, 1972). (See Figure 4.5c.) The animal's behavior suggests that in some way its nervous system computes two probabilities—the probability that the unconditioned stimulus will immediately follow any given occurrence of the potential conditioned stimulus and the probability that the unconditioned stimulus will occur at other times—and accepts the potential conditioned stimulus as a predictor only if the first probability is greater than the second.
3. *Conditioning is ineffective when the animal already has a good predictor.* If one conditioned stimulus already reliably precedes an unconditioned stimulus, a new stimulus, presented simultaneously with the existing conditioned stimulus, will not become a conditioned stimulus. This failure of conditioning is called the *blocking effect*; the already-conditioned stimulus blocks conditioning to the new stimulus (Kamin, 1969). (See Figure 4.5d.) A cognitive interpretation of this is that the animal has already solved the problem of predicting food and has no reason to find a new predictor.

Cognitive psychologists often emphasize that their use of terms such as *expectation* and *prediction* does not imply anything mystical. The terms describe the type of information that underlies the animal's behavior but say nothing about the physical form of that information, which presumably results from computations automatically

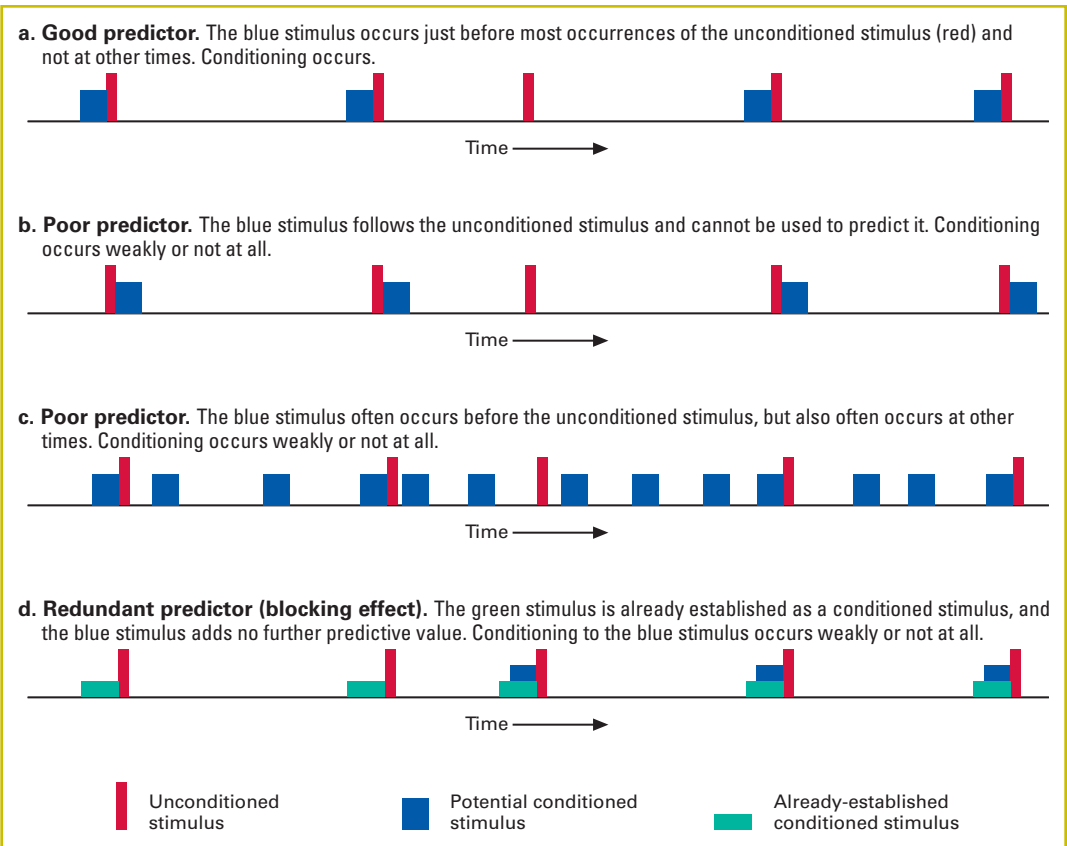


Mark Sherman / Bruce Coleman

Expectancy There is nothing inscrutable about this young tiger cat. The sound of the can being attached to the opener permits her to predict the arrival of food. Her response is not identical to her response to food itself, but one of rapt attention.

>>> 10 What are three conditions in which the pairing of a new stimulus with an unconditioned stimulus does not result in classical conditioning? How do these observations support the idea that classical conditioning is a process of learning to predict the onset of the unconditioned stimulus?

FIGURE 4.5 Conditions in which pairing of a potential conditioned stimulus with an unconditioned stimulus does or does not result in conditioning The top time line illustrates conditions in which good conditioning occurs to the blue stimulus. The three lower time lines illustrate conditions in which poor or no conditioning occurs even though the potential conditioned stimulus (blue) regularly precedes the unconditioned stimulus (red). Strong conditioning occurs when the potential conditioned stimulus is a reliable and non-redundant predictor of the unconditioned stimulus.



made by neurons in the brain. To illustrate that such computations can occur automatically, cognitive psychologists have programmed computers to behave in ways that simulate the processes of expectation and prediction (Commons & others, 1991). In fact, it is not difficult to design a machine that manifests all three of the conditioning characteristics just listed and in that sense is capable of learned prediction.

Conditioned Fear, Hunger, and Sexual Arousal

From a functional, evolutionary perspective, classical conditioning is a process by which individuals reflexively prepare for biologically significant events before they happen. A conditioned stimulus preceding a painful or startling event can elicit fear and bodily reactions that help brace the individual for that event; a conditioned stimulus preceding delivery of food can elicit hunger and bodily responses that help prepare the gastrointestinal tract for food; and a conditioned stimulus preceding an opportunity for sex can elicit high sex drive and bodily responses that help prepare the body for copulation.

Watson's Classic Demonstration of Fear Conditioning

11

How did Watson demonstrate that the emotion of fear can be conditioned?

John B. Watson was the first psychologist to show that the emotion of fear can be conditioned in human beings. Consistent with his behavioral perspective, Watson (1924) defined fear not as a feeling but as a set of observable responses: “a catching of the breath, a stiffening of the whole body, a turning away of the body from the source of stimulation, a running or crawling from it.” On the basis of this definition, Watson found two unconditioned stimuli for fear in human infants—sudden loud sound and sudden loss of support (as when a baby slips out of an adult’s hands). Other stimuli, he argued, come to elicit fear only as a result of conditioning.

In a classic demonstration, Watson and Rosalie Rayner (1920) conditioned an 11-month-old baby named Albert to fear laboratory rats. At first, Albert played happily



FIGURE 4.6 Little Albert with Watson, Rayner, and furry animals
Prior to the conditioning experience, 11-month-old Albert played happily with a live laboratory rat (*left*). After he was conditioned to respond fearfully to the rat, he also cried at the sight of other furry objects, including a rabbit, thereby exhibiting generalization (*right*).

with a rat that was placed in front of him. To condition the fear, the experimenters struck a steel bar with a hammer to produce a loud sound just after the rat was placed in front of Albert. After two such pairings Albert showed a moderate fear response to the rat alone, and after four more pairings he showed a strong fear reaction to the rat alone. Thus, in the terminology of classical conditioning, the rat had become a conditioned stimulus for fear through being paired with a loud sound, which was an unconditioned stimulus for fear. In subsequent tests, Watson and Rayner showed that Albert's conditioned fear generalized to other furry objects, such as a rabbit (see Figure 4.6).

You might (quite appropriately) wonder about the ethics of this experiment. The experiment most likely would not be approved by a modern ethics review committee and would not be done today. In fairness to Watson, however, there is no reason to believe that Albert's experience in this demonstration was any more frightening than are the countless startling events that everyone experiences in the normal course of growing up. Moreover, Watson was far more interested in how to eliminate unwanted fears than in how to produce them. In Chapter 17, you will see how Watson's research and ideas helped lead to clinical methods for extinguishing irrational fears through exposing the person to the fear-producing stimuli at gradually increasing magnitudes in a safe, relaxing context.

Fear, of course, is not the only emotional response that can be conditioned through Pavlovian procedures. When beer and car advertisers pair their products with scenes of beautiful people having wonderful times, they are trying to get you to drool with pleasure, like Pavlov's dogs, whenever you see their products.

Conditioned Hunger

A signal that reliably precedes food becomes a conditioned stimulus not just for salivation, but for a set of responses that help prepare the body for food. These responses include the secretion of digestive juices into the stomach and the secretion of certain hormones into the bloodstream (Woods & others, 2000). Through means discussed in Chapter 6, these reflexive responses induce a state of hunger. Thus, we have in humans the well-known appetizer effect. The taste of a small morsel of food, the smell of food, a dinner bell, a clock indicating that it is dinnertime, or any other signal that reliably precedes a meal can rather quickly cause us to feel much hungrier than we were just before the signal. Other animals, too, begin to act as if they are hungry, as indicated by their food-seeking behavior, when subjected to stimuli that have preceded food in the past (Pavlov, 1927/1960).

Conditioned Sexual Arousal

A number of experiments have demonstrated that sexual arousal can be conditioned in human subjects (Hoffman & others, 2004; Plaud & Martini, 1999). Typically, a

>>> 12
How can the appetizer effect be explained in terms of classical conditioning?

>>> 13
How has sexual arousal been conditioned in humans and other animals? What is the evidence, from experiments with non-human animals, that such conditioning promotes success in reproduction?



Conditioning young humans to love beer The unconditioned stimulus is the happy, sexually suggestive scene. The conditioned stimulus is the Budweiser label.

previously neutral stimulus is used as the conditioned stimulus (a slide showing a jar of pennies was used in one experiment) and an erotic, sexually arousing film clip is used as the unconditioned stimulus. Sexual arousal is measured in men by a strain gauge that indicates erection of the penis and in women by a device that responds to increased blood flow to the vagina. In such experiments, a stimulus that initially does not elicit arousal comes to elicit it after several pairings with the unconditioned stimulus.

Experiments on sexual conditioning in non-human animals have generally used males as subjects and either the sight of a sexually receptive female or physical access to such a female as the unconditioned stimulus. The general result is that a signal that reliably precedes the presentation of the female becomes

a conditioned stimulus for a set of responses that help prepare the male for courtship and mating (Domjan, 2003; Pfaus & others, 2001).

In one experiment with blue gourami fish, for example, males were conditioned, over several trials, to predict the arrival of a female when a signal light came on (Hollis & others, 1997). After conditioning, these fish responded to the signal by shifting rapidly from their normal aggressive mode of behavior to a courtship mode. When they were allowed access to a female after the signal, the conditioned males were much more successful in courtship than were the unconditioned males, primarily because the latter typically attacked the female before trying to court her. The biological payoff was that the conditioned males fathered many more offspring in these tests than did the unconditioned males.

Conditioned Drug Reactions

As you have just seen, bodily reactions associated with natural emotions and drives can be conditioned. In their early experiments, Pavlov's group showed that bodily reactions to drugs can also be conditioned. In one demonstration with a dog, they paired a tone with injection of a drug that elicited restlessness and vomiting. After repeated parings the dog began to exhibit those responses to the tone alone (Pavlov, 1927/1960). Since then, many experiments have shown that drug reactions can be conditioned. Researchers have found, for example, that for habitual coffee drinkers the mere smell and taste of coffee can produce increased alertness, apparently because of the previous pairings of that smell and taste with the effects of caffeine (Flaten & Blumenthal, 1999).

Conditioned Compensatory Reactions to Drugs

14

Why is the conditioned response to a drug-related stimulus often the opposite of the direct effect of the drug?

<<< Many drugs produce two effects: a direct effect followed by a compensatory reaction that counteracts the direct effect and tends to restore the normal bodily state. In such cases it is often found that only the compensatory reaction becomes conditioned. As a result, a stimulus that reliably precedes delivery of the drug produces a conditioned response that is opposite to the drug's direct effect. An example of such a drug is morphine. A direct effect of morphine is reduction in sensitivity to pain, but this is counteracted by reflexive responses in the body that tend to restore normal pain sensitivity. When rats are repeatedly injected with morphine in a distinctive environment, that environment becomes a conditioned stimulus for the counteractive response. When the rats are subsequently placed in that environment without morphine, the conditioned reaction occurs in the absence of the direct effect of the drug. The result is that the rats become temporarily hypersensitive to pain (Hinson & others, 1986).

The explanation of such effects rests on the fact that only responses that occur in a reflexive manner, involving the central nervous system (spinal cord and brain), can be conditioned (Siegel, 1999). The direct effects of some drugs are not

reflexes, and therefore cannot be conditioned, but the counteractive effects are reflexes. As an analogy, consider what would happen if a bell sound (conditioned stimulus) reliably preceded a shove from the front (unconditioned stimulus). The shove would tend to push you backward, but you would counter that with a reflexive movement forward. Only the reflexive movement forward would be conditioned, so if the bell were sounded without the shove, you might fall on your face—a reaction opposite to the direct (non-reflexive) effect of the shove. The body protects itself with counteractive reflexes to all sorts of interventions (such as shoves and drugs) that disrupt its normal functioning. The conditioning of such reflexes is normally useful because it allows the counteraction to begin even before the potentially harmful stimulus strikes (Dworkin, 1993).

Conditioned Reactions as Causes of Drug Tolerance

Shepard Siegel and his colleagues in Canada have shown that the phenomenon of drug tolerance depends at least partly on conditioning. *Drug tolerance* refers to the decline in physiological and behavioral effects that occur with some drugs when they are taken repeatedly. Because of tolerance, people who regularly take a drug have to increase their doses over time to continue to achieve the original effects. To some degree, tolerance is a result of long-term buildup of physiological systems in the body that help to counteract the drug's effects (Poulos & Cappell, 1991). However, it is also to some degree the result of conditioning. Because of conditioning, stimuli that normally precede drug intake cause the conditioned compensatory reaction to begin before the drug is actually taken, and that reaction counteracts the direct effect of the drug. For example, a conditioned increase in heart rate would counteract the effect of a drug whose direct effect is to slow the heart.

Siegel and his colleagues have found that many cases of “overdose” in heroin addicts are actually cases in which addicts took their usual drug doses in unusual environments (Siegel, 1984; Siegel & Ramos, 2002). When an addict takes a drug in the usual drug-taking environment, cues in that environment produce a conditioned compensatory reaction that allows the addict's body to tolerate a large dose of the drug. If the addict takes the same amount of the drug in a novel environment, where the conditioned cues aren't present, the full impact of the drug kicks in before a counteractive reaction begins, and the result is severe illness or death. Consistent with this interpretation, rats that had received many morphine injections in a specific highly distinctive cage were much more likely to survive a high dose of the drug if it was given to them in that same cage than if it was given in a different setting (Siegel, 1976; Siegel & Ramos, 2002). Similar effects have been shown in animal experiments using alcohol (Melchior, 1990) and various other drugs (Goudie, 1990).

15
How does classical conditioning contribute to the development of drug tolerance? Why is it dangerous for a drug addict to take his or her usual drug dose in an unusual environment?

Conditioned Reactions as Causes of Drug Relapse After Withdrawal

Another drug phenomenon that is at least partly explained by conditioned compensatory reactions is that of relapse by addicts who have undergone periods of drug withdrawal (Siegel, 1999). Addicts commonly withdraw from drugs during a stay at a treatment center that is quite unlike their home environment. After some period of drug-free existence at the center they no longer experience withdrawal symptoms or feel a craving for the drug. However, when they return to their own homes and neighborhoods, they are once again surrounded by many cues that, for them, are associated with drug use. These cues elicit compensatory drug reactions, which feel like withdrawal symptoms and elicit a strongly felt need for the drug—a felt need that is all too often irresistible.

In an effort to prevent such relapse, some drug treatment programs attempt to extinguish the effects of conditioning by repeatedly presenting addicts with stimuli associated with drug intake without the drug itself. Such programs have met with only partial success, however (Siegel & Ramos, 2002). It is impossible to identify and present, for each addict, all of the cues that, in his or her home environment,

16
How does classical conditioning help explain drug relapse after an addict returns home from a treatment center?

A strong conditioned stimulus for a heroin addict
The stimuli associated with drug preparation become strong conditioned stimuli for bodily responses that are opposite to the direct effect of the drug. Such responses help protect the body from the harmful effects of the drug, but they also produce intense drug craving and contribute to relapse in people who have gone through a period of drug withdrawal.



© Uwe Schmid / Corbis

have become conditioned stimuli for compensatory drug reactions. Moreover, spontaneous recovery from extinction reduces the long-term effectiveness of such treatments. An addict's best hope for overcoming a long-term addiction may be to move permanently, if possible, to an entirely new environment. During the Vietnam War many American soldiers became addicted to heroin in Vietnam. When they returned home, a surprisingly large number overcame their addiction immediately and never took heroin again (Robins & others, 1975). According to Siegel's analysis, this observation makes sense in view of the fact that they had left behind the cues that had become triggers for drug compensatory reactions and returned to an environment that contained none or few of those cues (Siegel, 1999).

SECTION REVIEW

Classical conditioning enables prediction of and preparation for unconditioned stimuli.

What Is Learned?

- According to the S-R theory, supported by early behaviorists, a link between the conditioned stimulus and response is learned.
- According to the S-S theory, a link between the conditioned and unconditioned stimulus is learned.
- The S-S theory implies expectancy and is supported by Rescorla's experiment.
- Consistent with the expectancy idea, conditioning occurs best when the conditioned stimulus is a reliable predictor of the unconditioned stimulus.

Conditioned Emotion and Motivation

- Examples are the conditioning of fear (in little Albert), the conditioning of hunger (the appetizer effect), and the conditioning of sexual arousal.
- In general, conditioned stimuli trigger responses that help prepare the individual for a biologically significant event.

Conditioned Drug Reactions

- With some drugs, repeated pairing with a conditioned stimulus causes that stimulus to elicit the same type of response as the drug.
- With some other drugs, the conditioned stimulus elicits a response that is *opposite* to the drug response. Such conditioned compensatory reactions contribute to drug tolerance and drug relapse. In the absence of conditioned stimuli, overdose can occur.

Operant Conditioning I: Fundamentals

We are pulled as well as pushed by events in our environment. That is, we do not just respond to stimuli; we also behave in ways that seem designed to *produce* or *obtain* certain environmental changes or stimuli. My dog rubs against the door to be let out. I flip a switch to illuminate a room, press keys on my computer to set up words on a screen, and say "Please pass the potatoes" to get potatoes. Most of my day seems to consist of behaviors of this sort, and I expect that most of my dog's day would, too, if there were more things that he could control. Surely, if Pavlov's dogs had had some way to control the delivery of food into their mouths, they

would have done more than salivate—they would have pushed a lever, bitten open a bag, or done whatever was required to get the food.

Such actions are called **operant responses** because they *operate* on the world to produce some effect. They are also called *instrumental responses* because they function like *instruments*, or tools, to bring about some change in the environment. The process by which people or other animals learn to make operant responses is called **operant conditioning**, or *instrumental conditioning*. Operant conditioning can be defined as a learning process by which the effect, or consequence, of a response affects the future rate of production of that response. In general, operant responses that produce effects that are favorable to the animal increase in rate, and those that produce effects that are unfavorable to the animal decrease in rate.

From the Law of Effect to Operant Conditioning: From Thorndike to Skinner

Although the labels “operant” and “instrumental” were not coined until the 1930s, this learning process was studied well before that time, most notably by E. L. Thorndike.

Thorndike’s Puzzle-Box Procedure

At about the same time that Pavlov began his research on conditioning, a young American student of psychology, Edward Lee Thorndike (1898), published a report on his own learning experiments with various animals, including cats. Thorndike’s training method was quite different from Pavlov’s, and so was his description of the learning process. His apparatus was a puzzle box—a small cage that could be opened from inside by some relatively simple act, such as pulling a loop or pressing a lever (see Figure 4.7).

In one experiment, Thorndike deprived cats of food long enough to make them hungry and then placed them inside the cage, one at a time, with food just outside it. When first placed inside, a cat would engage in many actions—such as clawing at the bars or pushing at the ceiling—in an apparent attempt to escape from the cage and get at the food. Finally, apparently by accident, the cat would pull the loop or push the lever that opened the door to freedom and food. Thorndike repeated this procedure many times with each cat. He found that in early trials the

>>> **17**
How did Thorndike train cats to escape from a puzzle box? How did this research contribute to Thorndike’s formulation of the law of effect?

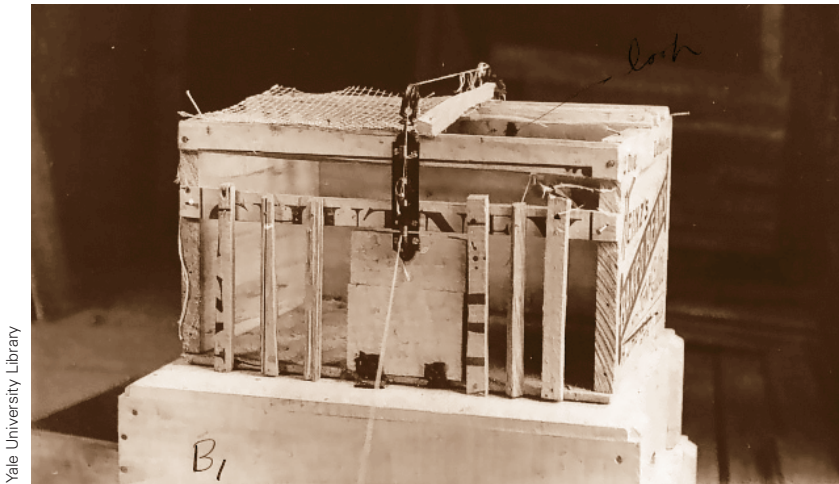


FIGURE 4.7 | One of Thorndike’s puzzle boxes Thorndike was a great psychologist, not a great carpenter! Shown here is a photograph of one of his actual puzzle boxes. A cat placed inside could open this box by pulling down on the loop hanging from the ceiling, which would pull up the bolt and allow the door to fall forward.



Edward Lee Thorndike After earning his Ph.D. in psychology with his puzzle-box experiments, Thorndike became a professor of educational psychology at Columbia University. Perhaps partly because of his early animal research, he maintained throughout his career that intelligence is a collection of separate learned skills, not a unitary characteristic of the person.

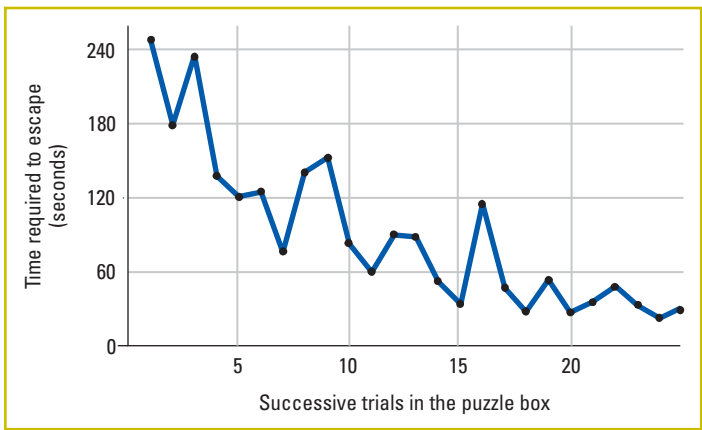


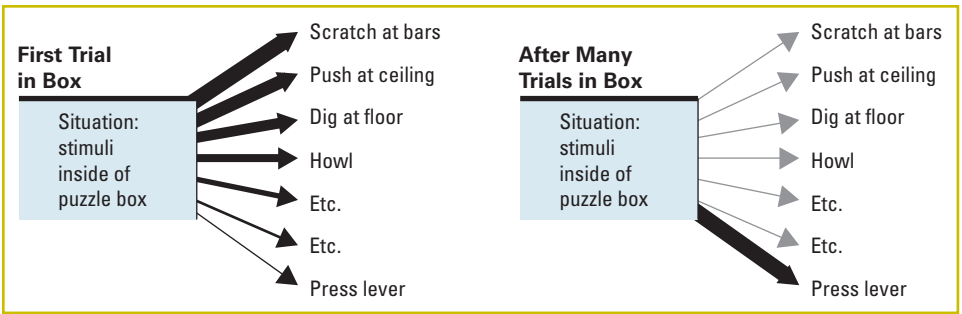
FIGURE 4.8 Typical learning curve for a cat in Thorndike's puzzle box As illustrated here for a single cat, Thorndike found that cats usually took less time to escape from the box on successive trials, although a great deal of variability occurred from trial to trial. (Adapted from Thorndike, 1898.)

Thorndike's Law of Effect

Thorndike's basic training procedure differed fundamentally from Pavlov's. Pavlov produced learning by controlling the relationship between two stimuli in the animal's environment, so that the animal learned to use one stimulus to predict the occurrence of the other. Thorndike, in contrast, produced learning by altering the consequence of some aspect of the animal's behavior. Thorndike's cats, unlike Pavlov's dogs, had some control over their environment. They could do more than merely predict when food would come; they could gain access to it through their own efforts.

Partly on the basis of his puzzle-box experiments, Thorndike (1898) formulated the **law of effect**, which can be stated briefly as follows: *Responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation.* In Thorndike's puzzle-box experiments, the *situation* presumably consisted of all the sights, sounds, smells, internal feelings, and so on that were experienced by the hungry animal in the box. None of them elicited the latch-release response in reflex-like fashion; rather, taken as a whole, they set the occasion for many possible responses to occur, only one of which would release the latch. Once the latch was released, the *satisfying effect*, including freedom from the box and access to food, caused that response to strengthen; so the next time the cat was in the same situation, the probability of that response's recurrence was increased (see Figure 4.9).

FIGURE 4.9 Thorndike's law of effect According to Thorndike, the stimulus situation (being inside the puzzle box) initially elicits many responses, some more strongly than others, but the satisfying consequence of the successful response (pressing the lever) causes that response to be more strongly elicited on successive trials.



Skinner's Method of Studying and Describing Operant Conditioning

18 How did Skinner's method for studying learning differ from Thorndike's, and why did he prefer the term *reinforcement* to Thorndike's *satisfaction*?

<<< The psychologist who did the most to extend and popularize the law of effect for more than half a century was Burrhus Fredric Skinner. Like Watson, Skinner was a confirmed behaviorist. He believed that principles of learning are the most fundamental principles in psychology, and he wanted to study and describe learning in ways that do not entail references to mental events. Unlike Watson, however, he

did not consider the stimulus-response reflex to be the fundamental unit of all of behavior. Thorndike's work provided Skinner with a model of how non-reflexive behaviors could be altered through learning.

As a graduate student at Harvard University around 1930, Skinner developed an apparatus for studying learning in animals that was considerably more efficient than Thorndike's puzzle boxes. His device, commonly called a "Skinner box," is a cage with a lever or another mechanism in it that the animal can operate to produce some effect, such as delivery of a pellet of food or a drop of water (see Figure 4.10). The advantage of Skinner's apparatus is that the animal, after completing a response and experiencing its effect, is still in the box and free to respond again. With Thorndike's puzzle boxes and similar apparatuses such as mazes, the animal has to be put back into the starting place at the end of each trial. With Skinner's apparatus, the animal is simply placed in the cage and left there until the end of the session. Throughout the session, there are no constraints on when the animal may or may not respond. Responses (such as lever presses) can easily be counted automatically, and the learning process can be depicted as changes in the rate of responses (see Figure 4.11). Because each pellet of food or drop of water is very small, the hungry or thirsty animal makes many responses before becoming satiated.

Skinner developed not only a more efficient apparatus for studying such learning but also a new vocabulary for talking about it. As a confirmed behaviorist, he wanted to describe learning in terms that refer to observable environmental events and behavioral acts and to avoid words that refer to mental events. It was he who coined the terms *operant response* to refer to any behavioral act that has some effect on the environment and *operant conditioning* to refer to the process by which the effect of an operant response changes the likelihood of the response's recurrence. Thus, in a typical experiment with a Skinner box, pressing the lever is an operant response, and the increased rate of lever pressing that occurs when the response is followed by a pellet of food exemplifies operant conditioning. Applying the same terms to Thorndike's experiments, the movement that opens the latch is an operant response, and the increase from trial to trial in the speed with which that movement is made exemplifies operant conditioning.

Skinner (1938) proposed the term **reinforcer**, as a replacement for such words as *satisfaction* and *reward*, to refer to a stimulus change that occurs after a response and that increases the subsequent frequency of that response. Skinner preferred this term because it makes no assumptions about anything happening in the mind; it merely refers to the effect that the stimulus change has on the animal's subsequent behavior. Thus, in a typical Skinner-box experiment, the delivery of a pellet of food or drop of water following a lever-press response is a reinforcer. In Thorndike's



Nina Leen / LIFE Magazine © Time Warner, Inc.

FIGURE 4.10 | B. F. Skinner and his operant-conditioning chamber To study operant behavior in rats and other animals, Skinner invented an apparatus widely known as the Skinner box. When the rat shown here presses the lever, it activates an electrical relay system that causes the delivery of a food pellet into a cup next to the lever. Each lever press can be automatically recorded to produce a cumulative record, such as that shown in Figure 4.11.

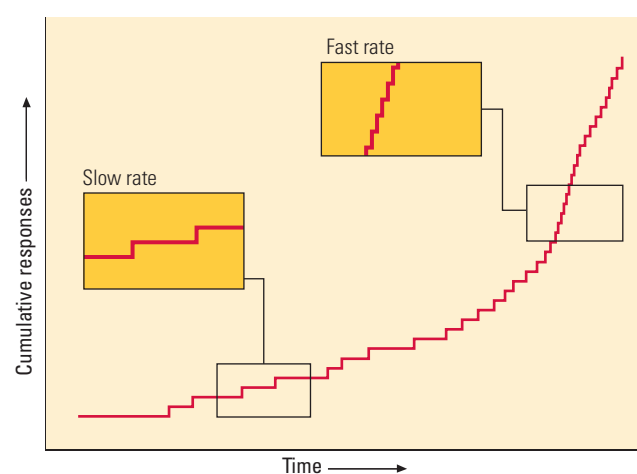


FIGURE 4.11 | Typical cumulative response curve for a rat learning to press a lever in a Skinner box This graph is called a cumulative response curve because the height of the curve at any point indicates the total (cumulative) number of responses that the rat has made up to that time. The graph is automatically produced, while the rat is responding, by a recording machine outside the Skinner box. A pen moves horizontally across a roll of paper at a constant rate, and each lever press made by the rat produces a slight vertical movement of the pen. Thus, the degree to which the curve slopes upward is a measure of the animal's response rate. Note that early in learning the response rate was very low and then it gradually increased to a fast, steady rate.

experiment, escape from the cage and access to the food outside it were reinforcers. Some stimuli, such as food for a food-deprived animal or water for a water-deprived one, are naturally reinforcing. Other stimuli have reinforcing value only because of previous learning, and Skinner referred to these as *conditioned reinforcers*. An example of a conditioned reinforcer for humans is money. Once a person learns what money can buy, he or she will learn to behave in ways that yield more of it.

Operant Conditioning Without Awareness

To Skinner and his followers, operant conditioning is not simply one kind of learning to be studied but represents an entire approach to psychology. In his many books and articles, Skinner argued that essentially all the things we do, from the moment we arise in the morning to the moment we fall asleep at night, can be understood as operant responses that occur because of their past reinforcement. In some cases we are clearly aware of the relationship between our responses and reinforcers, as when we place coins in a vending machine for a candy bar or study to obtain a good grade on a test. In other cases we may not be aware of the relationship, yet it exists and, according to Skinner (1953, 1966), is the real reason for our behavior. Skinner argued that awareness—which refers to a mental phenomenon—is not a useful construct for explaining behavior. We can never be sure what a person is aware of, but we can see directly the relationship between responses and reinforcers and use that to predict what a person will learn to do.

19

How were people conditioned without awareness to make a tiny thumb twitch, and how is this relevant for understanding the acquisition of motor skills?

<<< An illustration of conditioning without awareness is found in an experiment conducted many years ago, in which adults listened to music over which static was occasionally superimposed (Hefferline & others, 1959). Unbeknownst to the subjects, they could turn off the static by making an imperceptibly small twitch of the left thumb. Some subjects (the completely uninformed group) were told that the experiment had to do with the effect of music on body tension; they were told nothing about the static or how it could be turned off. Others (the partly informed group) were told that static would sometimes come on, that they could turn it off by a specific response, and that their task was to discover that response. The result was that all subjects in both groups learned to make the thumb-twitch response, thereby keeping the static off for increasingly long periods. But, when questioned afterward, none were aware that they had controlled the static with thumb twitches. Subjects in the uninformed group said that they had noticed the static decline over the session but were unaware that they had caused the decline. Most in the partly informed group said that they had not discovered how to control the static. Only one participant believed that he had discovered the effective response, and he claimed that it involved “subtle rowing movements with both hands, infinitesimal wriggles of both ankles, a slight displacement of the jaw to the left, breathing out, and then waiting”! While he was consciously making this superstitious response, he was unconsciously learning to make the thumb twitch.

If you think about it, the results of this experiment should not come as a great surprise. We constantly learn finely tuned muscle movements as we develop skill at playing the violin, riding a bicycle, hammering nails, or any other activity. The probable reinforcers are, respectively, the improved sound from the violin, the steadier movement on the bicycle, and the straight downward movement of the nail we are pounding; but often we do not know just what we are doing differently to produce these good effects. Our knowledge is often similar to that of the neophyte carpenter who said, after an hour’s practice at hammering, “The nails you’re giving me now don’t bend as easily as the ones you were giving me before.”

Principles of Reinforcement

Skinner and his followers identified and studied many behavioral phenomena associated with operant conditioning, including the ones described in this section.

Shaping of New Operant Responses

Suppose you put a rat in a Skinner box, and it never presses the lever; or you place a cat in a puzzle box, and it never pulls the loop; or you want to train your dog to jump through a hoop, but it never makes that leap. In operant conditioning, the reinforcer comes after the subject produces the desired response. But if the desired response never occurs, it can never be reinforced. The solution to this problem is a technique called *shaping*, in which successively closer approximations to the desired response are reinforced until the desired response finally occurs and can be reinforced.

Imagine that you want to shape a lever-press response in a rat whose initial rate of lever pressing is zero or so low that you don't have the patience to wait for the response to occur. To begin, you might present the reinforcer (such as a pellet of food) whenever the rat goes anywhere near the lever. As a result, the rat will soon be spending most of its time near the lever and occasionally will touch it. When that happens, you might provide the reinforcer only when the rat touches the lever, which will increase the rate of touching. Some touches will be more vigorous than others and produce the desired lever movement; when that has happened a few times, you can stop reinforcing any other response—your animal's lever pressing has now been shaped. Animal trainers regularly use this technique in teaching domestic or circus animals to perform new tricks or tasks (Pryor, 1985), and we all tend to use it, more or less deliberately and not always skillfully, when we teach new skills to other people. For example, when teaching a novice to play tennis, we tend at first to offer praise for any swing of the racket that propels the ball in the right general direction, but as improvement occurs, we begin to bestow our praise only on closer and closer approximations of an ideal swing.

>>> 20
How can operant conditioning be used to get an animal to do something that it currently doesn't do?

Extinction of Operantly Conditioned Responses

An operantly conditioned response declines in rate and eventually disappears if it no longer results in a reinforcer. Rats stop pressing levers if no food pellets appear, cats stop scratching at doors if nobody responds, and people stop smiling at those who don't smile back. The absence of reinforcement of the response and the consequent decline in response rate are both referred to as *extinction*. Extinction in operant conditioning is analogous to extinction in classical conditioning. Just as in classical conditioning, extinction in operant conditioning is not true "unlearning"

>>> 21
In what ways is extinction in operant conditioning similar to extinction in classical conditioning?



AP Photo/Tulsa World, James Gibbard

An effect of operant shaping
Through the process of rewarding gradual approximations to the desired behavior, animals can be trained to do things that they would never do without such training.

of the response. Passage of time following extinction can lead to spontaneous recovery of responding, and a single reinforced response following extinction can lead the individual to respond again at a rapid rate.

Schedules of Partial Reinforcement

In many cases, in the real world as well as in laboratory experiments, a particular response only sometimes produces a reinforcer. This is referred to as **partial reinforcement**, to distinguish it on the one hand from continuous reinforcement, where the response is always reinforced, and on the other hand from extinction, where the response is never reinforced. In initial training, continuous reinforcement is most efficient, but once trained, an animal will continue to perform for partial reinforcement. Skinner and other operant researchers have described the following four basic types of partial-reinforcement schedules:

22

How do the four types of partial-reinforcement schedules differ from one another, and why do responses generally occur faster to ratio schedules than to interval schedules?

- In a **fixed-ratio schedule** a reinforcer occurs after every n th response, where n is some whole number greater than 1. For example, in a *fixed-ratio 5 schedule* every fifth response is reinforced.
- A **variable-ratio schedule** is like a fixed-ratio schedule except that the number of responses required before reinforcement varies unpredictably around some average. For example, in a *variable-ratio 5 schedule* reinforcement might come after 7 responses on one trial, after 3 on another, and so on, in random manner, but the average number of responses required for reinforcement would be 5.
- In a **fixed-interval schedule** a fixed period of time must elapse between one reinforced response and the next. Any response occurring before that time elapses is not reinforced. For example, in a *fixed-interval 30-second schedule* the first response that occurs at least 30 seconds after the last reinforcer is reinforced.
- A **variable-interval schedule** is like a fixed-interval schedule except that the period that must elapse before a response will be reinforced varies unpredictably around some average. For example, in a *variable-interval 30-second schedule* the average period required before the next response will be reinforced is 30 seconds.

The different schedules produce different response rates, in ways that make sense if we assume that the person or animal is striving to maximize the number of reinforcers and minimize the number of unreinforced responses. *Ratio* schedules (whether fixed or variable) produce reinforcers at a rate that is directly proportional to the rate of responding, so, not surprisingly, such schedules typically induce rapid responding. With *interval* schedules, in contrast, the maximum number of reinforcers available is set by the clock, and such schedules result, not surprisingly, in relatively low response rates that depend on the length of the fixed or average interval.

23

How do variable-ratio and variable-interval schedules produce behavior that is highly resistant to extinction?

Behavior that has been reinforced on a variable-ratio or variable-interval schedule is often very difficult to extinguish. If a rat is trained to press a lever only on continuous reinforcement and then is shifted to extinction conditions, the rat will typically make a few bursts of lever-press responses and then quit. But if the rat has been shifted gradually from continuous reinforcement to an ever-stingier variable schedule and then finally to extinction, it will often make hundreds of unreinforced responses before quitting. Rats and humans who have been reinforced on stingy variable schedules have experienced reinforcement after long, unpredictable periods of no reinforcement, so they have learned (for better or worse) to be persistent. Skinner (1953) and others (Rachlin, 1990) have used this phenomenon to explain why gamblers often persist at the slot machine or dice game even after long periods of losing: they are hooked by the variable-ratio schedule of payoff that characterizes nearly every gambling device or game. From a cognitive perspective, we could say that the gambler keeps playing because of his or her knowledge that the very next bet or throw of the dice might be the one that pays off.



© Richard Smith / Corbis



David Simson / Stock Boston

Partial reinforcement in daily life Winning at slot machines occurs on a variable-ratio schedule, which produces a rapid and steady style of play. Success at reaching a previously busy telephone number occurs on a variable-interval schedule, which results in a slow and steady rate of redialing.

Distinction Between Positive and Negative Reinforcement

In Skinner's terminology, **reinforcement** refers to any process that increases the likelihood that a particular response will occur. Reinforcement can be positive or negative. **Positive reinforcement** occurs when the *arrival* of some stimulus following a response makes the response more likely to recur. The stimulus in this case is called a **positive reinforcer**. Food pellets, money, words of praise, and anything else that organisms will work to obtain can be used as positive reinforcers. **Negative reinforcement**, in contrast, occurs when the *removal* of some stimulus following a response makes the response more likely to recur. The stimulus in this case is called a **negative reinforcer**. Electric shocks, loud noises, unpleasant company, scoldings, and everything else that organisms will work to get away from can be used as negative reinforcers. The one example of negative reinforcement discussed so far was the experiment in which a thumb-twitch response was reinforced by the temporary removal of unpleasant static. Notice that *positive* and *negative* here do not refer to the direction of change in the response rate; that increases in either case. Rather, the terms indicate whether the reinforcing stimulus appears (positive) or disappears (negative) as a result of the operant response.

>>> 24
How does negative reinforcement differ from positive reinforcement?

Distinction Between Reinforcement and Punishment

In Skinner's terminology, **punishment** is the opposite of reinforcement. It is the process through which the consequence of a response *decreases* the likelihood that the response will recur. As with reinforcement, punishment can be positive or negative. In **positive punishment**, the arrival of a stimulus, such as electric shock for a rat or scolding for a person, decreases the likelihood that the response will occur again. In **negative punishment**, the removal of a stimulus, such as taking food away from a hungry rat or money away from a person, decreases the likelihood that the response will occur again. Both types of punishment can be distinguished from extinction, which, you recall, is the decline in a previously reinforced response when it no longer produces any effect.

>>> 25
How does punishment differ from reinforcement, and how do the two kinds of punishment parallel the two kinds of reinforcement?

To picture the distinction between positive and negative punishment, and to see their relation to positive and negative reinforcement, look at Figure 4.12. The terms are easy to remember if you recall that *positive* and *negative* always refer to the arrival or removal of a stimulus and that *reinforcement* and *punishment* always refer to an increase or decrease in the likelihood that the response will recur.

The figure also makes it clear that the same stimuli that can serve as positive reinforcers when presented can serve as negative punishers when removed, and the

FIGURE 4.12 Two types of reinforcement and two types of punishment Reinforcement (whether positive or negative) increases the response rate, and punishment (whether positive or negative) decreases the response rate. The terms *positive* and *negative* refer to whether the reinforcing stimulus arrives or is removed when the response is made.

		Response rate	
		Increases	Decreases
Response causes stimulus to be	Presented	Positive reinforcement (Lever press→food pellet)	Positive punishment (Lever press→shock)
	Removed	Negative reinforcement (Lever press→turns off shock)	Negative punishment (Lever press→removes food)

same stimuli that can serve as positive punishers when presented can serve as negative reinforcers when removed. It is easy to think of the former as “desired” stimuli and the latter as “undesired,” but Skinner avoided such mentalistic terms. He argued that the only way to tell whether a stimulus is desired or undesired is by observing its reinforcing or punishing effects, so the mentalistic terms add nothing to our understanding. For instance, an adult who scolds a child for misbehaving may think that the scolding is an undesired stimulus that will punish the behavior, but, in fact, it may be a desired stimulus to the child who seeks attention. To Skinner, the proof is in the pudding. If scolding causes the undesired behavior to become less frequent, then scolding is acting as a positive punisher; if scolding causes the behavior to become more frequent, it is acting as a positive reinforcer.

SECTION REVIEW

Operant conditioning reflects the impact of a behavior’s consequences on its recurrence.		
Work of Thorndike and Skinner • An operant response is an action that operates on the environment. • Thorndike’s puzzle box experiments led him to postulate the law of effect. • Skinner developed an efficient way to study operant conditioning. He defined <i>reinforcer</i> as a stimulus change that follows an operant response and increases the frequency of that response. • Operant conditioning can occur without awareness.	Variations in Availability of Reinforcement • Shaping occurs when successive approximations to the desired response are reinforced. • Extinction is the decline in response rate that occurs when an operant response is no longer reinforced. • Partial reinforcement can occur on various schedules of reinforcement. The type of schedule affects the response rate and resistance to extinction.	Reinforcement vs. Punishment • Reinforcement increases response rate; punishment decreases response rate. • Reinforcement can be either positive (e.g., praise is given) or negative (e.g., pain goes away). • Punishment can be either positive (e.g., a reprimand is given) or negative (e.g., computer privileges are taken away).

Operant Conditioning II: What Is Learned?

What does an animal learn in operant conditioning? It does not just learn to make a response more frequently. It learns something about the conditions in which the response will be rewarded and about the nature of the reward.

Learning When a Response Will Be Rewarded

A rat trained in a Skinner box has learned to make a lever-press response in a particular context, which includes being inside the Skinner box. The rat does not go foolishly about making the lever-pressing movement in its home cage or in other places where the response has never been reinforced. In his original formulation

of the law of effect, Thorndike (1898) emphasized the importance of the *situation* in which the animal is trained, saying, “Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that when it recurs, they will be more likely to recur.” The set of stimuli inside a puzzle box or a Skinner box is an example of a Thorndikian situation (look back at Figure 4.9). Only in the presence of those stimuli is the response reinforced; therefore, the response becomes likely to occur when those stimuli are present.

Discrimination Training in Operant Conditioning

Through *discrimination training*, an operant response can be brought under the control of a more specific stimulus than the entire inside of a Skinner box. Discrimination training in operant conditioning is analogous to discrimination training in classical conditioning. The essence of the procedure is to reinforce the animal's response when a specific stimulus is present and to extinguish the response when the stimulus is absent. Thus to train a rat to respond to a tone by pressing a lever, a trainer would alternate between reinforcement periods with the tone on (during which the animal gets food pellets for responding) and extinction periods with the tone off. After considerable training of this sort, the rat will begin pressing the lever as soon as the tone comes on and stop as soon as it goes off. The tone in this example is called a **discriminative stimulus**.

A discriminative stimulus can be thought of as a cue signaling the availability of a reinforcer; it is present when a particular response will be reinforced and absent otherwise. A discriminative stimulus in operant conditioning is similar to a conditioned stimulus in classical conditioning in that it promotes a particular response as a result of the subject's previous experience, but it does so in a different way. It *sets the occasion* for responding, rather than reflexively eliciting the response.

Operant discrimination training, like the analogous procedure in classical conditioning, can be used to learn about the sensory abilities of animals and human infants, who cannot describe their sensations in words. In one experiment, for example, researchers trained 1-day-old human babies to turn their heads in one direction, using a sip of sugar water as the reinforcer, whenever a tone was sounded and in the other direction whenever a buzzer was sounded (Siqueland & Lipsitt, 1966). Thus the babies learned to make two different responses to two different discriminative stimuli. This demonstrated, among other things, that the newborns could hear the difference between the two sounds. (More is said about such experiments in Chapter 11.) Operant discrimination training can also be used to learn about animals' understanding of concepts.

Discrimination and Generalization as an Index of Concept Understanding

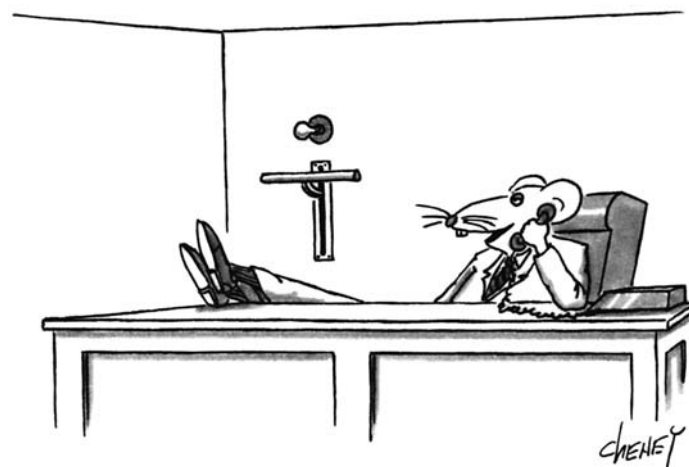
In classical conditioning, as you may recall, animals that have learned to respond to a conditioned stimulus will respond also to new stimuli that they perceive as similar to the original conditioned stimulus, a phenomenon referred to as *generalization*. Generalization also occurs in operant conditioning. After operant discrimination training, animals will respond to new stimuli that they perceive as similar to the discriminative stimulus. Such generalization can be used to test an animal's understanding of a concept.

Consider an experiment conducted by Richard Herrnstein (1979), who operantly conditioned pigeons to peck a key for grain, using slides depicting natural scenes as discriminative stimuli. Herrnstein divided the slides into two categories—those that had at least one tree or portion of a tree somewhere in the scene and those

>>>

26

How can an animal be trained to produce an operant response only when a specific cue is present?



“Oh, not bad. The light comes on, I press the bar, they write me a check. How about you?”

© The New Yorker Collection. 1993 Tom Cheney from cartoonbank.com. All rights reserved.

>>>

27

How was discrimination training used to demonstrate that pigeons understand the concept of a tree?



Michael Thompson / Comstock



Naoki Okamoto / Stock Market

FIGURE 4.13 Tree pictures similar to those used to study concepts in pigeons. Pigeons that had been trained to peck whenever a slide contained a tree, or part of a tree, pecked when they saw slides such as these and refrained from pecking when they saw similar slides that did not include a tree.

It's a flower Researchers have developed various means to test animals' ability to categorize complex objects. In the experiment illustrated here, Edward Wasserman (1995) trained pigeons to peck a different one of four keys depending on whether the slide shown was of a car, cat, chair, or flower. The birds proved to be quite accurate at classifying objects that they had never seen before.



American Scientist

that didn't (see Figure 4.13). The pigeons received grain (on a variable-ratio schedule of reinforcement) for pecking the key whenever a "tree" slide was shown and nothing when a "non-tree" slide was shown. In the first phase, 80 slides were presented each day, 40 of which contained trees and 40 of which didn't. By the end of 5 days of such training—that is, after five presentations of all 80 slides—all the birds were successfully discriminating between the two categories of slides, pecking when the slide contained a tree and not pecking otherwise.

Now, the question is: What did the pigeons learn in Herrnstein's experiment? Did they learn each slide as a separate stimulus, unrelated to the others, or did they learn a rule for categorizing the slides? Such a rule might be stated by a person as follows: "Respond whenever a slide includes a tree or part of a tree, and don't respond otherwise." To determine whether the pigeons had acquired such a rule, Herrnstein tested them with new slides, which they had never seen before, under conditions in which no grain was given. He found that the pigeons immediately pecked at a much higher rate when a new slide contained a tree than when it did not. In fact, the pigeons were as accurate with new slides as they were with slides that had been used during training. The birds apparently based their responses on a **concept** of trees (Herrnstein, 1990). A concept, as the term is used here, can be defined as a rule for categorizing stimuli into groups. The pigeons' tree concept, in this case, must have guided their decision to peck or not to peck.

How might one describe the pigeons' tree concept? That question is not easily answered. It is *not* the case, for example, that the pigeons simply learned to peck at slides that included a patch of green. Many of the non-tree slides had green grass, and some of the tree slides were of fall or winter scenes in New England, where the trees had red and yellow leaves or none at all. In some slides, only a small portion of a tree was apparent or the tree was in the distant background. So the method by which the pigeons distinguished tree slides from others cannot be easily stated in stimulus terms, although ultimately it must have been based on the birds' analysis of the stimulus material. Similar experiments have shown that pigeons can acquire concepts pertaining to such objects as cars, chairs, the difference between male and female human faces, and even abstract symbols (Loidolt & others, 2003; Wasserman, 1995; Zentall, 2003). The point is that, even for pigeons and certainly for

humans, sophisticated analysis of the stimulus information occurs before the stimulus is used to guide behavior.

Learning to Expect a Certain Reward

As they did for classical conditioning, some early behaviorists (for example, Guthrie, 1952) supported a stimulus-response (S-R) theory of operant conditioning. According to this theory, operant conditioning entails the strengthening of a bond between the reinforced response and stimuli that are present just before the response is made (the discriminative stimuli). Thus, for a rat learning to press a lever in a Skinner box, the learned connection could be described as: *stimuli inside Skinner box* → *lever press*. Or if a more specific discriminative stimulus such as a tone is used, it could be described as: *tone* → *lever press*. According to this view, the reinforcer (such as a pellet of food) is involved in learning only insofar as it helps stamp in the connection between the discriminative stimuli and the response.

Other theorists, however, taking a more cognitive perspective, have argued, with evidence, that during operant conditioning the animal learns much more than the S-R relationship. The animal also learns the S-S relationship between the discriminative stimuli and the reinforcing stimulus (Mowrer, 1960) and the R-S relationship between the response and the reinforcing stimulus (Mackintosh & Dickinson, 1979). In other words, a rat in a Skinner box learns that the discriminative stimuli within the box signal that food is available there (S-S relationship) and that pressing the lever will make the food appear (R-S relationship). In an early cognitive theory, Edward Tolman (1959) described operant conditioning as the learning of relationships between means and ends. The animal learns to expect that a particular response, made at the appropriate time (when the discriminative stimuli are present), will produce a certain consequence.

>>> 28
How do cognitive theories of operant conditioning differ from the S-R theory?

Evidence That Animals Learn Means-End Relationships

A *means-end relationship* is the animal's knowledge or belief that a particular response, in a particular situation, will have a particular effect. Thus, according to Tolman, if a rat is reinforced with a specific kind of food for pressing a lever when a tone is on, the rat acquires the knowledge that pressing the lever when the tone is on will produce that food. In the future, the occurrence of the tone does not automatically produce a lever-press response, as the S-R theory would hold, but, rather, activates the animal's knowledge that a lever press will bring a certain food—will be the means to an end. The animal can then press the lever or not, depending on its current interest in the food. Thus, in Tolman's view, the rat's behavior is best understood in terms of this sequence:

- tone → knowledge that lever press will now bring a certain kind of food
- decision to press lever or not, depending on whether the food is desired

In support of this cognitive view, animals that have learned an operant response have been found to behave in ways that are consistent with the premise that they have learned a means-end relationship (Dickinson, 1989; Dickinson & Balleine, 2000). For example, in one experiment, some hungry rats were trained to press a lever for sugar water and others were trained to press a lever for dry food pellets as a reinforcer (Dickinson & Dawson, 1987). Later all the rats were tested under extinction conditions (no reinforcer given), after having been deprived of either food or water. The interesting result was that when tested in the new drive state (thirst), those that had been reinforced with sugar water (which would satisfy thirst as well as hunger) pressed at a high rate, and those that had been reinforced with dry pellets (which would not satisfy thirst) pressed at a low rate (see Figure 4.14). The most direct way to explain this difference is to assume that the rats had acquired knowledge of the kind of reinforcer they would receive for pressing the lever and were able to use that knowledge to decide whether or not the reinforcer would satisfy their current drive state.

>>> 29
How can the view that operant conditioning involves means-end knowledge be experimentally tested? What were the results of one such test?

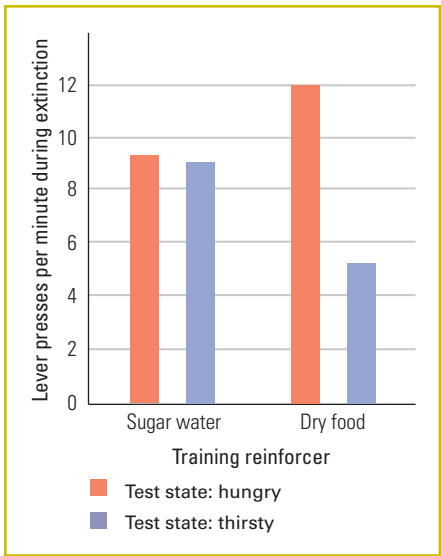


FIGURE 4.14 Evidence that rats learn what reinforcer a response produces Rats that were trained to press a lever for sugar water when hungry continued to press the lever at as high a rate (with no reinforcer given) when they were thirsty as when they were hungry (two left-hand bars). In contrast, rats that were trained to press a lever for dry food when they were hungry pressed the lever much less (with no reinforcer given) when they were thirsty than when they were hungry (two right-hand bars). These results suggest that the rats had a conception of the reinforcer that the response would produce and knew whether it would satisfy both thirst and hunger or only hunger. (Adapted from Dickinson & Dawson, 1987.)

Evidence That Animals Keep Track of the Relative Value of Rewards

30

How are reward contrast effects explained from a cognitive perspective?

Further evidence that animals learn to expect a certain reward for making a certain response is found in the existence of *reward contrast effects*, which involve shifts in the response rate when the value of the reward changes (Flaherty, 1996). Suppose that one group of rats is trained to press a lever for a highly valued reward (such as good-tasting food) and another group is trained to press for a less valued reward (food that doesn't taste so good). As you might predict, rats in the first group would typically respond at a faster rate than those in the second. So far, this is consistent with S-R theory. Presumably, the S-R connection built by the strong reinforcer is stronger than that built by the weak reinforcer, resulting in faster responses from the first group. Now suppose that the rats receiving the strong reinforcer are suddenly shifted to the weak one. According to the S-R model, they should continue, at least for a while, to respond at a higher rate than the other rats, because of the stronger S-R connection built during the first phase of the experiment. But what actually happens is the opposite. Rats that are shifted from a strong to a weak reinforcer show a sharp drop in their response rate, to a slower rate than that of rats that had been in the weak-reinforcer condition all along. This illustrates the **negative contrast effect**. Conversely, rats that are shifted from a weak to a strong reinforcer show an increase in their response rate, to a higher rate than that of rats that had been in the strong-reinforcer condition all along. This illustrates the **positive contrast effect**.

Not surprisingly, reward contrast effects also occur for people. If you have been getting \$1 per page for typing term papers, you will probably be delighted to discover that your pay is now \$2 per page, and you may type with renewed vigor; but if you have been getting \$4 per page, that same discovery (that you are now getting only \$2) may lead you to quit. Clearly, our experience of reward value is relative to what we are used to receiving. From a cognitive perspective, these contrast effects—in rats as in humans—are explained by assuming that the individual (a) has learned to expect a certain reward and (b) is able to compare the actual reward with the expected one. If the comparison is favorable, the animal increases its response rate; if unfavorable, the animal decreases its rate. The animal is constantly out to do better, and if a particular response leads to less reinforcement than it did before, the animal might do better by spending less time at it and more time looking for reinforcers elsewhere. In fact, animals that have experienced a downshift in reward size do engage in increased exploration, as if looking for the larger reward (Pecoraro & others, 1999).

When Rewards Backfire: The Overjustification Effect in Humans

In humans, rewards can affect the likelihood of engaging in a particular form of behavior by changing a person's understanding of the meaning of that behavior. Consider an experiment conducted with nursery-school children (Lepper & Greene, 1978). Children in one group were rewarded with attractive "Good Player" certificates for drawing with felt-tipped pens. This had the immediate effect of leading them to spend more time at this activity than did children in the other group (who were not rewarded). Later, however, when certificates were no longer given, the previously rewarded children showed a sharp drop in their use of the pens—to a level well below that of children in the unrewarded group.

Many subsequent experiments, with people of various ages, have produced similar results. The drop in performance following a period of reward is particularly likely to occur when the task is something that is initially enjoyed for its own sake, and the reward is given in such a manner that it seems to be designed deliberately to motivate the participants into engaging in the task (Lepper & Henderlong, 2000). This decline is called the *overjustification effect*, because the reward presumably provides an unneeded justification for engaging in the behavior. The result, according to the usual cognitive interpretation, is that the person comes to regard the task as work (the kind of thing one does for external rewards such as money, certificates, or an improved résumé) rather than play (the kind of thing that one does for its own sake). When the participants come to regard the task as work, they stop doing it when they no longer receive an external reward for it, even though they would have otherwise continued to do it for fun. The overjustification effect suggests that some rewards used in schools might have negative long-term effects. For example, rewarding children for reading might in some cases cause them to think of reading as work rather than fun, which would lead them to read less on their own. The broader point is that one must take into account the cognitive consequences of rewards in predicting their long-term effects.

>>> 31
Why might a period of reward lead to a subsequent decline in response rate when the reward is no longer available?



When does play become work? Baseball can be great fun for children, but when the focus is on winning trophies and pleasing parents and coaches, what was previously play can become work.

SECTION REVIEW

Operant conditioning entails learning about conditions and consequences.

Learning When to Respond

- If reinforcement is available only when a specific stimulus is present, that stimulus becomes a discriminative stimulus. Subjects learn to respond only when it is present.
- Learners generalize to stimuli that they perceive as similar to the discriminative stimulus, but can be trained to discriminate.
- Discrimination training has been used to study perception in infants and concepts in pigeons.

Learning What to Expect

- According to the cognitive view, subjects in operant conditioning learn a means-end relationship, i.e., that a certain response has a certain effect. This view is supported by an experiment in which the motivational state of rats was varied between training and testing.
- Means-end learning helps to explain reward contrast effects and overjustification effects. In these cases, expectations about reward influence the rate of operant responding.

Activities Designed for Learning: Play, Exploration, and Observation

Classical and operant conditioning have long been popular research topics in psychology, partly because they truly are ubiquitous learning processes and partly because they are relatively easy to study in the laboratory. By tightly controlling an animal's environment and motivational state, a researcher can, more or less at will, induce predictable changes in the animal's behavior. Indeed, the very word *conditioning* implies an active researcher (or trainer) and a passive animal. The researcher *conditions* the animal. The idea that learning is imposed by external conditions rather than controlled by the learner accords closely with the philosophy of early behaviorism.

In nature, however, animals (especially mammals) are active learners. They move about in ways that are well designed for learning. Through *play* young animals learn how to control their own behavior in effective ways. Through *exploration* animals of all ages keep track of significant changes in their environment. In addition, at least some animals acquire useful information through *observing* others of their kind. Play, exploration, and observation are behavioral tendencies, or drives, that came about through natural selection precisely because they promote learning. Unlike operant conditioning as studied in laboratories, these activities occur most fully when the animal is free of any strong, immediate need state, such as hunger or thirst, and is free to move about at will.

Play as a Vehicle for Learning How to Do Things

One of the defining features of play is that it serves no useful immediate end. People and other animals engage in play for its own sake, not for some reward outside of play. In that sense, play is quite different from operant behavior. A cat seriously preying on a mouse is engaged in operant behavior; its intent is to kill and eat the mouse. In contrast, a cat *playing* at preying on a mouse may repeatedly stalk, pounce on, and paw the mouse without killing it, or at least without deliberately killing it. In play, the mouse is not food but a toy. Young mammals of nearly all species spend a good deal of time playing.

How Play Is Distinguished from Non-Play

32

How is an animal's play distinguishable from its serious (non-play) behavior?

<<< Most actions classed as play in animals are similar to behaviors that are seen in serious contexts, such as fighting an enemy, fleeing from a predator, or stalking prey. The playful versions of such actions generally differ from the serious versions in the following ways (Power, 2000; Symons, 1978):

- Play often occurs in contexts where the corresponding serious behavior would be inappropriate or impossible. There is no real enemy to fight, or predator to escape from, or prey to stalk. The ball of yarn stalked and pounced on by a playing kitten is not real prey to be eaten.
- In play, the drive state appropriate to the corresponding serious behavior is absent. The playing animal “fights” without anger, “flees” without fear, and “preys” without hunger. In general, the drive to play emerges when other drives are relatively satisfied. The kitten playfully stalking a ball of yarn is not hungry.
- Play typically involves a great deal of repetition, not seen in serious behavior. Bear cubs might repeatedly chase one another, taking turns being the pursuer and the pursued, like children playing tag. After stalking and pouncing on the ball of yarn, the playful kitten pushes the yarn away and stalks and pounces again.
- Playful behaviors often occur in sequences that differ from what would occur in their serious form. Indeed, behavioral elements drawn from several different se-

rious contexts may be intermixed randomly in a single bout of play. The kitten may alternately fight with, flee from, and stalk the ball of yarn.

- During social play, especially during playful fighting and chasing, animals often exhibit special play signals. The signals apparently serve to remind playmates that the actions, which may appear aggressive, are only play and no harm is intended. An example of such a signal is the relaxed open-mouth display, or “play face,” exhibited by most species of primates, including humans (discussed and illustrated in Chapter 3).

Play as Practice of Species-Typical Skills

What good is play? It is easy to see that play has some costs. It consumes energy, which must be replenished by increased food intake, and it involves some risks. Although young mammals (including human children) are more cautious in play than they superficially appear to be, injuries and deaths do occasionally result from vigorous play. Such play also produces noises that can attract predators. Why would evolution by natural selection have produced a drive in young mammals to engage in such apparently purposeless and costly activity? The most widely accepted answer to that question is that play evolved to promote the practice of survival skills in young mammals. This theory was first proposed and elaborated upon by the German philosopher and naturalist Karl Groos near the end of the nineteenth century.

Groos was strongly influenced by the writings of Charles Darwin and had a sophisticated, modern understanding of instincts. He recognized that animals, especially mammals, must to varying degrees *learn* to use their instincts. Young mammals come into the world with biological drives and tendencies (instincts) to behave in certain ways, but to be effective such behaviors must be practiced and refined. In his book *The Play of Animals*, Groos (1898) argued that play evolved as a means for young animals to practice the skills they need for survival throughout life. He divided play into categories related to the type of skill it promotes, including movement play (running, leaping, climbing, swinging in trees, and so on); hunting play; fighting play; and nursing play (playful care of infants).

Research findings since Groos's time have generally been consistent with his thesis. To a considerable degree, it is possible to predict what a given species of animal will play at by knowing what skills that species must develop in order to survive and reproduce. Lion cubs and the young of other predators play at stalking and chasing (Schaller, 1972); zebra colts, which are preyed upon by lions and such, play at fleeing and dodging (Fontaine, 1994); primates play at swinging from branch to branch in trees. Among species in which males fight one another for access to females (most mammalian species), young males engage in much more rough-and-tumble fighting play than do young females (Meany & others, 1985). And, at least among some species of primates, young females, but not young males, engage in playful care of infants (Goodall, 1986; Lancaster, 1971).

Also consistent with Groos's thesis is the general observation that animals that appear to have the most to learn play the most (Power, 2000). Young animals, which have not yet developed their skills, play much more than do older animals. Mammals depend more on learning and behavioral flexibility than do other animal classes, and mammals are much more playful than are other classes of animals. Among mammals, carnivores (dog-like and cat-like animals) and primates are the most playful and appear to have to learn the most to survive. Young carnivores must learn to feed themselves by capturing other animals, and that takes a great deal of practice; primates are big-brained species that depend heavily on learning

>>>

33

What is Groos's theory about the evolutionary function of animals' play, and what is some evidence supporting that theory?



A. & M. Shah / Animals Animals

Young predators at play These lion cubs are playfully practicing maneuvers that will serve them well as they grow older.

to carry out all of their life-sustaining activities. Humans are the most playful animals of all, and humans have far more to learn than do any other species. (Human play is discussed in Chapter 12.)

Exploration as a Vehicle for Learning About the Environment

34

How does exploration differ from play in its evolutionary function?

<<< Play is not the only drive that came about in evolution to promote active learning. The other great one is curiosity, or the drive to explore. Groos (1898) considered exploration to be a category of play, but most students of play and exploration today consider the two to be distinct (Power, 2000). Learning can be divided at least roughly into two broad categories—learning *to do* (skill learning) and learning *about* (information learning). Play evolved to serve the former and exploration evolved to serve the latter.

Exploration is a more primitive, widespread, and robust category of behavior than is play. Animals whose activities are rather rigidly controlled by their genetic makeup, so that they don't have much to learn in the *to do* category, must nevertheless learn *about* their environment. They must learn where food, shelter, mates, and other necessities for life and reproduction are located. Fish don't play, at least not in any way that is reliably identified as such by researchers, but they do regularly approach and explore novel objects (Wilson & others, 1993). Even insects explore to find food and other necessities of life, although they do so in rather inflexible ways (Gordon, 1995). Mammals of all species, regardless of age, explore novel environments and objects with which they are confronted.

The Nature of Mammalian Exploration

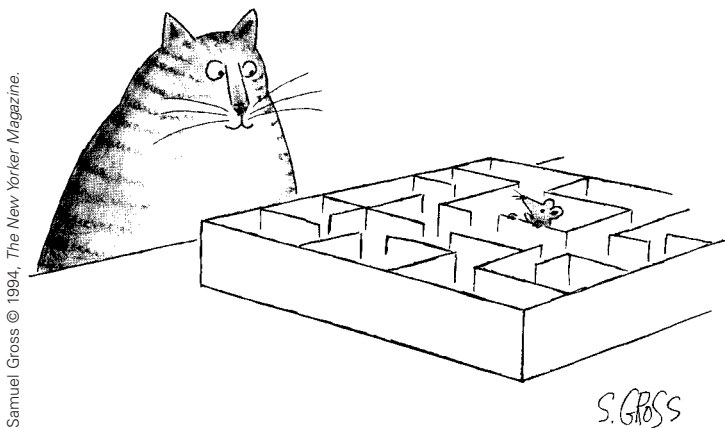
35

How do rats explore a novel environment? How did Tolman and subsequent researchers show that rats learn useful information in their exploration?

<<< Exploration, unlike play, is often mixed with a degree of fear. Exploration is elicited by novel stimuli, and novel stimuli often induce fear until they are fully explored. In fact, one purpose of exploration, in animals and people, is to determine whether or not an unfamiliar object or place is safe.

Explorers are often caught in a balance between curiosity, which drives them toward the unfamiliar terrain or novel object, and fear, which drives them away. A rat placed in a novel test arena with various objects in it will first cower in a corner. Then it will take a few steps out from the corner, along a wall, and dash back. Then it will venture a bit farther along the wall before dashing back again. Gradually, the rat will get bolder. It will eventually approach and explore—by smell, sight, and touch—the entire arena and all the objects within it. Once the rat is thoroughly familiar with the arena, it will reduce its movement but will continue periodically to tour the arena as if looking to see if anything has changed—a behavior referred to as *patrolling*. During its patrolling, the animal periodically rears up on its hind legs to get a better view. If a new object has been placed in the arena, the rat will attend to that rather than to old objects and will explore it at first gingerly and then more boldly (Inglis & others, 2001; Renner & Seltzer, 1991). Similar behaviors have been described in many other species of mammals.

Some of the earliest research on exploration came from studies of rats in mazes. Researchers learned that rats' movements through mazes are governed not just by their drive for the food in the goal box, but also by their drive to explore all of the maze's alleys (Dember & Fowler, 1958; Winfield & Dennis, 1934). Rats that have already learned the most direct route to the goal will often persist in exploring roundabout routes and dead-end alleys. Rats that are prevented from entering a particular alley, by a block placed at its entrance, will typically explore that alley first



Samuel Gross © 1994, The New Yorker Magazine.

"Well, you don't look like an experimental psychologist to me."



A curious monkey Sometimes naturalists find that the animals they study are as curious about them and their equipment as they (the naturalists) are about the animals. Here a squirrel monkey explores a camera.

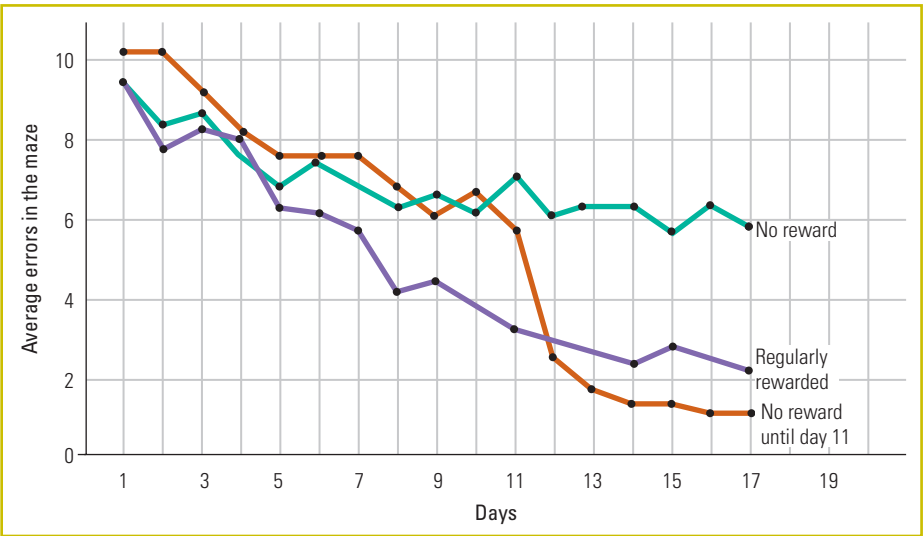
as soon as the block is removed. Not surprisingly, the hungrier a rat is, the more directly it will run to the food-containing goal box, and the less hungry it is, the more time it will spend exploring the other alleys in a maze; but even very hungry rats will often spend some time exploring (Inglis & others, 2001).

Evidence That Animals Acquire Useful Information Through Exploration

In a now classic experiment, Edward Tolman and C. H. Honzik (1930) showed that rats that are allowed to explore a maze, with no reward in the goal box, learn as much about the location of the goal box as do rats that regularly receive a reward there. These researchers tested three groups of rats in a complex maze under different reward conditions. Group 1 received one trial per day in the maze with no food or other reward in the goal box. As expected, this group showed little improvement from day to day in the time they took to reach the goal box (the goal box contained no “goal” for them). Group 2 received one trial per day with food in the goal box. As expected, this group improved considerably from day to day in their rate of movement to the goal box. The most interesting group was group 3. Rats in this group received one trial per day with no reward for 10 days, like group 1, but, beginning on the 11th day, they received one trial per day with a food reward, like group 2. These rats improved dramatically between days 11 and 12. On day 11, they were no better than the other unrewarded group (group 1), but on day 12, after just one experience with the reward, they were as fast at reaching the goal box as the rats that had been rewarded all along (see Figure 4.15).

On the basis of this and other experiments, Tolman (1948) argued that rewards affect what animals *do* more than what they *learn*. Animals learn the locations of distinctive places in

FIGURE 4.15 | Latent learning of a maze Each rat received one trial per day in the maze, with or without a food reward in the goal box. The group that received its first reward on day 11 performed as well on day 12 (and thereafter) as the group that had received a reward every day. From this, Tolman and Honzik concluded that the rats had learned the spatial layout of the maze even without a reward, but the rats did not use that knowledge until the changed conditions made it worthwhile for them to do so. (From Tolman & Honzik, 1930.)



their environment through exploration, whether or not they have ever found rewards there, but they do not run directly to those places unless they have found rewards there. Tolman used the term **latent learning** to refer to learning that is not immediately demonstrated in the animal's behavior. In the experiment just described, the rats in group 3 learned the spatial layout of the maze in the first 10 trials, but that learning remained *latent*, not manifested in their behavior, until the addition of a reward gave the rats a reason to run straight to the goal box.

More recent experiments have shown that rats very quickly learn about potential hiding places through their exploration (Leonard & McNaughton, 1990; Renner, 1988). Rats that had an opportunity to explore a novel arena that contained one or more hiding places ran much more quickly to a hiding place, when they were deliberately frightened in a later test, than did rats that had not previously explored the arena. In nature it is likely that most learning about the environment is latent. Animals learn about many aspects of their environment through exploration, but only an unpredictable portion of that knowledge becomes useful at a later time and affects the animal's behavior.

Learning by Watching Others

In everyday life, people learn a great amount by watching other people. Imagine what life would be like if such skills as driving a car or performing surgery were learned purely by trial and error! Fortunately, people learn such skills partly by observing closely and mimicking the behavior of those who have already mastered them. On a grander scale, learning by watching others seems to be a prerequisite for human culture. The skills and rituals acquired by each generation are passed on to the next not so much because the older generation deliberately trains the younger (though that is part of it) but more because members of the younger generation intently observe their elders to learn to behave as they do. Learning by watching others is called **observational learning**.

Observational Learning by Animals Other Than Humans

36

How does observation of skilled performers facilitate the learning of new operant tasks by animals? How does imitation differ from stimulus enhancement and goal enhancement, and what evidence suggests that chimpanzees can learn through imitation?

Observational learning is especially apparent in humans, but it is found in varying degrees in other creatures as well. For example, kittens can learn more quickly to press a lever for food if they have seen their mother do so than if they have not (Chesler, 1969), and dogs are better at moving around a barrier to get food if they have seen a human being do it first than if they haven't (Pongracz & others, 2001). A common view today is that such help occurs largely through stimulus enhancement and goal enhancement (Byrne & Russon, 1998; Zentall, 2003).

Stimulus enhancement refers to an increase in the salience or attractiveness of the object that the observed individual is acting upon, and *goal enhancement* refers to an increased drive to obtain rewards similar to what the observed individual is receiving. Thus a kitten that sees its mother pressing a lever for food pellets may become attracted to the lever (stimulus enhancement) and motivated to eat food pellets (goal enhancement). In this case, stimulus enhancement increases the likelihood that the kitten will press the lever, goal enhancement increases the reward value of the pellets, and the two combined increase the rate at which the kitten learns through operant conditioning to press the lever for pellets.

Stimulus enhancement and goal enhancement are cognitively simpler than imitation. *Imitation* refers to instances in which the observer reproduces the specific actions of the observed individual (Byrne & Russon, 1998). Imitation is complex because it requires that the learner remember the novel actions that were observed and then map them onto its own movement control system. Debate exists as to whether or not any mammals other than primates can imitate, but at least some primates clearly can. The most thoroughly studied primate imitators (other than humans) are chimpanzees.

Wild chimpanzees living in different groups, geographically isolated from one another, have different cultural traditions, which pass from generation to genera-

tion. For example, chimpanzees in some colonies crack hard-shelled nuts by placing a nut on a carefully selected rock and hitting it with another carefully selected rock (Mercader & others, 2002). The young in these colonies learn this difficult skill by observing their elders and practicing elements of it over a period of several months (Inoue-Nakamura & Matsuzawa, 1997). In other colonies, however, chimpanzees have never been observed to crack nuts, even though plenty of nuts and stones are available (Whiten & others, 1999). Apparently, in those colonies the discovery that nuts can be cracked was never made, so it could not be passed along. Researchers studying wild chimpanzees at seven different field stations identified 39 different behaviors, ranging from tool design to mating displays, that are distinct to specific groups and that seem to arise from cultural tradition rather than from constraints imposed by the environment (Whiten & others, 1999). In laboratory studies, captive chimpanzees have been found to observe both human and chimpanzee “tutors” closely to learn new, sometimes complex, sequences of actions to obtain rewards (Bjorklund & Bering, 2003; Hirata & Morimura, 2000; Whiten, 1998).

Bandura’s Classic Research on Observational Learning by People

Albert Bandura (1977)—the psychologist who pioneered the study of observational learning in humans—has emphasized that people observe others to learn not just specific motor skills (such as driving a car and performing surgery) but also more general modes or styles of behaving. When you enter a new situation, you probably look around to see what others are doing before doing much yourself. When you do begin to act, you may mimic rather precisely some of the actions you have observed—such as the maneuvers required to coax coffee from the unfamiliar coffeemaker in the room. But beyond that (unless you are a clown), you probably don’t imitate many of the exact actions of others. Rather, you adopt a general style of behavior that fits with what seems to be acceptable in the setting.

Bandura demonstrated both of these functions of observational learning—acquiring specific actions and learning general styles of behavior—in experiments with kindergarten children. In one experiment, one group of children observed an adult behaving very aggressively toward a large inflated Bobo doll (Bandura, 1969). The aggression included specific verbal insults and such physical acts as banging the doll with a mallet, hurling it down, kicking it, and bombarding it with balls. A second group watched the adult model behave in a gentle manner toward the doll, and a third group was not shown any model. Later, when each child was allowed to play in the room with a variety of toys, including the Bobo doll, those in the first group behaved more aggressively, and those in the second group behaved more gently, than those in the third group. The children in the first group not only mimicked many of the aggressive actions that they had observed in the adult model but also improvised many new aggressive actions of their own and directed them toward other toys as well as toward the Bobo doll. The children had learned not only specific aggressive actions but also the more general message that aggressive play was appropriate in this particular setting.

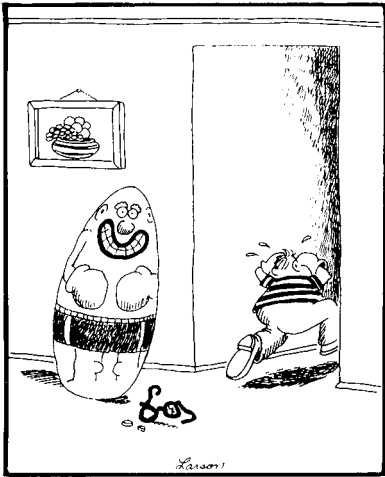
Bandura’s theory of observational learning is an explicitly cognitive one. Just as Tolman posited that rats actively explore a maze to acquire information about its contents and spatial layout, Bandura (1977, 1986) proposed that people actively observe the behavior of other people to gain knowledge about the kinds of things that people do in particular situations. As is emphasized in later chapters (especially Chapters 13 and 14), observational learning in humans goes way beyond direct



Tetsuro Matsuzawa, Primate Institute, Kyoto University. From Matsuzawa, T. (1994). Field experiments on use of stone tools in the wild. In R. Wraugham, F. de Waal, and P. Heltne (Eds.), *Chimpanzee cultures* (pp. 361–370). Cambridge, MA: Harvard University Press.

A not completely attentive observer
Chimpanzees learn through observation to crack nuts with rocks. This infant is too young to work seriously at nut cracking but may be learning something about it through watching the older master.

>>> **37**
How did Bandura demonstrate two functions of observational learning in experiments with children?



“The Far Side” cartoon is reprinted by permission of Chronicle Features, San Francisco.

imitation. People are strongly motivated to conform to the norms and expectations of others in their social group, and values as well as styles of behavior are acquired by observation.

SECTION REVIEW

In the natural environment, learning is promoted by play, exploration, and observation.		
Play • Play is common in young mammals and allows them to practice skills needed for survival. • Play often resembles serious survival activities, but differs in context, motivation, repetitiveness, sequencing, and presence of play signals.	Exploration • Exploration is more primitive than play, occurring in many more species and at all ages. It promotes learning about the environment. • Curiosity motivates exploration of novel objects and places, but is balanced by fear. • Latent learning experiments show that exploration alone, without external reward, produces useful knowledge.	Observational Learning • Stimulus enhancement and goal enhancement are simpler forms of observational learning than is imitation. • Cultural differences among chimpanzee groups apparently result from imitative ability. • Bandura demonstrated that children learn specific actions and general modes of behavior through observation.

Specialized Learning Abilities: Filling the Blanks in Species-Typical Behavior Patterns

Thus far we have been examining learning processes and activities that are quite general in the sense that they apply to many species of animals and operate in a wide variety of contexts. Animals can learn many different things through classical conditioning, operant conditioning, play, exploration, and observation. But natural selection has also endowed animals with specialized learning abilities that have quite limited domains of operation. These may be best thought of as adjuncts to particular instinctive, species-typical behavior patterns. Each such learning ability helps to mesh some aspect of the animal's instinctive behavior with particular variable characteristics of the animal's environment.

In Chapter 3, the white-crowned sparrow's ability to learn its parents' song was cited as an example of specialized learning designed to supplement a specific instinctive tendency (the instinct to sing). In Chapter 11, evidence that human language is acquired through unique, specialized learning abilities will be discussed. Now we shall look at the learning of food aversions and preferences and then, more briefly, at a few other examples of specialized learning abilities.

Special Abilities for Learning What to Eat

For some animals, learning what to eat is a relatively minor problem. Koalas, for instance, eat only the leaves of eucalyptus trees. Through natural selection, koalas evolved a food-identifying mechanism that tells them that eucalyptus leaves are food and everything else is not. That simplifies their food choice, but if eucalyptuses vanish, so will koalas. Other animals are more flexible in their diets. Most flexible of all are omnivorous creatures, such as rats and humans, which treat almost all organic matter as potential food and must *learn* what is safe to eat. Such animals have evolved special mechanisms for learning to identify healthful foods and to avoid potential poisons.

Food-Aversion Learning

If rats become ill after eating a novel-tasting food, they subsequently avoid that food (Garcia & others, 1972). In experiments demonstrating this, researchers in-

duce illness by adding a toxic substance to the food or by administering a drug or a high dose of x-rays to the animals after they have eaten. Similarly, people who by chance get sick after eating an unusual food often develop a long-term aversion to the food (Bernstein, 1991; Logue, 1988). For years as a child, I hated the taste and smell of a particular breakfast cereal, because once, a few hours after I ate it, I happened to develop a bad case of stomach flu. I knew, intellectually, that the cereal wasn't the cause of my illness, but that didn't help. The learning mechanism kicked in automatically and made me detest that cereal.

Some psychologists choose to describe such cases of food-aversion learning in terms of classical conditioning. In that description, the feeling of illness or nausea induced by the x-ray treatment or drug is the unconditioned stimulus for a reaction of aversion or revulsion, and the taste and smell of the food become conditioned stimuli for that reaction. But John Garcia, the researcher who pioneered the study of food-aversion learning, argues that such learning is quite different from standard cases of classical conditioning (Garcia & others, 1989).

One special characteristic of food-aversion learning has to do with the optimal delay between the conditioned and unconditioned stimuli. In typical cases of classical conditioning, such as the salivary reflex studied by Pavlov, conditioning occurs only when the unconditioned stimulus follows immediately (within a few seconds) after the conditioned stimulus. But food-aversion learning has been demonstrated even when the x-rays were administered as much as 24 hours after the animals had eaten the food (Etscorn & Stephens, 1973). In fact, food-aversion learning fails to occur if the gap between tasting the food and the induction of illness is less than a few minutes (Schafe & others, 1995).

Another special characteristic has to do with the sorts of stimuli that can serve as conditioned stimuli for such learning. In typical cases of classical conditioning, almost any kind of detectable stimulus can serve, but in food-aversion learning the stimulus must be a distinctive taste or smell (and taste generally works better than smell). Rats that become ill after eating a particular food subsequently avoid any food that tastes or smells like what they had eaten, even if it looks different, but they do not avoid a food that looks like what they had eaten if it tastes and smells different (Garcia & others, 1968, 1989).

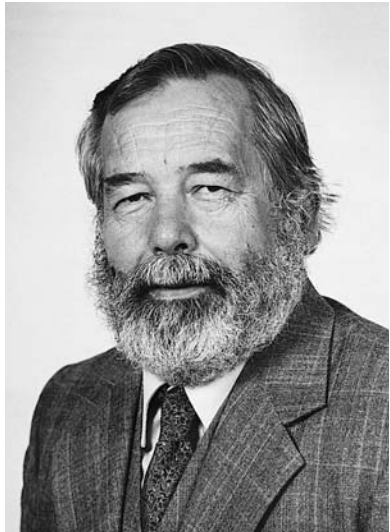
These distinguishing characteristics of food-aversion learning make excellent sense when considered in the light of the function that such learning serves in the natural environment. In general, poisons and spoiled foods do not make an individual ill immediately, but only after many minutes or several hours. Moreover, it is the chemical quality of a food, detectable in its taste and smell, not the visual quality, that affects health. For example, a food that has begun to rot and makes an animal sick may look quite like one that has not begun to rot, but its taste and smell are quite different. Thus, to be effective, a learning mechanism for food aversion must tolerate long delays and be tuned especially to those sensory qualities that correspond with the food's chemistry.

Food-Preference Learning

The other side of the coin from learning to avoid harmful foods is that of learning to choose foods that satisfy a specific nutritional requirement. Just as rats can learn to associate the taste of a food with subsequent illness and thereafter avoid that food, they may also (to a lesser extent) be able to associate a taste with a subsequent improvement in health and thereafter prefer that food.

A number of experiments have shown that when rats are deprived of a mineral (such as calcium) or a vitamin that is essential for health, they will learn to prefer the flavor of a new food that contains that mineral or vitamin (Rozin & Schull, 1988; Tordoff, 2002). In one series of experiments, researchers deprived rats of thiamine (one of the B vitamins, essential for health) for a period of time and then offered them a choice of foods, only one of which contained thiamine (Overmann, 1976; Rozin & Kalat, 1971). Each food had a distinct flavor, and thiamine—which itself has no flavor—was added to a different food for different rats. The result was

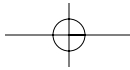
>>> **38**
What are two ways in which food-aversion learning differs from typical examples of classical conditioning? How do these differences make sense in terms of the function of such learning?



Courtesy of UCLA Media

John Garcia In the 1960s, Garcia discovered that food-avoidance learning violates certain principles of conditioning that had been accepted by mainstream psychologists as general laws of learning. He pursued this work despite the “better judgment” of his thesis advisers and even though the editor of a leading psychology journal refused to publish his early findings on the grounds that they could not be true. Garcia was a stubborn, charismatic, Spanish American son of migrant workers. One historian of psychology (Bolles, 1993) notes that Garcia was fond of saying *Viva yo*, which translates literally into “Long live me” but actually means something like “To hell with you guys.”

>>> **39**
How might rats learn which food contains a needed vitamin or mineral?



that, within a few days of experience with the foods, most rats strongly preferred the thiamine-containing food.

How did the rats “figure out” which food contained the thiamine? Close inspection of their eating patterns suggests a possible answer (Rozin & Kalat, 1971). When first presented with the choices, a rat usually ate just one or two of them. Then, typically after several hours, the rat would switch to a different food or two. Such behavior—eating just one or two foods at a time—seems ideally suited for isolating particular foods that lead to an increase or a decrease in health. If the rat had sampled all the foods at once, it would have had no basis for knowing which one had affected its health.

40

What observations with human babies are similar to results of food-selection experiments with rats? What precautions should be kept in mind in interpreting those observations?

<<< There is some evidence that humans, given the opportunity, can also learn which foods promote their health. In the 1920s, a pediatrician named Clara Davis performed a bold experiment on food selection with human infants—so bold, in fact, that it would probably not pass the ethics review board of a modern research institution (which helps explain why the experiment has never been repeated). The infants, whose mothers consented to their being in the experiment, lived in the pediatrics ward of a hospital for 6 months or longer beginning at 8 months of age. At each meal, a tray with about a dozen different foods was placed before each infant and the infants were allowed to choose their own foods and feed themselves. The choices were all natural foods—including cereals, fruits, ground meats, fish, eggs, and vegetables—but no single food contained all the required nutrients. The result was that the infants showed strong preferences for particular foods for periods up to a week, but then changed their preferences such that, over time, they ate a balanced diet (Davis, 1928). At the beginning of the experiment one of the babies had rickets, a disease caused by lack of vitamin D. He self-selected cod-liver oil, which is rich in vitamin D, until his rickets was cured, and then he stopped selecting it.

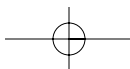
As this experiment has never been repeated, we must be cautious in interpreting it. Perhaps it was just a coincidence that the infant with rickets chose to eat cod-liver oil. Perhaps the infants changed food preferences periodically simply because, over time, they became bored with the food they had been eating. Maybe their choices had nothing to do with nutritional need. However, it is interesting that their behavior was quite similar to that of rats in food selection experiments. They ate few foods at a time, changed preferences over time, and self-selected a nutritionally balanced diet. Should this experiment tempt you to assume that your own appetite system will choose a healthy diet in all conditions, please keep in mind that all of the choices in Davis's experiment were natural and unsweetened. Even rats lose much of their ability to learn which food is good for them if one or more of the deficient foods is laced with sugar (Beck & Galef, 1989). From an evolutionary perspective, that is not surprising. The food-learning system in rats and humans evolved long before refined sugars were invented.

The Role of Social Learning in Food Selection

41

What evidence, with rats and people, points to the importance of social learning in food selection?

<<< In addition to learning from their own experiences with foods, rats learn what to eat from one another. Newly weaned wild rats generally limit their diets to foods that older rats in the colony regularly eat. Through this means, they can avoid even tasting a food that older animals have learned is poisonous (Galef & Clark, 1971) and can choose, from the beginning, a nutritious food that older animals have learned to prefer (Beck & Galef, 1989). Similar results have been found with kittens (Wyrwicka, 1996). Even in adulthood, rats are strongly influenced by one another's food choices. Bennett Galef (1990, 2002) has found that rats in a colony sniff near the mouth of a rat that has recently eaten and then show a strong preference for the food they had smelled on the demonstrator rat's breath. Through this and other means, adult rats introduced into a new colony acquire the colony's food preferences (Galef & Whiskin, 1997). The tendency to eat what others of one's kind have been eating has been demonstrated in many other species of animals as well (Galef & Giraldeau, 2001).



We humans, presumably, don't learn food preferences by smelling one another's breath, but we are certainly influenced by our observations of what those around us eat. In one experiment, children between 1 and 4 years old were more willing to taste a new food if they saw an adult eat it first than if they had never seen anyone eat it (Harper & Sanders, 1975). Other research suggests that children are most open to new foods from about 1 to 2 years of age, which is when they are most likely to be closely watched and fed by adults, and are least willing to try new foods between about 4 and 8 years of age, a time when they are often on their own but have not yet learned to distinguish foods from poisons and thus are most vulnerable to eating something harmful (Cashdan, 1994).

Summary of Rules for Learning What to Eat

Suppose that you were a wise teacher of young omnivorous animals and wanted to equip your charges with a few rules for food selection that could be applied no matter what food was available. Two that you would probably come up with are these: (1) When possible, eat what your elders eat. Such food is probably safe, as evidenced by the fact that your elders have most likely been eating it for some time and are still alive. (2) When you eat a new food, remember its taste and smell. If you don't feel sick within a few hours, continue choosing foods of that taste and smell, but if you do feel sick, avoid such foods.

Notice that these rules don't specify exactly what to eat but, instead, specify *how to learn* what to eat. The first rule describes a specific variety of observational learning, and the second describes a specific, efficient variety of associative learning. As you have just seen, rats do in fact behave in accordance with these rules, and young humans may also. Of course, we assume that these rules have been imparted not by a wise teacher of young omnivores but by natural selection, which has shaped the brain to operate in accordance with the rules.

>>> 42
In sum, what has natural selection imparted to young omnivores about food selection?



Mitch York / Tony Stone

Observational learning has its limits
Children acquire the food preferences of their culture by observing their elders, but sometimes it takes a while.

Other Examples of Special Learning Abilities

Food selection is by no means the only domain in which special learning abilities have apparently come about through evolution. Here are some other well-studied examples.

Innate Biases in Fear-Related Learning

Do you remember the demonstration by Watson and Rayner (1920), in which a young child was conditioned to fear a white rat by pairing it with a loud noise? Several years later, Elsie Bregman (1934), a graduate student working with Thorndike, tried to repeat that demonstration with one important modification. Instead of using a rat as the conditioned stimulus, she used various inanimate objects, including wooden blocks and pieces of cloth. Despite numerous attempts, with 15 different infants as subjects, she found no evidence of conditioning. What are we to make of this apparent discrepancy? One possibility, suggested by Martin Seligman (1971), is that people are biologically predisposed to acquire fears of situations and objects, such as rats and snakes, that posed a threat to our evolutionary ancestors and are less disposed to acquire fears of other situations and objects.

More recently, Susan Mineka and her colleagues showed that rhesus monkeys are not afraid of snakes when first exposed to them but very easily learn to fear them (see Figure 4.16). In one experiment, monkeys raised in the laboratory did not react fearfully to snakes until they saw a monkey that had been raised in the wild do so. After that, they showed strong fear reactions themselves (Mineka &

>>> 43
What is some evidence that people and monkeys are biologically predisposed to learn to fear some things more easily than other things?

FIGURE 4.16 | A biologically prepared learned reaction. Monkeys that have never been harmed by snakes nevertheless learn quickly to fear them through watching the fearful reactions of other monkeys.



Susan Minaka

others, 1984). In subsequent experiments, Michael Cook and Mineka (1989, 1990) used film splicing to produce videotapes in which a monkey was shown reacting fearfully in the presence of various objects, including toy snakes, flowers, and a toy rabbit. Through observing the tapes, monkeys that previously feared none of these objects developed a fear of toy snakes (and real snakes) but not of flowers or toy rabbits.

From an evolutionary perspective, this learning bias makes a good deal of sense. In some regions (depending on the kinds of snakes that are present), snakes are dangerous, and in other regions they are not. In places where snakes are harmless, an inflexible instinctive fear of them would be maladaptive. Thus, the learning mechanism may have evolved because it allows monkeys living in areas where snakes are dangerous to learn quickly to fear and avoid them, while it allows monkeys living elsewhere to go about their business relatively oblivious to snakes. We humans also vary greatly in the degree to which we fear snakes. Research suggests that we learn to fear snakes—and other objects, such as spiders and angry faces, that posed threats to our evolutionary ancestors—more readily than we learn to fear equally dangerous objects that were not present in our early evolutionary history, such as electrical outlets, sharp knives, and automobiles (Mineke & Öhman, 2002; Seligman, 1971).

Imprinting in Precocial Birds: Learning to Identify One's Mother

44

What aspects of a young fowl's ability to follow its mother depends on learning, and how is that learning guided by inborn biases?

<<< Some of the earliest evidence for specialized learning abilities came from studies of young precocial birds. Precocial birds are those species—such as chickens, geese, and ducks—in which the young can walk almost as soon as they hatch. Because they can walk, they can get separated from their mother, and because of that they have acquired, through natural selection, an efficient means to recognize their mother and remain near her. In the nineteenth century, Douglas Spalding (1873/1954) observed that when newly hatched chicks saw him, rather than their mother, move past their nest shortly after they hatched, they followed him (Spalding) as if he were their mother. They continued to follow him for weeks thereafter, and once attached in this way, they would not switch to following the mother hen. Some 60 years later, Konrad Lorenz (1935) made essentially the same discovery with newly hatched goslings. He labeled the phenomenon **imprinting**, a term that emphasizes the very sudden and apparently irreversible nature of the learning process involved.

One interesting feature of imprinting is the rather restricted **critical period** during which it can occur. Spalding (1873/1954) found that if chicks were prevented from seeing any moving object during the first 5 days after hatching and then he



Nina Leen / Life Magazine, © Time Warner, Inc.

Konrad Lorenz and followers Konrad Lorenz conducted research on imprinting and many other aspects of behavior in ducks and geese. These geese, which were hatched by Lorenz in an incubator, followed him everywhere as if he were their mother.

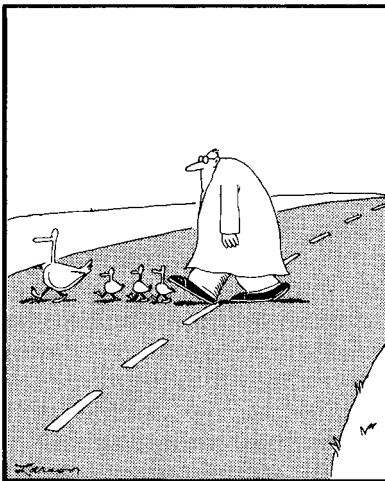
walked past them, they did not follow him but, rather, showed “great terror” and ran away. In more detailed studies, Eckhard Hess (1958, 1972) found that the optimal time for imprinting mallard ducklings is within the first 18 hours after hatching.

Although early studies suggested that young birds could be imprinted on humans or other moving objects as easily as on their mothers, later studies proved otherwise. Given a choice between a female of their species and some other object, newly hatched birds invariably choose to follow the former. Experiments with chicks indicate that this initial preference centers on visual features of the head. Newly hatched chicks will follow a box with a chicken head attached to it as readily as they will a complete stuffed chicken and more readily than any object without a chicken head (Johnson & Horn, 1988). The experience of following the object brings the imprinting mechanism into play, and this mechanism causes the chicks to be attracted thereafter to all the features of the moving object (Bateson, 2000). Under normal conditions, of course, the moving object is their mother, and imprinting leads them to distinguish their mother from any other hen. In sum, we have here a learning process for which the timing (the critical period), the stimulus features (characteristics typical of a mother bird of the species), and the behavioral response (following) are all genetically prepared in ways that promote its specific adaptive function.

Specialized Place-Learning Abilities

Many animals have specialized abilities for learning and remembering specific locations that have biological significance to them. As one example, Clark's nutcrackers (a species of bird inhabiting the southwestern United States) bury food in literally thousands of different sites, to which they return during the winter when the food is needed (Gould-Beierle & Kamil, 1999). Experiments have shown that the birds' ability to find each location depends on their memory of visual landmarks, such as stones, near the site (Kamil & Balda, 1985; Shettleworth, 1983). In other experiments, various bird species that hide seeds have been found to remember spatial locations better than do species that don't hide seeds and to have an enlargement of an area of the brain, called the hippocampus, that is crucial for spatial memory (Papini, 2002; Shettleworth & Westwood, 2002).

A quite different example of specialized place learning is the ability of Pacific salmon to return to their hatching grounds. Salmon that hatch in small streams in the northwestern United States migrate into the Pacific Ocean, where they swim



“The Far Side” cartoons by Gary Larson are reprinted by permission of Chronicle Features, San Francisco.

When imprinting studies go awry . . .

45 What special place-learning abilities have been found in (a) birds that hide food and (b) Pacific salmon? How do all examples of specialized learning mechanisms influence thought about the concept of intelligence?



© W. Perry Conway / Corbis

A seed-hiding bird Clark's nutcrackers, like many bird species, hide seeds in many different sites for the winter. Their ability to remember each hiding place is an example of a specialized learning ability.

around for 5 years or more, carrying a memory of the exact smell of the water in which they hatched. Then, when they are ready to spawn, they use their sense of smell to find their way back to the same stream from which they had come (Hasler & Larsen, 1955; Navitt & others, 1994).

So, in certain very specific ways, species of birds and fish appear to be “smarter” than chimpanzees or people. The more we understand about animal behavior, the more it becomes apparent that intelligence is a relative concept. To make the concept meaningful, we have to ask, “Intelligent at what?” Animals appear extraordinarily intelligent when we observe them in nature dealing with the kinds of learning tasks and problems for which they have been biologically equipped by natural selection. The same animals appear relatively stupid when we observe them on early trials in artificial test apparatuses, such as Thorndike's puzzle boxes or Skinner's operant conditioning chambers. The intelligence of animals comes not from a general ability to reason but from specialized learning abilities that have evolved over thousands of generations in the wild. This will be an interesting point to keep in mind when reading about human intelligence in Chapters 10 and 11. Might it be that we, too, have not one general intelligence but various specialized intelligences that evolved to solve different kinds of problems faced by our evolutionary ancestors?

SECTION REVIEW

Specialized learning abilities have evolved related to species-typical behaviors.

Choosing Food	Objects of Fear	Imprinting on Mother	Place Memory
• Rats and people avoid foods that they had eaten some minutes or hours before becoming ill. Such food avoidance learning differs in significant ways from general classical conditioning. • Rats, and possibly humans, can learn to prefer foods associated with health improvement. • Observation of what others eat influences food choice.	• We (and other species) are innately biased to learn to fear objects or situations that were threatening in the species' evolutionary past. • In experiments, monkeys learned to fear snakes, but not flowers or toy rabbits, by observing others' fearful reactions.	• Ducklings and goslings follow the first moving object they see within a critical period, and continue to follow it. • Certain characteristics of imprinting help to ensure that, under normal conditions, the young of these species will learn to identify and follow their own mothers.	• Birds that hide food in many locations have special abilities to remember where. • Pacific salmon return to streams where they hatched years before, using memory of the stream's smell.

Concluding Thoughts

In reviewing this or any chapter it is useful to think not just about the relationships among ideas within each major section, but also about the relationships among ideas across sections. One way to do that, for the present chapter, is to think about the three different perspectives on learning that are referred to at various places in the chapter: the *behavioral*, *cognitive*, and *evolutionary* perspectives. A *perspective* is a point of view, a framework, a set of ground rules and assumptions that scientists

bring to the topic studied. The perspective helps determine the kinds of questions asked, the kinds of evidence regarded as important, the kinds of studies conducted, and the vocabulary used to describe the observations. Here are some thoughts about each of the perspectives referred to in this chapter:

1. The behavioral perspective Behaviorists such as Watson and Skinner held strongly to two assumptions: (1) the assump-

tion that behavior is shaped by the environment and (2) the assumption that all aspects of behavior, including learning, are best described in terms of observable stimuli and responses without reference to unseen mental events. These assumptions led behaviorists to focus heavily on classical and operant conditioning. Conditioning, from the behaviorists' viewpoint, is something that is done to the animal (or person) by the environment. The environmental conditions that produce learning, in these cases, can be described in terms of relationships among stimuli in the environment or between responses and stimuli (including reinforcing stimuli), and learning can be quantified in terms of immediate changes in behavior (increased frequency of conditioned responses). Borrowing from Pavlov's terms for describing classical conditioning, and adding a parallel set of terms for operant conditioning, behaviorists brought to psychology a rich, objective vocabulary for talking about learning and many learning-related phenomena. That vocabulary is still very much a part of psychology today.

2. The cognitive perspective Among the pioneers of this perspective were psychologists, such as Tolman, who began as behaviorists but found that approach too limiting. They argued that you can go only so far in understanding learning (or anything else in psychology) without talking about mental processes. For example, you can establish the principle of stimulus generalization in classical and operant conditioning, but you can't predict the degree to which an individual will generalize from one stimulus to another unless you understand something about the individual's mental concepts. Depending on concepts, a person or animal can perceive two stimuli as similar even if they aren't physically similar. Using such cognitive

constructs as expectancies, predictions, and means-end relationships, cognitive psychologists have also helped solve the problem of what is learned in classical and operant conditioning. In addition, they have expanded the realm of learning research to include learning that stems from exploration and observation, which does not always manifest itself immediately in the animal's behavior. To be scientifically useful, however, cognitive constructs must make predictions that can be tested in studies of observable behavior, and most cognitive research involves such tests.

3. The evolutionary perspective This is the perspective that most clearly unites the two chapters on adaptation—the preceding one on evolution and the present one on learning. While behaviorism and cognitivism have roots in philosophy, which has traditionally tried to understand human behavior and the human mind in terms of general principles that have wide applicability (such as principles of mental associations and the law of effect), the evolutionary perspective grew out of biology, which recognizes the diversity of life processes. The view that learning mechanisms are products of natural selection implies that they should be specially designed to solve biologically important problems related to survival and reproduction. In this chapter the evolutionary perspective manifested itself most clearly in research having to do with the value of conditioning in helping animals to predict and control their environment in biologically useful ways, the role of play in motivating animals to practice life-sustaining skills, and the specialized learning processes that have evolved to serve particular needs related to survival and reproduction.

Further Reading

B. F. SKINNER (1978). *Reflections on behaviorism and society*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner—who wanted to be a novelist before he went into psychology—is always fun to read, and there is no better way to begin than with this collection of some of his essays. The titles include “Human Behavior and Democracy,” “Why I Am Not a Cognitive Psychologist,” “The Free and Happy Student,” “The Force of Coincidence,” and “Freedom and Dignity Revisited.” You will find here Skinner's basic philosophy about psychology and his suggestions for using behavioral learning principles to improve society.

JOHN D. BALDWIN & JANICE I. BALDWIN (2001). *Behavior principles in everyday life* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

This is an easy-to-read introduction to the principles of classical conditioning, operant conditioning, and observational learning, with examples taken from everyday human existence. Readers will discover many ways by which their own behavior is influenced by conditioned stimuli and reinforcers and may learn how to use learning principles to modify their own behavior in desired ways.

MICHAEL DOMJAN (2003). *The principles of learning and behavior* (5th ed.). Belmont, CA: Wadsworth.

This is one of several excellent mid-level undergraduate textbooks on learning. It deals with learning primarily from the behavioral and cognitive perspectives but also integrates into the chapters examples of species-typical and domain-specific learning abilities. Many of the ideas you have read about

in the chapter just completed are discussed in greater detail in this book.

ROBERT C. BOLLES & MICHAEL BEECHER (Eds.) (1988). *Evolution and learning*. Hillsdale, NJ: Erlbaum.

This collection of essays, each by one or more specialists in the subject, shows how the traditions of behaviorism and ethology have merged and begun to provide rich detail about species-typical learning processes. The book begins with historical chapters about the relationship of learning theory to evolutionary theory and then turns to contemporary research on learning in such biologically important domains as feeding, defending against predators, and sexual behavior.

FPO

PHYSIOLOGICAL MECHANISMS OF BEHAVIOR

Behavior is a product of the body's machinery, especially the nervous system. The nervous system receives information about the body's internal and external environments, integrates that information, and controls the body's movements. In this unit, Chapter 5 examines the structure of the nervous system and its principles of operation, and Chapter 6 is concerned with the neural and hormonal mechanisms underlying motivation and emotion.

part 3

The Neural Control of Behavior

NEURONS: CELLS THAT CREATE THE MIND

- Three Basic Varieties of Neurons, and Structures Common to Them
- How Neurons Send Messages down Their Axons
- How Neurons Are Influenced by Other Neurons: Synaptic Transmission

METHODS OF MAPPING THE BRAIN'S BEHAVIORAL FUNCTIONS

- Methods Used for Studying the Human Brain
- Methods Used for Studying the Brains of Non-Human Animals

FUNCTIONAL ORGANIZATION OF THE NERVOUS SYSTEM

- Peripheral Nerves: The Nervous System's Interface with the World
- The Spinal Cord: A Conduit and an Organizer of Simple Behaviors
- Subcortical Structures of the Brain
- The Cerebral Cortex
- Hierarchical Organization in the Control of Movement: A Summary

ASYMMETRY OF HIGHER FUNCTIONS OF THE CEREBRAL CORTEX

- Effects of Surgical Separation of the Hemispheres: Split Brain, Split Mind
- Language Areas of the Left Hemisphere

EFFECTS OF EXPERIENCE ON THE BRAIN

- If You Use It, It Will Grow
- Strengthening of Synapses as a Foundation for Learning

HOW HORMONES AND DRUGS ACT ON THE NERVOUS SYSTEM TO INFLUENCE BEHAVIOR

- Hormones
- Drugs

A human brain is, I must admit, somewhat disappointing in appearance. It is about the size and shape of a cantaloupe, but more gnarled. To the eye it seems quite dormant, even when viewed in a living person. Aristotle and many other ancient Greeks—who were among the first to try to figure out what the various parts of the body are for—were not much impressed by the brain. Noticing that the blood vessels leading into it are larger than those entering other organs, they suggested that the brain's main function was to cool the blood. They were much more impressed by the heart, an obviously dynamic organ, and proposed that the heart and blood are the source of feelings, thoughts, and all else that we today call “psychological.”

Not all of the ancients agreed with the heart theory of psychology, however. One who didn't was the Greek physician Hippocrates, whose observations of the effects of head injuries on people's thoughts and actions led him to focus on the brain. In the fourth century B.C.E., Hippocrates (1923) wrote: “From the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear. . . . Eyes, ears, tongue, hands and feet act in accordance with the discernment of the brain.” Hippocrates was right, of course, and that is why nearly every introductory psychology text, from William James's (1890/1950) classic on, contains a chapter about the nervous system, especially the brain. As psychologists and neuroscientists learn more about the brain, the nervous-system chapter becomes ever more meaningfully connected to the rest of the book, and material on the brain spills ever more copiously out of that chapter into the rest of the book. In a real sense, psychology is the study of what the nervous system does; indeed, psychology focuses on the most complex aspects of what the nervous system does.

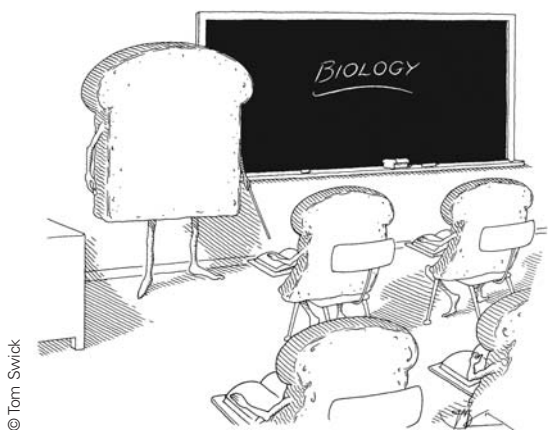
Every chapter of this book is at least indirectly about the nervous system, and this chapter is directly about it. You will read here about the mechanisms by which individual neurons act and communicate, the functions of many of the brain's major structures, how the brain is modified by experience, and how drugs and hormones act on the brain to alter behavior.

chapter 5

Neurons: Cells That Create the Mind

The brain may look inactive to the naked eye, but at a microscopic level it is the most dynamic organ of the body. It makes up about 2 percent of an average person's body weight, but consumes about 20 percent of the person's metabolic energy (Leonard, 2002). It is by far the most complex and compact computing machine in the known universe. The human brain contains roughly 100 *billion* nerve cells, or **neurons**, and roughly 100 *trillion* points of communication, or synapses, between neurons. These are all more-or-less constantly active, and their collective activity monitors our internal and external environments, creates all of our mental experiences, and controls all of our behavior.

The seeming magic of the nervous system—its abilities to analyze sensory data, create mental experiences, and control movements in adaptive ways—lies not in the individual neurons, but in the organization of the multitudes of them. Yet each neuron is itself a complex decision-making machine. Each neuron receives information from multiple sources, integrates that information, and sends its response out to many other neurons or, in some cases, to muscle cells. In this section we examine the structure and functions of individual neurons as a prelude to examining the larger structures and functions of the nervous system as a whole.



"The body is made up of millions and millions of crumbs."

Three Basic Varieties of Neurons, and Structures Common to Them

The overall layout of the human nervous system is sketched in Figure 5.1. The *brain* and *spinal cord* (which extends from the brain down through the bones of the spinal column) make up the **central nervous system**. Extensions from the central nervous system, called *nerves*, make up the **peripheral nervous system**.

Don't confuse the terms *neuron* and *nerve*. A neuron, as just defined, is a single cell of the nervous system. A **nerve** is a bundle of many neurons—or, more precisely, a bundle consisting of the axons (defined below) of many neurons—within the peripheral nervous system. Nerves connect the central nervous system to the body's sensory organs, muscles, and glands. Note, also, that despite their names, the central and peripheral nervous systems are not two separate systems, but are parts of an integrated whole.

1

What are three types of neurons, and what is the function of each?

<<< Neurons come in a wide variety of shapes and sizes and serve countless different specific functions, but at the broadest level of analysis, they can be grouped into three categories according to their functions and their locations in the overall layout of the nervous system (see Figure 5.2): (1) **Sensory neurons** carry information from sensory organs (including the eyes, ears, nose, tongue, and skin), through nerves, into the central nervous system. (2) **Motor neurons** carry messages out from the central nervous system, through nerves, to operate muscles and glands. (3) **Interneurons** exist entirely within the central nervous system and carry messages from one set of neurons to another. Interneurons bring together, organize, and integrate messages from various sources. They vastly outnumber the other two types. The human nervous system contains a few million sensory and motor neurons and roughly 100 billion interneurons. Our interneurons make sense of the input that comes from sensory neurons, generate all our mental experiences, and initiate and coordinate all our behavioral actions through their connections to motor neurons.

2

What are the main parts common to all or most neurons, and what is the function of each part?

<<< Although the shapes vary widely, most neurons share the same basic parts, which are labeled for all three types of neurons in Figure 5.2 and illustrated more fully for a motor neuron in Figure 5.3.

- The **cell body** is the widest part of the neuron. It contains the cell nucleus and other basic machinery common to all bodily cells.

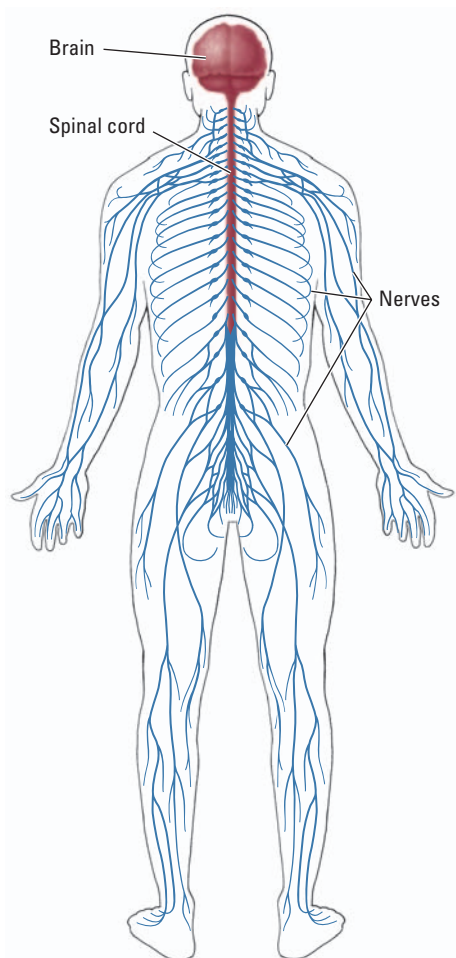


FIGURE 5.1 | Human nervous system The central nervous system consists of the brain and the spinal cord, which runs through the bones of the spinal column down the center of the back. The peripheral nervous system consists of the entire set of nerves, which connect the brain and spinal cord to the sensory organs, muscles, and glands.

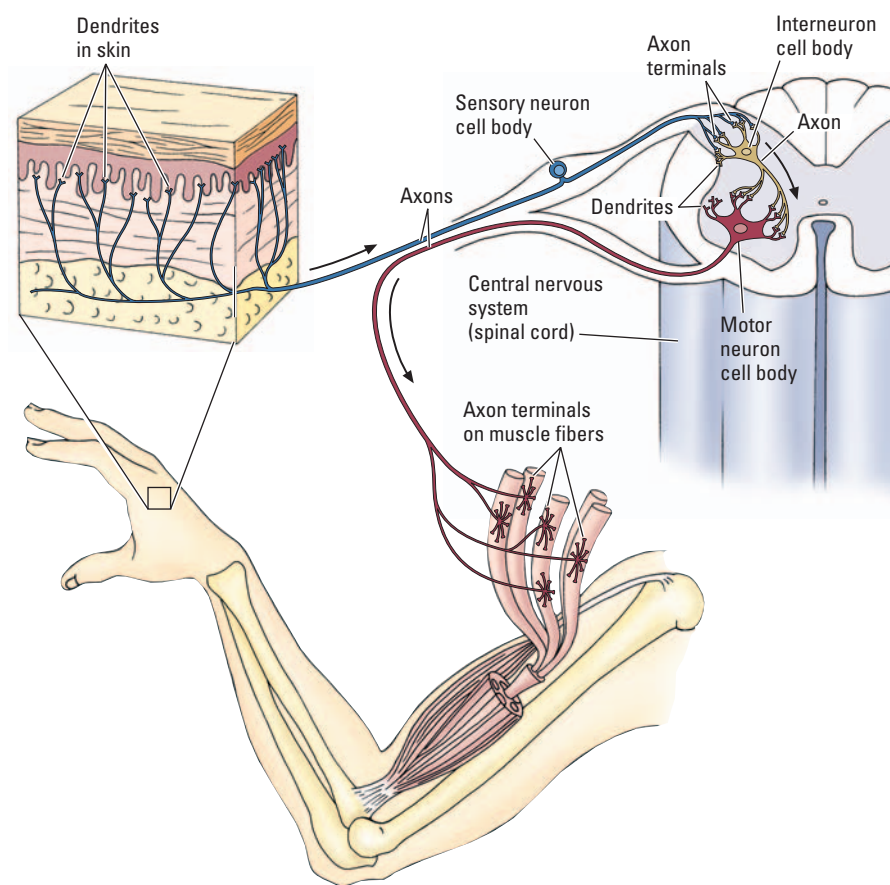


FIGURE 5.2 | The three classes of neurons This illustration shows the positions in the nervous system of the three types of neurons. On the right is the central nervous system (more specifically, a cross section of the spinal cord), and on the left are muscles and skin. Motor neurons send messages from the central nervous system to muscles and glands. Sensory neurons send messages into the central nervous system from sensory organs, such as the skin. And interneurons, located entirely within the central nervous system, carry messages between neurons.

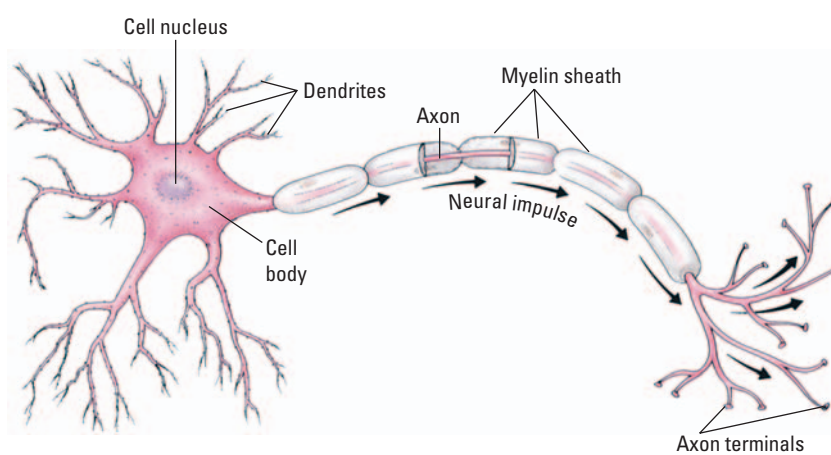
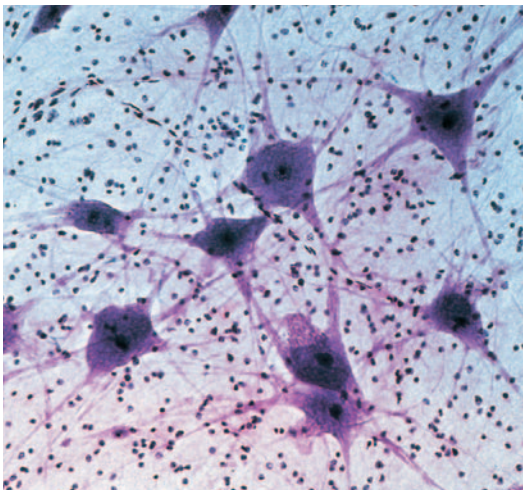


FIGURE 5.3 | A motor neuron The parts common to many neurons can be seen in this diagram of a motor neuron. The neuron receives input from other neurons on its dendrites and cell body and sends its own output down the axon to the axon terminals. The myelin sheath is not part of the neuron; it is formed of separate glial cells that are wrapped around the axon.



Manfred Kage / Peter Arnold

Motor neurons in the spinal cord
In this photograph taken through a microscope, you can see the cell bodies and some of the branching dendrites of several motor neurons.

- **Dendrites** are thin, tube-like extensions that branch extensively and function to receive input to the neuron. In motor neurons and interneurons, the dendrites extend directly off the cell body and generally branch extensively near the cell body, forming bush-like structures. These structures increase the surface area of the cell and thereby allow for receipt of signals from many other neurons. In sensory neurons, dendrites extend from one end of the axon, rather than directly from the cell body. They extend into a sensory organ and respond to sensory signals, such as sound waves in the ear or touch on the skin (shown for skin in Figure 5.2).
- The **axon** is another thin, tube-like extension from the cell body. Its function is to carry messages to other neurons or (in the case of motor neurons) to muscle cells. Although microscopically thin, some axons are extremely long. You have axons of sensory neurons extending all the way from your big toe into your spinal cord and then up to the base of your brain—a distance of about 5 feet or more. Most axons form many branches some distance away from the cell body, and each branch ends with a small swelling called an **axon terminal**. Axon terminals are designed to release chemical transmitter molecules onto other neurons or, in the case of motor neurons, onto muscle cells or glandular cells. The axons of some neurons are surrounded by a casing called a **myelin sheath**, formed from special non-neural cells that are wrapped tightly around the axon.

How Neurons Send Messages Down Their Axons

Neurons exert their influence on other neurons and muscle cells by firing off all-or-none impulses, called **action potentials**. In motor neurons and interneurons, action potentials are triggered at the junction between the cell body and the axon, and they travel rapidly down the axon to the axon terminals. In sensory neurons they are triggered at the dendritic end of the axon (shown in Figure 5.2) and travel through or past the cell body to the axon terminals. Action potentials are described as “all or none” for two reasons: (1) Each action potential produced by a given neuron is the same strength as any other action potential produced by that neuron, and (2) each action potential retains its full strength all the way down the axon.

Although each action potential is all or none, a neuron can convey varying degrees of urgency in its message by varying its rate of producing action potentials. A given neuron might fire off action potentials at a rate anywhere from zero per second to as high as 1,000 per second. By varying its rate of action potentials, a neuron varies the strength of its effect on other neurons or muscle cells.

The Resting Neuron Has a Constant Electrical Charge Across Its Membrane

To understand how action potentials are propagated, you have to know something about the functioning of the **cell membrane** that encloses each neuron. The membrane is a porous “skin” that permits certain chemicals to flow into and out of the cell, while blocking others. You can think of the neuron as a membrane-encased tube filled with a solution of water and dissolved chemicals called *intracellular fluid*. It is bathed on the outside by another solution of water and dissolved chemicals called *extracellular fluid*.

3

How does the resting potential arise from the distribution of ions across the cell membrane?

<<< Among the various chemicals dissolved in the intracellular and extracellular fluids are some that have electrical charges. These include *soluble protein molecules* (A^-), which have negative charges and exist only in the intracellular fluid; *potassium ions* (K^+), which are more concentrated in the intracellular than the extracellular fluid; and *sodium ions* (Na^+) and *chloride ions* (Cl^-), which are more concentrated in the extracellular than the intracellular fluid. For the reason described in Figure 5.4, more negatively charged particles exist inside the cell than outside. This imbalance results in an electrical charge across the membrane, with the inside typically about -70 millivolts (a millivolt [mV] is one-thousandth of a volt) relative to the

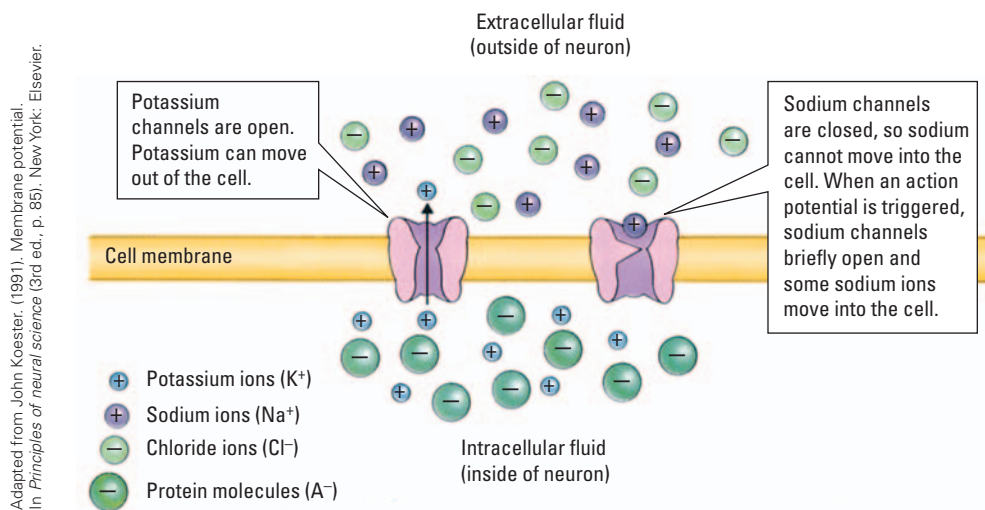


FIGURE 5.4 | The resting potential Illustrated here is a portion of a neuron's cell membrane with dissolved ions on each side. Negatively charged protein molecules (A^-) exist only inside the cell. Potassium ions (K^+) exist mostly inside the cell. Sodium ions (Na^+) and chloride ions (Cl^-) exist mostly outside the cell. Because channels in the membrane that are permeable to potassium remain open, some potassium ions diffuse out, resulting in a surplus of positive charges outside the cell and a deficit of positive charges inside. For this reason, the resting membrane has an electrical charge across it of about 70 mV, with the inside negative compared to the outside.

outside. This charge across the membrane of an inactive neuron is called its **resting potential**. Just as the charge between the negative and positive poles of a battery is the source of electrical energy in a flashlight, so the resting potential is the source of electrical energy that makes an action potential possible.

The Action Potential Derives from a Brief Change in Membrane Permeability

The action potential is a wave of change in the electrical charge across the axon membrane, which moves rapidly from one end of the axon to the other. Figure 5.5 depicts an electrical recording of the action potential, over time, at a given location on the axon.

The action potential is initiated by a change in the structure of the cell membrane: Thousands of tiny channels that permit sodium ions to pass through open up (look again at Figure 5.4). Two forces tend to drive sodium ions into the cell when the channels are open: (1) a concentration force, which occurs simply because more sodium ions exist outside the cell than inside; and (2) an electrical force, which occurs because like charges repel each other, so the positive electrical environment outside the cell pushes the positive sodium ions inward. As a result of these two forces, enough sodium moves inward to cause the electrical charge across the membrane to reverse itself and become momentarily positive inside relative to outside. This sudden shift constitutes the *depolarization phase* of the action potential (the rising part of the wave shown in Figure 5.5).

As soon as depolarization occurs, the channels that permitted sodium to pass through close, but channels that permit potassium to pass through remain open. Because potassium ions are more concentrated inside the cell than outside, and because they are repelled by the temporarily positive environment inside the cell, they are pushed outward. In this process, enough positively charged potassium ions move out of the cell to reestablish the original resting potential. This constitutes the *repolarization phase* of the action potential. As shown in Figure 5.5, the entire action potential, from depolarization to repolarization, takes less than a millisecond (one-thousandth of a second) to occur at any given point on the axon.

With each action potential, a small amount of sodium enters the cell and a small amount of potassium leaves it. To maintain the original balance of these ions across the membrane, each portion of the membrane contains a chemical mechanism, referred to as the *sodium-potassium pump*, that continuously moves sodium out of the cell and potassium into it.

How do the two phases of the action potential (depolarization and repolarization) result from the successive opening and closing of two kinds of channels in the cell membrane?

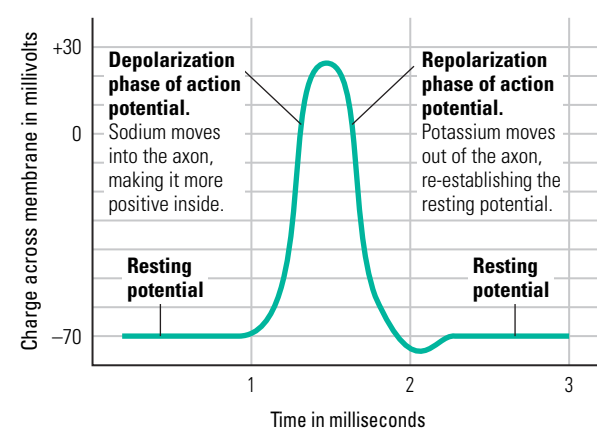


FIGURE 5.5 | Electrical change during an action potential This graph depicts the change, with time, in the electrical charge across a given point on the axon membrane as an action potential passes through that point.

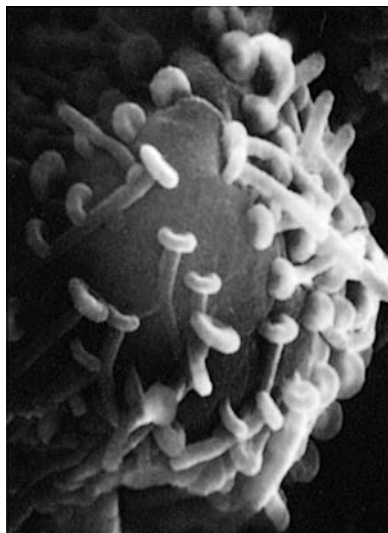
The Action Potential Regenerates Itself from Point to Point Along the Axon

Action potentials are triggered at one end of an axon by influences that tend to reduce the electrical charge across the membrane. In sensory neurons these influences derive from sensory stimuli acting on the dendrites; in motor neurons and interneurons they derive from effects of other neurons that act eventually on the axon at its junction with the cell body. The axon's membrane is constructed in such a way that depolarization (reduction in charge across the membrane) to some critical value causes the sodium channels to open, thereby triggering an action potential. This critical value (for example, to -65 millivolts inside, compared with a resting potential of -70 millivolts inside) is referred to as the cell's *threshold*. Once an action potential occurs at one location on the axon, it depolarizes the area of the axon just ahead of where it is occurring, thus triggering the sodium channels to open there. In this way the action potential keeps renewing itself and moves continuously along the axon. When an axon branches, the action potential follows each branch and thus reaches each of the possibly thousands of axon terminals.

5

How is an axon's conduction speed related to its diameter and to the presence or absence of a myelin sheath?

<<< The speed at which an action potential moves down an axon is affected by the axon's diameter. Large-diameter axons present less resistance to the spread of electric currents and therefore conduct action potentials faster than thin ones. Another feature that speeds up the rate of conduction in many axons is the presence of a myelin sheath (look back at Figure 5.3). The cells that form the sheath insulate the axon's membrane, so ions can move through it only at spaces (nodes) between adjacent cells. Each action potential skips down the axon, from one node to the next, faster than it could move as a continuous wave. The thickest and most thoroughly myelinated axons in the nervous system can conduct action potentials at a velocity of about 100 meters per second. Thus it takes about one-hundredth of a second for an action potential to run along that kind of axon from the central nervous system to a muscle about 1 meter away (a toe or finger muscle, for example). Very thin axons without myelin sheaths, in contrast, may conduct at rates as slow as 1 or 2 meters per second. When you poke your finger with a pin, you feel the pressure of the pin before you feel the pain. That is partly because the sensory neurons for pressure are large and myelinated, while those for pain are thin and mostly unmyelinated.



E. R. Lewis

FIGURE 5.6 | Axon terminals This electron micrograph shows the terminals of many axons forming synapses on a portion of the cell body of a single neuron. Synaptic vesicles, filled with neurotransmitter molecules, reside within the button-like swelling of each axon terminal. In the central nervous system, the cell bodies and dendrites of motor neurons and some interneurons are blanketed with thousands of such terminals.

How Neurons Are Influenced by Other Neurons: Synaptic Transmission

Neurons generate action potentials at rates that are influenced by all the information that is sent to them from other neurons. The cell body and dendrites of a typical motor neuron or interneuron are blanketed by tens of thousands of axon terminals, which come from the branching axons of hundreds or thousands of different neurons (see Figure 5.6). The junction between each axon terminal and the cell body or dendrite of the receiving neuron is referred to as a *synapse*. When an action potential reaches an axon terminal, it causes the terminal to release packets of a chemical substance, called a *neurotransmitter*, or *transmitter*, which alters the receiving neuron in ways that influence its production of action potentials.

Synapses can be categorized into two basic types, *fast* and *slow*. The classic, most fully studied synapses are the fast type. We shall look at them first.

Fast Synapses Quickly Excite or Inhibit Post-Synaptic Neurons

The best-understood fast synapses are those between an axon terminal and either a muscle cell or the dendrite or cell body of another neuron. The axon terminal is separated from the membrane of the cell that it influences by a very narrow gap, the *synaptic cleft*. The membrane of the axon terminal that abuts the cleft is the *pre-synaptic membrane*, and that of the cell on the other side of the cleft is the *post-synaptic membrane*. Within the axon terminal are hundreds of tiny globelike

vesicles, each of which contains several thousand molecules of a chemical neurotransmitter. All these structures are depicted in Figure 5.7.

When an action potential reaches an axon terminal, it causes some of the vesicles to spill their neurotransmitter molecules into the cleft. The molecules then diffuse through the fluid in the cleft, and some become attached to special receptors on the post-synaptic membrane. Each molecule of transmitter can be thought of as a key, and each receptor can be thought of as a lock. A molecule key entering a receptor lock opens a gate in the channel, allowing ions to pass through. If the post-synaptic cell is a muscle cell, this flow of ions triggers a biochemical process that causes the cell to contract. If the post-synaptic cell is a neuron, one of two opposite effects occurs, depending on the type of synapse (Byrne, 2003). At an **excitatory synapse** (as shown in Figure 5.7), the transmitter opens sodium (Na^+) channels. The movement of the positively charged sodium ions into the cell causes a slight depolarization of the receiving neuron (the neuron becomes less negative inside), which tends to increase the rate of action potentials triggered by that neuron. At an **inhibitory synapse**, the transmitter opens either chloride (Cl^-) channels or potassium (K^+) channels. The movement of negatively charged chloride ions into the cell or of positively charged potassium ions out of the cell causes a slight *hyperpolarization* of the receiving neuron (the neuron becomes even more

How do neurotransmitters at excitatory and inhibitory fast synapses affect the rate at which action potentials are produced in the post-synaptic neuron?

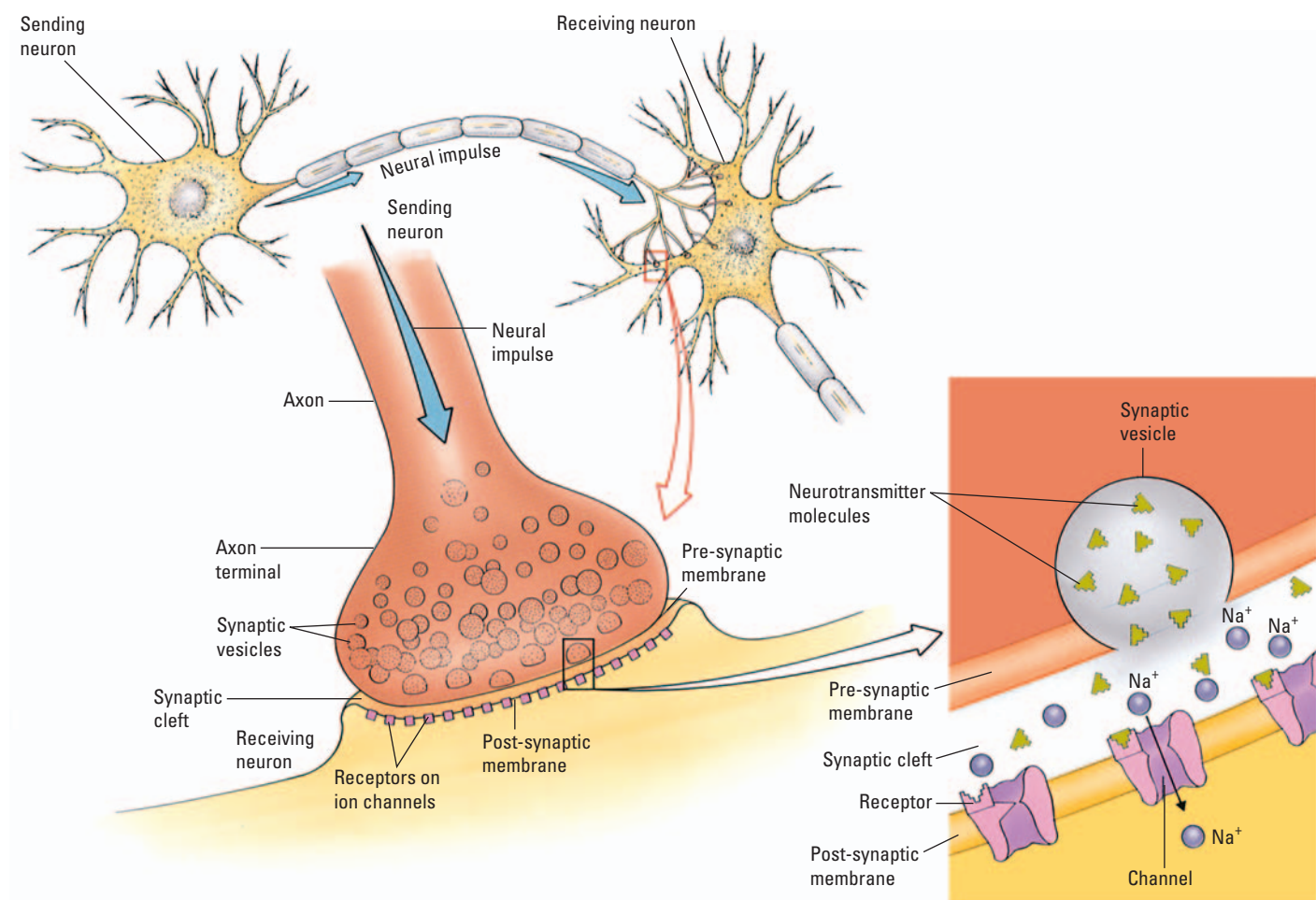


FIGURE 5.7 | Transmission across the synapse When an action potential reaches an axon terminal, it causes some of the synaptic vesicles to spill their transmitter molecules into the synaptic cleft. Some of the molecules diffuse across the cleft and bind at special receptors on the post-synaptic membrane, where they open channels that permit ions to flow through the membrane. At an excitatory synapse (as in this example), channels permeable to sodium ions (Na^+) open, allowing an influx of positive charges into the receiving neuron. At an inhibitory synapse, channels permeable to chloride ions (Cl^-) open, allowing an influx of negative charges.

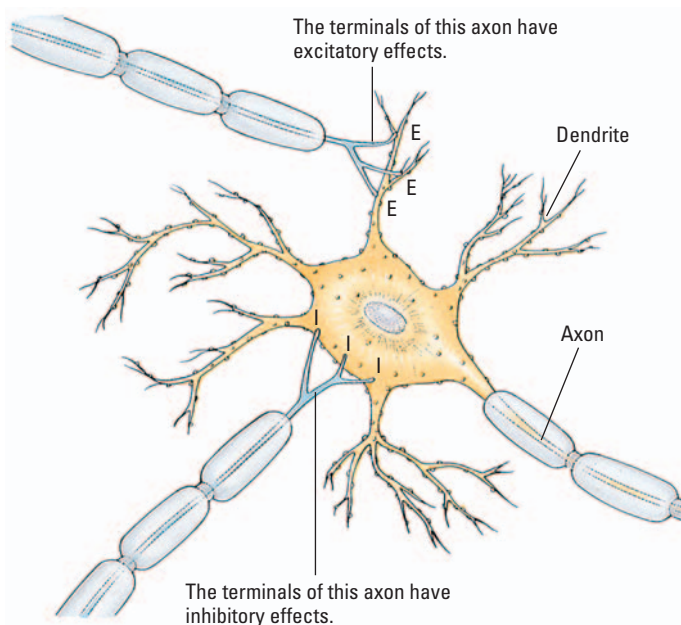


FIGURE 5.8 | Excitatory and inhibitory synapses

Neurons in the central nervous system receive many synaptic connections from other neurons, some excitatory and some inhibitory. All synapses from any single neuron are either excitatory or inhibitory.

negative inside than it was before), which tends to decrease the rate of action potentials triggered by that neuron.

All the chemical substances that have been identified as fast synaptic transmitters are small molecules (containing fewer than 10 carbon atoms) that can diffuse quickly across a synaptic gap. Many of them fall into the chemical class called *amino acids*, which are also the basic building blocks of protein molecules. The amino acid *glutamate* is the transmitter at most of the brain's excitatory fast synapses, and the amino acid *GABA* (*gamma-aminobutyric acid*) is the transmitter at most of the brain's inhibitory fast synapses (Greengard, 2001).

Post-Synaptic Neurons Integrate Their Excitatory and Inhibitory Inputs

At any given moment, a neuron may receive input at dozens, hundreds, or thousands of its fast synapses. Some of these synapses are excitatory and some inhibitory (see Figure 5.8). At each excitatory synapse the transmitter causes a slight depolarization, and at each inhibitory synapse the transmitter causes a slight hyperpolarization. These effects spread passively through the dendrites and cell body and affect the electrical charge across the membrane of the axon at its junction with the cell body.

Recall that whenever the axon membrane is depolarized below the critical value, action potentials are triggered. The greater the degree of depolarization below that value, the greater the number of action potentials per second. Thus the rate of action potentials in the post-synaptic neuron's axon depends on the net effect of the depolarizing and hyperpolarizing influences from excitatory and inhibitory synapses.

Slow Synapses Alter Post-Synaptic Neurons in Prolonged Ways

7

How does slow synaptic transmission differ from fast transmission, and what kinds of functions are served by slow transmission?

<<< At slow synapses, pre-synaptic neurons exert a wide variety of long-term effects on post-synaptic neurons. Instead of directly opening ion channels for brief periods (a few milliseconds), slow neurotransmitters trigger sequences of biochemical events in post-synaptic neurons that take time to develop and can alter the neuron's functioning for periods ranging from several hundred milliseconds up to hours, weeks, or longer (Greengard, 2001).

Within the central nervous system are dozens, and maybe hundreds, of different types of slow-acting synapses, which differ in the kinds of effects they have. At some synapses, slow-acting transmitters (like fast-acting transmitters) open ion channels and thereby produce either depolarization or hyperpolarization of the post-synaptic membrane, but they do this in an indirect way that lasts much longer than does the effect of a fast-acting transmitter. At other synapses, slow-acting transmitters cause changes in the post-synaptic neuron that do not directly alter its electrical charge, but rather alter the way the neuron responds to fast-acting transmitters. For instance, a slow-acting transmitter might stimulate the growth of new receptors on the post-synaptic membrane, which make the post-synaptic cell more responsive to the effects of one or another fast-acting transmitter. Transmitters that alter the cell in long-lasting ways such as this are sometimes referred to as **neuromodulators**. They work by modulating the post-synaptic neuron's response to other transmitters.

Researchers still have much to learn about the functions and mechanisms of neuromodulators and other slow-acting transmitters. New discoveries in this area are of great interest to psychologists, because slow-acting transmitters are responsible for changes in a person's psychological state (discussed in Chapter 6). Through their actions in various parts of the brain, slow-acting transmitters pro-

duce states of sleep, arousal, motivation (such as hunger or sexual drive), and emotion (such as fear, anger, or joy). Slow-acting transmitters also trigger the more-or-less permanent changes in neurons that provide the neural foundation for learning (discussed later in this chapter).

The nervous system contains many more types of slow-acting than fast-acting transmitter substances. Researchers have identified over 50 substances as definite slow-acting transmitters and have identified many more as probable slow-acting transmitters (Kolb & Whishaw, 2003). Many of these substances fall into a chemical class referred to as *neuropeptides*. These are relatively large molecules consisting of chains of amino acids. One subset of neuropeptide transmitters, called *endorphins*, can reduce the experience of pain through their modulating effects on pain neurons in the spinal cord and brain (discussed in Chapter 7). Another chemical class of transmitters that work primarily (if not entirely) through slow-acting means consists of the *biogenic amines*, which include *dopamine*, *norepinephrine*, and *serotonin* (Greengard, 2001). As discussed in Chapter 6, these are particularly involved in producing states of sleep, arousal, motivation, and emotion.

SECTION REVIEW

Neurons—individual nerve cells—“fire” electrochemically and communicate at synapses.		
Basic Neural Anatomy • Interneurons exist entirely within the central nervous system (brain and spinal cord). • Sensory and motor neurons connect the central nervous system to sensory organs, muscles, and glands by way of the peripheral nervous system. • A typical neuron has a cell body, dendrites (which receive input to the neuron), and an axon (which carries the neuron’s output).	Basic Functioning of the Neuron • The resting potential is an electrical imbalance that exists across the neural cell membrane when the neuron is not firing; the inside of the cell is negative relative to the outside. • An action potential involves a very brief reversal of polarization across the cell membrane, followed by a return to the resting state. These changes sweep down the axon to the axon terminals. • Action potentials are all-or-none—they do not vary in strength in a given neuron. A neuron’s rate of action potentials does vary, however.	Synapses • Neurotransmitter released from an axon terminal crosses the synaptic cleft to affect another cell. • Fast synapses have brief, immediate effects—either excitatory or inhibitory—on the postsynaptic neuron. • Slow synapses, of which there are many varieties, produce sustained effects, including effects that underlie drives and moods.

Methods of Mapping the Brain’s Behavioral Functions

With its billions of neurons and trillions of synapses, the brain is a daunting subject of study. It would be hopeless, with present technology, to try to work out all of the details of its wiring, as one might with a human-made machine such as a radio or a computer. Fortunately, though, for those of us who want to understand as much as we can of it, patterns exist in the brain’s trillions of connections.

Neurons in the central nervous system are organized into nuclei and tracts. A **nucleus** is a cluster of cell bodies in the central nervous system (not to be confused with the cell nucleus, which exists within each cell), and a **tract** is a bundle of axons that course together from one nucleus to another (see Figure 5.9). A tract in the central nervous system is analogous to a nerve in the peripheral nervous system. Tracts are referred to collectively as *white matter*, because the myelin sheaths around axons make tracts appear relatively white; nuclei, which are darker, are referred to collectively as *gray matter*.

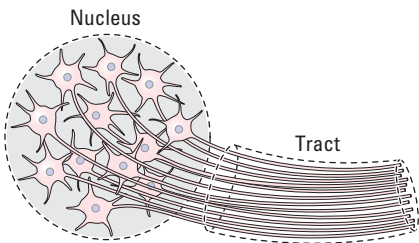


FIGURE 5.9 | Diagram of a nucleus and tract A nucleus is a cluster (usually thousands or millions) of cell bodies of neurons, and a tract is a group of axons that run parallel with one another.

In general, neurons whose cell bodies occupy the same nucleus and whose axons run in the same tract have identical or similar functions, and groups of nuclei located near one another have functions that are closely related to one another. Because of this pattern of organization, we can speak of the general functions of relatively large anatomical structures within the brain.

Psychologists and other neuroscientists have developed a number of methods of identifying the functions of specific brain areas. The methods fall into three general categories: (1) observing behavioral deficits that occur when a part of the brain is destroyed or is temporarily inactivated; (2) observing behavioral effects of artificially stimulating specific parts of the brain; and (3) recording changes in neural activity that occur in specific parts of the brain when a person or animal is engaged in a particular mental or behavioral task. What follows are brief descriptions of the most common methods used within each of these three categories. We'll look first at methods used in studying the brains of humans and then at more intrusive methods used in studying the brains of non-human animals.

Methods Used for Studying the Human Brain

8

How do researchers identify functions of areas of the human brain through (a) studying the effects of brain damage, (b) using a magnetic field to interrupt normal brain activity, (c) recording electrical activity that passes through the skull and scalp, and (d) creating images that depict patterns of blood flow?

<<< To identify the functions of specific areas of the human brain, researchers can study patients who have suffered damage to the brain, and they can use safe, non-invasive methods to alter or record ongoing brain activity.

Observing Effects of Localized Brain Damage

At the height of his career, the Russian composer V. G. Shebalin suffered a stroke (a rupturing of blood vessels in the brain) that damaged a portion of his left cerebral cortex (the highest outer-most part of the brain). From then on, he had great difficulty expressing himself in words or understanding what others were saying, but he continued to create great music. His *Fifth Symphony*, composed after his stroke, was described by the composer Dmitri Shostakovich as "a brilliant work, filled with highest emotions, optimistic and full of life" (Gardner, 1974).

As illustrated by this case, brain damage often leaves a person deficient in some areas of mental functioning yet fully capable, or nearly so, in others. By studying many people who have damage in the same general area of the brain, psychologists can draw reasonable inferences about the behavioral and mental functions of that area. The logic here is that if some abilities are missing and other abilities remain when a part of the brain is missing, then that brain part must contribute in some essential way to the missing abilities but not to the abilities that remain. As you will discover later in this chapter and elsewhere in the book, much has been learned about the human brain through systematic studies of people who have suffered from brain damage.

Observing Effects of Magnetic Interference with Normal Brain Activity

A relatively new procedure for localizing functions in the human brain is **transcranial magnetic stimulation**, or **TMS** (Pascual-Leone & Walsh, 2003). A pulse of electricity sent through a small copper coil induces a magnetic field around the coil. If the coil is held just above a person's scalp, the magnetic field passes through the scalp and skull and induces an electric current in the neurons immediately below the coil. Repetitive pulses cause the neurons to lose temporarily their ability to fire normally. The effect is comparable to that of producing a small area of damage to the brain, but the effect is temporary and reversible. For instance, if the coil is held over an area of the brain that is crucial for fluent speech, the person becomes unable to speak fluently while the current is on but then resumes fluent speech as soon as the current is turned off. Because the magnetic field affects only that part of the brain that lies immediately below the skull, TMS can be used only for mapping the functions of the outermost part of the brain, the cerebral cortex.

Although TMS is usually used to study the effects of temporary *inactivation* of an area of the brain, it can also be used to study the effects of temporary *activation* of a brain area. A single pulse of current can cause the sets of neurons just below the coil to send a burst of action potentials through their axons. When the coil is held over an area of the brain responsible for controlling the body's muscle movements, the result is a brief muscle twitch somewhere in the body—such as in the thumb of one hand. In this way, TMS can be used to produce a map showing the functional connections between specific areas of the cerebral cortex and the muscles controlled by those brain areas.

Recording Brain Activity Electrically, Through the Scalp

The constant activity of the brain's billions of neurons produces continuous electrical “chatter,” which to some degree penetrates the overlying skull and scalp. By placing electrodes on a person's scalp, researchers can detect and amplify these signals. The resulting record of brain activity is referred to as an **electroencephalogram** [e-**lec**-trow-en-**sef**-uh-low-gram] (abbreviated **EEG**). As discussed in Chapter 6, patterns in the EEG can be used as an index of whether a person is highly aroused, or relaxed, or asleep, and can be used to identify various stages of sleep.

With modern computer technology, it is possible to record and analyze brain activity simultaneously from many electrodes, each at a different location on the scalp, to see how brain activity varies from place to place as well as from instant to instant (see Figure 5.10). In some of these studies, subjects are asked to respond to stimuli that are presented to them as their brain activity is being recorded. For instance, a person might be presented with sets of letters flashed on a screen and instructed to push a button if and only if the letters spell out an English word. The brief change in the EEG record immediately following the stimulus is referred to as an *event-related potential*, or *ERP*. By testing the person repeatedly in such a task and averaging the EEG records obtained, it is possible to produce an average ERP for each electrode location. Comparison of the average ERPs recorded at different scalp locations reveals the pattern of activity in the brain as the person detects and responds to the stimulus. For the task just described, areas of the brain that are involved in seeing the letters would respond first, and areas that are involved in reading and making judgments would respond a split second later.

Viewing Brain Activity with Imaging Methods Sensitive to Blood Flow

Some methods for localizing brain activity rely on the fact that increased neural activity in any area of the brain is accompanied by increased blood flow to that area. Like other bodily tissues, the brain is permeated by tiny blood vessels. When a portion of the brain becomes more active, blood vessels there immediately enlarge, so more blood enters that portion. The blood carries oxygen and glucose, which are the sources of energy required to sustain the increased neural activity. Using technically complex methods, researchers can create three-dimensional pictures, referred to as *neuroimages*, that depict the relative amount of blood flowing through each part of the brain. The assumption, now well validated, is that increased blood flow reflects increased neural activity (Attwell & Iadecola, 2002).

The first of these neuroimaging methods (developed in the 1970s) was **positron emission tomography** [tah-mah-graf-ee], or **PET**. This method involves injecting a radioactive substance into the blood (in an amount that is not dangerous to the subject) and a method of measuring the radioactivity that is emitted from each portion



AJ Photo / Photo Researchers, Inc.

FIGURE 5.10 |
Electroencephalography The brain waves depicted on the screen reflect the electrical activity of neurons existing below each the areas of the scalp where electrodes have been placed.



© Lester Lefkowitz / Corbis

FIGURE 5.11 | Functional magnetic resonance imaging (fMRI) When the subject is in the scanner, the researchers will be able to communicate with him using an intercom system and a visual projection system. The image of the brain depicts, with colors of the rainbow, the amount of blood flow in each part of the brain, which indicates the amount of neural activity in each part.

of the brain. Another method (first used extensively in the 1990s) is **functional magnetic resonance imaging**, or **fMRI**. This method involves the creation of a magnetic field around a person's head, which causes hemoglobin molecules that are carrying oxygen in the blood to give off radio waves of a certain frequency, which can be detected and used to assess the amount of blood in each part of the brain. With either method, the person's head must be surrounded by a set of sensors, and with fMRI also by magnetic coils (see Figure 5.11). However, the person can communicate with the experimenter through a microphone and can respond to visual stimuli presented on a screen inside the scanning device. With either PET or fMRI, a computer is used to generate a three-dimensional image of the brain that depicts variations in blood flow as variations in color.

Unlike the EEG, both PET and fMRI can depict activity anywhere in the brain, not just on the surface near the skull. These methods also produce a more fine-grained picture of the spatial locations of activity than is possible with EEG. Today fMRI is generally preferred over PET, partly because it shows better spatial resolution.

At any given time, the brain is doing many things, and all portions of it are always active to some degree. To determine, with either PET or fMRI, which brain areas are most directly involved in a particular task, researchers need to employ an appropriate control condition. By subtracting the amount of activity measured in each brain area in the control condition (when the person is not performing the specific task) from the amount measured in the same areas in the experimental condition (when the person is performing the task), the researcher can determine which areas show the greatest increase in activity during the task. (For an example of such an experiment, and to see sample PET images, you might look ahead to Figure 5.27 on p 164.)

Methods Used for Studying the Brains of Non-Human Animals

9

How do researchers damage, stimulate, and record from neurons in specific areas of animal brains to learn about the functions of those brain areas?

<<< With non-human animals, researchers can localize brain functions using methods that are more intrusive than those used with humans. They can destroy, stimulate, or record neural activity in small, well-localized areas anywhere in the brain in order to assess the behavioral functions of those areas.

Observing Effects of Deliberately Placed Brain Lesions

Areas of damage (referred to as *lesions*) can be produced in the brains of rats or other laboratory animals by either electrical or chemical means. To produce a lesion electrically, a thin wire electrode, insulated everywhere except at its tip, is surgically inserted into the brain with the help of a *stereotaxic instrument* (see Figure 5.12), and enough current is sent through the electrode to destroy the neu-

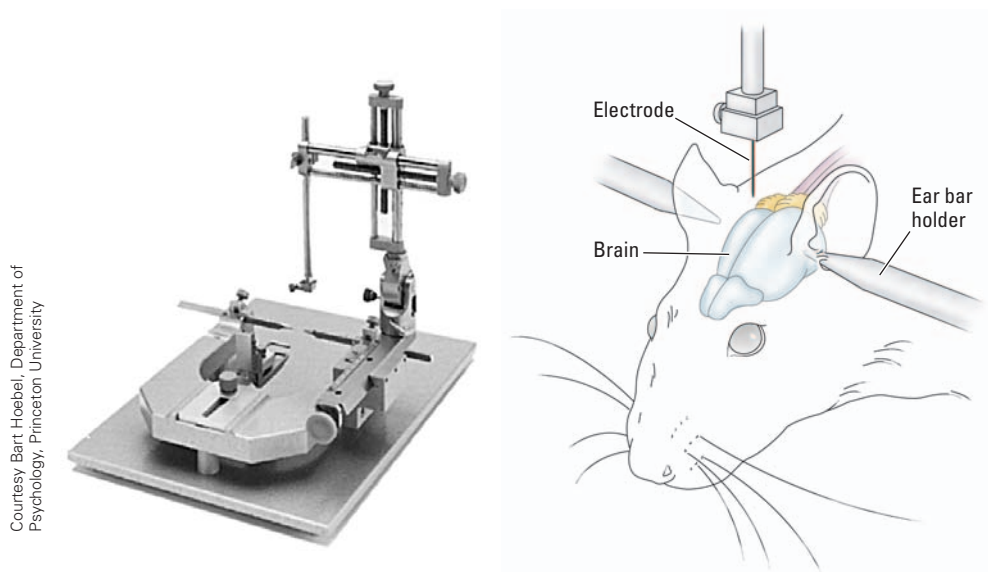


FIGURE 5.12 | Method for making lesions in or stimulating a rat's brain A stereotaxic instrument (left) is used to insert an electrode into a precise location in the anesthetized (unconscious) animal's brain (right). To produce a lesion, an electric current that is strong enough to destroy neurons near the tip is sent through the electrode, and then the electrode is removed. To prepare the animal for electrical brain stimulation, the electrode is cemented in place so that it can be stimulated through wire leads during behavioral tests after the animal has recovered from surgery.

rons adjacent to its tip. To produce a lesion chemically, a tiny tube called a *cannula* is inserted into the brain, again using a stereotaxic instrument, and a small amount of a chemical is injected through the cannula, destroying neurons whose cell bodies are located near the cannula's tip. The lesions produced by such means can be as small as one-fourth of a millimeter in diameter. By varying the precise location of the damage in different groups of animals, and then comparing the groups in behavioral tests, researchers can identify precise areas of the brain that are crucial for specific types of behaviors.

In most such studies, the lesions are made in deep, primitive areas of the brain, areas whose functions are similar for all mammals, including humans. Such studies have helped researchers identify specific brain nuclei that are crucial for basic motivational and emotional states, such as hunger, sexual drive, and fear (all discussed in Chapter 6).

Effects of Stimulating Specific Areas of the Brain

Stimulation of specific brain areas can also be accomplished either electrically or chemically. To stimulate neurons electrically, a wire electrode is lowered into the brain using the same surgical method as that used to produce a lesion. The electrode is then cemented in place (as shown in Figure 5.12) and can be activated at any time after surgery, through a wire connection or by radio waves. The electrical current used for stimulation is much weaker than that used for producing a lesion; it is strong enough to induce action potentials in, but not destroy, neurons near the electrode tip. To stimulate neurons chemically, a cannula is permanently implanted in the brain, and shortly before behavioral testing a tiny amount of a neurotransmitter substance or other chemical known to activate neurons is injected through it. Electrical or chemical stimulation in certain deep areas of the brain can cause an animal to exhibit drive states or emotional states that are bound in time to the period of stimulation (discussed in Chapter 6). For instance, stimulation in one area causes a previously sated animal to eat, and stimulation in another area causes a previously calm animal to behave as if frightened, as long as the stimulation continues.

Electrical Recording from Single Neurons

Electrodes can be used not just to destroy or stimulate specific brain areas, but also to record neural activity in specific areas as the animal engages in some behavioral task. Extremely thin microelectrodes, with very fine points that can penetrate into the cell bodies of single neurons, can be permanently implanted in the brain. In some experiments many such electrodes are inserted at once, each able to record the activity of a different single neuron.

Experiments using this technique have revealed some remarkable correlations between an animal's behavior and the rate of action potentials produced by individual neurons. For example, certain neurons in one area of the brain (the hippocampus) fire rapidly when, and only when, the animal (usually a rat) faces in a specific direction within the environment in which it is tested (Best & others, 2001). If the entire test environment (such as an enclosed box with a maze in it) is rotated, the preferred orientations of these neurons rotate too. Thus a cell that originally fired whenever the animal faced north will fire whenever the animal faces south after the apparatus is turned 180 degrees. Apparently, these neurons, called *place cells*, help animals keep track of the directions they are facing within familiar environments. If the area of the brain in which these cells are located is destroyed, animals lose their ability to find their way through mazes or to locate objects easily in environments that they have explored.

SECTION REVIEW

A number of techniques allow the brain's behavioral functions to be mapped.		
Some Fundamentals • Neurons in a given nucleus and tract have the same or similar functions, and groups of neighboring nuclei generally have closely related functions. • To study brain function, we can observe behavior when a specific brain area cannot function or is artificially stimulated. We can also measure neural activity.	Human Studies • One method is systematic study of the effects of existing localized brain damage. • A non-invasive technique, TMS temporarily disrupts or briefly triggers activity in specific cortical areas by means of a magnetic field. • An EEG records gross electrical activity in areas of the brain just beneath the skull from electrodes placed on the scalp. • PET and fMRI produce images that depict changes in neural activity in each area of the brain by measuring changes in blood flow.	Animal Studies • Precisely placed lesions in the brain can be created electrically or chemically to see how damage affects behavior. • Electrical or chemical stimulation of a specific brain area can help to reveal its functions. • Microelectrodes permit the recording of electrical activity in single neurons under varying conditions.

Functional Organization of the Nervous System

So far in this chapter we have examined the basic workings of individual neurons and have overviewed the main methods for learning about the functions of large groups of neurons in the brain. Now it is time to tour the whole nervous system. We shall look into each of its major divisions with an eye toward understanding its role in the broad task of governing the person's behavior.

As we proceed, it will be helpful to keep in mind that the nervous system is hierarchically organized. Indeed, it contains two distinct but interacting hierarchies. One, the *sensory-perceptual hierarchy*, is involved in data processing. It receives sensory data about a person's internal and external environment, and it analyzes those data to make decisions about the person's bodily needs and about threats and opportunities in the outside world. The flow of information in this hierarchy is primarily from bottom (sensory receptors) to top (perceptual centers in the brain). The other hierarchy, the *motor-control hierarchy*, is involved in control of movement. The flow of information here is primarily from top to bottom. At the top of this hierarchy are executive centers that make decisions about the activities that the person as a whole should engage in, and at lower levels are centers that translate those decisions into specific patterns of muscle movement. Most of the large

anatomical divisions of the nervous system are involved in both hierarchies, so we shall examine both at once. We'll pay most attention here, though, to the motor hierarchy; the sensory-perceptual system is discussed in detail in Chapters 7 and 8. Let's start at the bottom and work our way upward.

Peripheral Nerves: The Nervous System's Interface with the World

The peripheral nervous system, as you may recall, consists of the entire set of nerves, which connect the central nervous system to the body's sensory organs, muscles, and glands. Nerves are divided into two classes that correspond to the portion of the central nervous system from which they protrude. **Cranial nerves** project directly from the brain, and **spinal nerves** project from the spinal cord. Like most other structures in the body, nerves exist in pairs; there is a right and a left member in each pair. Humans have 12 pairs of cranial nerves and 31 pairs of spinal nerves. With their many branches, these nerves extend to all portions of the body.

Three pairs of cranial nerves contain only sensory neurons and five pairs contain only motor neurons, but the other four pairs of cranial nerves and all 31 pairs of spinal nerves contain both sensory and motor neurons. The total number of sensory and motor neurons combined is "only" about five or six million, a tiny fraction of the roughly 100 billion neurons of the nervous system as a whole. Yet these neurons are crucially important. Without them, the rest of the nervous system would be useless, because it would have no way of receiving sensory input or of controlling the person's actions.

Sensory Neurons Provide Data Needed for Governing Behavior

As noted earlier, sensory neurons are activated, at their dendritic ends, by the effects of sensory stimuli (such as light in the eye, chemicals on the tongue or in the nose, sound pressure waves in the ear, or pressure on the skin). They send their action potentials all the way into the central nervous system by way of their very long axons. The rates and patterns of action potentials in sensory neurons are the data that perceptual areas of the central nervous system use to figure out the state of the external and internal environment. Without such input, the central nervous system would have no information on which to base its behavior-controlling decisions.

Sensory input from the specialized sensory organs of the head—the eyes, ears, nose, and tongue—enters the brain by way of cranial nerves. Sensory input that comes from the rest of the body—from the skin, muscles, tendons, and various internal organs—enters the central nervous system by way of all of the spinal nerves and some of the cranial nerves. The sensations conveyed by these inputs, which include touch and pain, are referred to collectively as **somatosensation**. *Soma* means "body," and somatosensation is the set of sensations that derive from the whole body as opposed to those that come just from the special sensory organs of the head. (A full discussion of the processes of sensation and perception is found in Chapters 7 and 8.)

Motor Neurons Are the "Final Common Path" for Control of Behavior

Motor neurons, as noted earlier, have their cell bodies in the central nervous system and send their long axons out, by way of cranial or spinal nerves, to terminate on muscles or glands. Ultimately, all of the behavioral decisions of the nervous system are translated into patterns of action potentials in the axons of motor neurons, and those patterns determine our behavior. In the words of the pioneering neurophysiologist Charles Sherrington (1906), motor neurons are the "final common path" of the nervous system. Through them, and only through them, can the

nervous system control behavior. All of the brain's calculations—including those that we experience consciously as perceptions, thoughts, emotions, desires, and intentions—would be useless if they could not act on muscles and glands. The 100 billion neurons of the central nervous system are all involved, ultimately, in controlling the two or three million motor neurons, which in turn control behavior.

The Motor System Includes Skeletal and Autonomic Divisions

10 How do the autonomic and skeletal motor systems differ from one another in function? How do the sympathetic and parasympathetic divisions of the autonomic system differ from one another in function?

<<< Motor neurons act on two broad classes of structures. One class consists of the *skeletal muscles*, the muscles that are attached to bones and that produce externally observable movements of the body when contracted. The other class consists of the *visceral muscles* and *glands*. Visceral muscles are muscles that are not attached to bones and do not move the skeleton when they contract. They form the walls of such structures as the heart, arteries, stomach, and intestines. Glands are structures that produce secretions, such as the salivary glands and sweat glands. Neurons that act on skeletal muscles make up the **skeletal** portion of the peripheral motor system, and those that act on visceral muscles and glands make up the **autonomic** portion.

Whereas skeletal motor neurons initiate activity in the skeletal muscles, autonomic motor neurons typically *modulate* (modify) rather than initiate activity in the visceral muscles. Skeletal muscles are completely inactive in the absence of neural input, but visceral muscles have built-in, non-neural mechanisms for generating activity. The heart continues to beat and the muscular walls of such structures as the intestines and arteries continue to contract in response to local influences, even if all the nerves to these organs are destroyed. Most visceral muscles and glands receive two sets of neurons, which produce opposite effects and come from two anatomically distinct divisions of the autonomic system: the *sympathetic* and *parasympathetic* (see Figure 5.13).

The **sympathetic division** responds especially to stressful stimulation and helps prepare the body for possible “fight or flight.” Among its effects are (a) increased heart rate and blood pressure, (b) the release of energy molecules (sugars and fats) from storage deposits to permit high energy expenditure, (c) increased blood flow to the skeletal muscles (to help prepare them for action), and (d) inhibition of digestive processes (which helps explain why a heated argument at the dinner table can lead to a stomachache). Conversely, the **parasympathetic division** serves regenerative, growth-promoting, and energy-conserving functions through effects that include the opposites of those just listed for the sympathetic division. If you are relaxed while reading this book, your parasympathetic activity probably predominates over your sympathetic, so your heart is beating at a slow, normal rate and your digestion is working fine. If you are cramming for a test that is coming up in an hour or so, the opposite may be true.

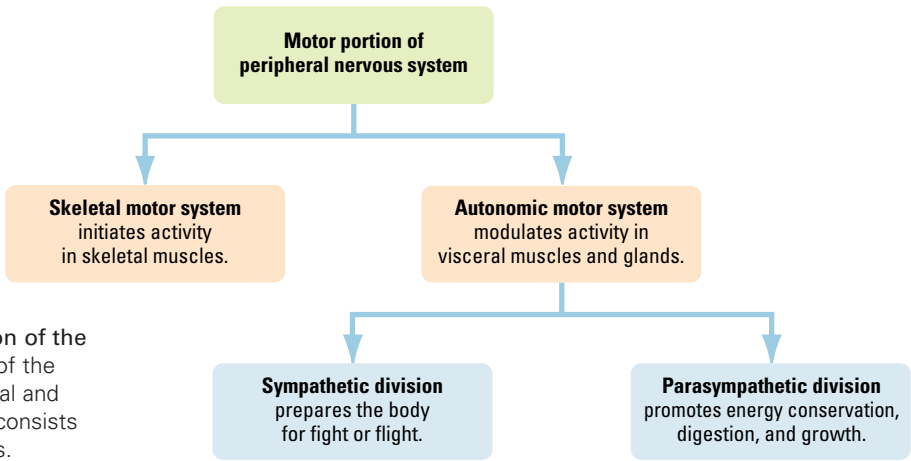


FIGURE 5.13 Divisions of the motor portion of the peripheral nervous system The motor portion of the peripheral nervous system consists of the skeletal and autonomic systems, and the autonomic system consists of the sympathetic and parasympathetic systems.

The Spinal Cord: A Conduit and an Organizer of Simple Behaviors

The spinal cord (depicted in Figure 5.1, on p. 135) connects the spinal nerves to the brain, and it also organizes some simple reflexes and rhythmic movements.

The Spinal Cord Contains Pathways to and from the Brain

The spinal cord contains *ascending tracts*, which carry somatosensory information brought in by the spinal nerves up to the brain, and *descending tracts*, which carry motor control commands down from the brain to be transmitted out by spinal nerves to muscles. A person whose spinal cord is completely severed will be completely paralyzed and lacking in sensation in those parts of the body that are innervated by nerves that come from below the place of injury. The closer the place of injury is to the head, the greater is the number of spinal nerves cut off from the brain and the greater is the extent of paralysis and insensitivity. Thus if the spinal cord is cut through just below the brain, the paralysis and insensitivity will include the arms, trunk, and legs; but if the cut is farther down, only the legs may be paralyzed.

The Spinal Cord Organizes Simple Reflexes

Some reflexive behaviors do not require the brain; they are organized by the spinal cord alone. Such reflexes are most clearly demonstrated in animals whose spinal cords have been surgically separated from the brain. (Experiments of this sort may seem cruel, but the knowledge gained from them has been extremely valuable in helping people who have spinal cord injuries.) Animals that have had this operation—referred to as *spinal animals*—still have both a brain and a spinal cord, but these structures no longer communicate with each other.

If the paw of a spinal cat is pricked with a pin, the animal does not hiss or show facial signs of pain, as a normal cat would, because the stimulus input cannot reach the pain and vocalization centers of the brain. The cat cannot feel sensations from below the neck because feeling is mediated by the brain. Nevertheless, the animal's paw quickly withdraws from the pin. This reflex is called the *flexion reflex* because it involves contraction of the flexor muscles—the muscles that bend the limb at each joint, causing it to be pulled inward (flexed) toward the body. The adaptive advantage of the flexion reflex is obvious: It quickly and automatically moves the limb away from potentially damaging stimuli. In the intact, normal cat, this reflex occurs quickly, even before the cat shows evidence of feeling pain. It occurs quickly precisely because it occurs at the level of the spinal cord and does not require that messages be sent up to the brain and then back down again.

The Spinal Cord Contains Pattern Generators for Locomotion

If you have ever witnessed a freshly decapitated chicken flapping its wings and running around the barnyard, you know that the spinal cord is capable of generating sustained, organized movements that are not immediate responses to stimuli. The spinal cord contains networks of neurons that stimulate one another in a cyclic manner and thereby produce bursts of action potentials that wax and wane in a regular, repeating rhythm. These networks, called *pattern generators*, activate motor neurons in the spinal cord in such a way as to produce the rhythmic sequence of muscle movements that results in walking, running, flying (in birds), or swimming (in fish). In some animals (but not in humans), the pattern generators become active when released from the brain's inhibitory control over them, which accounts for the flying and running motions of the headless chicken. Normally, in

>>> 11
What are three categories of functions of the spinal cord?



A courageous advocate for research Christopher Reeve, who had played Superman and other leading movie roles, became paralyzed in 1995, when he was thrown from a horse and broke his neck. He founded the Christopher Reeve Paralysis Foundation and became a leading advocate for research on spinal cord regeneration. The foundation supports both basic and applied research aimed at helping paralyzed people regain control of their bodies. Reeve died in 2004.

intact animals, pattern generators are controlled by neurons descending from the brain. They can be either inhibited, producing a motionless animal, or activated to varying degrees, producing varying rates of locomotion (Pearson & Gordon, 2000).

Subcortical Structures of the Brain

We now leave the spinal cord and enter the brain itself. The lower, more primitive parts of the brain are referred to as *subcortical* structures because of their position beneath the cerebral cortex, the topmost part of the brain. Working our way upward from the bottom, we begin our tour of the brain with the brainstem.

The Brainstem Organizes Instinctive Behavior Patterns

12

How is the brainstem similar to and different from the spinal cord? What role does the brainstem play in the control of behavior?

<<< As it enters the head, the spinal cord enlarges and becomes the **brainstem**. The parts of the brainstem, beginning closest to the spinal cord and going upward toward the top of the head, are the **medulla**, **pons**, and **midbrain** (see Figure 5.14). The brainstem is functionally and anatomically quite similar to the spinal cord, but is more elaborate. The spinal cord is the site of entry of spinal nerves, and the brainstem is the site of entry of most (10 of the 12 pairs) of the cranial nerves. The spinal cord contains ascending (sensory) and descending (motor) tracts connecting nerves to higher parts of the central nervous system, and so does the brainstem. Also like the spinal cord, the brainstem has some neural centers that organize reflexes and certain instinctive behavior patterns.

The medulla and pons organize reflexes that are more complex and sustained than spinal reflexes. They include *postural reflexes*, which help an animal maintain balance while standing or moving, and certain so-called *vital reflexes*, such as those that regulate breathing rate and heart rate in response to input signaling the body's metabolic needs. The midbrain contains neural centers that help govern most of an animal's species-typical movement patterns, such as those involved in eating, drinking, attacking, or copulating (Klemm, 1990). Also in the midbrain are neurons that act on pattern generators in the spinal cord to increase or decrease the speed of locomotion (Pearson & Gordon, 2000).

An animal (such as a cat) whose central nervous system is cut completely through just above the brainstem, referred to as a *brainstem animal*, is a fascinating creature to watch. It can produce most of the instinctive behaviors that a normal animal can produce (Klemm, 1990; Schmidt, 1986). It can walk, run, jump, climb, groom itself, attack, produce copulatory movements, chew, swallow, and so on. Unlike a normal animal, however, it makes these responses only when provoked by immediate stimuli; it does not behave in either a spontaneous or a goal-directed manner. If placed on

a pole, for example, the animal automatically climbs, but it does not itself *choose* to climb a pole that has food at the top or avoid one that does not. The animal behaves like a machine that responds to certain triggers rather than like an intelligent, decision-making mammal. Such behavior indicates that the midbrain and the structures below it contain neural systems that organize instinctive patterns of movement but do not contain neural systems that permit deliberate decisions to move or refrain from moving in accordance with the animal's long-term interests or needs.

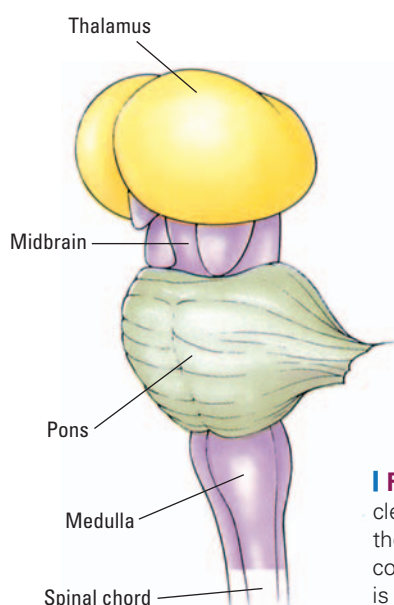


FIGURE 5.14 | Brainstem and thalamus This figure makes clear why the medulla, pons, and midbrain are collectively called the *brainstem*. They form a stem-like continuation from the spinal cord, to which other brain structures are connected. The thalamus is attached to the top of the brainstem.

The Thalamus Is a Relay Station for Sensory and Motor Pathways

Directly atop the brainstem is the **thalamus** [**thal**-uh-muss] (see Figure 5.14). This structure, seated squarely in the middle of the brain, is most conveniently thought of as a relay station that connects various parts of the brain with one another. Most of the sensory tracts that ascend through the brainstem terminate in special nuclei in the thalamus; those nuclei, in turn, send their output to specific areas in the cerebral cortex (discussed in Chapter 7). The thalamus also has nuclei that relay messages from higher parts of the brain to movement-control centers in the brainstem.

The Cerebellum and the Basal Ganglia Help to Coordinate Skilled Movements

We move now to two portions of the brain that are anatomically distinct but closely related in function—the cerebellum and the basal ganglia (see Figure 5.15). *Cerebellum* means “little brain” in Latin, and the **cerebellum** [sair-uh-bell-um] indeed looks something like a smaller version of the rest of the brain, riding piggy-back on the rear of the brainstem. The **basal ganglia** [bay-sul-gang-gee-uh] are a set of interconnected structures lying on each side of the thalamus. Damage to either the cerebellum or the basal ganglia can greatly interfere with a person’s ability to produce learned, skilled, well-coordinated movements.

Damage to the cerebellum is especially associated with loss in ability to behave in ways that require rapid, well-timed sequences of muscle movements, such as pitching a baseball, leaping over a hurdle, playing a musical instrument, or typing a series of words at a computer (Hauk & Munguini, 2003). In contrast, damage to the basal ganglia is especially associated with loss of ability to coordinate slower, deliberate movements, such as walking or reaching out to pick up an object (Packard & Knowlton, 2002).

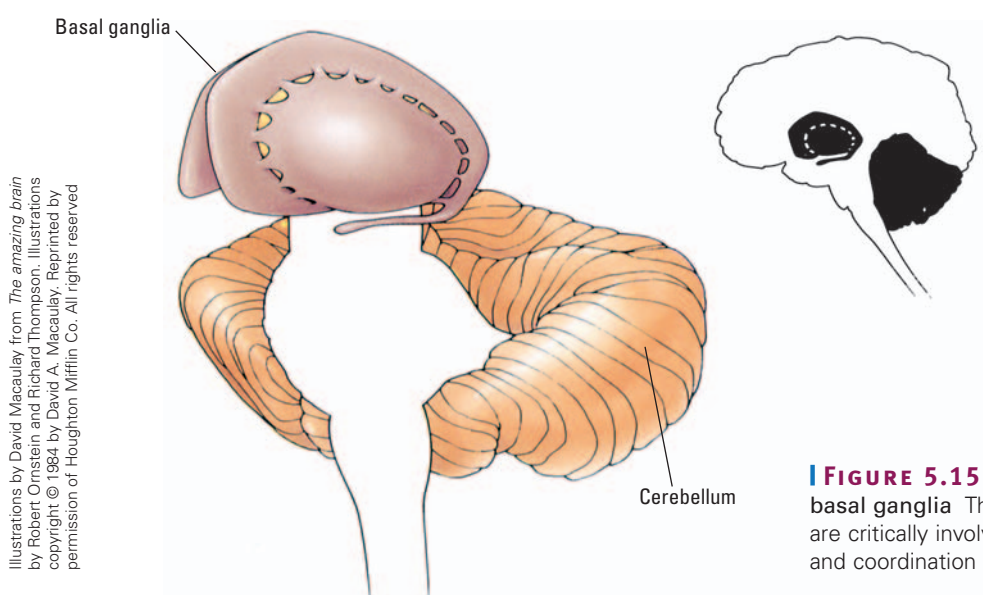
Both structures are specialized to use sensory information to guide movements, but they apparently use that information in different ways. The basal ganglia appear to use sensory information primarily in a *feedback* manner. In other words, sensory input pertaining to an ongoing movement (such as the sight of how the hand is moving) feeds back into the basal ganglia and is used to adjust the movement as it progresses. The cerebellum, in contrast, uses sensory information primarily in a *feed-forward* manner (Ohyama & others, 2003). That is, it uses sensory information to program the appropriate force and timing of a movement before the movement is initiated. That is why the cerebellum is especially crucial for movements that occur too rapidly to be modified once they are in progress.

>>> 13

What are the functional similarities and differences between the cerebellum and the basal ganglia?



The cerebellum in action The cerebellum is involved in well-coordinated, precisely timed movements, which occur too fast to be controlled by sensory feedback.



Illustrations by David Macaulay from *The amazing brain* by Robert Ornstein and Richard Thompson. Illustrations copyright © 1984 by David A. Macaulay. Reprinted by permission of Houghton Mifflin Co. All rights reserved

FIGURE 5.15 Cerebellum and basal ganglia These two structures are critically involved in the initiation and coordination of movements.

These characterizations are useful, but they do not describe the full range of functions of these large brain structures. People with damage in either the cerebellum or the basal ganglia can show a wide range of motor deficits, depending on the precise location of damage. Moreover, neuroimaging studies (using PET and fMRI) show that portions of the cerebellum, basal ganglia, and certain motor planning areas of the cerebral cortex become active not just when people are producing movements but also when they are imagining themselves producing movements (Houk & Mugnaini, 2003; Kolb & Whishaw, 2003). Perhaps when divers, gymnasts, or pianists “visualize” their performance before they act, what they are doing, partly, is warming up specific neurons in these motor areas, thereby setting up the neural programs that will eventuate in their best performance.

The Limbic System and the Hypothalamus Play Essential Roles in Motivation and Emotion

14

Why is the limbic system so named, and what functions does it perform?

<<< The term *limbic* comes from the Latin word *limbus*, meaning “border.” The **limbic system** can be thought of as the border dividing the evolutionarily older parts of the brain, below it, from the newest part (the cerebral cortex), above it. The limbic system consists of several distinct structures—including the **amygdala** [uh-mig-duh-luh] and **hippocampus** [hip-oh-camp-us]—that interconnect with one another in a circuit wrapped around the thalamus and basal ganglia (see Figure 5.16). These structures are involved in the regulation of basic drives and emotions (discussed in Chapter 6). The limbic system is believed to have evolved originally as a system for the sophisticated analysis of olfactory input (Thompson, 1985), and its connections with the nose remain strong. This may help explain the special influence that smells—such as the aroma of good food or perfume, or the stench of vomit—can have on drives and emotions. But the limbic system also receives input from all the other sensory organs. In addition, it is intimately connected to the basal ganglia, and that connection is believed to help translate emotions and drives into actions.

The limbic system is involved in more than drives and emotions. One part of it, the hippocampus, is crucial for keeping track of spatial location (the direction-sensitive place cells, noted earlier in the chapter, are located there) and for encoding certain kinds of memories (discussed in Chapter 9).

15

What are three ways by which the hypothalamus controls the body’s internal environment?

<<< The **hypothalamus** [hype-oh-thal-uh-muss] is a small but vitally important structure. Its name derives from its position directly underneath the thalamus (*hypo* in this case means “underneath”). The hypothalamus is not technically part of the limbic system, but is intimately connected to all the structures of that system. Its primary task is to help regulate the internal environment of the body. This

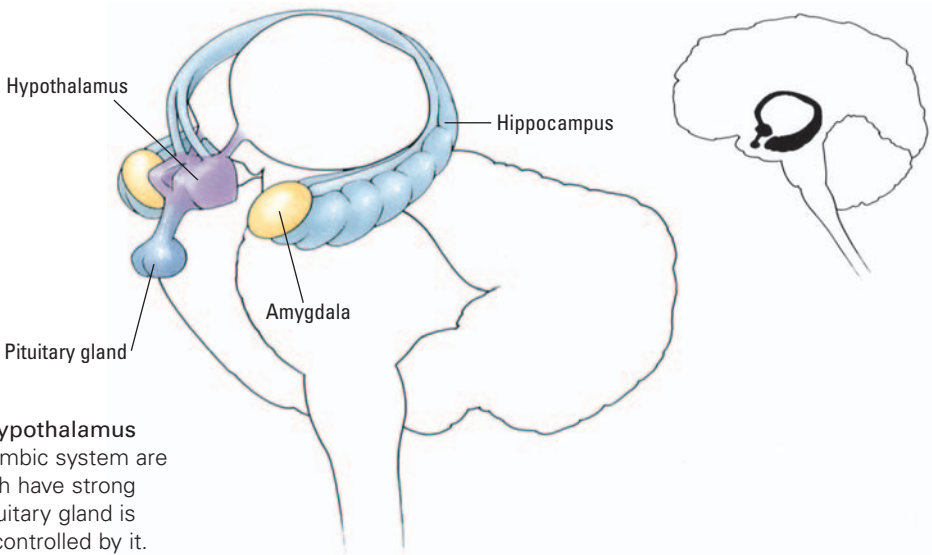


FIGURE 5.16 | Limbic system and hypothalamus
The most conspicuous structures of the limbic system are the hippocampus and the amygdala, which have strong connections to the hypothalamus. The pituitary gland is strongly tied to the hypothalamus and is controlled by it.

Illustrations by David Macaulay from *The amazing brain* by Robert Ornstein and Richard Thompson. Illustrations copyright © 1984 by David A. Macaulay. Reprinted by permission of Houghton Mifflin Co. All rights reserved

it accomplishes by (a) influencing the activity of the autonomic nervous system, (b) controlling the release of certain hormones (to be described later), and (c) affecting certain drive states, such as hunger and thirst. In addition, through its connections with the limbic system, the hypothalamus helps regulate emotional states, such as fear and anger. You will read in Chapter 6 about the role of the hypothalamus in drives and emotions. If I had to give up a cubic millimeter (a tiny speck) of tissue from some part of my brain, the last place I would want it taken from is the hypothalamus. Depending on just which part was taken, I could be left without one or more of my basic drives, without a normal cycle of sleep and wakefulness, or without the ability to regulate my body metabolism.

The Cerebral Cortex

We now move up to the anatomically outermost and evolutionarily newest part of the brain, the **cerebral cortex**. *Cerebrum* is the Latin word for “brain” (the term is now sometimes used to refer to all parts of the brain other than the brainstem and cerebellum). *Cortex* is the Latin word for “bark,” and in anatomical usage it refers to the outside layer of any structure—in this case, the brain. It is by far the largest part of the human brain, accounting for approximately 80 percent of its total volume (Kolb & Whishaw, 2003). Its surface area is much greater than it appears because it folds inward in many places. If the cortex were unfolded and spread out as a single sheet, it would be only 2 to 3 millimeters thick and would occupy a surface area equivalent to a square that is half a meter long on each side (Kolb & Whishaw, 2003).

The entire folded cerebral cortex is divided into left and right *hemispheres*, and each hemisphere is further divided into four lobes, or segments, demarcated at least partly by rather prominent folds. The lobes, whose positions you can see in Figure 5.17, are, from back to front, the **occipital** [ock-sip-it-ul], **temporal** [temp-or-ul], **parietal** [pah-rye-ut-ul], and **frontal lobes**.

A human brain Only one-third of the surface area of the brain’s cortex is visible from the outside. The remaining two-thirds is hidden in the folds and fissures.



Republished with permission of Globe Newspaper Co., Inc.

>>> 16
What are the four lobes of the cortex, and what are three functional categories of areas that exist within these lobes?

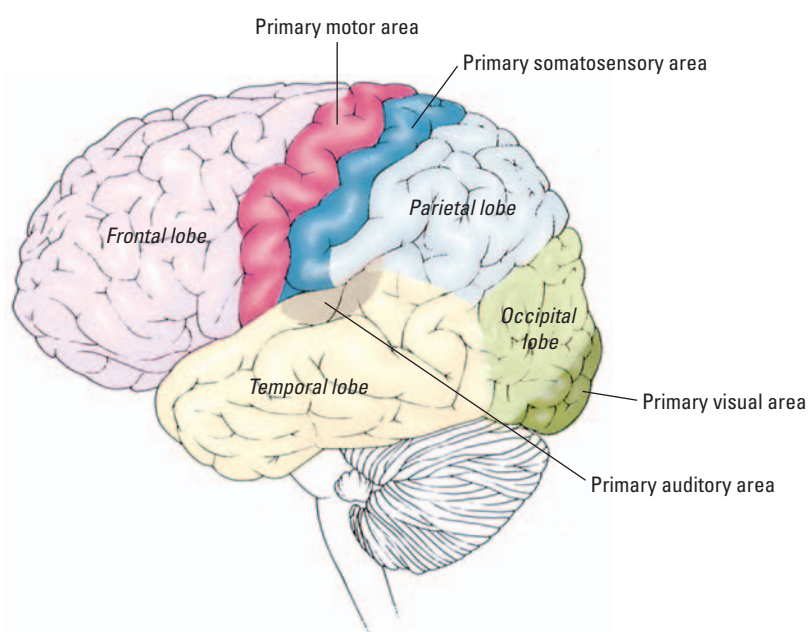


FIGURE 5.17 | Cerebral cortex This figure shows the four lobes of the cortex, as well as the locations of the primary motor area and the primary sensory areas for vision, hearing, and somatosensation.

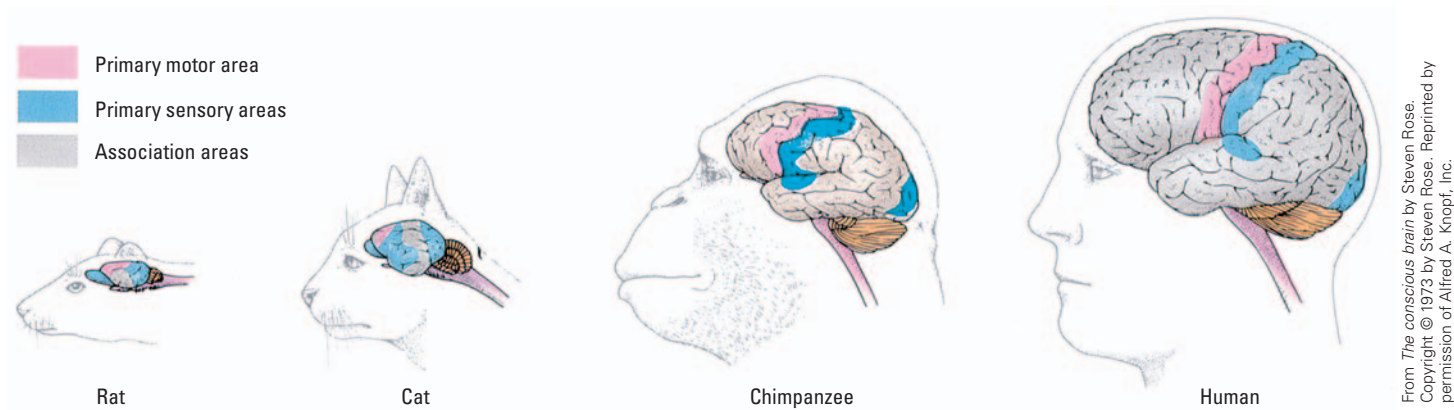


FIGURE 5.18 | Comparison of the brains of four mammals. All the brains contain the same structures, but in the chimpanzee and human, proportionately much more cortical space is devoted to association areas than in the rat and cat.

The Cortex Includes Sensory, Motor, and Association Areas

Researchers who study the functions of the cortex divide it into three categories of functional regions, or areas. One category consists of the **primary sensory areas**, which receive signals from sensory nerves and tracts by way of relay nuclei in the thalamus. As shown in Figure 5.17, primary sensory areas include the visual area in the occipital lobe, the auditory area in the temporal lobe, and the somatosensory area in the parietal lobe. A second category is the **primary motor area**, which sends axons down to motor neurons in the brainstem and spinal cord. As shown in Figure 5.17, this area occupies the rear portion of the frontal lobe, directly in front of the somatosensory area. The third category consists of the remaining parts of the cortex, which are called **association areas**. These areas receive input from the sensory areas and lower parts of the brain and are involved in the complex processes that we call perception, thought, and decision making. As shown in Figure 5.18, the amount of association cortex, relative to the other two categories, increases dramatically as one goes from simpler mammals, such as the rat and the cat, to more complex ones, such as the chimpanzee and the human.

The Primary Sensory and Motor Areas Are Topographically Organized

17 What does it mean to say that cortical sensory and motor areas in the cortex are topographically organized?

<<< The primary sensory and motor areas of the cortex are organized in such a way that adjacent neurons receive signals from or send signals to adjacent portions of the sensory or muscular tissue to which they are ultimately connected. This fact is referred to as the *principle of topographic organization*. For example, neurons that are near one another in the visual cortex receive signals from receptor cells that are near one another in the retina of the eye. Similarly, neurons that are near one another in the somatosensory cortex receive signals from adjacent areas of the skin, and neurons that are near one another in the primary motor cortex send signals to adjacent sets of muscle fibers. It is possible to map onto the somatosensory cortex the parts of the body from which each portion of somatosensory cortex receives its signals, and onto the motor cortex the part of the body to which each portion of motor cortex sends its signals.

The resulting maps, depicted in Figure 5.19, show a distorted view of the human body. This is because the amount of cortex devoted to each part of the body corresponds not to the size of the body part but to the degree of sensitivity of that part (in the case of the sensory map) or the fineness of its movements (in the case of the motor map). As you can see, huge areas of the human primary motor cortex are devoted to control of the fingers and vocal apparatus, where fine control is needed. In other animals, other body parts have greater representation, depending on the range and delicacy of their movements. In a cat, for example, large portions of the somatosensory and primary motor areas of the cortex are devoted to the whiskers, and in a spider monkey—a creature that uses its tail as a fifth arm and hand—large areas are devoted to the tail (Walker, 1973).

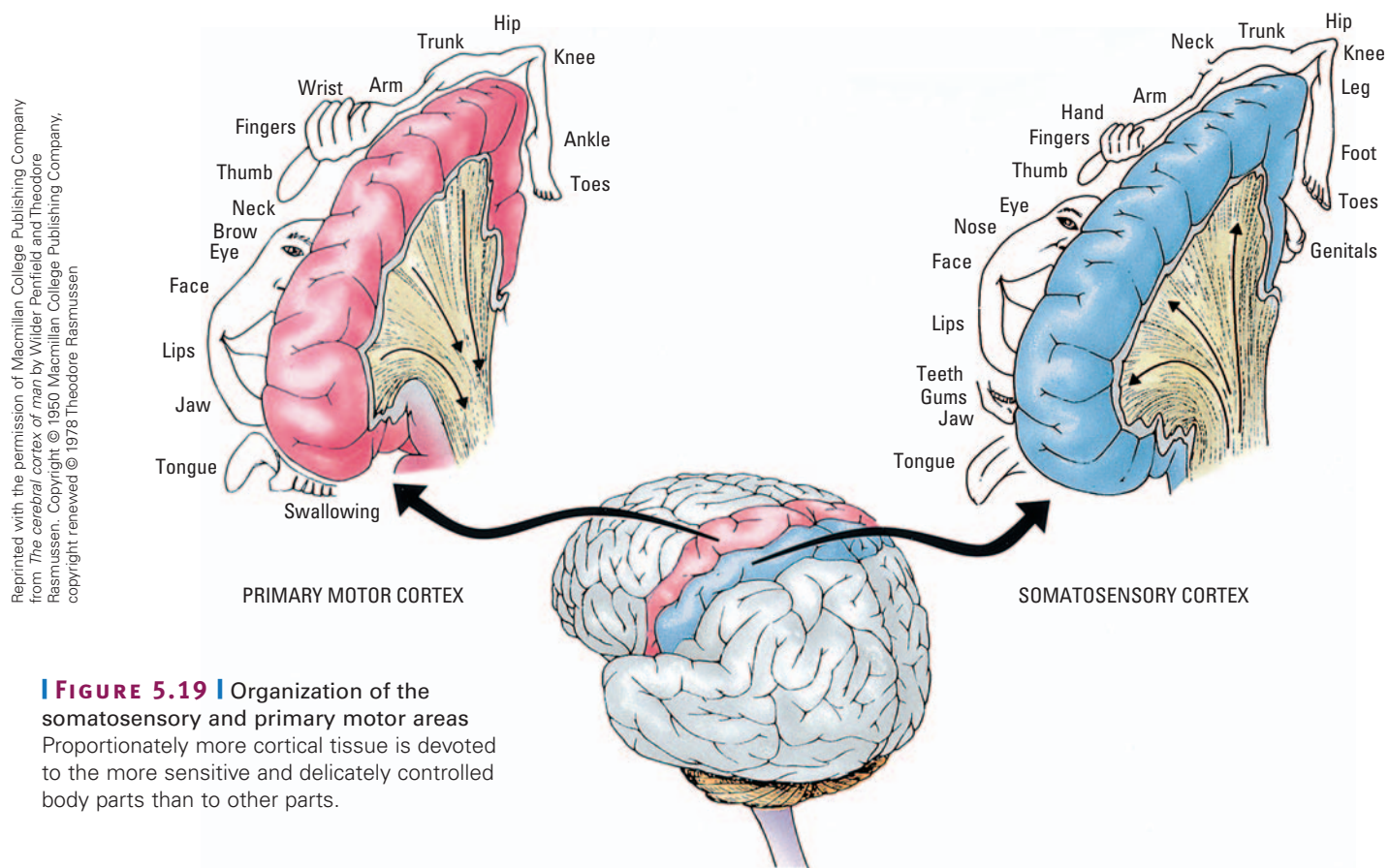


FIGURE 5.19 Organization of the somatosensory and primary motor areas. Proportionately more cortical tissue is devoted to the more sensitive and delicately controlled body parts than to other parts.

The primary motor cortex is part of the chain of command in controlling movements, but it is not at the top of that chain and is not involved in all types of movements. It receives input from the basal ganglia and cerebellum and is specialized to fine-tune the signals going to small muscles, such as those of the fingers and tongue, which must operate in a finely graded way. Experiments conducted many years ago, in which monkeys made well-controlled hand movements to obtain food rewards, revealed that each movement was preceded first by a burst of activity in the basal ganglia and then by a burst in the primary motor cortex (Evarts, 1979; Kornhuber, 1974). This is part of the evidence that the primary motor cortex generally comes later than the basal ganglia and the cerebellum in the chain of command. In other experiments, monkeys whose primary motor cortex had been entirely destroyed behaved normally in most respects but were unable to make delicate hand movements, such as those needed to lift a small piece of food out of a narrow hole (Passingham & others, 1983).

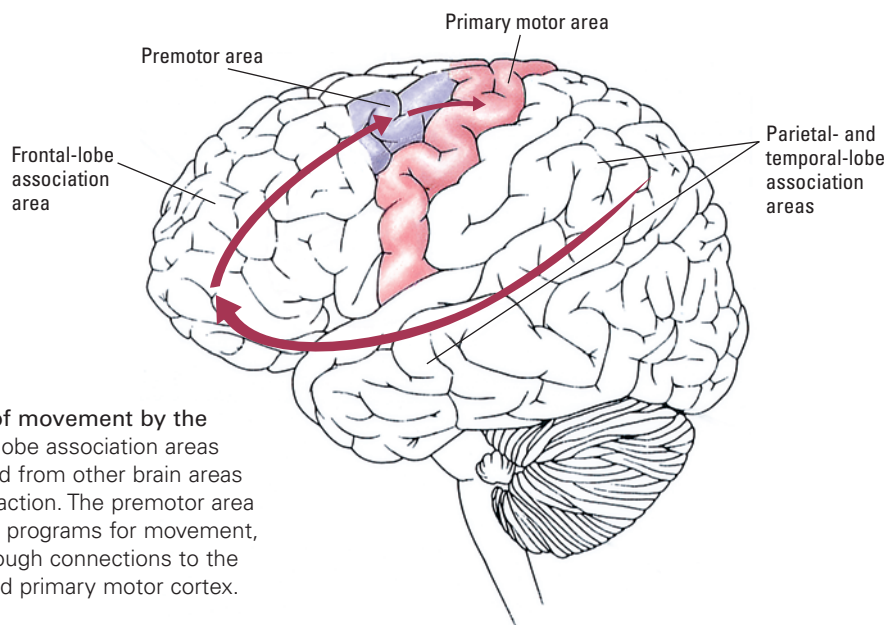
18 What is some evidence that the primary motor cortex comes relatively late in the chain of command preceding an action and that its function is to refine the more delicate parts of the action?

The Frontal Lobe Is Involved in the Planning of Movements

Directly in front of the primary motor area lies another cortical area devoted to motor control, the **premotor area** (see Figure 5.20). This area is involved in choosing the specific movement or pattern of movements to be made at a given time. It bases that choice partly on information sent to it from other areas of the frontal lobe, which are involved in behavioral planning, and it executes that choice by sending output to the cerebellum, basal ganglia, and motor cortex, which refine the program further. Like the cerebellum and basal ganglia, it becomes active during the mental rehearsal of coordinated movements as well as during the actual production of such movements (Cisek & Kalaska, 2004).

In recent experiments, Etienne Koechlin and his colleagues (2003) asked human subjects to respond to visual stimuli (referred to as task stimuli) with one of two

19 What are the roles of (a) the premotor area and (b) other frontal lobe association areas in the control of behavior? What evidence supports these conclusions?



Ghez, C. (1985). Voluntary movement. In E. R. Kandel & J. H. Schwartz (Eds.), *Principles of Neural Science* (3rd ed.). Norwalk, CT: Appleton & Lange

FIGURE 5.20 Control of movement by the cerebral cortex The frontal-lobe association areas integrate information received from other brain areas and make a general plan for action. The premotor area converts this plan into neural programs for movement, which are then executed through connections to the cerebellum, basal ganglia, and primary motor cortex.

different hand movements. Several rules were established as to which response the subjects should make to which stimulus, and a cue was presented before each trial telling them which rule they should follow. As they performed these tasks, their brains were imaged by fMRI. The results were that certain areas of the frontal lobe in front of the premotor area were most active in response to the instructional cues, and the premotor area itself was most active in response to the task stimuli. This finding is consistent with the view that frontal areas of the frontal lobe are involved in planning general strategies of action, based on instructions or other contextual information, and that the prefrontal cortex translates such strategies into specific patterns of movement, taking immediate stimuli into account.

To get an idea of the general flow of information in the cortex involved in the control of movement, notice the arrows in Figure 5.20. Association areas in the rear parts of the cortex, especially in the parietal and temporal lobes, analyze information that comes to them from sensory areas. These areas, in turn, send output to the association areas of the frontal lobe, which also receive information about the internal environment through strong connections with the limbic system. Combining all this information, the frontal association areas set up general plans for action that can be put into effect through connections to the premotor cortex and through downward links to the basal ganglia and cerebellum.

Consistent with this model of cortical control, damage to the frontal-lobe association areas does not, as a rule, reduce one's ability to extract information from the environment, but it sharply reduces one's ability to use that information effectively to control behavior. Depending on the location and extent of the injury, it can destroy either short-range planning, such as working out the series of movements needed to solve a particular puzzle, or long-range planning, such as organizing one's day, week, or life (Kolb & Whishaw, 2003).

Hierarchical Organization in the Control of Movement: A Summary

20 How are the movement-control functions of the nervous system summarized as a hierarchical, top-down flow of information? How is the hierarchy illustrated by an imaginative tour through the nervous system of a person who decides to eat some peanuts?

<<< We have ascended the nervous system, from bottom to top, and glimpsed the functions of each of its divisions in the overall task of controlling behavior. The parts, as I said at the outset, work together in a hierarchical manner. To review the movement-control functions of the various parts of the nervous system and to visualize their hierarchical organization, look at Figure 5.21. Structures are organized in this figure according to their general roles in controlling movement, not accord-

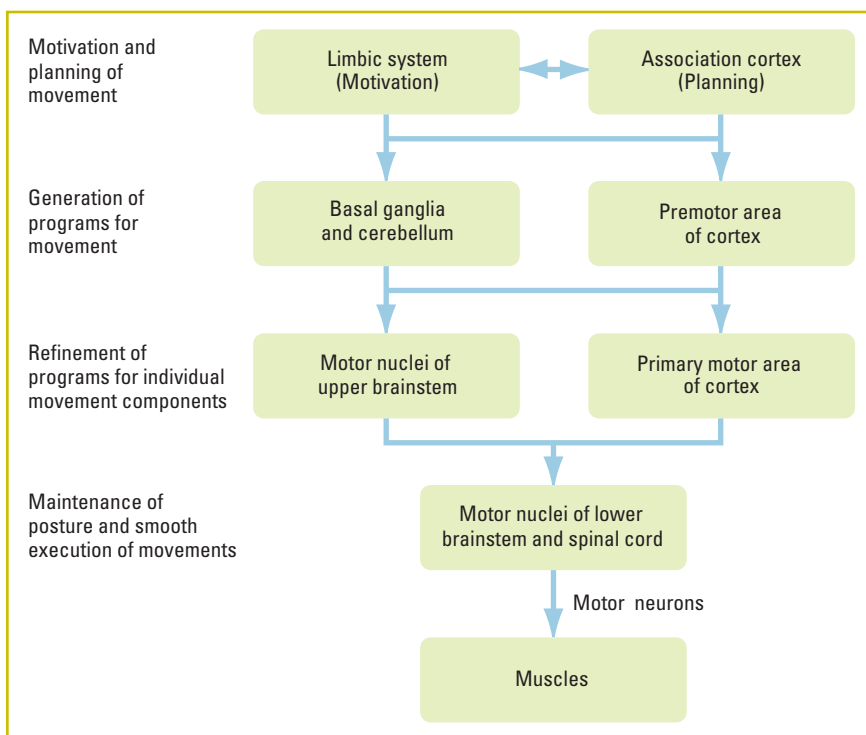
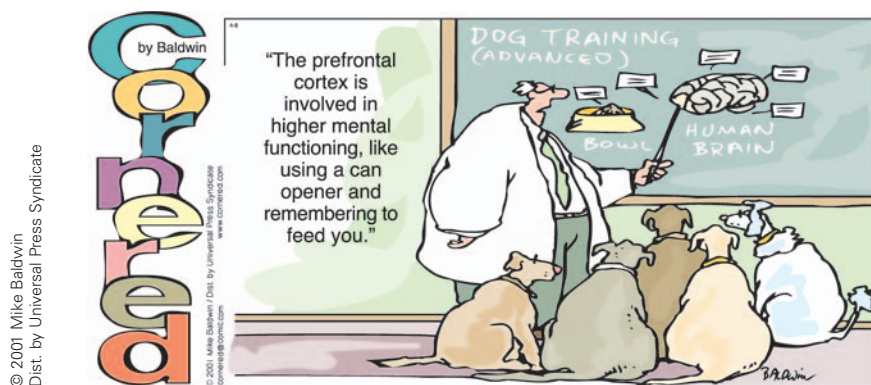


FIGURE 5.21 | Hierarchy of motor control This figure summarizes the broad functions of various structures of the nervous system in the control of movement. The structures shown higher up are involved in the more global aspects of an action, and those shown farther down are involved in the finer details of carrying it out. Notice that both subcortical and cortical structures exist at each of the top three levels of the hierarchy. Although this portrayal does not show all the known pathways, it provides a useful way to think about the neural flow of information from the planning of an action to its execution. (Based partly on information in Ghez & Krakauer, 2000, and Schmidt, 1986)

ing to their anatomical positions. The highest structures are involved in motivation and planning, and the lower structures are involved in refining and executing the plans, turning them into action. Notice that subcortical structures (shown on the left side of the diagram) and cortical structures (shown on the right side) are involved at each of the top three levels in the hierarchy.

To illustrate the hierarchy further, imagine what might occur in the nervous system of a person who has not eaten in a while and sees some peanuts. At the top of the hierarchy, the limbic system (which most directly monitors the internal state of the body) senses that food is needed and sends a message of “hunger” to the cortical association areas with which it is connected. These areas, which share the top of the hierarchy with the limbic system, analyze information coming to them from the visual cortex and determine that some peanuts lie in a bowl across the room. Other information is also considered by the association areas, including memories about the taste of peanuts, about how to eat them, and about the propriety of eating them in this room at this time. Such information, integrated by association areas in the frontal lobes, leads to a decision to cross the room, pick up some peanuts, and eat them.



At the second level, the basal ganglia, cerebellum, and premotor cortex receive the program of planned action from the limbic system and association cortex. They also receive somatosensory input about the exact position of parts of the body and visual input about the exact location of the peanuts. They use this information to refine the motor program, that is, to work out the specific timing and patterning of the movements to be made.

At the third level, the motor program is conveyed through two pathways for further refinement. The program for larger movements, such as walking toward the peanuts, is sent directly down to a set of motor nuclei in the upper part of the brainstem. The program for delicate movements, such as those needed for removing the peanuts from their shells, is conveyed to the motor cortex, which, in turn, sends its output down to the brainstem and spinal cord.

Finally, at the fourth level of the hierarchy, in the lower brainstem and spinal cord, are motor neurons—the final common pathway in control of behavior. Skeletal motor neurons send their messages out to the skeletal muscles of the body, to trigger all the movements needed to reach the peanuts, pick them up, and put them in the mouth. Autonomic motor neurons (of the parasympathetic division) send their messages out to the salivary glands and muscular walls of the gastrointestinal system to prepare the digestive tract for the receipt of peanuts.

A Word of Caution

The hierarchy just described is useful as a first approach to understanding the nervous system, and it accurately reflects the kinds of behavioral deficits that occur when different parts of the nervous system are damaged. However, there is a possible danger in this portrayal: It can seduce us into believing that we know more than we actually do know.

21 What is the difference between knowing where a brain function occurs and knowing how it occurs?

<<< Specifically, the knowledge that certain parts of the brain are essential for certain aspects of behavioral control can be mistaken for knowledge about *how* those processes are accomplished. But the discovery of “where” does not explain “how.” In illustrating the hierarchy, I spoke of “decisions” made in association areas of the cortex and of programs for action developed and refined by other brain areas. What do such statements mean? They mean only that individuals who suffer damage in one part of the brain lose the ability to make reasonable choices for action and that those who suffer damage in another part retain the ability to make reasonable choices but lose the ability to carry them out in a coordinated manner. Such statements don’t address the far more difficult question of how the association cortex makes decisions or how various other structures develop and refine programs for action.

SECTION REVIEW

The nervous system’s functioning requires coordination of different divisions.			
Peripheral Nervous System • This division of the nervous system is necessary for perception; sensory neurons inform the central nervous system about stimuli. • It is also necessary for action. Skeletal motor neurons contract skeletal muscles. Autonomic motor neurons (both sympathetic and parasympathetic) act on internal organs.	Spinal Cord • The spinal cord acts as a conduit, carrying somatosensory information to the brain and motor commands from the brain. • It also organizes spinal reflexes and contains pattern generators that produce rhythmic movement sequences, such as those of walking.	Subcortical Structures • The brainstem is similar to the spinal cord but more elaborate. • The thalamus is a major sensory and motor relay station. • The cerebellum and basal ganglia are crucial for the production of learned, coordinated movement. • The hypothalamus and limbic system are critically involved in motivation and emotion as well as other functions.	Cerebral Cortex • The cortex is divided into two hemispheres, each of which has four lobes. • Primary sensory and primary motor areas are organized topographically. • Association areas of the cortex are crucial for thought and planning.

Asymmetry of Higher Functions of the Cerebral Cortex

Nearly every part of the brain exists in duplicate. We have a right and a left member of each anatomical portion of the brainstem, thalamus, cerebellum, and so on. The part of the brain in which the right-left division is most evident, however, is the cerebral cortex. Each half of the cortex folds inward where it would abut the other half, forming a deep, fore-to-aft midline fissure that divides the cortex into distinct right and left hemispheres. The two hemispheres are connected, however, by a massive bundle of axons called the **corpus callosum** [cuh-loh-sum] (see Figure 5.22).

The two hemispheres are quite symmetrical in their primary sensory and motor functions. Each does the same job, but for a different half of the body. Most of the neural paths between the primary sensory and motor areas of the cortex and the parts of the body to which they connect are crossed, or *contralateral*. Thus, sensory neurons that arise from the skin on the right side of the body send their signals to the somatosensory area of the left hemisphere, and vice versa. Similarly, neurons in the primary motor area of the left hemisphere send their signals to muscles on the right side of the body, and vice versa. Such symmetry breaks down, however, in the association areas.

>>> 22
In what ways are the two hemispheres of the cerebral cortex symmetrical, and in what ways are they asymmetrical?

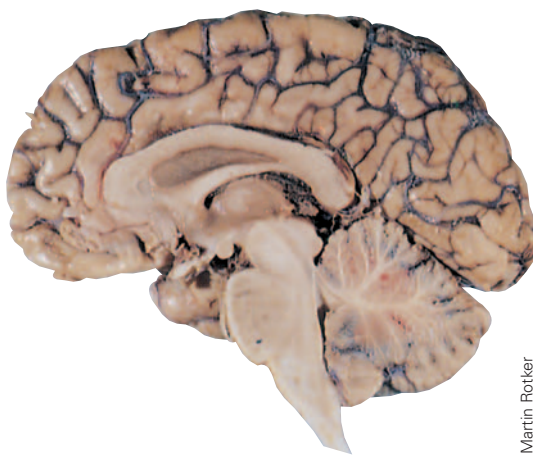
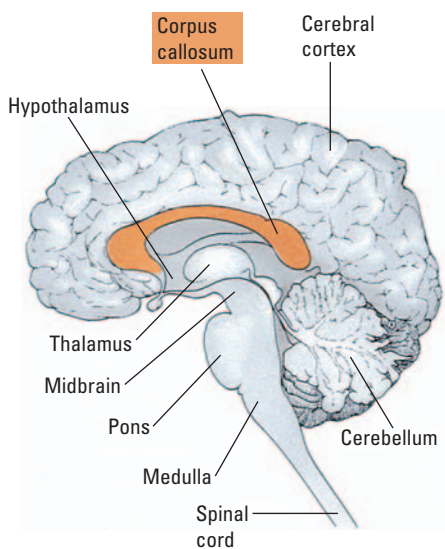


FIGURE 5.22 | Corpus callosum The corpus callosum is a huge bundle of axons that connects the right and left hemispheres of the cerebral cortex. You can see it here in an inside view of the right hemisphere of a human brain that has been cut in half. This photograph and the matched diagram also give you an inside view of other brain structures.

The most obvious distinction between the two cortical hemispheres in humans is that large areas in the left are specialized for language and comparable areas in the right are specialized for nonverbal, visuospatial analysis of information. The earliest evidence of this difference came from observing people who had suffered strokes or other injuries that affected just one hemisphere. In general, damage to the left hemisphere results in deficits in using and comprehending language, and damage to the right results in deficits in such tasks as recognizing faces, reading maps, and drawing geometric shapes, all of which depend on perceiving spatial relationships.

Effects of Surgical Separation of the Hemispheres: Split Brain, Split Mind

Dramatic further evidence of the separate abilities of the two hemispheres appeared in the 1960s, when Michael Gazzaniga and his colleagues began to study people who, as a last-resort treatment for epilepsy, had undergone an operation in which the two hemispheres were separated by cutting the corpus callosum.

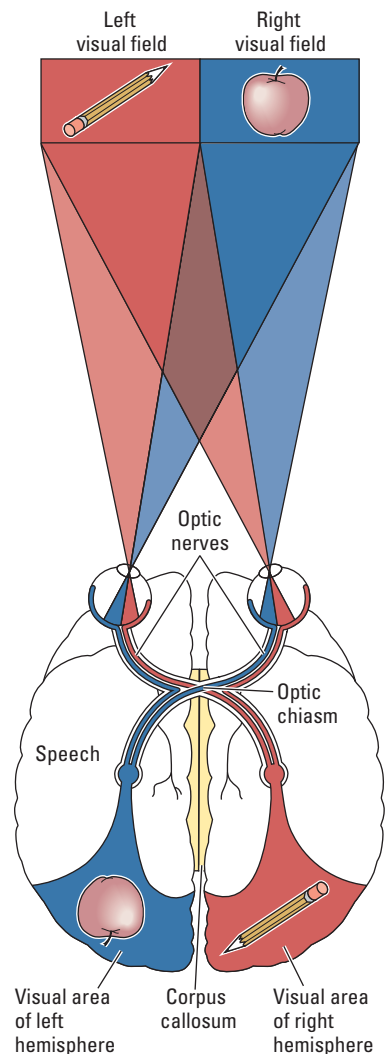
Early, more casual observations had revealed no remarkable deficits in people who had undergone this operation. The operation was generally successful in reducing or eliminating epileptic seizures, and, after a period of recovery, there is usually no drop in measured IQ, in ability to carry on a conversation, or even in ability to coordinate the two sides of the body in skilled tasks. But Gazzaniga (1967, 1998) showed that under special test conditions, in which information is provided to just one hemisphere or the other, these people behave as though they have two separate minds with different abilities.

How Each Hemisphere Can Be Tested Separately After the Split-Brain Operation

23

How is it possible to test each hemisphere separately in people whose corpus callosum has been cut? How do such tests confirm that the left hemisphere controls speech and the right hemisphere has superior spatial ability?

<<< Split-brain studies take advantage of the crossed sensory and motor connections between the brain and peripheral parts of the body. Recall that each hemisphere most directly controls movement in, and receives somatosensory information from, the opposite half of the body. In addition, as illustrated in Figure 5.23, connections from the eyes to the brain are such that input from the right visual field (the right-hand half of a person's field of view) goes first to the left hemisphere and input from the left visual field goes first to the right hemisphere. In the normal brain, all information that goes to either hemisphere subsequently travels to the other through the corpus callosum. But the split-brain operation destroys those connections. Therefore, with the testing apparatus shown in Figure 5.24 and with split-brain patients as subjects, it is possible to (a) send visual information to just one hemisphere by presenting the stimulus in only the opposite half of the visual field, (b) send tactile (touch) information to just one hemisphere by having the subject feel an object with the opposite hand, and (c) test the knowledge of just one hemisphere by having the subject respond with the opposite hand.



Split-Brain Evidence for Left-Hemisphere Language and Right-Hemisphere Spatial Ability

In a typical experiment, Gazzaniga flashed pictures of common objects to either the right or the left visual field of a split-brain patient. When a picture was flashed in the right field (projecting to the left hemisphere), the patient described it as well as someone with an intact brain would; but when a picture was flashed in the left field (projecting to the right hemisphere), the patient either claimed to see nothing or made a random guess. Then Gazzaniga asked the same person to reach under a barrier with one hand or the other and identify, by touch, the object that had been flashed. The fascinating result was that the person could reliably identify with the left hand (but not with the right) the same object that he or she had just vocally denied having seen (Gazzaniga, 1967). For example, if the object flashed to the right hemisphere was a pencil, the subject's left hand picked out the pencil from a set of objects even while the subject's voice was continuing to say that nothing had been flashed. Such findings are consistent with the conclusion that only the left hemisphere can generate speech and that neither hemisphere has direct access to what the other one knows.

Research of this sort also revealed large individual differences among split-brain patients in the degree to which the right hemisphere can comprehend speech. Some show essentially no right-hemisphere comprehension. They cannot participate in experiments such as that just described, because their right hemispheres don't understand such instructions as "Pick up the object that you saw on the screen." Some have reasonably good right-hemisphere comprehension, although

FIGURE 5.23 Neural pathways from the eyes to the right and left hemispheres of the cortex. Neurons of the optic nerves either cross or don't cross at the optic chiasm. The pattern of crossing is such that neurons on the right side of either retina, which receive input from the left visual field, go to the right hemisphere of the brain, and vice versa.

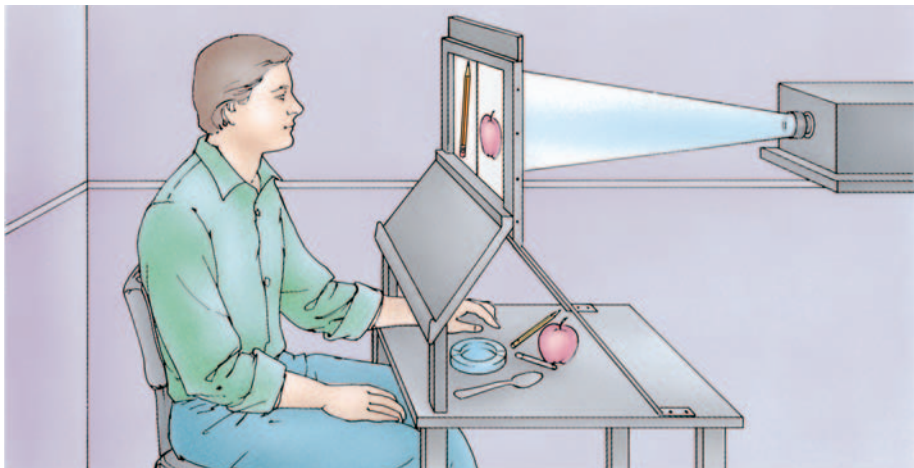


FIGURE 5.24 | Testing apparatus for split-brain subjects With this apparatus, it is possible to flash a stimulus in either visual field (or in both at once) and to ask the person to identify objects by touch with either hand. With the image of a pencil flashed in his left visual field, this split-brain subject will be able to identify the pencil by touch with his left hand. Vocally, however, he will report having seen an apple on the screen, as that was flashed in his right visual field, which projects to his left, speaking hemisphere. (Adapted from Gazzaniga, 1967)

their right hemisphere is still unable to generate speech; and a few show a reversal, with the right hemisphere superior to the left in language comprehension and production. Other tests, on people whose corpus callosum has not been cut, indicate that about 4 percent of right-handed individuals and 15 percent of left-handed individuals have their speech centers located in the right hemisphere rather than the left (Rasmussen & Milner, 1977).

Other experiments with split-brain patients revealed that although the right hemisphere is unable to speak, it is much better than the left in solving visuospatial problems. When asked to arrange puzzle pieces to match a particular design or to copy three-dimensional pictures, each subject performed far better with the left hand (controlled by the right hemisphere) than with the right hand (controlled by the left hemisphere), even though all of the subjects were right-handed (see Figure 5.25). In more recent experiments, split-brain patients were asked to judge small differences between visual stimuli presented as pairs to either hemisphere. The results showed that the left hemisphere is as good as the right in judging non-spatial differences, such as differences in brightness or in color, but not as good as the right in judging spatial differences, such as differences in size or in the slant of two lines (Corballis, 2003).

How Split-Brain Patients Cope as Well as They Do

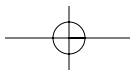
How do people who have had the split-brain operation get along in the world as well as they do? What keeps their two hemispheres from going off in opposite directions and creating continuous conflict between the two halves of the body? In some instances, especially shortly after the surgery, conflict does occur. One man described a situation in which, while dressing in the morning, his right hand was trying to pull his pants on while his left hand was trying to pull them back off (Gazzaniga, 1970); apparently, his right hemisphere wanted to go back to bed. But such conflicts are rare, and when they do occur the left hemisphere (right hand) usually wins.

The patient's ability to coordinate the two hemispheres probably involves several mechanisms. First, only the cerebral cortex and some parts of the limbic system are divided when the corpus callosum is cut. Motor centers that control movements of the larger muscles, such as those of the legs and arms (but not the fingers) lie in the lower, undivided parts, and some sensory information also passes from one hemisphere to the other by way of those lower routes (Reuter-Lorenz & Miller, 1998). The intact connections also apparently allow each hemisphere to inhibit the motor output of the other, so the more competent hemisphere can take control of any given task (Reuter-Lorenz & Miller, 1998). In addition, under normal conditions, when the eyes can move around and things can be felt with both hands, the two hemispheres can receive the same or similar information through their separate channels. Finally, each hemisphere apparently learns to communicate indirectly with the other by observing and responding to the behavior that the

	Example	Left hand	Right hand
1			
2			
3			

Adapted from Gazzaniga, M. S. (1967, August). The split brain in man. *Scientific American*, 24-27. Copyright © 1967 by Scientific American, Inc. All rights reserved.

FIGURE 5.25 | Evidence for right-hemisphere superiority in spatial representation Although the split-brain patient who produced these drawings was right-handed, he could copy geometric figures much better with his left hand (controlled by his right hemisphere) than with his right hand.



other produces, a process that Gazzaniga (1967) labeled *cross-cuing*. For example, the right hemisphere may perceive something unpleasant and precipitate a frown, and the left may feel the frown and say, “I’m displeased.”

Split-Brain Insight into Consciousness: The Left-Hemisphere Interpreter

24

How do studies of split-brain patients tend to confirm and extend an idea about the nature of consciousness that was developed long ago by Sigmund Freud?

<<< One of the most fascinating findings from studies of split-brain patients is that they rarely express surprise or confusion concerning the seemingly contradictory actions that their two hemispheres generate. When asked to explain some behavior triggered by the right hemisphere, the person (speaking from the left hemisphere) usually generates quickly and confidently a seemingly logical (but often false) explanation.

For instance, in one experiment a split-brain patient was presented simultaneously with a picture of a chicken claw to the left hemisphere and a picture of a snow scene to the right hemisphere (Gazzaniga, 2000). Then he was shown (in a way that could be seen by both hemispheres) a set of pictured objects and was asked to point with both hands to the one most closely related to what he had seen on the screen. He immediately pointed with his right hand to the chicken (clearly related to the chicken claw) and with his left hand to the snow shovel (clearly related to the snow scene). When asked why he was pointing to these two objects, his left, speaking hemisphere responded, “Oh, that’s simple. The chicken claw goes with chicken, and you need a shovel to clean out the chicken shed.” He seemed completely satisfied by this explanation. In other experiments, Gazzaniga and his colleagues found that they could induce states of annoyance or of pleasure in split-brain patients by flashing annoying or pleasant scenes to their right hemispheres. When asked to explain their feelings, the patients always came up with plausible (but clearly false) answers. For example, they might comment on some aspect of the experimental equipment—or of the experimenter’s behavior—that was either annoying them or pleasing them.

Such observations have led Gazzaniga (2000) to posit that one of the natural functions of the left hemisphere is to interpret, or try to make logical sense of, everything that the person does. You might think of this *left-hemisphere interpreter* as analogous to the public relations department of a government. Its role is to tell stories, both to the self and to others, designed to make sense of the seemingly contradictory and irrational things that the person does. The idea of such an interpreter in the human brain or mind is by no means new. It was the centerpiece of a theory of consciousness proposed by Sigmund Freud (1912/1932) nearly 100 years ago (discussed in Chapter 15). According to Freud, we do things because unconscious decision-making processes in our mind make us do them. But one part of our mind observes what we do and tells a running story about it; that story constitutes our conscious understanding of our actions and the reasons for them. The split-brain studies indicate that the neural mechanism for generating such stories is located in the left hemisphere and is intimately connected with the brain areas that generate speech.

Language Areas of the Left Hemisphere

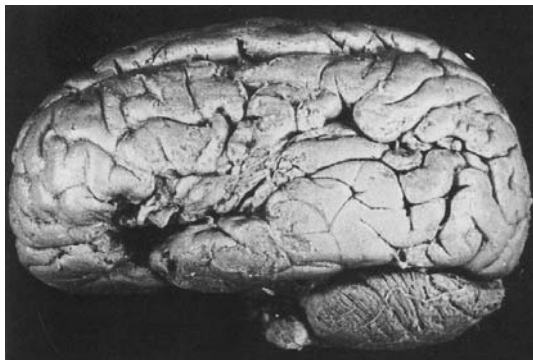
Perhaps the most distinctively human behavioral ability is that of producing and understanding a complex, grammar-based language (discussed in more detail in Chapter 11). Much of the left hemisphere of the human cortex is devoted in one way or another to language. Damage anywhere within large portions of the left hemisphere disrupts language ability, and the nature of the disruption depends on just where the destruction occurs.

25

What are the differences between Broca’s and Wernicke’s aphasias in (a) language production, (b) language comprehension, and (c) areas of the brain damaged?

<<< Any loss of language ability resulting from brain damage is called **aphasia** [uh-fay-zyuh]. Aphasias have been classified into a number of types, depending on the specific nature and degree of loss (Dronkers & others, 2000). The most well known and fully studied of these are two that were first described by nineteenth-century neurologists—one by Paul Broca and the other by Carl Wernicke.

Musée Dupuytren, Courtesy of Assistance
Publique, Hôpitaux de Paris



The preserved brain of a Broca's patient In 1861, a man who had great difficulty speaking after suffering a stroke and who identified himself merely as "Tan" checked into a hospital and was examined by Paul Broca. After Tan's death, Broca examined his brain and found the massive damage, shown here, to the region of the left hemisphere now called "Broca's area."

Effects of Damage to Broca's Area

Broca (1861/1965) observed that people who had brain damage that included an area of the left frontal lobe just anterior to the primary motor area, now called *Broca's* [**broke**-uh's] *area* (see Figure 5.26), suffered from a type of aphasia in which speech becomes labored and *telegraphic*—that is, the minimum number of words are used to convey the message. The speech consists mostly of nouns and verbs, and a sentence is rarely more than three or four words long. If you ask a person with this disorder what he or she did today, the answer might be, "Buy bread store." This disorder is now called **Broca's aphasia** or, more descriptively, *non-fluent aphasia*. The lack of fluency in Broca's aphasia suggests that neurons in Broca's area are crucial for setting up the motor programs that are involved in fluent speech production. This conclusion is also consistent with the observation that Broca's area lies very near to the region of the primary motor cortex that controls movements of the tongue and other speech muscles (compare Figure 5.26 with Figure 5.19).

Research since Broca's time indicates that people with damage to Broca's area also have some difficulties understanding language. They understand most of what they hear, but often fail to understand grammatically complex sentences. They become confused, for example, by sentences in which the agent and object of an action are reversed from the usual word order and the meaning cannot be inferred from the individual word meanings alone (Grodzinsky, 2000; Zurif, 1990). Thus they easily understand *The boy pushed the girl* (a simple, active sentence) or *The apple was eaten by the boy* (which can be understood just from word meanings, because apples don't eat boys), but *The girl was pushed by the boy* leaves them unsure as to who pushed whom. Research suggests that people with an intact Broca's area understand that last sentence by transforming it mentally into its simpler, active equivalent—*The boy pushed the girl*. People lacking Broca's area apparently fail to produce such transformations.

In sum, studies of people with damage to Broca's area suggest that neurons in and around this area are crucial for at least two seemingly distinct language functions: (1) articulating words and sentences in a fluent manner and (2) transforming grammatically complex sentences that are heard into simpler ones in order to extract the meaning.

Effects of Damage to Wernicke's Area

Wernicke (1874/1977) observed that people who had damage to a certain area of the left temporal lobe, now called *Wernicke's* [**wur**-nick-ee's] *area* (see Figure 5.26), suffered from a type of aphasia quite different from that described by Broca. These patients, unlike Broca's aphasics, had difficulty understanding the meanings of words that they heard and also had difficulty finding the appropriate words to express the meanings that they wanted to convey. The speech of these patients is almost the opposite of that of Broca's aphasics. It is rich in the little words that serve primarily to form the grammatical structure of a sentence—the articles (*a*, *the*), prepositions (such as *of*, *on*), and conjunctions (*and*, *but*). However, it is markedly

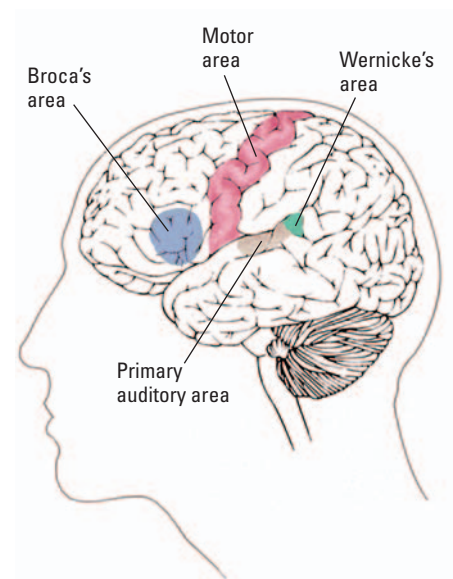


FIGURE 5.26 | Left-hemisphere language areas Damage to Broca's area leads to a loss of the ability to generate fluid, grammatically complete sentences but not to a loss of the ability to supply the words needed to convey the main content of the sentence. Damage to Wernicke's area has the opposite effect on speech production and greatly impairs speech comprehension.

deficient in the nouns, verbs, and adjectives that give most of the meaning to a sentence. One such patient, trying to describe a simple picture, said: “Nothing the keesereez the, these are davereez and these and this one and these are living. This one’s right in and these are . . . uh . . . and that’s nothing, that’s nothing” (Schwartz, 1987). The inability to come up with the correct names of objects and actions leads to a heavy use of pronouns and nonsense words as substitutes. The speech retains its grammatical structure but loses its meaning.

This disorder is now called **Wernicke’s aphasia**, or, more descriptively, *fluent aphasia*. Theories to explain it generally center on the idea that neurons in and near Wernicke’s area are crucially involved in translating the sounds of words into their meanings and in locating, through connections to other cortical association areas, the words needed to express intended meanings (Dronkers & others, 2000). The location of this area near to the primary auditory area (see Figure 5.26) is consistent with the role of translating sounds into meaningful words (Geschwind, 1972).

Identifying Language Areas Through Neuroimaging

With PET or fMRI it is possible to determine which areas of the brain increase in activity when a person engages in a language-related task. Such studies tend in some ways to confirm the theories about the functions of Broca’s and Wernicke’s areas that were derived from studies of people with brain damage and in other ways to challenge those theories.

26 How was PET used to identify brain areas involved in word perception and production?

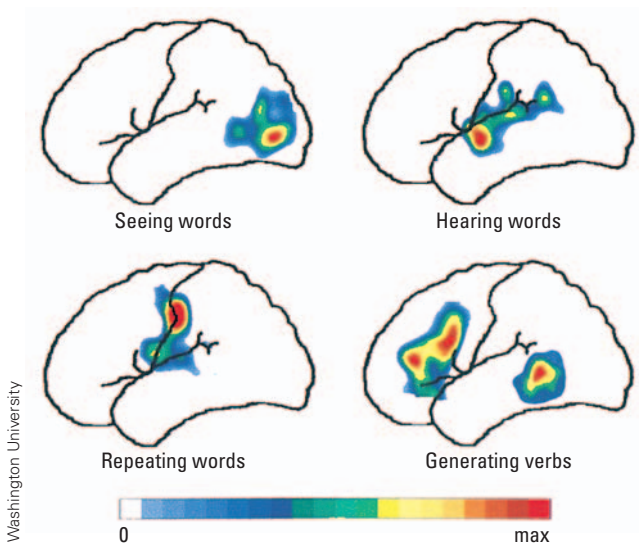
<<< One of the earliest and most-often-cited of such studies was conducted by Steven Petersen and his colleagues (1989). These researchers used PET to image the brains of people as they carried out four types of language-related tasks that varied stepwise in level of complexity. At the first level (simplest task), the subjects simply gazed at a spot marked by crosshairs in the center of a video screen. At the second level, they continued to gaze at the crosshairs while they either saw (superimposed on the crosshairs) or heard (through earphones) a series of common English nouns. The third level was just like the second, except that now they were asked to speak aloud each word that they saw or heard. The fourth level was like the third, except that instead of simply repeating each noun, they were asked to think of and say aloud a verb that represented an action appropriate to the noun (for example, in response to *hammer*, they might say “pound”).

In order to identify the brain areas brought into play by each type of task, the researchers had the computer calculate, for each small area of the brain, the difference between the average amount of activity during that task and the average amount during the task that ranked one level below that task. The results are depicted in Figure 5.27 and can be summarized as follows:

- Viewing or hearing words, without having to act on them in any way, resulted (as expected) in high activity in the relevant sensory areas—visual areas of the occipital lobe for viewing and auditory areas of the temporal lobe for hearing.
- Repeating aloud words that were seen or heard resulted in high activity in areas of the primary motor cortex that are involved in control of the vocal apparatus.
- Generating appropriate verbs in response to seen or heard nouns resulted in high activity in an area of the frontal lobe that encompassed Broca’s area and in a portion of the temporal lobe somewhat behind Wernicke’s area.

Notice that these results are not completely consistent with the theories about Broca’s and Wernicke’s areas developed from the brain-damage studies. Those theories would predict that Broca’s area should be involved in speaking words that have just been seen or heard as well as in speaking words that are mentally generated, and that Wernicke’s area, rather than the spot

FIGURE 5.27 Brain activity during word-processing tasks, as revealed by PET Each figure shows the amount of activity in each part of the left hemisphere during the performance of the indicated task, relative to the amount of activity in that part during the performance of the task one level below it in the hierarchy. The colors violet, blue, green, yellow, and red, in that order, represent increasing amounts of activity.



behind it, should be involved in generating words with the appropriate meaning. At present, the neuroimaging studies are leading to new theories that many regions of the left hemisphere, not just Broca's and Wernicke's areas, are involved in language (Bookheimer, 2002; Dronkers & others, 2000).

SECTION REVIEW

The right and left hemispheres are specialized for different higher functions.	
Split-Brain Studies • Split-brain patients enable researchers to study the functioning of each hemisphere independent of the other. • Results indicate that only the left hemisphere produces language in most people, while the right hemisphere is superior in visuospatial tasks. • Split-brain studies also suggest that the verbal left hemisphere interprets a person's own behavior, making "sense" of it even when the left and right hemispheres produce contradictory actions.	Language and the Left Hemisphere • Much of the left hemisphere is devoted to language use. • Aphasia is any language deficit due to brain damage. Deficits observed in Broca's and Wernicke's aphasia are used to infer functions of the corresponding brain areas. • The results of neuroimaging studies are only partly consistent with aphasia research, which is leading to new theories and more extensive mapping of language areas in the brain.

Effects of Experience on the Brain

People sometimes speak of the "wiring" of the nervous system, or even the "hardwiring" of it. This metaphor is useful for some purposes, but it has its limits. The nervous system is not hardwired like a computer or other human-made device. Neurons are soft, pliable, living cells. They can change their sizes, shapes, excitabilities, and patterns of connections in ways that help adapt their possessors to life's ever-changing circumstances. Every day you grow millions of new synapses and lose millions of others, and at least some of that change depends on the unique experiences you had that day.

If You Use It, It Will Grow

Like muscles, regions of the brain tend to grow when used and to atrophy when not used. Here is some of the evidence for that statement.

Effects of Deprived and Enriched Environments on the Brain

Early evidence that experience can change the structure of the brain arose from experiments, conducted in the 1960s and later, in which rats were housed in either *enriched* or *deprived* environments (Greenough & Black, 1992; Rosenzweig & others, 1972). The enriched environments were large cages in which rats lived with others and had many objects to explore. ("Enriched" is a relative term. These environments were enriched compared to the typical environment of a laboratory rat but not compared to, say, a garbage dump, where a wild rat might live.) The deprived environments were small cages in which each rat lived alone and had no objects except food and a water bottle. After weeks in these environments, the brains of the two groups showed many differences. The brains of the enriched group had thicker cerebral cortexes, larger cortical neurons, more acetylcholine (a prominent neurotransmitter in the cortex), and thicker, more fully developed synapses than did those of the deprived group. These changes occurred even when rats were not placed in the differing environments until well into adulthood.

>>> 27
What brain changes have been observed in rats and mice caged in enriched environments?

A playground for mice In mice, as well as in rats, experience in an enriched environment enhances growth in the brain.



The researchers who performed these early experiments assumed that the brain growth they observed must derive solely from modifications of existing neurons and possibly the addition of new glial cells (the non-neural cells in the brain that provide structural and nutritional support to neurons). It was believed at that time that the mammalian brain is incapable of producing new neurons after birth. In the late 1990s, however, researchers using new techniques found ample evidence that new neurons are constantly being generated in the brain, including the adult human brain (Kempermann & Gage, 1999). Generation of new neurons is most apparent in the hippocampus, a structure known to be involved in learning and memory. New hippocampal neurons are generated more rapidly in rats and mice housed in enriched environments than in those housed in deprived environments (Brown & others, 2003; Prickaerts & others, 2004). There is also evidence of generation of new neurons in association areas of the cerebral cortex, but that evidence is less certain (Gould & Gross, 2002).

Restructuring of the Cortex During Skill Development

28

What evidence shows that practice at a skill alters neural connections so that more neurons become devoted to that skill?

<<< As an animal or person develops skill at a task, ever more neurons in the brain are recruited into the performance of that skill. In one of the first clear demonstrations of this phenomenon, Gregg Recanzone and his colleagues (1992) trained monkeys to discriminate between subtly different rates of vibration applied to a particular patch of skin on one finger. The monkeys received banana pellets for making a certain response each time the vibration rate increased even slightly above 20 cycles per second. Other, “untrained” monkeys received the same vibrations to the skin but were not required to discriminate among them for a food reward. Subsequently, the researchers mapped the somatosensory area of the cortex of all the monkeys by touching points on the skin with a thin probe while recording the activity of cortical neurons. They found that in the trained monkeys the area of the cortex that received input from the “trained” spot of skin was, on average, two to three times larger than the equivalent area in untrained monkeys. Apparently, the brain reorganization resulted not from the skin stimulation per se but from the monkeys’ use of that stimulation to guide their behavior. Other studies have revealed comparable brain changes in visual and auditory sensory areas when animals are trained to discriminate among subtly different sights or sounds (Bakin & others, 1996; Zohary & others, 1994).

Research using PET and fMRI neuroimaging reveals that effects like these occur for people, too. In one such study with stringed-instrument players (six violinists, two cellists, and a guitarist), unusually large areas of the somatosensory cortex responded to stimulation of the fingers of the left hand—the same fingers that the musicians had used for years in fingering the strings of their instruments (Elbert & others, 1995). In other studies, blind people who read Braille showed unusually

large areas of cortical activation in response to stimulation of the index finger that they used in reading Braille (Pascual-Leone & Torres, 1993; Sadato & others, 1996).

Blind people develop other skills, not just reading Braille, that help them compensate for their blindness. At least some of these skills make use of regions of the occipital cortex that in sighted people are devoted to vision. For instance, research has shown that regions of the right occipital lobe that in sighted people are involved in the visual analysis of three-dimensional space become devoted, in the blind, to the task of identifying the places from which sounds are coming (Weeks & others, 2000). Blind people also generally develop superior verbal memory to compensate for their inability to look up information easily or to find objects by looking. In one recent fMRI study, blind and sighted people were given lists of nouns to memorize as their brains were scanned (Amedy & others, 2003). The blind subjects, but not the sighted ones, showed marked activation of portions of the occipital cortex during this task, and they also showed superior memory. Moreover, those blind subjects who scored best on the memory test showed the most activity in the occipital cortex.

Spatial Learning and Growth of the Hippocampus

As described in Chapter 4, some bird species hide seeds in multiple locations and retrieve them in the winter, and these birds generally have larger hippocampi than do related species that do not hide seeds. Researchers working with one of these species, the mountain chickadee, have shown that enlargement of the hippocampus depends at least partly on experience (Clayton, 2001). When caged chickadees are allowed to hide and retrieve seeds, their hippocampi grow, and when they are then prevented from hiding and retrieving seeds for a period of time, their hippocampi shrink again. The hippocampus is involved in many forms of memory, but is especially involved in memory for spatial locations.

A study of London cabbies suggests that extensive spatial learning can increase hippocampal size in humans, too. Cab drivers in big cities develop remarkable spatial abilities, and this is especially true of London cab drivers, who, to get a license, must go through rigorous training, which typically takes about 2 years, and pass a test of their ability to find the shortest route between any two locations in that large city. Brain scans revealed that the posterior (rear) part of the hippocampus (the part most involved in spatial memory) is significantly larger in London cab drivers than in otherwise similar people who do not drive cabs (Maguire & others, 2000). They also revealed a significant positive correlation between years of cab-driving experience and growth in the hippocampus: In general, the longer a man had been driving a cab, the larger was his posterior hippocampus.

>>> 29
What evidence, with birds and with humans, indicates that spatial learning can result in growth in the hippocampus?

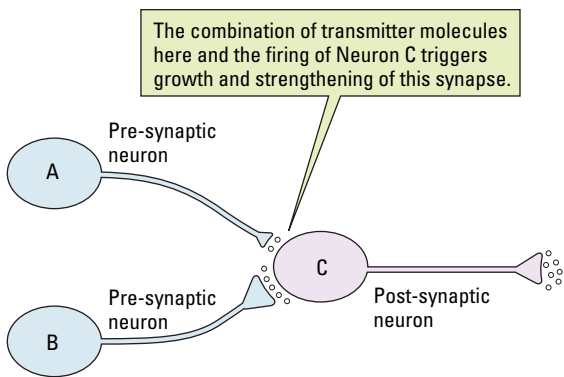
Strengthening of Synapses as a Foundation for Learning

Learning undoubtedly involves many types of changes in the brain. But at the cellular level the type of change that has been most clearly linked to learning is the strengthening of synaptic connections between already existing neurons.

The Hebbian Synapse: Neurons That Fire Together Wire Together

Many years ago, the Canadian psychologist Donald Hebb (1949) theorized that some synapses in the brain have the property of growing stronger (more effective) whenever the post-synaptic neuron fires immediately after the pre-synaptic neuron fires (see Figure 5.28). Through this means, Hebb suggested, neurons could acquire the capacity to respond to input that they previously didn't respond to. This could provide a basis for classical conditioning and other forms of learning. In the 1970s, Timothy Bliss and Terje Lømo (1973) discovered a phenomenon called *long-term potentiation*, or *LTP*, which strongly supports Hebb's theory.

FIGURE 5.28 **A Hebbian synapse** The synapse between neuron A and neuron C is initially weak, so neuron A is ineffective in stimulating neuron C. However, if neuron C fires (because of the firing of neuron B) immediately after neurotransmitter molecules have been released from neuron A onto neuron C, the synapse between neuron A and neuron C will grow stronger. If this happens a sufficient number of times, the synapse may become sufficiently strong that neuron A will be able to trigger action in neuron C even when neuron B is inactive. This type of synapse is called *Hebbian* because its existence was first postulated by Donald Hebb. Recent research into the phenomenon of long-term potentiation confirms the existence of such synapses.



30

How has the discovery of long-term potentiation tended to confirm Hebb's theory about synaptic strengthening?

<<< In the laboratory, LTP is produced by artificially stimulating, with a burst of electrical pulses, a bundle of neurons entering a particular region of an animal's brain. This results in the strengthening of the synapses that those neurons form with post-synaptic neurons, so that weak stimulation of the same bundle elicits a stronger response in the post-synaptic neurons than it would have before. This *potentiation* (strengthening) is *long-term*: It lasts for hours or even weeks, depending on the conditions. Subsequent research has shown that, at least in some brain areas, LTP works in the following manner.

Imagine a weak synapse between neuron A and neuron C (depicted in Figure 5.28). When neuron A becomes active, some of the neurotransmitter molecules it releases become bound to conventional, fast-acting receptors on the post-synaptic membrane, where they produce a depolarization that is too slight to play a significant role in triggering action potentials. Other transmitter molecules at the same synapse, however, become bound temporarily to special LTP-inducing receptors on the post-synaptic membrane. If neuron C then fires an action potential (due to input from other neurons, such as B in Figure 5.28), the combination of that firing and the presence of transmitter molecules in the LTP-inducing receptors triggers a series of biochemical events that strengthen the synapse (Bi & Poo, 2001). The pre-synaptic terminal becomes larger, able to release more transmitter substance than it could before, and the post-synaptic membrane develops more conventional receptor sites than it had before (Zakharenko & others, 2001). In addition, these processes may also trigger the growth of new synaptic connections from neuron A to neuron C. As a result of such changes, firing in neuron A produces more depolarization in neuron C than it did before and therefore plays a greater role in triggering action potentials than it did before.

Evidence That Long-Term Potentiation Is a Basis for Learning

31

What evidence shows that long-term potentiation is involved in learning?

<<< Evidence that LTP is actually involved in learning comes from many experiments showing that interference with the brain's normal capacity for such potentiation interferes with the animal's ability to learn. In one experiment with mice, for example, a drug that prevents LTP was injected into a portion of the amygdala that is known to be crucial for fear learning. After this treatment, the researchers tried, through classical conditioning, to train the mice to fear a tone that was paired with an electric shock. (As you may recall from Chapter 4, classical conditioning occurs when a *conditioned stimulus*, which does not initially elicit a response, is paired with an *unconditioned stimulus*, which does elicit a response. After a sufficient number of pairings, the conditioned stimulus elicits a response even in the absence of the unconditioned stimulus.) The LTP-inhibited mice failed to learn such fear. They responded to the shock as did normal mice, but, unlike normal mice, they did not subsequently show fear responses to the tone (Maren, 1999).

Given such evidence that LTP is essential for learning, what would happen if the capacity for LTP were increased above the normal level? Joe Tsien (2000) and his



Peter Murphy

FIGURE 5.29 | A genetically altered “smart” mouse being tested for object recognition. Mice, like other mammals, explore new objects more than familiar ones. In this test, each mouse was exposed to two objects: one that it had explored in a previous 5-minute session and one that was new. Genetically altered Doogie mice explored the new object (in this case, the one on the right) more than the old one even when several days had elapsed since the initial session. In contrast, unaltered mice explored the new object more than the old one only when a shorter period of time had elapsed. Tsien and his colleagues took this as evidence that the altered mice remembered the original object for a longer period of time than did the unaltered mice. (Photograph from Tsien, 2000.)

colleagues found a way to answer that question. The post-synaptic receptors that are involved in initiating LTP come in two forms—a strong form, which is highly effective in triggering LTP, and a weak, less effective form. Through genetic engineering, Tsien and his colleagues produced a strain of mice—which they named “Doogie,” after the boy genius of the old television show *Doogie Howser, M.D.*—that had many more of the strong receptors and fewer of the weak ones than do normal mice. As predicted, brain neurons in the Doogie mice showed more LTP in response to electrical stimulation than did those in the normal, unaltered mice. Remarkably, but just as predicted, the Doogie mice also showed better memory than the unaltered mice in a series of three widely different tests: maze learning, classical conditioning of a fear response, and object recognition (see Figure 5.29). In each case, the altered mice behaved similarly to the unaltered mice during the learning experience but showed significantly better memory when tested a day or more later (Tang & others, 1999; Tsien, 2000).

The results of this study raise many questions that have not yet been answered. Will similar results be found with other animals and other forms of learning? Could memory in people be improved through methods that increase LTP? If a simple genetic change can improve memory by increasing LTP, then why hasn’t evolution already produced that change through natural selection? Perhaps LTP beyond a certain level, or even memory beyond a certain level, is maladaptive in the conditions of everyday life.

SECTION REVIEW

The brain physically changes in response to experience.	
Growth and Reorganization - Enriched environments, which provide enriched experience, lead to brain growth. Rats in such environments develop thicker cortexes with larger neurons and more and stronger synaptic interconnections. - Skill learning causes larger portions of the brain to become involved in performing that particular skill. - The hippocampus, an area critical to spatial memory, grows as a result of spatial learning (as demonstrated in London taxi drivers).	Long-Term Potentiation (LTP) - LTP strengthens synaptic connections in ways that mediate learning. - In line with Hebb’s theory, the coordinated firing of a pre-synaptic and post-synaptic neuron strengthens the synaptic connections of the first onto the second. - LTP involves enlargement of axon terminals, generation of new receptor sites on post-synaptic membranes, and, in some cases, growth of new synapses.

How Hormones and Drugs Interact with the Nervous System

When the ancient Greeks argued that the heart is the seat of thought, feeling, and behavioral control, they were not without reasons. Like the brain, the heart has long protrusions (blood vessels, in this case) that connect it with other parts of the body. Blood vessels are easier to see than nerves, and because they can be found in all the sense organs and muscles, as well as in other tissues, early theorists believed that blood vessels were the conduits of sensory and motor messages. Today we know that the circulatory system does indeed play a vital communicative role in the body. A slower messenger system than the nervous system, it carries chemicals that affect both physical growth and behavior. Among these chemicals are hormones, which are secreted naturally into the bloodstream, and drugs, which may enter the bloodstream artificially through various routes.

Hormones

Hormones are chemical messengers that are secreted into the blood. They are carried by the blood to all parts of the body, where they act on specific *target tissues*. Dozens of hormones have been identified. The classic hormones—the first to be identified and the best understood—are secreted by special hormone-producing glands called *endocrine glands* (see Figure 5.30). But many other hormones are secreted by organs that are not usually classed as endocrine glands, such as the stomach, intestines, kidneys, and brain.

Comparison of Hormones and Neurotransmitters

32

How are hormones similar to and different from neurotransmitters?

<<< Hormones and neurotransmitters appear to have a common origin in evolution (Snyder, 1985). The earliest multicellular animals evolved a system of chemical messengers that allowed the cells to communicate and coordinate their actions, enabling an animal to behave as a unit. As organisms grew more complex, the chemical communication system differentiated along two distinct routes. One route involved the nervous system and was designed for rapid, specific, point-to-point communication. Here the primitive chemical messengers evolved into neurotransmitters. The other route involved the circulatory system and was designed for relatively slow, diffuse, widespread communication. Here the primitive chemical messengers evolved into hormones.

The main difference between a hormone and a neurotransmitter is the distance that each must travel from its site of release, through fluid, to its site of action. Neurotransmitter molecules released from an axon terminal must diffuse across the synaptic cleft—a distance of about 20 nanometers (20 billionths of a meter)—in order to affect the post-synaptic cell. In contrast, hormone molecules often must travel through the entire circulatory system before they bind to their target cells and exert their effects.

One line of evidence for a common origin of hormones and neurotransmitters lies in their chemical similarity. In fact, some hormones are chemically identical to some neurotransmitters. The chemical *norepinephrine*, for example, is a hormone when secreted into the blood by the adrenal gland (shown in Figure 5.30) but is a neurotransmitter when released by sympathetic motor neurons of the autonomic nervous system on visceral muscles and glands. It is also a neurotransmitter in certain pathways in the brain. In

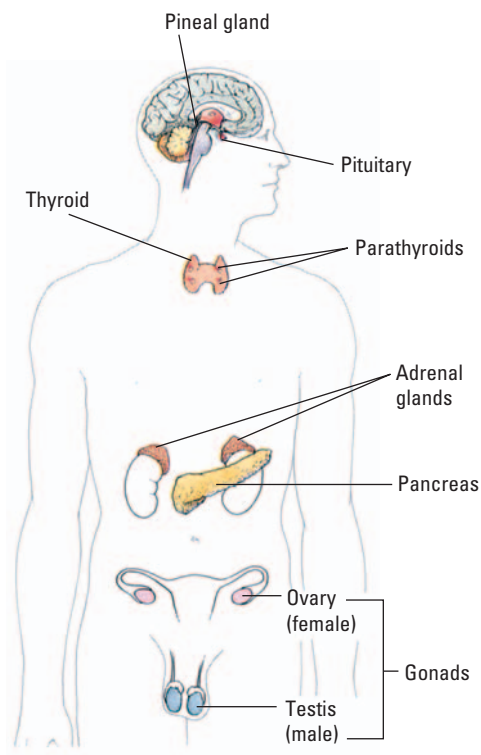


FIGURE 5.30 | Endocrine glands These are some of the glands that secrete hormones into the bloodstream. The pituitary, which is controlled by the brain, secretes hormones that, in turn, control the production of hormones by the thyroid, adrenals, and ovaries or testes.

each of these roles, norepinephrine helps arouse and alert the body. As a hormone and a sympathetic transmitter, it has effects such as increasing the heart rate. As a transmitter in the brain, it helps produce a state that we experience psychologically as high arousal or alertness.

Other evidence for a common origin of hormones and neurotransmitters lies in the existence of **neurohormones**. Like neurotransmitters, these chemicals are produced by neurons and released from axon terminals in response to action potentials. But they are classed as hormones because they are released not into a synaptic cleft but into a bed of capillaries (tiny blood vessels), where they are absorbed into the bloodstream. As you will see, some neurohormones promote the secretion of other hormones, and in that way they provide a means by which the nervous system controls the activity of the endocrine system. To compare the four modes of chemical communication just discussed, see Figure 5.31.

How Hormones Affect Behavior

Hormones influence behavior in many ways. They affect the growth of peripheral bodily structures, including muscles and bones, and in that way influence behavioral capacity. Hormones also affect metabolic processes throughout the body and thereby influence the amount of energy that is available for action. Of greatest interest to psychologists is the fact that hormones also act in the brain in ways that influence drives and moods.

Some effects of hormones are long-term or irreversible, and some of these occur before birth. For example, essentially all the anatomical differences between newborn boys and girls are caused by the hormone testosterone, which is produced by the male fetus but not by the female fetus. These anatomical differences are evident in the brain as well as in the genitals, and the brain differences provide one basis for sex differences in behavior throughout life (discussed in Chapter 6). At puberty, the increased production of sex hormones—especially testosterone in the male and estrogen in the female—stimulates a new set of growth processes that further differentiate males and females anatomically and thereby influence their behavior.

The short-term effects of hormones range in duration from a few minutes to many days. In response to stressful stimulation, for example, the adrenal cortex

33
What are some examples of long-term and short-term effects of hormones?

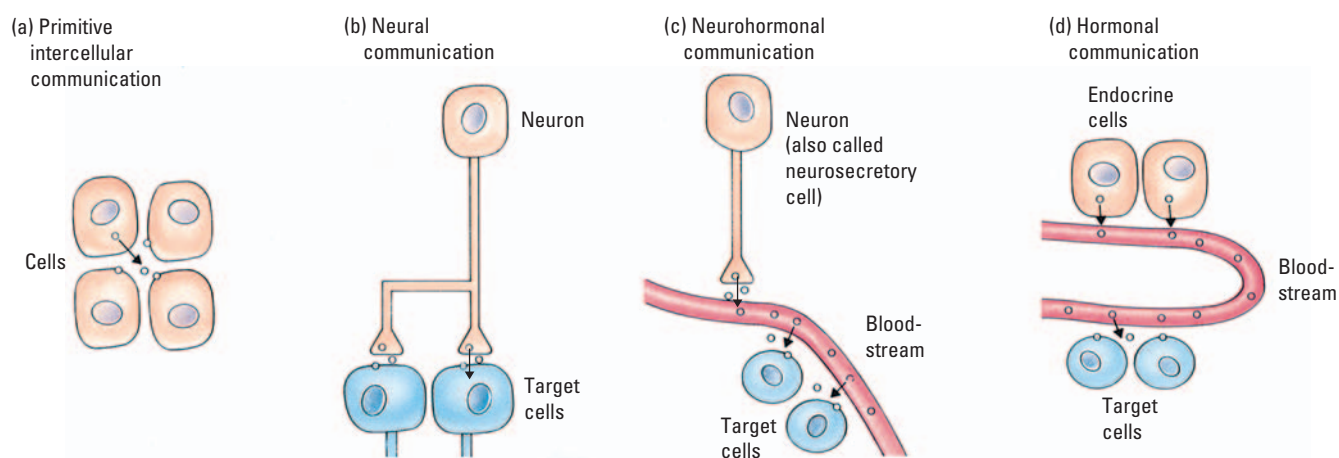


FIGURE 5.31 Four modes of chemical communication between cells The primitive mode (a) of intercellular chemical communication evolved to form three more complex modes of such communication. These are a neural mode (b), in which chemicals from a neuron diffuse across a small synaptic gap to their target cell (another neuron or muscle cell); a neurohormonal mode (c), in which chemicals from neurons enter the bloodstream and are carried to target cells; and a hormonal mode (d), in which chemicals from an endocrine gland enter the bloodstream and are carried to target cells. (Based on figure in Snyder, 1985.)

(the external layer of the adrenal gland) secretes various hormones, including cortisol, which are sometimes referred to as “stress hormones.” These hormones produce a variety of effects throughout the body that help the animal in a stressful situation. For example, they release sugar and fat molecules into the blood to supply extra energy for possible “fight or flight,” and they suppress inflammation caused by wounds. These hormones are also taken up by neurons in certain parts of the brain, especially in the hippocampus (McEwen & others, 1986). Just what role they play in the brain is not yet clear, but some evidence suggests that they help the animal calm down after a period of stress and promote the formation of long-term memories related to the stressful experience (McEwen, 1989).

How Hormones Are Controlled by the Brain

34

How does the brain control the release of hormones from the two lobes of the pituitary and thereby control the release of other hormones as well?

<<< The pituitary, which sits at the base of the brain, is sometimes called the master endocrine gland because it produces hormones that, in turn, stimulate the production of other hormones in other glands, including the adrenal cortex and the gonads (ovaries in the female and testes in the male). But we might more accurately say that the brain is the master endocrine gland, because through neurohormones it controls the pituitary.

To visualize the intimate relationship between the brain and the pituitary, look at Figure 5.32. The rear part of the pituitary, the *posterior lobe*, is in fact a part of the brain. The posterior lobe consists mainly of modified neurons, referred to as *neurosecretory cells*, which extend down from the hypothalamus and secrete neurohormones into a bed of capillaries. Once these hormones enter the capillaries, they are transported into the rest of the circulatory system to affect various parts of the body.

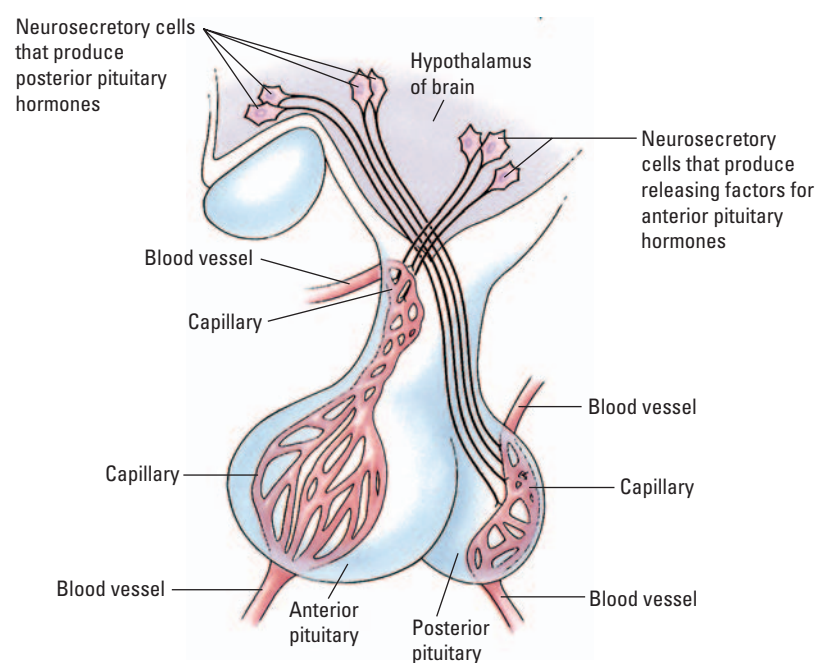


FIGURE 5.32 How the hypothalamus controls pituitary hormones

Specialized neurons in the hypothalamus, called neurosecretory cells, control the activity of the pituitary gland. Some neurosecretory cells secrete hormones directly into capillaries in the posterior pituitary, where they enter the general bloodstream. Others secrete hormones called *releasing factors* into a special capillary system that carries them to the anterior pituitary, where they stimulate the release of hormones manufactured there.

The remainder of the pituitary, the *anterior lobe*, is not part of the brain (no neurons descend into it) but is intimately connected to the brain by a specialized set of capillaries, as shown in Figure 5.32. Neurosecretory cells in the brain's hypothalamus produce *releasing factors*, neurohormones that are secreted into the capillary system and are carried to the anterior pituitary, where they cause the synthesis and release of pituitary hormones. Different releasing factors, produced by different sets of neurosecretory cells in the hypothalamus, act selectively to stimulate the production of different anterior pituitary hormones.

Consider the sequence of hormonal events, triggered by the brain, that is diagrammed in Figure 5.33: (1) A shadowy figure is seen at night, and the brain interprets it as fearsome. (2) The association cortex sends a neural message to the viewer's hypothalamus that causes it to secrete *corticotropin-releasing factor*. (3) The specialized capillary system transports this releasing factor to the anterior pituitary, where it stimulates the release of another hormone, *corticotropin*, into the bloodstream. (4) From the bloodstream, corticotropin enters the adrenal cortex, where it causes the release of adrenal cortical hormones, including cortisol. (5) These adrenal hormones are carried throughout the body to help prepare it for a possible emergency. At the same time, many other brain-controlled effects are also occurring to deal with the possible emergency suggested by the sight of the shadowy figure. They range from the activation of the sympathetic portion of the autonomic nervous system to the development of a plan to escape.

Drugs

The main difference between a hormone and a drug is that the former is produced in the body and the latter is introduced from the outside. Like hormones, drugs are carried by the blood and taken up in target tissues in various parts of the body.

How Drugs Get into the Brain

To be carried to potential sites of action, a drug first must get into an individual's bloodstream. Sometimes drugs are introduced directly into the blood by intravenous (into-a-vein) injection, but more commonly they are administered by routes that bring them into contact with capillaries, where they are gradually absorbed into the blood. Drugs are absorbed by capillaries in the intestines if taken orally, by capillaries in the lungs if taken by inhalation, by capillaries in the rectum if taken by rectal suppository, and by capillaries under the skin or in the muscles if taken by subcutaneous or intramuscular injection. The preferred route of administration depends largely on the properties of the drug and the desired time course of action. A drug that is destroyed by digestive juices in the stomach, for example, cannot be taken orally. A drug that is injected intravenously will act more quickly than one that is administered by any other route.

If a drug is going to act in the brain, it must pass from the blood into the extracellular fluid that surrounds neurons in the brain. The capillaries in the brain are much less porous than those in other tissues. In addition, they are tightly surrounded by the fatty membranes of a certain variety of glial cells. The tight capillary walls and the surrounding glial cells form the **blood-brain barrier**, which helps protect the brain from poisons. To act in the brain, a drug (or hormone) must be able to pass through this barrier. In general, fat-soluble substances pass through easily, and non-fat-soluble substances may or may not pass through, depending on other characteristics of their chemistry.

How Drugs Can Alter Synaptic Transmission

Many different drugs are used in psychiatry and neurology to alter a person's mood or behavioral capacities. Nearly all such drugs work by enhancing or blocking synaptic transmission somewhere in the nervous system (Stone, 1995). In normal synaptic transmission, neurotransmitter molecules are released from the

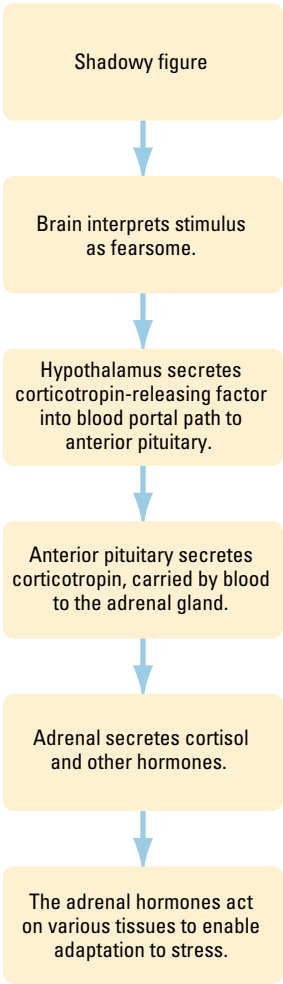


FIGURE 5.33 | Brain-pituitary-adrenal response to a fearful stimulus This is one example of a brain-mediated hormonal response to sensory stimulation.

35

What are three ways in which drugs can alter activity at a synapse?

<<< Drugs can influence activity at a synapse in at least three ways: (1) They can act on the pre-synaptic neuron to either facilitate or inhibit the release of the transmitter, thereby affecting the amount of it that enters the cleft. (2) They can act in the cleft to either facilitate or inhibit the processes that normally terminate the action of the transmitter once it has been released, either prolonging or shortening the amount of time that the transmitter remains in the cleft and exerts its effects. (3) They can act directly on post-synaptic binding sites, either producing the same effect as the transmitter or blocking the transmitter from producing its normal effect.

Of the three, the post-synaptic effect is best understood. To visualize how this effect occurs, recall the comparison of neurotransmitter molecules to keys and binding sites on post-synaptic membrane channels to locks (and look at Figure 5.34). A drug molecule that diffuses into a synapse may act as a substitute key, producing the same effect as the transmitter would, or it may act as a mis-shapen key, filling the keyhole without turning the lock, thereby preventing the transmitter from having its normal effect. Many different types of neurotransmitters exist in the nervous system, and a particular drug may alter the activity of just one or several of them. Thus drugs can be more or less specific, affecting either a small or a large class of synapses, and hence can be more or less specific in the effects they produce on behavior.

How Drugs Act at Different Levels of the Behavior-Control Hierarchy

One way to imagine the many kinds of effects that drugs can have on behavior is to think of them in relation to the hierarchy of behavioral control depicted in Figure 5.21 (on p. 157). Drugs that act at a particular level of the hierarchy have behavioral effects that are consistent with the functions of that level, as described in the figure.

36

How can the effects of curare, L-dopa, and psychoactive drugs be interpreted in terms of the hierarchical model of behavioral control?

<<< *Curare* is a drug that acts at the lowest level, at synapses between motor neurons and skeletal muscle cells. This poison, long used by certain South American Indians on the tips of their arrows for hunting, paralyzes the animal that is hit. Curare produces paralysis by blocking the post-synaptic binding sites for acetyl-

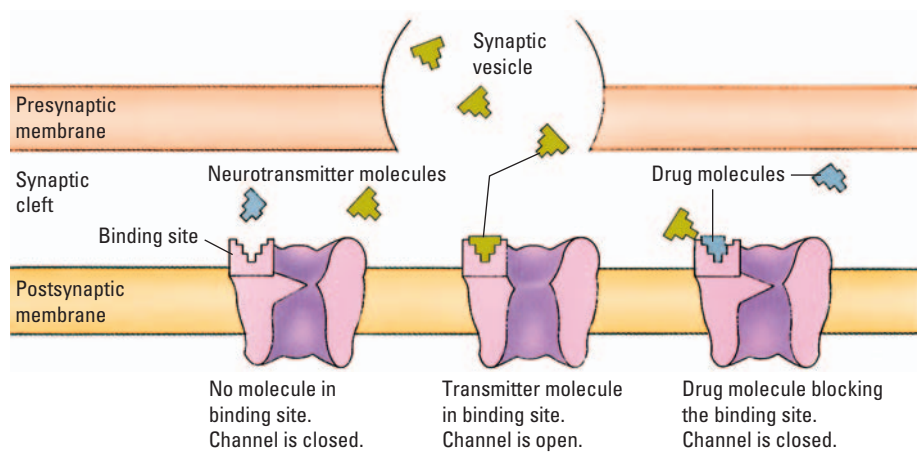


FIGURE 5.34 Lock-and-key view of drug effects in a synapse Each binding site on the membrane of the post-synaptic cell can be thought of as a keyhole, and each neurotransmitter molecule as a key that fits into it and opens the lock, causing the ion channel in the cell membrane to open. A drug molecule can serve as a substitute key and produce the same effect as the neurotransmitter, or it can serve as a misshapen key (as shown here) that occupies the keyhole without activating the lock and thus prevents the neurotransmitter from acting.

choline, which is the transmitter released by skeletal motor neurons onto muscle cells. Because the muscle cells can no longer respond to acetylcholine, they can no longer contract, so the animal is unable to move. Acetylcholine is also a transmitter at many places within the central nervous system, but curare does not act in those places because it cannot pass the blood-brain barrier.

A drug that acts somewhat higher up in the movement-control hierarchy is *L-dopa*, which is used to treat Parkinson's disease. Parkinson's disease is caused by the degeneration of certain neurons whose axons originate in the brainstem and terminate in the basal ganglia. The disease is characterized by muscle tremors and severe difficulty in initiating and coordinating movements. These symptoms are manifested when the degenerating neurons fail to release a sufficient quantity of the neurotransmitter *dopamine* onto neurons in the basal ganglia. It might seem logical to treat these symptoms by introducing additional dopamine into the body, but dopamine cannot cross the blood-brain barrier. However, *L-dopa*, a precursor in the synthesis of dopamine, can cross the barrier. Once in the brain, *L-dopa* is converted to dopamine.

Drugs that work at the highest levels of the behavior-control hierarchy are called *psychoactive drugs*, because they influence processes that we think of as psychological. These are drugs that alter mood or level of arousal or alter one's thoughts and perceptions. Table 5.1 shows one way of categorizing psychoactive drugs and what is known about the mechanism of action of some of them, including such everyday examples as caffeine and nicotine. Chapter 17 describes some of the clinical uses of psychoactive drugs in the treatment of severe anxiety, depression, and other mental disorders.

TABLE 5.1 |
Some categories of psychoactive drugs

1. Behavioral stimulants and antidepressants

Drugs that increase alertness and activity level and elevate mood, including:

- *Amphetamines and cocaine* Often abused because of the psychological “high” they produce, these drugs increase the release of norepinephrine and dopamine into synapses and prolong their action there.
- *Clinical antidepressants* These counteract depression but do not produce the sleeplessness or euphoria that amphetamines and cocaine produce. One class, the *monoamine oxidase inhibitors*, increases the synaptic activity of monoamine neurotransmitters (serotonin, norepinephrine, and dopamine) by inhibiting the enzyme that breaks them down (discussed in Chapters 16 and 17).
- *Caffeine* Found in coffee, tea, and cocoa beans, this drug promotes arousal partly by activating the peripheral sympathetic nervous system. In the brain, it increases neural activity by inhibiting the postsynaptic action of adenosine, which functions as an inhibitory neuromodulator.
- *Nicotine* Found in tobacco, this drug increases activity in many neurons in the brain by acting on some of the same postsynaptic receptor sites as does the neurotransmitter acetylcholine (through the substitute-key mechanism). In this way, it also activates portions of the peripheral sympathetic nervous system, causing, as one effect, increased heart rate.

2. Tranquilizers and central-nervous-system depressants

Drugs that counteract anxiety and/or decrease alertness and activity level, including:

- *Benzodiazepines* Prescribed to counteract anxiety, these drugs (including diazepam, sold as Valium) act on postsynaptic receptors to make them more responsive to GABA, which is an inhibitory transmitter. Increased responsiveness to GABA causes a decrease in neural activity in those parts of the brain where it is present.

- *Alcohol and barbiturates* These often-abused drugs depress neural activity throughout the brain by a variety of means that are not well understood. Low doses may produce relief from anxiety, and higher doses produce sedation, sleep, unconsciousness, coma, and death, in that order. Their antianxiety effects are believed to work in the same way as those of benzodiazepines.

3. Opiates

These include *opium*, a crude extract from the opium plant; *morphine*, the most active ingredient in opium; *codeine*, an extract from opium that is chemically similar to morphine but less active; and *heroin*, which is synthesized by chemical modification of morphine. All are potent pain reducers and can cause feelings of euphoria. They produce at least some of their effects by activating neurons that normally respond to *endorphins*, a class of slow-acting transmitters that are part of the body's natural system for relieving pain (discussed in Chapter 7).

4. Antipsychotic drugs

Prescribed mostly to treat schizophrenia, most of these drugs (including chlorpromazine and haloperidol) are believed to work by decreasing activity at synapses where *dopamine* is the transmitter (discussed in Chapters 16 and 17).

5. Hallucinogenic drugs

These induce hallucinations and other sensory distortions. Some, such as *mescaline* and *psilocybin*, exist naturally in plants and have a long history of use among various peoples, including Native Americans, in religious rites. Others, such as *LSD* (lysergic acid diethylamide), are synthetic compounds. Most are structurally similar to neurotransmitters, and they are believed to act either by mimicking or blocking transmitters that they resemble. Both LSD and psilocybin resemble the transmitter *serotonin*.

Source: Compiled largely from McKim, W. A. (2000) *Drugs and behavior: An introduction to behavioral pharmacology* (4th ed.) Upper Saddle River, NJ: Prentice-Hall.

SECTION REVIEW

Hormones and drugs in the bloodstream can affect the nervous system and thus behavior.

Hormones

- Hormones are chemical messengers secreted into the bloodstream. They are chemically similar (sometimes identical) to neurotransmitters but travel farther and exert their effects on many different target issues.
- Hormones can influence behavior by affecting bodily growth, metabolism, and brain activity, including brain activity responsible for drives and moods.
- Hormonal effects can be long-term or even permanent, as in the case of prenatal testosterone’s effects on brain development. They can also be short-term, as in the case of stress hormones’ effects in preparing for an emergency.
- The pituitary gland controls hormone production by other glands, but is itself controlled by the brain.

Drugs

- Drugs can be introduced into the bloodstream in several ways. To act directly on the brain, they must pass through the blood-brain barrier.
- Drugs influence behavior largely by influencing synaptic transmission. They may stimulate neurotransmitter release, mimic neurotransmitter effects, or block binding sites.
- Drugs can affect different levels of the behavior-control hierarchy. Curare works at the bottom level, L-dopa works at a middle level, and psychoactive drugs such as alcohol work at the top.

Concluding Thoughts

Hippocrates was right: The brain is the organ of the mind. The mind is a set of processes (feeling, thinking, initiating action, and so on) carried out by physical activities in the brain and interfered with by lesions in the brain. The brain, of course, does not work in isolation from the rest of the body or the environment. It needs input from sensory nerves, it is affected by chemicals carried in the blood, and it acts through motor nerves and (to a smaller degree) hormones. Yet the brain is the center of all that we call the mind: It contains the mechanisms needed to analyze all inputs and organize all outputs.

In reviewing this chapter, so full of terms and details, you may find it useful to keep the following broad points in mind:

1. The value of a functionalist perspective As you review the structures described in this chapter—ranging from the little ones, such as *synaptic vesicles* and *neurohormones*, to the big ones, such as the *limbic system* and *autonomic nervous system*—ask yourself, for each: What is it for? That is, what role does it play in the workings of the larger machine that is the human being? How is it related to other parts of the machine, and how can variations in it affect human behavior? The structures are a lot easier to remember, and certainly more interesting, if you think of them in terms of their roles in a larger system rather than as isolated entities.

2. Uses of the hierarchical model The hierarchical model described in this chapter (summarized in Figure 5.21) provides a way to organize thinking about the nervous system. It is a useful memory scheme because it allows us to see each part in relation to the whole. It summarizes, in a very general way, the effects of damage to different parts of the nervous system. It also summarizes, again in a general way, the effects of drugs that act in different parts of the nervous system. As you review the discussion of the central nervous system, and the later discussion of drugs, tie the bits and pieces together into the hierarchical model.

3. Brain science in relation to the rest of psychology As more is learned about the brain, knowledge about it becomes relevant to broader areas of psychology. In later chapters, you will read about the brain in relation to psychology’s attempt to understand basic processes of motivation, sensation, memory, and thought. Still later, you will read of brain-based theories of mental disorders and of drugs that are believed to alleviate specific mental disorders through their interactions with neurotransmitters. Your review of this chapter will be more effective if you try to anticipate the ways that each topic discussed here might be relevant later. Ask yourself: Why might a *psychologist* want to know about this structure or process? You may be surprised at how often you come up with a good answer.

Further Reading

FLOYD BLOOM, CHARLES NELSON, & ARLYNE LAZERSON (2001). *Brain, mind, and behavior* (3rd ed.). New York: Worth.

This is a beautifully illustrated, lucidly written introduction to the brain and its functions of creating the mind and controlling behavior. The book begins with chapters on the basic workings and development of the brain and proceeds through chapters on the brain's role in sensation, movement control, circadian rhythms, emotions, memory, thinking, and mental disorders.

SALLY SPRINGER & GEORG DEUTSCH (1997). *Left brain, right brain* (5th ed.). New York: W. H. Freeman.

Research on right- and left-hemisphere differences, including split-brain studies,

is always interesting. This book provides a well-documented account of such research in a form that can be understood by the beginning student.

DANIEL WEGNER (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

An underlying assumption of the chapter you have just read is that events in our nervous system make us do what we do. Where, then, does free will come in? Wegner's thesis in this enjoyable book is that the experience of willing a movement is not the cause of a movement, but derives from brain activity that occurs milliseconds after the nervous system has begun to produce the movement. Wegner explores the func-

tion of this feeling and presents examples of situations in which our actual movements are not the same as those that we consciously feel that we have willed.

WILLIAM MCKIM (2003). *Drugs and behavior: An introduction to behavioral pharmacology* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Each major class of drugs that alters mood or behavior is discussed here in a separate chapter. McKim describes clearly not just the behavioral effects and physiological mechanisms of each drug but also the drug's role in human history and culture.

FPO

Mechanisms of Motivation and Emotion

GENERAL PRINCIPLES OF MOTIVATION

- Varieties of Drives
- Reward Mechanisms of the Brain
- Drives as States of the Brain

HUNGER: AN EXAMPLE OF A REGULATORY DRIVE

- Neural and Hormonal Control of Appetite
- Roles of Sensory Stimuli in Control of Appetite
- Problems of Obesity

SEX: AN EXAMPLE OF A NON-REGULATORY DRIVE

- Hormonal Influences on Sexual Drive
- Brain Development and Sexual Orientation in Humans

THE SLEEP DRIVE

- Description of Sleep as a Physiological and Behavioral State
- Theories About the Functions of Sleep
- Brain Mechanisms Controlling Sleep

FOUNDATIONS FOR UNDERSTANDING EMOTIONS

- The Nature and Value of Emotions
- Effects of Bodily Responses on Emotional Feelings
- Brain Mechanisms of Emotion

The kaleidoscope that makes a day or a year of mental life has both fast-moving and slow-moving components. Sensations, perceptions, thoughts, and muscle movements flit through our consciousness and behavior at speeds measured in milliseconds. But slower changes, measurable in minutes or hours, modulate and help direct these rapid changes. The slower changes are referred to as *behavioral states*; they include variations in motivation, emotion, and level of arousal.

Are you hungry right now, or angry, or sleepy? Even as you attempt to study this chapter, your mental state affects your capacity to pay attention, and it may direct your attention to some things over others. If you are hungry, your thoughts of food may be capturing most of your attention. If you are sleepy, your reaction even to the most interesting ideas in the chapter may be “oh hmmm ... zzzzzzzz.” Clearly, your mental state affects your momentary thoughts and actions. But what affects your mental state? What exactly makes you hungry, or angry, or sleepy? This is a fundamental question, one that links psychology tightly to the study of the brain and its interactions with the body's internal environment as well as the external environment. This chapter is about the physiological underpinnings of motivation and emotion. You will read first about general principles of motivation, from a physiological perspective, and then about hunger, sexual drive, sleep, dreams, and emotionality—in that order. Social and cultural influences on motivation and emotion, which are only touched on here, are discussed more fully in later chapters.

General Principles of Motivation

To *motivate*, in the most general sense of the term, is to set in motion. In psychology, the term **motivation** is often used to refer to the entire constellation of factors, some inside the organism and some outside, that cause an individual to behave in a particular way at a particular time. Motivation defined this way is a very broad concept, almost as broad as all of psychology. Every chapter in this book deals with one or another facet of motivation. Genes, learning, physiological variables, perceptual and thought processes, developmental variables, social experiences, and personality characteristics are all constructs that psychologists describe as contributors to motivation.

A more precise label for the specific topic of our present discussion is **motivational state** or **drive**. These terms are used interchangeably to denote an internal condition that orients an individual toward a specific category of goals and that can change over time in a reversible way. Different drives have different



© Picturequest / Stock Connection / Studio M

Motivation The predatory drive of this cat is activated by the goldfish—the incentive.

goals. *Hunger* orients one toward food, *sex* toward sexual gratification, *curiosity* toward novel stimuli, and so on. For the most part, drives are thought of in psychology as hypothetical constructs. The psychologist does not observe a state of hunger, thirst, or curiosity inside the animal but infers the existence of that state from the animal's behavior. An animal is said to be hungry if it behaves in ways that bring it closer to food, to be sexually motivated if it behaves in ways that bring it into contact with a sexual partner, and to be curious if it seeks out and explores new environments. To say that the drive varies over time is to say that the animal will work harder, or accept more discomfort, to attain the goal at some times than at others. The assumption is that something inside the animal changes, causing it to behave differently, at different times in the same environment.

1 How do drives and incentives (a) complement one another and (b) influence one another in their contributions to motivation?

<<< But the inside interacts constantly with the outside. Motivated behavior is directed toward *incentives*, the sought-after objects or ends that exist in the external environment. Incentives are also called *reinforcers* (the term used in Chapter 4), *rewards*, or *goals*. The motivational state that leads you to stand in line at the cafeteria is presumably hunger, but the incentive for doing so is the hamburger you intend to purchase. Drives and incentives complement one another in the control of behavior; if one is weak, the other must be strong to motivate the goal-directed action. Thus, if you know that the cafeteria's hamburger tastes like cardboard (weak incentive), you are likely to wait in line for it only if your hunger drive is strong; but if the cafeteria serves a really great hamburger (strong incentive), you are likely to wait even if your hunger drive is weak.

Drives and incentives not only complement each other but also influence each other's strength. A strong drive can enhance the attractiveness (incentive value) of a particular object: If you are very hungry, even a hamburger that tastes like cardboard might seem quite attractive. Conversely, a strong incentive can strengthen a drive: The savory aroma of a broiling hamburger wafting your way as you wait in line might increase your hunger drive, which might in turn induce you to eat something that previously wouldn't have interested you if, by the time you get to the grill, all the hamburgers are gone.

© The New Yorker Collection 2004, Glen LeVie from Cartoonbank.com. All rights reserved.



Varieties of Drives

In general, drives motivate us toward goals that promote our survival and reproduction. Some drives promote survival by helping us maintain the internal bodily conditions that are essential for life.

Drives That Help Preserve Homeostasis

In a now classic book entitled *The Wisdom of the Body* (1932/1963), the physiologist Walter B. Cannon described simply and elegantly the requirements of the tissues of the human body. For life to be sustained, certain substances and characteristics within the body must be kept within a restricted range, going neither above nor below certain levels. These include body temperature, oxygen, minerals, water, and energy-producing food molecules. Physiological processes, such as digestion and respiration, must continually work toward achieving what Cannon termed **homeostasis**, the constancy of internal conditions that the body must actively maintain. Most relevant for psychology, Cannon pointed out that maintenance of homeostasis involves the organism's outward behavior as well as its internal processes. To stay alive, individuals must find and consume foods, salts, and water and must maintain their body temperature through such means as finding shelter. Cannon theorized that the basic physiological underpinning for some drives is a loss of homeostasis, which acts on the nervous system to induce behavior designed to correct the imbalance.

Following Cannon, psychologists and physiologists performed experiments showing that animals indeed do behave in accordance with the needs of their bodily tissues. For example, if the caloric (energy) content of its food is increased or decreased, an animal will compensate by eating less or more of it, keeping the daily intake of calories relatively constant. As another example, removal of the adrenal glands causes an animal to lose too much salt in its urine (because one of the adrenal hormones is essential for conserving salt). This loss of salt dramatically increases the animal's drive to seek out and eat extra salt, which keeps the animal alive as long as salt is available (Stricker, 1973).

The force of homeostasis in human behavior was dramatically and poignantly illustrated by the clinical case of a boy, referred to as D. W., who when 1 year old developed a great craving for salt (Wilkins & Richter, 1940). His favorite foods were salted crackers, pretzels, potato chips, olives, and pickles; he would also take salt directly from the shaker. When salt was denied him, he would cry until his parents gave in, and when he learned to speak, *salt* was one of his first and favorite words. D. W. survived until the age of 3½, when he was hospitalized for other symptoms and placed on a standard hospital diet. The hospital staff would not yield to his demands for salt, and he died within a few days. An autopsy revealed that his adrenal glands were deficient; only then did D. W.'s doctors realize that his salt craving came from physiological need. His strong drive for salt and his ability to manipulate his parents into supplying it, even though they were unaware that he needed it, had kept D. W. alive for 2½ years—powerful evidence for “the wisdom of the body.”

Limitations of Homeostasis: Regulatory and Non-Regulatory Drives

Homeostasis is a useful concept for understanding hunger, thirst, and the drives for salt, oxygen, and an appropriate body temperature, but not for understanding many other drives. Consider sex, for example. People are highly motivated to engage in sex, and sex serves an obvious evolutionary function, but there is no tissue need for it. No vital bodily substance is affected by engaging in sexual behavior; nobody can die from lack of sex (despite what an overly amorous someone may have

>>> **2**
How is the concept of homeostasis related to that of drive? How is this relationship demonstrated in the case of a little boy who craved salt?



Photo Edit

Ahh, homeostasis When the body's water level is reduced from prolonged athletic exertion, nothing is more satisfying than replenishing it. Thirst is a typical example of a regulatory drive.

>>> **3**
What is the distinction between regulatory and non-regulatory drives, and how can mammalian drives be classified into five categories based on function?

told you). Psychologists find it useful to distinguish between two general classes of drives—regulatory and non-regulatory. A **regulatory drive** is one, like hunger, that helps preserve homeostasis, and a **non-regulatory drive** is one, like sex, that serves some other purpose.

A Functional Classification of Mammalian Drives

One way to think about the whole set of drives that we share with other mammals is to categorize them in accordance with their evolutionary functions—their roles in promoting survival and reproduction. In this regard, you may find it useful to distinguish among the following five categories of mammalian drives:

1. *Regulatory drives.* As already noted, these are drives that promote survival by helping to maintain homeostatic conditions in the body. *Hunger* and *thirst* are prime examples.
2. *Safety drives.* These are drives that motivate an animal to avoid, escape, or fend off dangers such as precipices, predators, or enemies. The most obvious safety drive is *fear*, which motivates individuals to flee from danger. Another is *anger*, which is manifested when fight (or at least threat) rather than flight is needed to ensure one's safety. I will argue later in this chapter that *sleep* is also a safety drive. It evolved at least partly as a means of keeping animals tucked quietly away during that part of each 24-hour day when they would be most in danger if they were moving around.
3. *Reproductive drives.* The most obvious of these are the sexual drive and the drive to care for young once they are born. When they are at a peak, these drives can be extraordinarily powerful. Animals (including people) will risk their lives to mate and to protect their offspring. As discussed in Chapter 3, sexual jealousy, including the anger associated with it, also serves the function of reproduction to the degree that it promotes the fidelity of one's sexual partner.
4. *Social drives.* Most mammals, and especially humans, require the cooperation of others to survive. The social drives include the drives for friendship and for acceptance and approval by the social groups of which one is a part. In humans, these drives can be as powerful as the regulatory, safety, and reproductive drives. People will risk their lives for friendship and for social approval. The social drives figure prominently in the discussions that run through Chapters 12, 13, and 14 of this book.
5. *Educative drives.* These consist primarily of the drives to play and to explore (curiosity). As discussed in Chapter 4, the young of nearly all mammals practice life-sustaining skills through play, and mammals of all ages acquire useful information about their environment by exploring novel objects and territories. When other drives are not too pressing, the drives for play and exploration come to the foreground.

Human Drives That Seem Not to Promote Survival or Reproduction

4 What are two possible explanations of the universal human drives for art, music, and literature?

<<< Not all of human motivation is easily understood in terms of survival and reproduction. For instance, humans everywhere like to produce and experience art, music, and literature (including oral stories and poetry). What motivates these activities? Do we have special, evolved aesthetic drives? If so, what adaptive functions prompted the natural selection of such drives?

At present, these questions are much debated and there is no firm answer. My own view is that the pursuits of art, music, and literature are natural extensions of our drives for play and exploration. These pursuits can exercise perceptual and motor skills, imagination, and creative thinking in ways that may be useful in future real-life situations and can also provide us with ideas for governing our own lives. Like other playful and exploratory activities, these pursuits help our minds to grow during periods when there are no more pressing survival needs that must be fulfilled.

A somewhat different (but not incompatible) view, presented by Steven Pinker (1997), is that art, music, and literature appeal to us not because we have special drives for them but because they tap into many of our already existing drives and proclivities, which evolved for other purposes. For example, in describing the appeal of fiction, Pinker (1997, p. 539) writes: “When we are absorbed in a book or movie, we get to see breathtaking landscapes, hobnob with important people, fall in love with ravishing men and women, protect loved ones, attain impossible goals, and defeat wicked enemies.” In this example, a book or movie appeals to our drives for sex, love, social esteem, parenting, achievement, and aggression. To suggest that art, music, and literature may be vicarious means of satisfying other drives rather than drives in and of themselves is not to diminish them. These pursuits enrich our lives immensely; they extend us beyond evolution’s narrow dictates of mere survival and reproduction.

Of course, some things that people become motivated for are truly harmful. Drug addictions and compulsive gambling are artificial drives, created by human inventions, which can ruin people’s lives. How these tap artificially into our natural drive mechanisms is a topic to which we shall return soon.



Serge de Sazo / Photo Researchers

Early art We have no idea whether or not the person who produced this drawing—about 18,000 years ago in one of the Lascaux Caves, in France—had a concept of or word for *art*. To the modern eye, however, this is a work of art. It may represent a universal human proclivity.

Reward Mechanisms of the Brain

As I noted earlier, motivated behavior involves the pursuit of *rewards* (also known as incentives, goals, or reinforcers). Let us now look more closely at the concept of reward and at research into how rewards act on the brain to promote and reinforce the behaviors that produced them.

Three Components of Reward: Liking, Wanting, and Reinforcement

Psychologically, the term *reward* has three interrelated but in some ways separable meanings. A reward is something that we *like*, something that we *want*, and something that serves as a *reinforcer* in learning (Berridge & Robinson, 2003).

Liking refers to the subjective feeling of pleasure, or satisfaction, that occurs when one receives a reward. We know this feeling from our own experience. We experience pleasure from good food when we are hungry, from water when we are thirsty, from drifting off to sleep when we are tired, and from sexual activity when we are sexually motivated. We also experience pleasure from pay, praise, the company of good friends, play, music, discoveries made through exploration, and our own assessment of a job well done. Most of the things we seek in life bring us pleasure when we obtain them. We have no way to know for sure that other animals also experience pleasure from the rewards they receive, but their behavior certainly suggests that they do (see Figure 6.1).



Image Source / Alamy

FIGURE 6.1 | Liking Many species of mammals show a similar facial reaction to good-tasting foods. The reaction includes a tongue protrusion, which looks as if the individual were lapping up the last bits of the food. This expression has been used as an objective index of “liking” in research involving food rewards in laboratory rats. (Berridge & Robinson, 2003.)

5 What are three interrelated components of the concept of reward?

Wanting refers to the desire to obtain a reward. This is the component of reward that links closely to the concept of motivation. Whereas pleasure occurs when a reward is received, wanting occurs before it is received. Wanting is typically measured by assessing the amount of effort an individual will exert, or the amount of pain the individual will bear, in order to obtain the reward. Usually objects that are wanted are also liked, but, as you will see later, it is possible to separate the two.

Reinforcement refers to the effects that rewards have in promoting learning. As discussed in Chapter 4 (in the section on operant conditioning), animals and people learn to attend to stimuli that signal the availability of rewards and, in the presence of those stimuli, to make responses that bring rewards. Somehow, through its effects on the brain, a reward helps to stamp in, or *reinforce*, the memory of stimuli and actions that occurred just before the reward was received. Such learning helps the individual to become more effective in finding and procuring the same type of reward in the future.

Studies of the brain have provided some clues to the mechanisms of each of these three components of reward.

Identification of Reward Neurons in the Brain

6

How did Olds and Milner identify reward pathways in the brain?

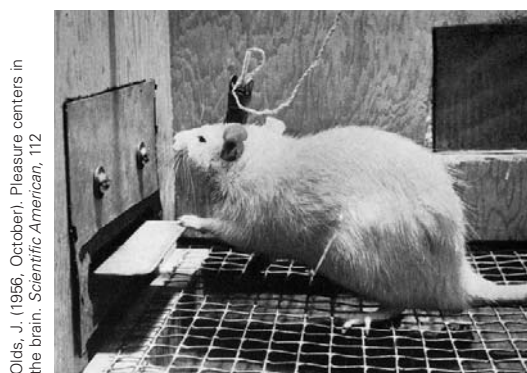
<<< The study of brain mechanisms of reward was initiated in the 1950s, when James Olds and Peter Milner made a remarkable discovery. These researchers observed, by accident at first, that when rats received electrical stimulation through thin wires implanted in certain brain areas, they behaved as if they were trying to get more of that stimulation. For example, if a rat happened to receive the stimulation while exploring a particular corner of the cage, the animal would return repeatedly to that corner. To see if such brain stimulation would serve as a reinforcer for learning, Olds and Milner tested rats in an apparatus in which they could electrically stimulate their own brains by pressing a lever (see Figure 6.2). With electrodes placed in certain brain areas, rats learned very quickly to press the lever and continued to press at high rates, sometimes for many hours without stopping (Olds & Milner, 1954).

Subsequent research showed that rats and other animals will work hardest and longest to stimulate a tract in the brain called the **medial forebrain bundle**. The neurons of this tract that are most crucial for this rewarding effect have their cell bodies in nuclei in the midbrain and synaptic terminals in a large nucleus in the basal ganglia called the **nucleus accumbens** [uh-cum-bens] (see Figure 6.3). The nucleus accumbens itself has connections to large areas of the limbic system and the cerebral cortex, and it is now understood to be a crucial center for the behavioral effects of rewards, in humans as well as in other mammals.

7

What is some evidence that the medial forebrain bundle and nucleus accumbens are essential pathways for the effects of a wide variety of rewards?

<<< When Olds and Milner inserted electrodes into the medial forebrain bundle, they were tapping artificially into the brain's natural reward system. Subsequent research, involving the recording of activity in the brain, has shown that the medial forebrain bundle and the nucleus accumbens become active in all sorts of situations in which an individual receives a reward, whether the reward is food, opportunity to copulate, novel objects to explore, or (in humans) a prize received



Olds, J. (1956, October). Pleasure centers in the brain. *Scientific American*, 112

FIGURE 6.2 | Lever pressing for electrical stimulation to the brain
Each time this rat presses the lever, it receives a brief pulse of electrical current through an electrode implanted in its brain. Some of Olds' rats continued to press the lever for 24 hours without rest, as often as 5000 times per hour. (From Olds, 1956.)



FIGURE 6.3 | A reward pathway in the brain The medial forebrain bundle is a neural tract consisting of neurons whose cell bodies are in nuclei in the midbrain and whose synaptic terminals are in the nucleus accumbens of the basal ganglia. Rats will quickly learn to press a lever to electrically stimulate the medial forebrain bundle or the nucleus accumbens.

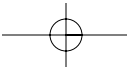
for winning a game (Breiter & others, 2001; Damsma & others, 1992; Fiorello & others, 2003). Moreover, damage to either of these brain structures destroys all sorts of motivated behaviors. Without a functioning medial forebrain bundle or nucleus accumbens, animals will not work to obtain rewards and will die unless they are given food and water artificially through tubes into their stomachs (Grossman, 1979; Stricker, 1982).

Separating the “Liking” and “Wanting” Systems

Many of the neurons of the medial forebrain bundle that terminate in the nucleus accumbens release **dopamine** [dope-uh-mean] as their neurotransmitter. This release appears to be essential for the “wanting” component of reward, but not for the “liking” component. Animals that have been well trained to press a lever for some reward, such as food, show a release of dopamine into the nucleus accumbens just before they begin to press the lever, but not after they receive the reward (Phillips & others, 2003). This pattern is consistent with the idea that dopamine helps motivate the animal to work for the reward (promotes “wanting”) but is not essential for the pleasure received from obtaining the reward.

More direct evidence that dopamine is involved in “wanting” but not “liking” comes from studies in which rats are treated with drugs that block the effect of dopamine in the nucleus accumbens. These animals continue to consume foods, copulate with sexual partners, and explore novel stimuli that are immediately present. They also continue to exhibit the facial “liking” expression (shown in Figure 6.1) when they taste a sugar solution. However, they do not continue to seek out or work for rewards that are not immediately present (Berridge & Robinson, 2003; Bevins, 2001; López & Ettenberg, 2002; Zhang & others, 2003). Their behavior suggests that they continue to enjoy the consumption of rewards, but are no longer concerned with (no longer behave as if they want) rewards that are absent. Conversely, drugs that increase the activity of dopamine in the nucleus accumbens increase the rate at which rats and other animals will work for food, but do not increase the facial “liking” response to sucrose or the animal’s consumption of food that is immediately available (Berridge & Robinson, 2003).

>>> 8 What is some evidence that different neurotransmitters in the nucleus accumbens are involved in the “wanting” and “liking” components of reward?



If dopamine is responsible for the “wanting” component of reward, what is responsible for the “liking” component? Some of the neurons of the medial forebrain bundle that terminate in the nucleus accumbens release not dopamine but a different transmitter, one that is in the **endorphin** [en-**dorf**-in] family. The term *endorphin* is short for *endogenous morphine-like substance*. (*Endogenous* means “created within the body.”) Endorphins are chemicals created within the body that have effects similar to those of morphine and other opiate drugs such as opium and heroin. As discussed in Chapter 7, endorphins are best known for their role in inhibiting the sense of pain. Recent experiments suggest, however, that endorphins released into the nucleus accumbens are also crucial for the immediate pleasure experienced when rewards are received or consumed. Injections, into the nucleus accumbens, of drugs that increase the effectiveness of endorphins increase the facial “liking” reaction to sucrose (Berridge & Robinson, 2003) and also increase the amount of immediately present food that an animal will eat (Zhang & Kelley, 2000). In humans, drugs that decrease the effectiveness of endorphins have been reported to decrease people’s perceived enjoyment of food and other rewards (Yeomans & Gray, 1996).

Role of Dopamine in Reinforcement for Learning

9 What evidence suggests that dopamine is crucial to the capacity of rewards to promote new learning?

<<< The learning component of reward is closely related to the “wanting” component. Animals learn to associate certain cues with the availability of a reward, and those cues lead the animal to search for or work for the reward, which is the behavioral indicator of “wanting.” The release of dopamine into the nucleus accumbens appears to be crucial not just for motivating animals to work for rewards, but also for their ability to learn to use cues to predict when and where rewards are available. One line of evidence for this contention is the observation that dopamine release promotes long-term potentiation (LTP) of neural connections within the nucleus accumbens (Reynolds & others, 2001). As discussed in Chapter 5, LTP is believed to be part of the cellular basis for learning throughout the brain.

Other evidence comes from studies in which the amount of dopamine released into the nucleus accumbens is directly measured as animals anticipate and receive rewards (Schultz, 1998). If food is occasionally, at unpredictable times, presented to a hungry rat, a burst of dopamine is released into the nucleus accumbens each time food is presented. If the situation is then changed so that a signal light comes on 2 seconds prior to each presentation of food, the animal soon learns to anticipate food each time the light comes on. In this new situation, after several trials, no burst of dopamine release occurs when the food is presented, but a burst occurs, instead, when the light comes on. The animal continues to eat and apparently enjoy the food, but no dopamine release accompanies that behavior.

This pattern of dopamine release is consistent with the idea that dopamine is involved in new learning (Schultz, 1998). When a reward is unexpected, dopamine release immediately after the reward helps to reinforce the remembered association between the reward and any stimulus or response that happened to precede it. When the cues and responses leading to a reward have already been well learned, however, there is no need for further reinforcement of that learning, and dopamine release in response to the reward ceases. Dopamine release now occurs in response to the signal preceding reward, because now the animal’s interest lies in learning how to predict when the signal will appear or how to make it appear.

Drug Addiction: Hijacking the Brain’s Reward System

10 How does an understanding of the brain’s reward system help us to understand drug addiction and compulsive gambling?

<<< When Olds and Milner delivered electrical impulses into the medial forebrain bundle in rats, they were short-circuiting the rats’ natural reward systems. When people take drugs such as cocaine, amphetamine, heroin, or opium they are doing essentially that same thing with their own brains. All these drugs, and other often-abused drugs as well, exert their euphoric and habit-producing effects through action on the brain’s reward pathways (Nestler & Malenka, 2004). In various ways, these drugs mimic or promote the effects of dopamine and endorphins in the nucleus accumbens.

Rats fitted with tubes and mechanisms for pumping drugs into their bloodstreams will self-administer cocaine and other such drugs, and become addicted, but will stop self-administering the drugs if the nucleus accumbens is destroyed or chemically blocked (Wise, 1996). Rats will work as hard to administer tiny amounts of cocaine or amphetamine through a cannula (tiny tube) directly into the nucleus accumbens as they will to administer much larger amounts into the bloodstream (Hoebel & others, 1983; Wood & Emmett-Oglesby, 1989).

Our understanding of the brain's reward mechanisms gives us a clue as to why such drugs are addictive. Not only do they produce an immediate sense of euphoria, but, even more significant for the problem of addiction, they strongly activate the dopamine-receiving neurons in the nucleus accumbens that are responsible for promoting reward-based learning. Normal rewards, such as food, activate these neurons only when the reward is unexpected; but cocaine and other abused drugs activate these neurons every time the drug is taken. The result may be a sort of super learning (Self, 2003). With each dose of the drug, the dopamine response acts to reinforce, once again, associations between any cues that are present in the environment and the feelings and behaviors of wanting and taking the drug. The result is the buildup of an extraordinarily strong craving and habit, which are triggered whenever cues that have been present during past drug taking are present.

It has often been observed that drug addicts gradually lose their "liking" of the drug (experience less pleasure) over time, even while their "wanting" of the drug increases (Kelley & Berridge, 2002). The loss in liking occurs, presumably, because of drug-induced changes in the brain that reduce the endorphin-mediated pleasure response. However, because the dopamine response is not reduced, the learned drug craving and habit continues to grow stronger with each dose (Kelley & Berridge, 2002; Nestler & Malenka, 2004). The craving itself, rather than any expected pleasure, becomes the main reason for taking the drug. Drug taking becomes a compulsion rather than something that one freely chooses to do for pleasure. (Another reason why drug addicts continue to take drugs, having to do with conditioned counteractive responses, is discussed in Chapter 4.)

A Brain-Based Theory of Compulsive Gambling

In North America somewhere between 1 and 2 percent of adults suffer from a compulsive, pathological drive to gamble (Grant & others, 2004), a drive that persists even though it may leave the person and his or her family in financial ruins. Compulsive gambling is in some ways similar to drug addiction (Holden, 2001).



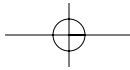
Janine Wiedel / Photo Library / Alamy

Wanting without liking After repeated use, a drug such as cocaine or heroine continues to promote dopamine release in the brain, and thus to reinforce "wanting," but the drug may no longer promote endorphin release, so the "liking" response no longer occurs.



David Young-Wolff / Photo Edit

Internet gambling Whether done on the Internet or at an actual casino, gambling can become compulsive. Because wins are essentially random and unpredictable, each win may result in a new burst of dopamine release, which helps to create the compulsion.



Gamblers claim to feel a euphoric high when they are playing their game and winning, and to experience withdrawal symptoms—such as sweating, restlessness, and sleeplessness—when they try to abstain. Every cue in the environment that has been previously associated with gambling elicits in them a strong, often irresistible urge to gamble. Our understanding of the brain's reward mechanisms gives us some clues about the origins of this compulsion.

Brain imaging studies (using the fMRI technique, discussed in Chapter 5) with healthy human subjects have revealed that games of chance, with monetary rewards, are powerful activators of the nucleus accumbens and other structures known to be part of the brain's reward system (Breiter & others, 2001; Knutsen & others, 2001). Because the payoff is never predictable, every instance of payoff results in a new burst of dopamine release in the nucleus accumbens, no matter how many times the person plays. Thus, gambling, like drug taking, overrides the brain's dopamine-conserving mechanism—the mechanism that shuts off the dopamine response when reward is predictable. Consciously, the person may know that the game pays off in a way that is unpredictable and uninfluenced by anything that he or she does, but the brain's primitive reward system nevertheless behaves as if it is constantly trying to learn how to predict and produce the reward. The repeated reinforcement, by dopamine, of associations between payoffs and the cues and behaviors that precede each payoff results in the buildup of an abnormally strong habit.

Drives as States of the Brain

The reward mechanisms just discussed are involved in all sorts of motivated behaviors. Their destruction produces an animal that fails to seek out or work for the objects and conditions that it needs for survival and reproduction. But what accounts for the differences among different drives? What makes an animal thirsty, or hungry, or sexually motivated, or desirous of sleep?

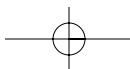
According to the **central-state theory of drives**, which is accepted by nearly all neuroscientists today, different drives correspond to neural activity in different sets of neurons in the brain. A set of neurons in which activity constitutes a drive is called a **central drive system**. Although the central drive systems for different drives must be at least partly different from one another, they may have overlapping components. For example, because hunger and sex are different drives, the neural circuits for them cannot be identical. However, those circuits may share components that produce behavioral effects that are common to both drives, such as increased alertness and general level of motor activity.

11

In theory, what characteristics must a set of neurons have to function as a central drive system? What characteristics of the hypothalamus seem to suit it to be a hub of such systems?

<<< What characteristics must a set of neurons have to serve as a central drive system? First, it must receive and integrate the various signals that can raise or lower the drive state. For hunger, these signals include chemicals in the blood, the presence or absence of food in the stomach, and the sight and smell of food in the environment. Second, a central drive system must act on all the neural processes that would be involved in carrying out the motivated behavior. It must direct perceptual mechanisms toward stimuli related to the goal, cognitive mechanisms toward working out strategies to achieve the goal, and motor mechanisms toward producing the appropriate movements. Look back at Figure 5.21 (p. 157), which depicts a hierarchical model of the control of action, with mechanisms involved in motivation and planning at the top. The central drive systems are part of that top level of the hierarchy. To affect behavior (for example, to cause a hungry person to cross a room for some peanuts), they must influence the activity of motor systems at lower levels of the hierarchy.

Researchers have sound reasons to believe that the hypothalamus is the hub of many central drive systems. Anatomically, this brain structure is ideally located to play such a role (see Figure 6.4). It is centered at the base of the brain, just above the brainstem, and is strongly interconnected with higher areas of the brain. It also has direct connections to nerves that carry input from, and autonomic motor



output to, the body's internal organs. It has many capillaries and is more sensitive to hormones and other substances carried by the blood than are other brain areas. Finally, through its connections to the pituitary gland, it controls the release of many hormones (as described in Chapter 5). Thus, the hypothalamus has all the inputs and outputs that central drive systems would be expected to have. And, as you will soon see, small disruptions in particular parts of the hypothalamus can have dramatic effects on an animal's drives.

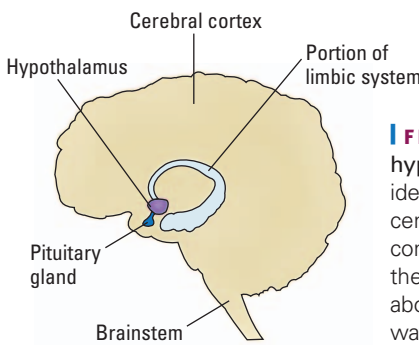


FIGURE 6.4 | Location of the hypothalamus The hypothalamus is ideally situated to serve as a hub for central drive systems. It has strong connections to the brainstem below, the limbic system and cerebral cortex above, and the endocrine system (by way of its tie to the pituitary gland).

SECTION REVIEW

Drives are reversible internal conditions that orient individuals toward specific types of goals.

Drives

- Drives (motivational states) complement and interact with incentives (the objects of drives).
- Regulatory drives (e.g., hunger) promote homeostasis, while nonregulatory drives (e.g., sex) serve other purposes.
- Mammalian drives can be classified by function into regulatory, safety, reproductive, social, and educative categories. Humans also exhibit drives toward art, music, and literature.
- Different drives involve different central drive systems. The hypothalamus is a hub of such systems.

Reward Mechanisms in the Brain

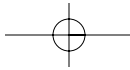
- Reward mechanisms mediate three interrelated components of reward—wanting, liking, and reinforcement.
- The medial forebrain bundle and nucleus accumbens contain essential pathways of the brain's reward system. Animals will work for electrical stimulation in these areas.
- The release of dopamine into the nucleus accumbens is associated with wanting; the release of endorphins into this area is associated with liking.
- The release of dopamine into the nucleus accumbens is also crucial to reinforcement; it promotes learning about how to predict and obtain a given reward.
- Addictive drugs and the random reward schedules used in gambling may override the mechanism that shuts down dopamine release, resulting in the overlearning of cues and habits.

Hunger: An Example of a Regulatory Drive

It is no accident that eating is one of life's great pleasures. Throughout our evolutionary history, the difficulty of finding adequate food was one of the major barriers, if not the major barrier, to survival. As a result, natural selection built into us (and other animals) powerful, robust hunger mechanisms that lead us to search for food, to eat when food is available, and to experience pleasure when we eat. Natural selection also built into us satiety mechanisms, which tend to keep us from overeating and becoming obese. But the satiety mechanisms are not as robust as the hunger mechanisms. In our evolutionary history, food scarcity was a much bigger problem than overabundance. Far more people died of starvation than of obesity (and this is still true in many parts of the world today). Things are different, however, for those of us who live in postindustrial parts of today's world. For us (and especially for those of us in North America), obesity has become a major health problem. In what follows, we will examine some of the mechanisms that control appetite and then look at the modern problem of obesity.

Neural and Hormonal Control of Appetite

The purpose of hunger and satiety is to regulate the amount of food materials in the body at an appropriate level for survival and well-being. Any regulatory system,



12

What is meant by feedback control, and how does the arcuate nucleus of the hypothalamus serve as a control center for appetite?



whether it is part of a human-made machine or part of an animal, makes use of *feedback control*. The substance or quality being regulated feeds back upon the controlling device and inhibits the production of more of that substance or quality when an appropriate level is reached. A home thermostat, which controls the operation of the heating system, is a good example. The thermostat is sensitive to temperature. When the temperature is low, a switch closes in the thermostat, which turns on the furnace, which provides heat. When the temperature rises above the set level, the switch opens and the furnace turns off. Thus heat, produced by the furnace, feeds back onto the thermostat to shut off the source of heat when enough of it is present.

The mammalian brain regulates food intake in a manner that is a bit like the operation of a home thermostat, but far more complicated. Sets of neurons in the brain's hypothalamus raise or lower the animal's drive to eat, and these neurons are themselves regulated by the body's deficit or surfeit of food materials. We might think of these neurons as the brain's "food-o-stat." When food materials are relatively low in the body, the food-o-stat cranks appetite up, which motivates the animal to consume more food. When food materials are plentiful in the body, various indicators of that plenitude feed back upon the food-o-stat and turn appetite off, or at least down a bit.

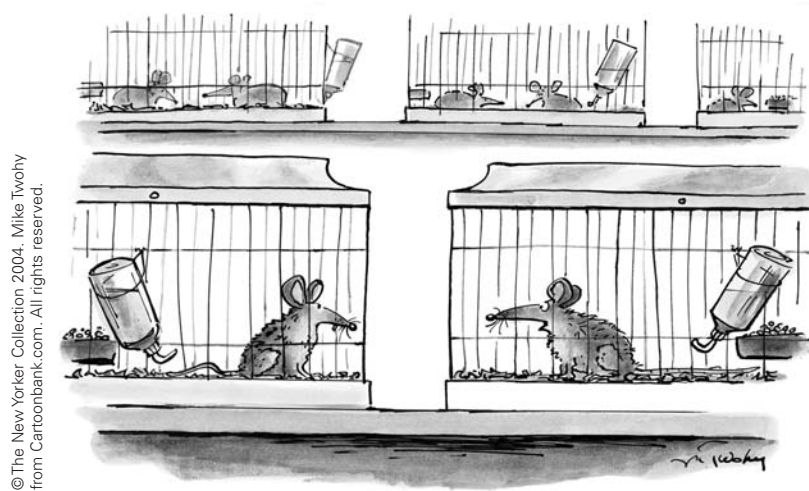
A Nucleus in the Hypothalamus Serves as an Appetite-Control Center

The neurons that comprise the food-o-stat exist in several closely interconnected portions of the hypothalamus, but are most concentrated in the *arcuate* [*arc*-you-ut] **nucleus**, which lies in the center of the lowest portion of the hypothalamus, very close to the pituitary gland (Korner & Leibel, 2003). This tiny brain area has been described as the "master center" of appetite control and weight regulation (Marx, 2003). It contains two classes of neurons that have opposite effects on appetite. One class, the appetite-stimulating neurons, connect to various parts of the brain and promote all the effects that are associated with increased hunger, including craving for food, increased attention to food-related cues, increased exploration in search of food, and heightened enjoyment of the taste of food. The other class, the appetite-suppressing neurons, have effects on various parts of the brain that are opposite to those of the appetite-stimulating neurons.

Both of these classes of arcuate neurons exert their effects on other brain areas through the release of slow-acting neurotransmitters. Unlike fast transmitters, slow transmitters (discussed in Chapter 5) have the capacity to alter neural activity for long periods of time—in this case for periods ranging from minutes to several hours. One of the neurotransmitters released by appetite-stimulating neurons is *neuropeptide Y*, which is the most potent appetite stimulator yet discovered. When injected into any of various regions in the hypothalamus, this chemical causes a previously sated animal to eat voraciously (Stanley & Gillard, 1994). The neurons of the arcuate nucleus are themselves acted upon by many different inputs that, in one way or another, reflect the need or lack of need for food.

Many Internal Signals Contribute to Short-Term Regulation of Appetite

Eating a large meal produces a number of physiological changes in the body. Among these are a slightly elevated body temperature (resulting from a heightened rate of metabolism), an increased blood level of glucose (a simple sugar molecule derived from the breakdown of carbohydrate foods), distention of the stomach and intestines (resulting from food inside those structures), and the release of certain hormones produced by endocrine cells in the stomach and intestines. There is evidence that all these changes can act either directly or indirectly on neurons in the arcuate nucleus and nearby areas of the hypothalamus to activate hunger-suppressing neurons and inhibit hunger-stimulating neurons (Havel & others, 2000; Korner & Leibel, 2003). When all these effects are operating properly, the result is a decline in appetite for several hours following ingestion of a meal.



© The New Yorker Collection 2004. Mike Twotchy from Cartoonbank.com. All rights reserved.

“What if these guys in white coats who bring us food are, like, studying us and we’re part of some kind of big experiment?”

One appetite-suppressing hormone that has received much recent attention is *peptide YY₃₋₃₆* (abbreviated **PYY**), which is produced by special endocrine cells in the large intestine. Food entering the intestines after a meal stimulates secretion of PYY into the bloodstream. In humans, blood levels of the hormone begin to increase 15 minutes after a meal is eaten, peak at about 60 minutes, and remain elevated for as long as 6 hours after a large meal (Batterham & others, 2003). Research with rodents shows that one of the target tissues of PYY is the arcuate nucleus, where the hormone excites appetite-suppressing neurons and inhibits appetite-stimulating neurons (Marx, 2003).

In both rats and humans, injection of extra PYY into the bloodstream reduces total food consumed over the next several hours. In one double-blind experiment with humans, PYY injection reduced the amount of food eaten at a buffet luncheon, by both lean and obese human volunteers, by an average of about 30 percent and also reduced reported level of appetite in both groups (Batterham & others, 2003). The same researchers also found that lean subjects had higher baseline levels of naturally produced PYY than did obese subjects, and exhibited a much greater increase in PYY following a meal (see Figure 6.5). This result suggests that insufficient PYY production may be a contributing cause of obesity. As you can well imagine, pharmaceutical companies are currently investigating the possibility that PYY, or some modified form of it, might be developed and marketed as a weight-control drug.

13 What is the evidence that the hormone PYY helps reduce appetite after a meal and that underproduction of PYY may contribute to obesity?

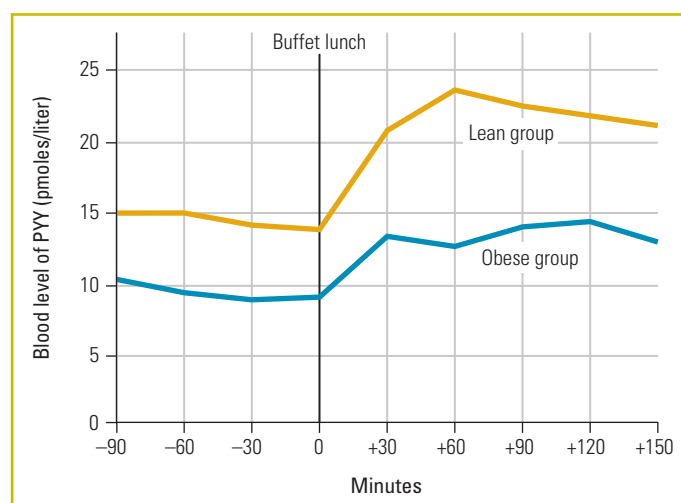


FIGURE 6.5 | Blood levels of the PYY in obese and lean human subjects Blood levels of the hormone PYY were found to be lower in obese subjects than in lean subjects. After eating a meal, lean subjects showed a much larger increase in PYY than did the obese subjects. Researchers believe that this difference may be a cause of obesity, as PYY acts on the brain to suppress appetite. (From Batterham & others, 2003.)

Leptin Contributes to the Long-Term Control of Appetite and Body Weight

14

How does the hormone leptin contribute to weight regulation, and why isn't leptin a good anti-obesity drug?

<<< In the short term, eating provides an immediate supply of building blocks (such as amino acids) and energy molecules (such as glucose) needed to grow, repair, and fuel the body. For the long term, eating more than is immediately needed adds to the amount of fat that is stored in special fat cells in various tissues of the body. Fat stores provide an extra source of energy that the body can call upon when food is not available in the environment. Too much fat, however, impedes movement and puts stress on all the body's organs. Not surprisingly, the hunger mechanism, when it is working optimally, is sensitive not just to the short-term indicators of amount of food recently eaten but also to the amount of fat stored in the body.

Fat cells in mice, humans, and other mammals secrete a hormone, called **leptin**, at a rate that is directly proportional to the amount of fat that is in the cells (Woods & others, 2000). Leptin is taken up into the brain and acts on neurons in the arcuate nucleus and other parts of the hypothalamus to reduce appetite. Animals that lack either the gene needed to produce leptin or the gene needed to produce the receptor sites for leptin in the hypothalamus become extraordinarily obese (see Figure 6.6) (Friedman, 1997). Some people—though very few—lack the gene needed to produce leptin. Such individuals are extremely obese, but they reduce their eating, lose weight rapidly, and keep it off, when given daily injections of leptin (Farooqi & others, 2002).

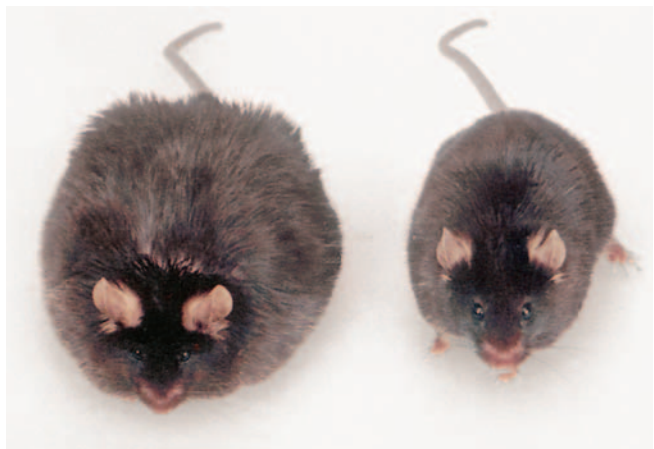


FIGURE 6.6 | Lacking leptin The mouse on the left, which lacks the gene needed to synthesize the weight-regulating hormone leptin, outweighs two genetically normal mice. Without leptin, the animal overeats and becomes obese.

When the hunger-suppressing effect of leptin was discovered, in the 1990s, there was much excitement about the possibility that injections of this hormone could help many people lose weight. Subsequent research, however, proved otherwise. The great majority of overweight people are not lacking in leptin. Their fat cells constantly produce very high levels of the hormone. Research shows that hunger is reduced by increased leptin up to a certain level, but most overweight people already have blood concentrations of leptin well above that level, and additional leptin has no effect (Marx, 2003). Other research suggests that many obese people feel chronically hungry not because they lack leptin but because their brains are relatively insensitive to the hormone (Friedman, 2003). A drug that helped restore leptin sensitivity might help them lose weight, but so far no such drug is in sight.

Roles of Sensory Stimuli in Control of Appetite

15

How do conditioned stimuli and the availability of many foods, with different flavors, contribute to appetite and obesity?

<<< As you well know, hunger is provoked not just by events inside us but also by sensory stimuli in the environment. In the natural environment, food is not always available. That was especially true in the environment of our evolutionary ancestors. Evolution has led us and other animals to be opportunists; we are more likely to experience hunger when food is available than when it isn't.

Through classical conditioning (discussed in Chapter 4), any cues that have been previously associated with the opportunity to eat—such as the sight or smell of good food, the sound of a dinner bell, or the sight of a clock showing it is dinner time—can bring on a sudden surge of appetite. Such conditioning is reflected not just in reports of increased hunger but also in the occurrence of reflexive physiological responses, such as the secretion of saliva and digestive juices, that help to prepare the body for food and that add further to the sense of hunger (Woods & others, 2000).

The taste of food also plays a major role in appetite. People and laboratory animals who eat one type of food until they are satiated experience renewed appetite when a different food, with a different taste, is placed before them. This phenom-

enon is referred to as *sensory-specific satiety*, and many experiments show that it is mediated primarily by the sense of taste (Raynor & Epstein, 2001). When people eat one food at a meal, their rating of the taste pleasantness of that food declines relative to their rating of the taste pleasantness of other foods. This effect begins immediately after eating the food and lasts typically for several hours. Experiments with animals show that the sight and smell of a new food can result in renewed activity in appetite-stimulating neurons in the hypothalamus after the animal has been sated on a different food (Rolls & others, 1986). Laboratory animals that can regularly choose from a variety of different-tasting foods eat more, and become fatter, relative to animals that have only one food choice, even if the nutritional content of the different foods is identical (Raynor & Epstein, 2001). People, too, eat more when there are more food choices (Raynor & Epstein, 2001).

Problems of Obesity

Human evolution occurred almost entirely in environments where food choices were far fewer, and far lower in fat and sugar, than are the foods available to people in modern cultures. Natural selection built into us excellent mechanisms to defend against weight loss in times of scarcity but rather poor mechanisms to defend against obesity in times of plenty. Obesity is a cultural disease of our time and place.

To assess obesity, the World Health Organization uses a measure called the *body mass index*, or **BMI**, defined as body weight in kilograms divided by the square of the person's height in meters. A BMI of 25 or more is considered *overweight*, and one of 30 or more is considered *obese*. Thus a person who is 1.7 meters (5 feet 7 inches) tall and weighs 73 kilograms (161 pounds) is deemed overweight ($BMI = 73/1.7^2 = 25.3$), and a person of that same height who weighs 88 kilograms (194 pounds) is considered obese ($BMI = 88/1.7^2 = 30.4$).

By these criteria, according to a survey taken in 1999–2000, 65 percent of adults in the United States are overweight and 31 percent are obese (Hill & others, 2003). By comparison, in a similar survey just 10 years earlier, 56 percent of adults were overweight and 23 percent were obese (Hill & others, 2003). Obesity is also rising rapidly in many other parts of the world. As obesity rises, so does the rate of diseases that are secondary to it, including Type 2 diabetes, coronary heart disease, stroke, and certain types of cancers (Marx, 2003). In the United States, obesity is rapidly overtaking smoking as the major underlying cause of death (Marx, 2003). The immediate causes of this obesity epidemic are clear: People eat more and exercise less than they used to.

Effects of Genes and Culture on Body Weight

Within the United States or any other Western culture, the determination of who does or does not become obese depends very much on genes and relatively little on the specific home environment (Barsh & others, 2000). The weights of adopted children correlate much more strongly with the weights of their biological parents than with those of their adoptive parents; identical twins usually have very similar weights even if raised in different homes; and pairs of biological siblings raised in different homes are, on average, nearly as similar to each other in weight as are pairs raised in the same home (Grilo & Pogue-Geile, 1991; Stunkard & others, 1986). This does not mean that body weight is little influenced by the environment. It simply means that the environmental conditions that promote obesity are fairly constant within Western cultures, so differences in weight have mostly to do with genetic differences in how individuals respond to those conditions.



Greg Ceo / Taxi / Getty

Overriding the satiety mechanism

People on a cruise typically gain weight, partly because of the large variety of food that is served. These people may become satiated with one type of food but still have an appetite for others.

>>> **16**
What is the evidence that within a culture, differences in body weight result mostly from differences in genes, but across cultures, environment plays a large role?

FIGURE 6.7 | Effect of culture on body weight If the only Pima Indians you ever saw were those living in Mexico and following traditional Pima practices (like these photographed at a holy festival), you might conclude that they are genetically predisposed to be slender. Obesity is almost nonexistent among the Mexican Pima. However, their close relatives living across the border in the United States are often obese. That difference lies not in genes but in the choices of foods available.



John Ammerino

Across cultures, environmental differences can have a large effect on body weight. As one example, obesity is very common among Pima Indians living in Arizona but is essentially absent among their genetic relatives living in Mexico (see Figure 6.7). The Mexican Pimas subsist mainly on grains and vegetables; their culture does not include the high-calorie foods available to the American Pimas (Gibbs, 1996). The genes that promote obesity in our culture apparently do so by increasing the person's attraction to high-calorie foods, by decreasing one or another of the feedback effects that high food intake or fat level has on the hunger mechanisms in the hypothalamus, and by decreasing the body's ability to burn up excess calories quickly (Barsh & others, 2000; Friedman, 2003). Where high-calorie foods are unavailable, the same genes generally don't lead to obesity.

Problems of Dieting

Weight gained is often very difficult to lose. Decreased food intake not only activates the hunger mechanisms in the brain but can also produce a decline in *basal metabolism* (the rate at which calories are burned up while the individual is at rest), so food is converted more efficiently to fat (Keesey & Corbett, 1984; Leibel & others, 1995). In one extreme case, a woman managed to reduce her weight from 312 pounds to a still-obese but much healthier 192 pounds through diet. For at least 18 months after that she maintained her new weight, without losing any more, by eating a total of 1000 to 1200 calories a day—less than half of what most women would have to eat to maintain that weight (Rodin & others, 1989).

Despite the odds, some people do lose significant amounts of weight and avoid regaining it. A recent study of approximately 3,000 highly successful dieters, who had lost an average of over 60 pounds per person and had kept it off for an average of 5 years at the time of the study, revealed that they succeeded primarily by avoiding high-fat foods and by greatly increasing their exercise (Butler, 2004; Wing & Hill, 2004). Regular exercise not only burns up calories immediately but also builds muscle, which, even when resting, burns up calories at a higher rate than do other body tissues (Van Itallie & Kissileff, 1990).

Researchers who study appetite and metabolism often have some sensible advice for people who want to lose weight. Here is my summary of what I have gleaned from their work:

- Don't try to lose weight rapidly. Don't go on a diet that you can't stay on for the rest of your life. Weight rapidly lost through starvation dieting is rapidly regained when the diet ends.

17

On the basis of the reports of successful dieters and the advice of appetite researchers, what can people do to maintain a lower weight?

- Instead of reducing food intake to a level that leaves you hungry, shift the kind of food you eat, away from high-calorie fats and sweets and toward gut-filling but low-calorie vegetables, fruits, and complex carbohydrates (whole-grain breads and cereals). Contrary to the claims of diet books that promote high-fat, low-carbohydrate diets, people who keep weight off for long periods usually do so by sharply reducing their fat intake (Butler, 2004; Wing & Hill, 2004).
- Make the sensory-specific satiety effect work for you instead of against you. Provide yourself, over time, with a luscious variety of vegetables and fruits and very few choices of high-calorie meats and dairy products; and omit the pastries, soda pop, potato chips, and French fries entirely.
- Eat each meal slowly, so as to enjoy it and to provide time for the food's satiety-producing effects to develop. If you still feel hungry after eating what you initially planned to eat, wait at least 15 minutes before deciding whether or not to eat more. By that time you may no longer feel hungry. It takes about 15 minutes for PYY and other satiety-producing consequences of a meal to begin to exert their effects on the brain's arcuate nucleus.
- If you have a sedentary job or are a student, develop some pleasurable hobbies that involve muscle-building exercise for at least a few hours a week. (I for one can't stand jogging and get quickly bored with weight lifting, but I love bicycling, kayaking, and cross-country skiing.) And when you need to get from one place to another, use your muscles rather than an automobile or an elevator to convey you whenever you can.

Through such changes in habits, many people can lose a significant amount of weight and keep it off for a lifetime, without feeling deprived at all.

SECTION REVIEW

Hunger is a regulatory drive controlled by neural, hormonal, and sensory factors.

Control Mechanisms of Hunger

- The arcuate nucleus of the hypothalamus is a feedback-based appetite control center, with both appetite-stimulating and appetite-suppressing neurons.
- Eating a large meal causes physiological changes, including the release of PYY, that influence the arcuate nucleus and nearby areas to reduce hunger.
- Leptin, a hormone produced by fat cells, helps to regulate body weight by acting on the hypothalamus to reduce appetite.
- Sensory stimuli also affect appetite, as illustrated by sensory-specific satiety and by the appetite-boosting power of learned cues that signal the availability of food.

Obesity

- Within a culture, genetic differences are the primary determinant of who becomes obese, but across cultures, environmental differences play a substantial role.
- Decreasing food intake activates hunger mechanisms in the brain and can reduce basal metabolism, making weight loss harder.
- Certain techniques help at least some people lose substantial amounts of weight and keep it off.

Sex: An Example of a Non-Regulatory Drive

Just as hunger is the most thoroughly studied regulatory drive, sex is the most thoroughly studied non-regulatory drive. As with hunger, most research on the physiological basis of the sex drive has been conducted with laboratory animals.

There are limits, of course, in the degree to which we can understand hunger, sex, or any drive in humans by studying it in other animals. Human culture, intellect, sensibility, and capacity for conscious self-control affect all human behavior in ways that cannot be studied in laboratory animals. People don't just eat; they *dine*,

>>> **18**
Why is caution advised in extrapolating from laboratory animals to humans in the study of drives, especially the sex drive?

Barnett, S. A. (1975). *The study of rat behavior* (rev. ed.). Chicago: University of Chicago Press



FIGURE 6.8 | Copulation in rats
Rats, like most other non-human mammals, have a stereotyped (unvarying) pattern of copulation, with clearly different postures for the female and the male.

which connotes all sorts of social, cognitive, and aesthetic influences. And people don't just copulate; they fall in love, compose romantic sonnets, gaze into each other's eyes over candlelit dinners, swear by the moon to be faithful, have affairs, suffer guilt, and engage in long, intimate discussions with their beloved. Keep in mind, as you read on, that our concern here is the basic physiological mechanisms that we humans share, more or less, with other mammals, not the whole range of issues concerning human sexuality. (Sex is discussed more from a social and cultural perspective in Chapter 12.)

Even when dealing with the copulatory act itself, humans differ quite sharply from rats and other laboratory animals. Among non-human mammals, including most other primates, copulation occurs in a stereotyped way, with one set of postures and movements for the female and a different set for the male (see Figure 6.8). Among humans, by contrast, the variety of ways to copulate is limited only by imagination. As you will discover when you read further, differences between humans and other species exist also in the hormonal regulation of the sexual drive, especially in females.

Hormonal Influences on Sexual Drive

Sex hormones affect sexual drive in humans and other mammals through their influences on the brain. Such influences are of two types: differentiating and activating. **Differentiating effects** occur before or (in some species) immediately after birth and cause the brain to develop in a male or female direction. They cause the *differences* between males and females in sexual drive and orientation. **Activating effects** occur later, around the time of puberty and after, when hormones work on the already-differentiated brain structures to prime, or *activate*, sexual drive. Neurons that are involved in sexual drive have special binding sites on them for sex hormones, and the hormones act there to enable the neurons to respond to sexually arousing stimuli in the environment. Here we shall look first at activating effects, and then at differentiating effects.

Hormones and the Onset of Sexual Drive Near Puberty

19

What roles are played by testosterone, estrogen, and DHEA in the developmental changes around the time of puberty?

<<< In both humans and other mammals, the production of sex hormones greatly increases at the onset of the developmental stage called *puberty*. In human males, increased *testosterone*, produced by the testes, stimulates such changes as beard growth and the male pattern of muscle development; in human females, increased estrogen, produced by the ovaries, stimulates such changes as breast growth. Less well known is the fact that, in humans as well as other animals, and in both sexes, the adrenal glands also produce sex hormones. In humans the primary adrenal sex hormone is *dehydroepiandrosterone*, mercifully abbreviated *DHEA*. This substance is commonly classed as a “male hormone”—that is, as an *androgen*—because it has effects in men similar to those of testosterone. However, it is produced as much in women as in men, and there is reason to believe that it is the primary hormone responsible for the onset of sexual feelings and attractions in young humans of both sexes (Herdt & McClintock, 2000; McClintock & Herdt, 1996).

Boys' and girls' adrenals begin to secrete DHEA at about 6 years of age, and the amount secreted rises continuously until the mid-teenage years, when it stabilizes at adult levels (see Figure 6.9). Research in both Western and non-Western cultures suggests that most people, of both sexes, recall their earliest clear feelings of sexual attraction to another person as occurring at about 10 years of age, well before the prominent physical changes brought on by testosterone or estrogen (Herdt & McClintock, 2000). If testosterone and estrogen brought on this developmental effect, then we would expect girls to experience sexual attraction at an earlier age than boys, since the rise in estrogen precedes the rise in testosterone by about 2 years, on average. But repeated studies find no difference between the sexes in the age of earliest sexual attraction, which is consistent with the idea that this feeling is brought on by DHEA, which increases at the same rate in the two sexes. DHEA,

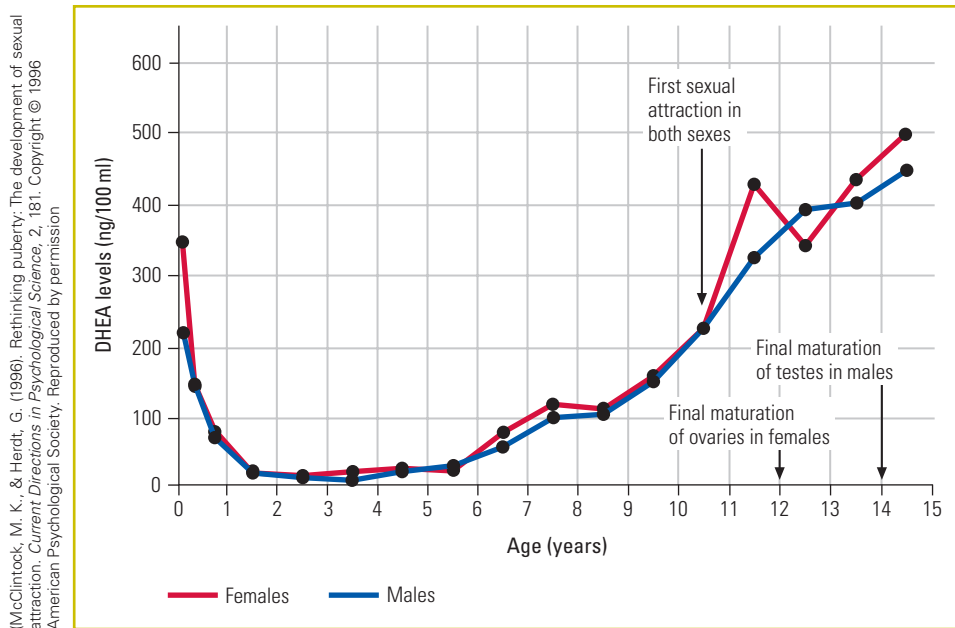


FIGURE 6.9 A sex hormone that increases at the same rate in girls and boys Shown here are the average blood levels of the adrenal androgen DHEA (dehydroepiandrosterone) in boys and girls as a function of age. This hormone is believed to activate the brain in a way that enables the young person, whether male or female and regardless of sexual orientation, to feel sexually attracted toward others, beginning at about age 10.

like other sex hormones, passes into the brain and alters activity in certain neural centers there, especially in the hypothalamus.

Activating Effects of Hormones on the Sex Drive in Male Mammals

In male mammals, including humans, the most crucial hormone for the maintenance of the sexual drive after puberty is not DHEA but testosterone. Castration (removal of the testes, which produce the main supply of testosterone) causes a marked decline in the sex drive—not all at once, but gradually (Feder, 1984). It takes days to occur in rats, weeks in dogs, sometimes months in monkeys. But the injection of testosterone into the bloodstream of castrated animals fully restores their drive.

The sex drive can also be restored in castrated male animals by implanting a tiny crystal of testosterone into an area of the hypothalamus called the *medial preoptic area* (see Figure 6.10 for its location) (Davidson, 1980). Neurons in the medial preoptic area contain many receptor sites for testosterone, and small lesions there abolish sexual behavior in male rats (Meisel & Sachs, 1994). Apparently, the medial preoptic area of the hypothalamus is a crucial part of the central drive system for sex in male rats other male animals that have been studied, and testosterone acts there in a rather prolonged way to enable neural activity to occur and sustain the drive.

Testosterone is also crucial for maintaining the sex drive in human males. Men castrated in an accident or for medical reasons almost always experience a decline (though often not a complete loss) in sex drive and behavior, and testosterone injections restore their drive, usually fully (Money & Ehrhardt, 1972). In other studies, testosterone injections administered to non-castrated men whose testes were producing abnormally low amounts of the hormone sharply increased their sexual behavior (Davidson & Myers, 1988; Davidson & others, 1979). This effect was more on drive than on sexual capability. Men with abnormally low levels of testosterone were usually capable of the mechanics of sexual behavior—including erection and ejaculation—but had little desire for it until injected with testosterone. The subjects in this research did not know when they were receiving testosterone and when they were not, so the results must have been due to the effects of the hormone and not to their expectations.

>>> 20 What is some evidence that testosterone is needed to maintain the male's sex drive and that, at least in some species, it does so by direct action in the hypothalamus?

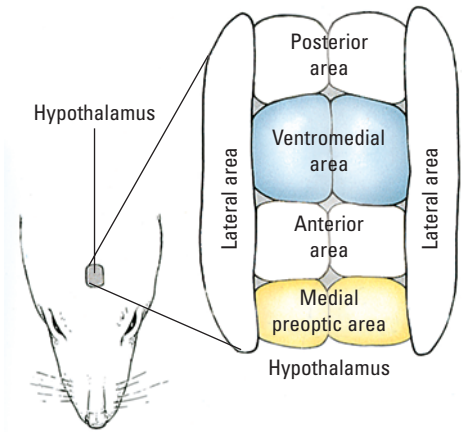


FIGURE 6.10 Hypothalamic areas where hormones activate sexual behavior In rats, testosterone promotes male sexual behavior by activating neurons in the medial preoptic area of the hypothalamus, and estrogen promotes female sexual behavior by activating neurons in the ventromedial area of the hypothalamus.

Activating Effects of Hormones on the Sex Drive in Female Mammals

21

What evidence indicates that ovarian hormones act directly on the brain to activate the sexual drive in female rats? How do female non-human primates and humans differ from one another, and from rats, in regard to the role of hormones in regulating sexual drive?

<<< After puberty, a female's ovaries begin to secrete estrogen and progesterone in a cyclic pattern over time, producing the cycle of physiological changes referred to as the *menstrual cycle* in humans and the *estrous cycle* in other mammals. In both humans and non-humans, this cycle controls ovulation (the release of one or more eggs so that pregnancy can occur). In most mammals, it also tightly controls the sex drive—which ranges from very strong at the time of ovulation to nonexistent at other times. Removal of the ovaries completely abolishes sexual behavior in most non-human female mammals, and injection of hormones can fully restore it. For some species an injection of estrogen alone is most effective, and for others (including rats) a sequence of estrogen followed 2 or 3 days later by progesterone (another female hormone) is most effective.

At least in rats, the *ventromedial area* of the hypothalamus (see Figure 6.10) plays a role in sexual behavior in the female analogous to that of the preoptic area in the male. Insertion of small amounts of estrogen and progesterone directly into this area brings on sexual behavior in rats whose ovaries have been removed, and lesions in this area abolish sexual behavior in otherwise intact females (Mani & others, 1994; Pleim & Barfield, 1988; Schwartz-Giblin & others, 1989). Apparently the cyclic variation in ovarian hormones acts on the ventromedial area to cause the cyclic waxing and waning of sexual drive.

Female monkeys and apes depend less on ovarian hormones for sexual behavior than do most other female mammals. Most primate females can, and sometimes will, copulate with a sexually active male at any time in their hormone cycle, though they are more likely to seek out and initiate sexual contact with a male at the time in their cycle when they are ovulating and the level of ovarian hormones is high (Wallen, 1990). In at least some species of primates, including rhesus monkeys, adrenal androgens apparently help maintain sexual drive and behavior during times in the estrous cycle when ovarian hormones are low (Wallen, 1990).

Human females show still greater liberation of sexual behavior from cyclic hormonal control than do other primates. Women can experience a high or low sex drive at any time in their hormone cycle. In fact, debate exists as to whether women's sex drive is affected in a consistent way at all by the hormone cycle. Some researchers have found that, on average, women report a higher sexual drive (Adams & others, 1978), and manifest greater attraction to men with highly masculine features (Gangestad & others, 2004), around the time of ovulation than they do at other times in their hormonal cycle. But these effects are subtle compared to effects of the hormone cycle on sexual drive in the females of most other species. In women, ovarian hormones apparently play a relatively small role in regulation of the sex drive, and adrenal hormones play a larger role. The adrenals of adult women produce both DHEA and testosterone. In clinical studies, women whose ovaries have been removed generally do not report a decline in sexual drive, but women whose adrenals have been removed generally do report such a decline, and treatment with testosterone reliably increases the sex drive reported by women with low sex drive (Bancroft, 1978; Guay, 2001; Sherwin & Gelfand, 1987).

Brain-Differentiating Effects of the Early Presence or Absence of Testosterone

22

What are some effects of the presence or absence of testosterone before birth on development of the genitals, the brain, and behavior?

<<< As noted in Chapter 3, the only initial difference between the two sexes, in all mammals, is that females have two X chromosomes and males have a small Y chromosome in place of the second X. A specific gene on the Y chromosome causes the growth of testes (the male gonads) from structures that would otherwise develop into ovaries (the female gonads) (Page & others, 1987). Before birth the testes begin to produce testosterone, which acts on the brain and other bodily structures to steer development in the male direction. The rudimentary genitals of the fetus develop into male structures (including the penis and scrotum) if testosterone is present,

and they develop into female structures (including the clitoris and vagina) if testosterone is absent. Early testosterone also promotes the development of brain pathways involved in the male sex drive and to inhibit the development of neural systems involved in the female sex drive (Gorski, 1996; Simerly, 2002). For an example of one such effect, see Figure 6.11.

In order to produce these brain-differentiating effects, testosterone must act within a critical period in the animal's development. In rats, this period runs from a few days before birth to a day or so after birth. In many other species, the critical period ends before birth. The critical period for testosterone's effect on the brain is later than that for its effect on the genitals. Thus, manipulation of hormones at the appropriate time can produce animals that have the genitals of one sex but the brain structures and behavior of the other sex (Feder, 1984; Ward, 1992).

Perhaps you are wondering why a male hormone, not a female hormone, plays the key role in early sexual differentiation in mammals. The answer is that the female hormones (progesterone and estrogen) are produced by pregnant females at high levels and get into the tissues of all fetuses, of both sexes. If female hormones promoted growth of female structures during fetal development, all mammalian infants would be born looking like females. In birds and reptiles—which develop in eggs outside the mother's body—early sexual differentiation is determined by the presence or absence of estrogen, not testosterone (Adkins-Regan, 1981).

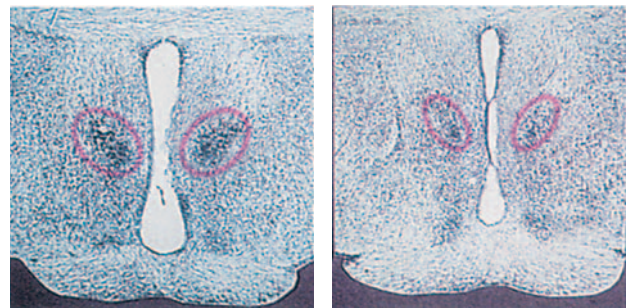
Brain Development and Sexual Orientation in Humans

It is quite possible that the brain mechanisms for male and female sexual motivation are less different from one another in humans than in most other mammalian species. As already noted, in people, unlike most other mammals, androgens may be the primary sex-drive hormones in both sexes. Moreover, men and women show similar patterns of physiological changes during sexual arousal and orgasm (Masters & others, 1992) and describe the subjective feelings they have during sex in similar ways (Vance & Wagner, 1976). But one general difference between men and women in sex drive is in the object of the drive: Most men are attracted to women, and most women are attracted to men. Might this difference result from a structural difference in the brain? If so, what creates that brain difference? The answers to these questions are not yet certain, but some interesting ideas about them have emerged from studies of people who are sexually attracted to members of their own sex rather than the opposite sex.

Using various criteria and survey methods, researchers estimate that 2 to 5 percent of men, and 1 to 2 percent of women, are exclusively homosexual in orientation (Rahman & Wilson, 2003). An apparently larger percentage, even harder to estimate precisely, are bisexual (attracted to members of both sexes), and bisexuality appears to be more common in women than in men (Rahman & Wilson, 2003).

Lack of Evidence for Social Learning Explanations of Sexual Orientation

In the past many psychologists argued that sexual orientation derives primarily from social learning, and a few still argue that today. But the existence of a significant homosexual population in every culture—including cultures where homosexuality is severely punished—presents a challenge to that hypothesis (LeVay & Hamer, 1994). If sexual orientation is learned through social experiences, then why would anyone developing in a society that glorifies heterosexuality and strongly discourages homosexuality grow up as a homosexual? Over the years psychologists have suggested many possible answers to that question aimed at preserving the learning hypothesis, but none has so far withstood the test of research.



Rosenzweig, Leithman, & Breedlov.
Biological Psychology, 1996

FIGURE 6.11 | A sex difference in the rat's hypothalamus Shown here is a cross section of the medial preoptic area of the hypothalamus of a male rat (*left*) and a female rat (*right*). The dark spots represent cell bodies of neurons, and the dense cluster of them represents the *sexual dimorphic nucleus* (marked by red circles). The nucleus receives its name from the fact that it is clearly different in the two sexes; it is about 5 times larger, on average, in the male than in the female (Gorski & others, 1980). The difference is due entirely to the presence or absence of testosterone early in the rat's development. Early treatment with testosterone in females and early deprivation of testosterone in males can reverse this sex difference. The nucleus is believed to be part of the neural circuitry controlling male sexual drive and behavior (DeJonge & others, 1989).

>>> 23

What evidence suggests that in humans sexual orientation (a) is not a result of environmental variables that in the past were hypothesized as causal, (b) appears early in development, (c) correlates with certain anatomical brain differences, (d) is influenced by differences in genes, and (e) is influenced by the prenatal environment, including (in males) an effect on that environment induced by one's older brothers?



Winslow Townson / AP Photo

Gay marriage The idea that homosexual couples should be allowed to enjoy the benefits and take on the responsibilities of marriage is still politically controversial. The fact that the controversy is even occurring, however, illustrates that the culture has taken strides toward accepting homosexuality as a normal human condition.

Interviews and surveys of thousands of homosexuals and heterosexuals have found no consistent relationship between sexual orientation and the kinds of experiences in childhood or adolescence that have at times been theorized to cause homosexuality (Bell & others, 1981; Dawood & others, 2000; Rahman & Wilson, 2003). Such studies reveal no evidence that style of parenting, absence of the father or mother, early seduction or rape by someone of the same or opposite sex, or degree of opportunity for one or another type of sexual experience in adolescence contributes significantly to the development of sexual orientation. Such studies do show, however, that whatever its cause or causes, sexual orientation is usually a deeply ingrained and early-emerging aspect of one's being. Homosexuals and heterosexuals alike generally report that their sexual orientation was present, though not necessarily understood or accepted, in their childhood thoughts and fantasies, typically by the age of about 10 or 11 (Bell & others, 1981; Herdt & Boxer, 1993; Pattatucci & Hamer, 1995). The feeling of strong attraction to one sex or the other often exists for years before it is expressed. Most psychologists are now convinced that sexual orientation is not something that most people can choose or change through willpower or psychotherapy but, rather, is something that people discover about themselves.

Brain Correlates of Sexual Orientation

In humans, as in other mammals, there are certain measurable anatomical differences in the brains of males and females. Some brain areas are larger in men than in women, and the reverse is true for some other brain areas. Postmortem studies indicate that in a few of these brain areas, the male homosexual brain is more like the female heterosexual brain than like the male heterosexual brain (Rahman & Wilson, 2003). One such area is a nucleus in the anterior hypothalamus, which is larger in heterosexual men than it is in homosexual men or in heterosexual women and is believed to be the human homologue of the sexually dimorphic nucleus depicted in Figure 6.11 (LeVay, 1991; Rahman & Wilson, 2003). Comparable studies of the brains of homosexual women have not yet been done to see if their brains, at these locations, are like those of heterosexual men. Nobody knows for sure, but such brain differences may provide a biological foundation for differences in sexual orientation, and the brain differences themselves may result from early developmental influences of genes and the prenatal environment.

Effects of Genes and Prenatal Environment on Sexual Orientation

Genetic differences among individuals apparently play some role in determining sexual orientation but are not solely responsible. Several studies indicate that roughly 50 percent of the identical twin brothers and sisters of homosexual men and women are also homosexual, compared with about 15 percent of the same-sex non-identical twins or non-twin siblings of homosexual men and women (Rahman & Wilson, 2003).

Other studies, which are still quite controversial, suggest that sexual orientation might be affected by any of various factors—ranging from prenatal stress to certain medications taken during pregnancy—that alter the amount of testosterone or other androgen available to the fetus's brain during a critical period in development (Ellis & Ames, 1987; Ellis & others, 1988; Meyer-Bahlburg & others, 1995). Such work suggests that a high level of androgen action during the critical period may promote the development of brain mechanisms that predispose sexual attraction toward women and not toward men and that a relative lack of androgen during the same period may promote the opposite. Genes might influence sexual orienta-

tion by influencing the degree to which androgen hormones can bind to the relevant brain areas during the critical developmental period.

The single most consistent, non-genetic influence on sexual orientation discovered to date is the *fraternal birth-order effect* on male homosexuality (Bogaert, 2003). The more older brothers a man has, the greater is the likelihood of his being homosexual. The effect is quite strong. One large study revealed that, on average, a couple's first son has a 2.0 percent chance of being homosexual and that this percentage increases to about 2.6 percent for a second son, 3.4 percent for a third son, 4.6 percent for a fourth son, and 6.0 percent for a fifth son (Blanchard, 2001). In contrast, the number of older sisters a man has plays no role in the likelihood of his being homosexual, and there is no birth-order effect on homosexuality in women.

The fraternal birth-order effect appears to have nothing to do with the amount or type of interaction that a boy has with his older siblings (Bogaert, 2003). The most prominent current hypothesis is that the effect results from a change in the prenatal environment that is induced by previously born sons (Blanchard, 2001; Rahman & Wilson,2003). Genes on the Y chromosome of a male fetus produce certain proteins that may help androgens to masculinize regions of the fetus's brain. These proteins are foreign to the mother's body, so if they get into her bloodstream they are likely to induce an immune reaction in her. With successive sons, the immune reaction could generate a sufficient supply of antibodies in the mother's bloodstream to inactivate one or more of those proteins within a developing fetus and thereby reduce the masculinizing effect of androgens in a portion of the fetus's brain. At present, such an explanation lies in the realm of reasonable speculation; there is no strong evidence supporting or refuting the hypothesis.

SECTION REVIEW

Sex is a non-regulatory drive powerfully influenced by hormones.

Hormones • Before birth, the presence or relative absence of testosterone promotes physical differentiation between males and females, including brain differences that will affect the sex drive. • Before puberty, the adrenal hormone DHEA triggers the beginning of sexual feeling and attraction in girls and boys. • At puberty, increased production of testosterone in males and estrogen in females stimulates sexual maturation. • After puberty, testosterone maintains the sex drive in male mammals (including men) while ovarian hormones do so in female mammals (except some primates). In human females, androgens produced by the adrenal glands maintain the sex drive.	Sexual Orientation • Social learning explanations for homosexuality do not appear to be supported by the evidence. • Sexual orientation may be related to prenatal brain development. • Genes play a role in sexual orientation. Variation in the prenatal environment may also play a role. • The more older brothers a man has, the greater the chance that he will be homosexual. This may derive from the mother's immune response to proteins produced by the Y chromosome in male fetuses.
--	--

The Sleep Drive

Sleepiness is clearly a drive. A sleepy person is motivated to go to sleep and will expend effort to reach a safe and comfortable place to do so. Achieving this goal and drifting off to sleep provides a sense of pleasure analogous to that which comes from eating when hungry or copulating when sexually motivated.

Sleepiness operates in some ways like a regulatory drive. As with hunger, thirst, or other regulatory drives, the longer one goes without satisfying the sleep drive, the stronger the drive becomes. But, unlike other regulatory drives, it is not clear what the sleep drive regulates, except sleep itself. Also, sleepiness is controlled not just by amount of sleep deprivation but also by time of day. Regardless of how

much sleep one has had recently, sleepiness tends to increase at night and decrease during daytime hours. Sleepiness is controlled not just by some consequence of sleep deprivation, but also by a biological clock that keeps time with the 24-hour day-night cycle.

The discussion in this section focuses on three questions: (1) What is sleep? (2) What are the functions of sleep? (3) What brain mechanisms control sleepiness and arousal?

Description of Sleep as a Physiological and Behavioral State

Sleep is a condition of relative unresponsiveness to the environment. Because people show little overt behavior and cannot answer questions when they are asleep, scientists who study this state must focus on physiological and subtle behavioral changes.

The most valuable index of sleep is based on the electroencephalogram (abbreviated EEG). As discussed in Chapter 5, the EEG is an amplified recording of the electrical activity of the brain that is picked up by electrodes pasted to the person's skull. The electrical signals can be stored and analyzed by a computer and can also be used to move pens up and down on a continuously moving roll of paper, producing records like those shown in Figure 6.12. The EEG recording is a gross index of the electrical activity of the brain, representing a sort of average of the activity of billions of neurons, with the greatest weight given to those lying closest to the recording site.

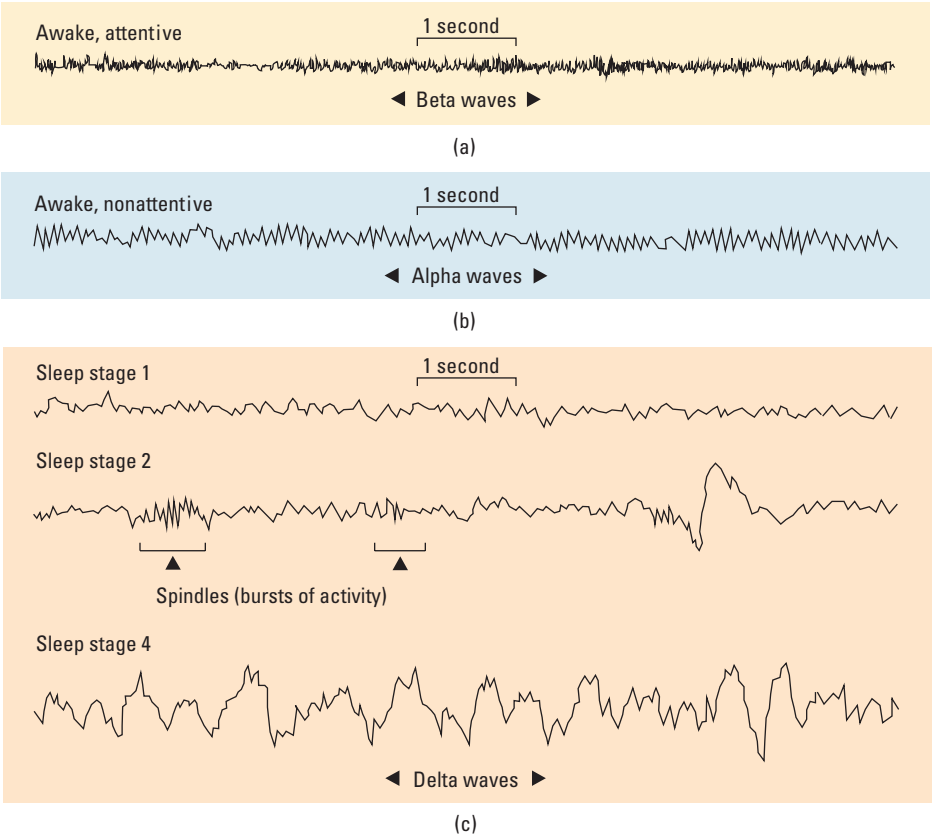
EEG Waves Accompanying Wakefulness and Stages of Sleep

24

How does a person's EEG change as the person goes from alert to relaxed to various sleep stages?

When a person is relaxed but awake, with eyes closed, and not thinking of anything in particular, the EEG typically consists of large, regular waves called *alpha waves*, which occur at a frequency of about 8 to 13 cycles per second (see Figure

FIGURE 6.12 EEG waves of waking and sleep stages In general, as one goes from an alert to a relaxed state, and then to ever-deeper stages of sleep, the EEG waves become slower in frequency (fewer waves per second) and higher in amplitude (as shown by their greater vertical extent in the EEG record). The brief bursts of rapid waves called *sleep spindles* that appear in stage 2 are the most distinctive markers of the onset of sleep. Sleep stage 3 is not shown here; it is arbitrarily defined as the period when 10 to 50 percent of the EEG consists of delta waves. REM sleep, also not shown, is characterized by beta waves that look like those of the awake, attentive state. (From Snyder, F., & Scott, J., 1972)



The psychophysiology of sleep. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology*. New York: Holt, Rinehart & Winston. Copyright © 1972 by Holt, Rinehart & Winston, Inc., reproduced by permission of the publisher.

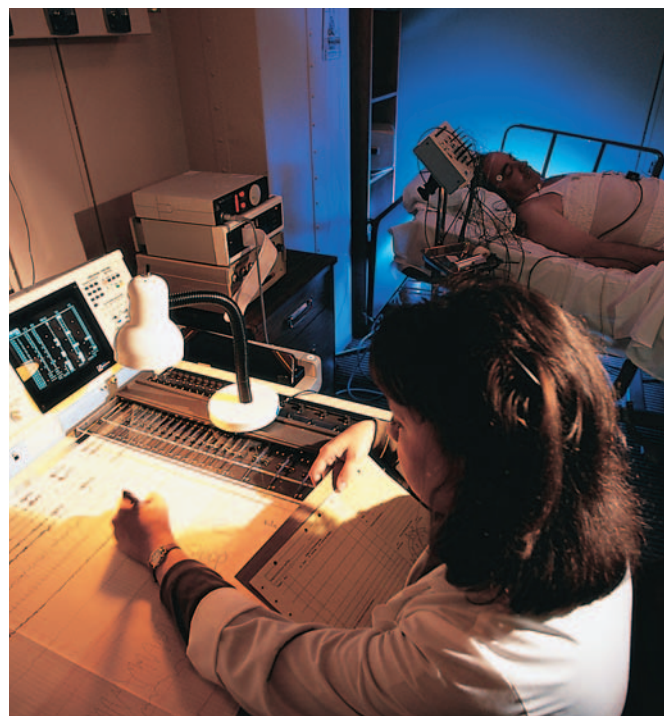
6.12b). These relatively slow waves stem from a synchronized pulsing of neurons in the thalamus and cerebral cortex that occurs in the absence of focused mental activity or emotional excitement. When a person concentrates on an external stimulus or tries to solve a problem or becomes excited, the EEG pattern changes to low-amplitude, fast, irregular waves called *beta waves* (see Figure 6.12a). The low amplitude of these waves indicates that neurons are firing in an unsynchronized manner, such that their contributions to the EEG tend to cancel one another out. Whereas alpha waves are analogous to the large, regular waves that occur on a pond undisturbed by anything but a steady wind, beta waves are more akin to the effect of a million pebbles tossed suddenly onto the surface of the pond. The crests of the ripples created by some pebbles would cancel out the troughs created by others, resulting in a chaotic, high-frequency, low-amplitude pattern of ripples.

When a person falls asleep, the EEG goes through a fairly regular sequence of changes, which are used by researchers to divide sleep into four stages, illustrated in Figure 6.12c. *Stage 1* is a brief transition stage, when the person is first falling asleep, and *stages 2* through *4* are successively deeper stages of true sleep. As sleep deepens, an increased percentage of the EEG is devoted to slow, irregular, high-amplitude waves called *delta waves*. These waves, like others, are controlled by neurons in the thalamus that respond in an oscillating manner and synchronize the activity of billions of neurons in the cerebral cortex (Steriade & others, 1993). Corresponding to this EEG change, muscle tension, heart rate, and breathing rate decline, and the person becomes increasingly hard to awaken.

Cyclic Repetition of Sleep Stages Through the Night

Having reached stage 4, a person does not remain there for the rest of the night. Instead, after about 80 to 100 minutes of total sleep time, sleep rapidly lightens, returning through stages 3 and 2, and then a new, quite fascinating stage of sleep appears for a period of about 10 minutes or more. During this new stage the EEG is unsynchronized, looking much like the beta waves of alert wakefulness. On the basis of the EEG alone, one might think that the person had awakened, but direct observation shows that the person is sound asleep, and the record of muscle tension shows that the muscles are more relaxed than at any other sleep stage. Yet, consistent with the unsynchronized EEG, other indices of high arousal are apparent: Breathing and the heart rate become more rapid and less regular; penile erection occurs in males (even in infants and young boys); twitching movements occur in the small muscles of the fingers and face; and, most indicative of all, the eyes move rapidly back and forth and up and down under the closed eyelids. These eye movements, which can be recorded electrically along with the EEG, give this stage of sleep its name, *rapid-eye-movement sleep*, abbreviated **REM sleep**. As you may have guessed, it is during REM sleep that most dreams occur. REM sleep is also sometimes called *emergent stage 1*, because, even though it is different from the original stage 1, it marks the onset of a new sleep cycle. Stages 2, 3, and 4 are referred to collectively as **non-REM sleep**.

In a typical night's sleep, a person goes through four or five sleep cycles, each involving gradual descent into deeper stages of non-REM sleep, followed by a rapid lightening of non-REM sleep, followed by REM sleep (Hobson, 1995). Each complete cycle takes about 90 minutes. As depicted in Figure 6.13, the deepest non-REM sleep occurs in the first cycle or two. With each successive cycle, less time is spent in the deeper stages of non-REM sleep (stages 3 and 4), and more time is spent in light non-REM sleep (stage 2) and REM sleep.



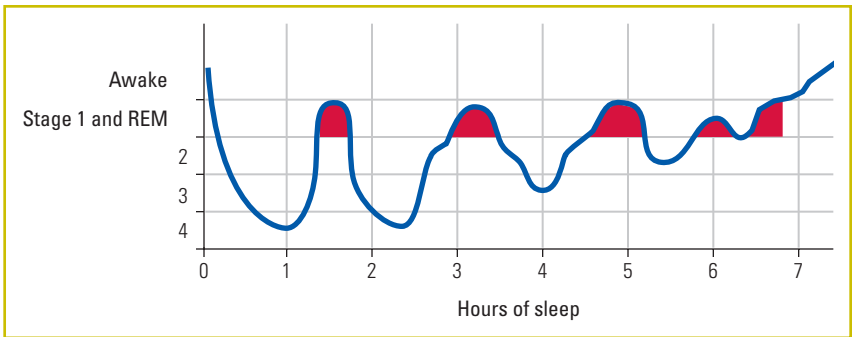
Will & Deni McIntyre / Photo Researchers

Keeping track of sleep With electrodes attached to the man's scalp and other areas of his body, the researcher monitors EEG, eye movements, breathing rate, and other muscle movements to keep track of his stages of sleep and wakefulness.

>>> **25**
How do REM and non-REM sleep differ, and how do they cycle through the night?

FIGURE 6.13 The cycle of sleep stages through a night People usually go through four or five sleep cycles per night, each ending with a period of REM sleep. With successive cycles the depth of slow-wave sleep becomes less, and the amount of time spent in REM sleep increases. (From Snyder, F., & Scott, J., 1972.)

The psychophysiology of sleep. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology*. New York: Holt, Rinehart & Winston. Copyright © 1972 by Holt, Rinehart & Winston, Inc., reproduced by permission of the publisher)



26

What are some general characteristics of dreams that people describe when aroused from REM sleep, and how do these differ from the “sleep thought” that people describe when aroused from non-REM sleep?

Dreams and Other Mental Activity During Sleep

When people are awakened during REM sleep, they usually (in about 90 percent of the cases) report a mental experience that researchers call a *true dream* (Foulkes, 1985). Such a dream is experienced as if it were a real event rather than something merely imagined or thought about. The dreamer has the feeling of actually seeing or in other ways sensing various objects and people and of actually moving and behaving in the dream environment. Moreover, a true dream usually involves a progression of such experiences, woven into a somewhat coherent though often bizarre story. The more time the sleeper spends in REM sleep before awakening, the longer and more elaborate is the reported dream. Studies show that essentially everyone dreams several times a night. Those who believe that they rarely dream, or who can recall only fragments of dreams upon normal awakening in the morning, describe vivid, detailed dreams if awakened during REM periods.

Analyses of the contents of hundreds of reported dreams reveal a number of generalities about them (Domhoff, 2003). Most dreams are about people, objects, and activities that are well known and meaningful to the dreamer, but very few dreams are repetitions of events that actually happened in the dreamer's daytime experience. Most dreams involve emotions, especially negative emotions. Dreams involving fear, worry, or embarrassment are more common than joyous dreams. Among college students, dreams of being lost, of being late for an examination, and of being inappropriately dressed (or undressed) in public are common.

People who are awakened during non-REM sleep report some sort of mental activity just before awakening roughly half the time (Foulkes, 1985; Hobson, 1995). Such reports are usually not of true dreams but of *sleep thought*, which lacks the vivid sensory and motor hallucinations of true dreams and is more akin to daytime thinking. Often the subject of sleep thought is some problem that had been of concern during the day. For example, a student who had been cramming for a math exam might report working on a calculus problem while sleeping. A major difference between sleep thought and daytime thought is that the former is usually ineffective. Although the sleeper may feel that he or she is solving a calculus problem, questions upon awakening indicate that no real progress was made (Hobson, 1987).

During sleep a person is less responsive to events in the environment than when awake, but not completely unresponsive. The eyes are closed, but all of the other sensory channels remain open. In one study using fMRI, people's brains were observed to respond to various sounds when they were asleep in much the same way as they did when they were awake (Portas & others, 2000). Both when asleep and when awake, the sound of the person's own name had effects on emotional centers in the limbic system that did not occur in response to sounds that were less meaningful to the person. The fact that a parent can sleep through a thunderstorm but become aroused by the whimpering of his or her child in the next room is further evidence that sounds are sorted out to some degree by meaning in the sleeping person's brain.



One characteristic of all mental activity during sleep is that it is quickly forgotten. Dreams, sleep thoughts, and sensory experiences during sleep are lost unless the person wakes up during them and thinks about them while awake. This is fortunate. Such forgetting prevents us from carrying around thousands of bizarre and confusing memories or of mixing up real events with dreamed ones.

Theories About the Functions of Sleep

Why must we sleep? Countless children have asked that question to protest being put to bed, and many scientists have asked it, too. Sleep must have come about in the course of evolution to serve some function or functions related to survival and reproduction; otherwise it would not be such a universal and compelling drive. Several theories have been proposed to explain the evolution of the sleep drive. The theories are not incompatible with one another, and they all appear to have some degree of validity.

The Preservation and Protection Theory

The *preservation and protection* theory of sleep derives primarily from comparison of sleep patterns across different species of animals. It posits that sleep came about in evolution to preserve energy and protect individuals during that portion of each 24-hour day when there is relatively little value and considerable danger in moving about. An animal needs only a certain number of hours per day to do the things that are necessary or useful for survival, and the rest of the time, according to this theory, it is better off asleep—quiet, hidden, and protected from predators and other possible dangers.

Support for this theory comes from evidence that variations in sleep time among different species do not correspond with differences in physical exertion while awake but do correspond with feeding habits and ways of achieving safety (Allison & Cicchetti, 1976; Webb, 1982). At one extreme, large grazing animals such as bison and horses average only 2 or 3 hours of sleep per 24-hour day. Because of their large size and because they eat grass and other vegetation, which are extremely low in calories, they must spend most of their time eating, and therefore

>>> **27**
What evidence supports the preservation and protection theory of sleep?

Kenneth M. Highfill / Photo Researchers



Asleep and protected According to the preservation and protection theory, a major function of sleep is to keep animals quiet and hidden during that portion of the day or night when it would be most dangerous and least profitable for them to be moving about.

they have little time to sleep. Moreover, because of their size and the fact that they cannot burrow or climb trees, such animals are not adept at finding safe nooks in which to sleep. Thus, they are safer awake.

Even among animals that are roughly the same size as each other, grazing animals sleep less than do meat-eaters. Sheep and goats, for example, sleep only 4 or 5 hours per 24 hours, while lions and tigers sleep 14 to 16 hours (Campbell & Tobler, 1984). Sheep and goats must spend more time eating than lions and tigers, and, because they are more preyed upon, the former are at much greater risk when asleep than are the latter. At the extreme in sleep time are opossums and bats, which average about 20 hours of sleep each 24-hour day. These two species need little time to obtain food (such as high-calorie insects or grubs), and they are adapted to hide in out-of-the-way places. According to the preservation and protection theory, they sleep so much because they have no need to be awake for long and are protected from predators while asleep.

In addition to explaining differences in total amount of sleep, the preservation and protection theory also explains differences in the time of day at which different species

sleep. Animals that rely heavily on vision generally forage during the day and sleep at night. Conversely, animals such as mice and rats that rely more on other senses, and are preyed upon by animals that use vision, generally sleep during the day and forage at night. The theory also offers an explanation for the fact that infants in most species of mammals sleep much more than adults. Infants who are being cared for by adults do not need to spend time foraging, and sleep protects them from wandering away into danger. Their sleep also gives their caregivers an opportunity to rest or attend to other needs.

It is interesting to speculate, in this vein, about the evolutionary conditions behind the 8-hour nighttime sleep pattern that characterizes adult humans throughout the world. Humans are highly visual creatures who need light to find food and do other things necessary for survival. At night it may have been best for us, during most of our evolution, to be tucked away asleep in a cave or other hiding place, so as not to be tempted to walk about and risk falling over a cliff or being attacked by a nocturnal predator. Only during the past few centuries—an insignificant speck of evolutionary time—have lights and other contrivances of civilization made the night relatively safe for us. According to this line of thinking, our pattern of sleep might be in part a vestigial trait, a carryover from a period when the night was a time of great danger. To the degree that nighttime is still more dangerous than daytime, our pattern of sleep may continue to serve an adaptive function.

The Body-Restoration Theory

The *body-restoration theory* of sleep function is the theory that most people intuitively believe. It is the theory that your parents probably repeated to you as their reason for sending you to bed at a certain hour. According to this view, the body wears out during the day and sleep is necessary to put it back in shape.

Scientific support for this theory includes the observation that sleep **is** a time of rest and recuperation. The muscles are relaxed, metabolic rate is down, and growth hormone, which promotes body repair, is secreted at a much higher rate than during wakefulness (Douglas, 2002; Siegel, 2003). Also consistent with the restoration theory is the observation that prolonged, complete sleep deprivation in rats results in breakdown of various bodily tissues, leading, within about 3 weeks, to death (Everson, 1993; Everson & others, 1989).

The theory also offers an explanation for the general tendency of small mammals to sleep longer than large ones. Small mammals need to maintain a higher

28

What evidence supports the body-restoration theory of sleep, and what are some limitations of the theory?

overall level of metabolism than do large mammals, because they lose body heat more rapidly, and higher metabolism leads to greater wear and tear on bodily tissues (Siegel, 2003). The theory does not, however, explain the large differences in sleep time between grazing animals and meat-eating animals that have similar body sizes and metabolic rates. It also does not explain the failure of researchers to find consistent positive correlations, either across species or within species, between the amount of time an animal sleeps and the average amount of energy it expends through vigorous activity during the day.

The fact that all vertebrate animals sleep at least an hour or two out of every 24 hours, regardless of the degree to which they are at risk while sleeping, suggests that some amount of sleep is needed for body repair. But the body-restoration theory does not provide a satisfactory explanation of the large differences among species in sleep time, and it offers no explanation for the fact that some animals sleep during the day while others sleep at night.

The Brain-Maintenance and Memory-Consolidation Theories of REM Sleep

If sleep in itself serves purposes of energy conservation, protection from danger, and bodily restoration, then what is the function of REM sleep? Why is restful non-REM sleep interrupted regularly by these periods of increased brain activity and energy expenditure? This question has generated much debate and research.

One long-standing theory is that REM sleep provides regular exercise to groups of neurons in the brain (Hobson, 1988). Synapses can degenerate if they go too long without being active (Edelman, 1987), so neural activity during REM sleep may help preserve important circuits. One line of evidence for this theory is that the longer a person or animal sleeps, the greater is the proportion of sleep time spent in REM sleep. With longer sleep periods there may be more need to interrupt non-REM sleep with exercise. The theory also helps explain why REM sleep occurs to a much greater degree in fetuses and infants than in adults, regardless of species (see Figure 6.14). In fact, the peak of REM sleep in humans occurs in 30-day-old fetuses, which spend almost 24 hours a day in this state. Perhaps as their brains are developing in the relative isolation of the womb, they need to exercise sensory and motor pathways, and REM sleep is their means for doing that

29
What evidence supports the theories that REM sleep promotes the maintenance of brain circuits and the consolidation of newly learned skills?

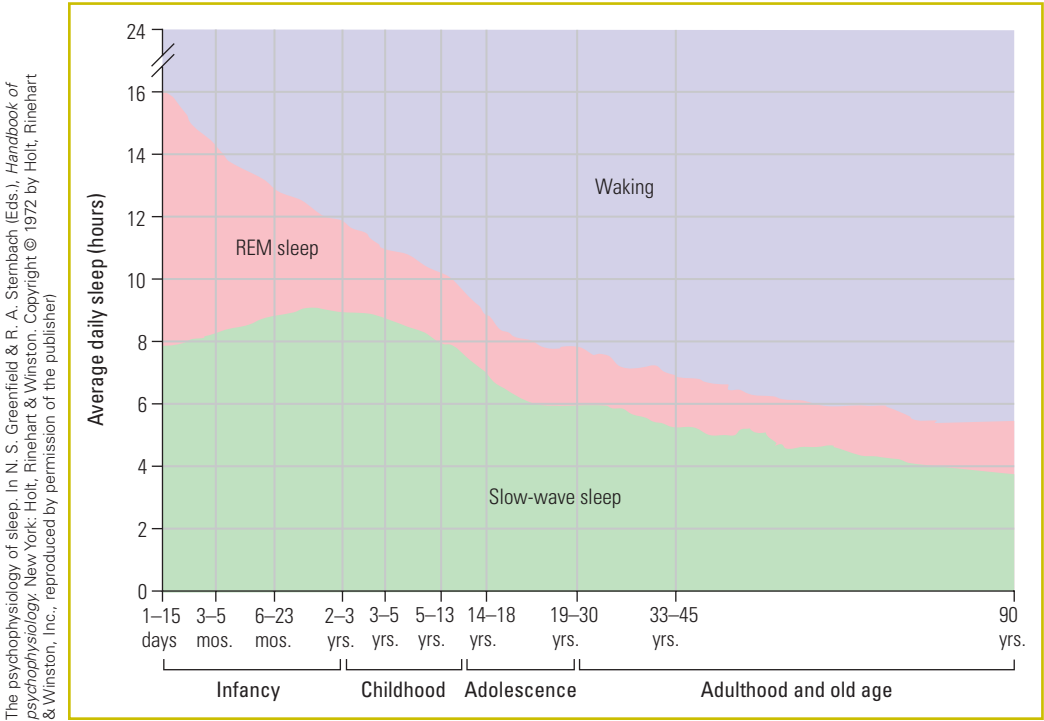


FIGURE 6.14 | Changes in sleep over the course of life As shown here, both the total daily sleep time and the percentage of sleep time spent in REM sleep decrease as a person gets older. If the curves were extended to the left, they would show that prior to birth, REM sleep occupies most of each 24-hour day. (Snyder, F., & Scott, J., 1972.)



Sweet dreams? We have no idea whether human fetuses, such as this one of 7 months, experience dream sensations or not, but we do know that they spend most of their time in REM sleep.

(Hobson, 1988). In the fetus, REM sleep is accompanied by body movements such as kicking and twisting, which are apparently triggered by the bursts of activity in motor areas of the brain, so muscles as well as brain circuits are exercised. By the time of birth a neural inhibitory system matures, which inactivates most motor neurons during REM sleep and thus prevents most movements that would otherwise occur. However, the motor neurons to the eyes and to various internal structures, such as the heart, remain uninhibited, so eye movements and increased heart rate persist as observable effects of the brain's activity.

Closely related to the theory that REM sleep helps to maintain existing brain circuits is the theory that such sleep plays a role in consolidating the effects of new learning. As discussed in Chapter 5, memory consolidation involves the selective strengthening of existing synapses and the manufacture of new ones. According to some sleep researchers, the physiological changes observed in the brain during REM sleep are ideal for promoting such growth (Hobson, 2002).

In many experiments, laboratory animals that were trained in a task and then tested several hours later performed better if they were permitted REM sleep during the period between training and testing than if they were deprived of such sleep (Smith, 1995; Maquet, 2001). Similar experiments with people have produced mixed results, depending apparently on the kind of task. There is little evidence that REM sleep improves subsequent recall for verbally learned factual information, but there is considerable evidence that it improves performance on perceptual-motor tasks—tasks that are more similar to those used with laboratory animals (Stickgold & others, 2001).

In one experiment, college students practiced the skill of locating, as quickly as possible, specific images hidden in visually textured backgrounds (Karni & others, 1994). After this training session, the students slept in the laboratory for about 8 hours. Some were deprived of REM sleep by being awakened whenever they entered that sleep stage; others were deprived of a comparable amount of non-REM sleep by being awakened regularly during non-REM sleep; and still others were allowed to sleep undisturbed through the night. The results were quite dramatic. Those who were deprived of non-REM sleep or were not sleep-deprived at all performed the perceptual task better right after the sleep period than they had just before the sleep period, but those deprived of REM sleep showed no such improvement. This effect occurred only for the newly learned task; lack of REM sleep had no effect on performance of a similar task that had been learned several days earlier. Subsequent experiments have shown that improved performance of such tasks occurs even after a 60- to 90-minute nap, as long as the nap includes at least some time spent in REM sleep (Mednick & others, 2003). Learned motor skills, such as tapping one's fingers in a particular sequence, comparable to playing a series of notes on a piano, also improve dramatically when a period of REM sleep follows initial practice (Walker & others, 2003).

Dreams Viewed as Side Effects of Brain Activity in REM Sleep

30

How might dreams be explained as inevitable consequences of the state of the brain during REM sleep?

<<< Nobody knows if dreams themselves have a function. One view prevalent among researchers today is that dreams don't serve any special purpose, but are side effects of physiological changes in REM sleep that do serve a purpose (Antrobus, 2000; Hobson, 2002, 2004). Neurons in visual and motor areas of the brain become active during REM sleep, and hallucinations of sights and movements may be an inevitable consequence of such activity. Neurons involved in memory retrieval and emotions also become active, and these may bring familiar images and feelings into the sleeping person's mind. In research done many years ago, electrical stimulation in portions of the cerebral cortex produced dreamlike hallucinations in people who were awake (Penfield & Perot, 1963). A similar phenomenon may well occur in REM sleep. In addition to producing hallucinations, the brain continues in REM sleep to engage in some degree of thought, just as it does in non-REM sleep. But now the thought becomes wrapped up in trying to make sense of the hallucinations. The result is the weaving of a story connecting

one hallucination to the next—hence, the dream. Because of reduced mental capacity during sleep, the story is less logical than one the awake brain would develop, but it still contains some degree of logic.

Sometimes the side-effect theory just described is interpreted as an argument against the psychoanalytic view that dream analysis can be useful for understanding a person's mind. But that interpretation seems unjustified. Even if dreams are triggered by random events in the brain, the actual images, emotions, and story lines that constitute the dream are not random. They certainly contain elements based on the dreamer's experience, and because they occur at a time of reduced mental capacity, ideas or feelings that are normally suppressed by higher mental processes could emerge and perhaps be useful in psychoanalysis (Reiser, 1991). In fact, in one recent experiment, people were more likely to dream about a particular person if they were asked to suppress thoughts about that person just before going to sleep than if they were asked to think actively about that person just before going to sleep (Wegner & others, 2004). This finding is at least consistent with the idea that dream analysis might reveal ideas and concerns that a person is actively suppressing during the day.

Individual Variation in the Sleep Drive, and Effects of Failure to Satisfy That Drive

The sleep drive varies from person to person. Some need more sleep than the typical 8 hours a night to function well, and others need less (Douglas, 2002). At the extreme are rare people, referred to as *nonsomniacs*, who sleep far less than most of us and yet do not feel tired during the day. A study of nonsomniacs conducted many years ago by Ray Meddis (1977) suggests that such people are generally vigorous and healthy. One of Meddis's subjects was a 70-year-old nurse who reported that for most of her life she had slept about 50 minutes per night. She was very active during the day and usually spent the night in quiet activities, such as reading or painting. To verify her nonsomnia, Meddis observed her continuously for a prolonged period in the sleep lab. She slept not at all the first 3 days and nights in the lab, remaining cheerful and talkative throughout. Finally, on the fourth night, she slept a total of 99 minutes and awoke feeling fully rested.

The fact that nonsomnia is compatible with physical and psychological health adds to the evidence that only a relatively small amount of sleep is needed for body repair and growth of new synapses in the brain. Yet, most of us do need roughly 8 hours of sleep to function well. We need that sleep because we have a sleep drive that overwhelms our mind and makes us tired, miserable, and relatively ineffective at mental tasks when we fail to meet it. The drive for that much sleep may have evolved for reasons other than body repair and brain growth, but that doesn't mean we can ignore it. It is important to distinguish nonsomnia, which is very rare, from *insomnia*, which is relatively common. An insomniac is someone who has a normal drive for sleep but who, for some reason (such as worry), has great difficulty sleeping at night. Unlike a nonsomniac, an insomniac feels tired all day as a result of not sleeping. And so do most people who voluntarily reduce their sleep.

Many laboratory studies have been conducted in which people with normal sleep drives have voluntarily stayed awake for periods of 3 or 4 days or even longer. After about 72 hours awake, some people begin to experience symptoms such as distorted perceptions and extreme irritability (Borbély, 1986). Sleepiness waxes and wanes during such studies, in accordance with the inner clock that controls it.

>>>

31

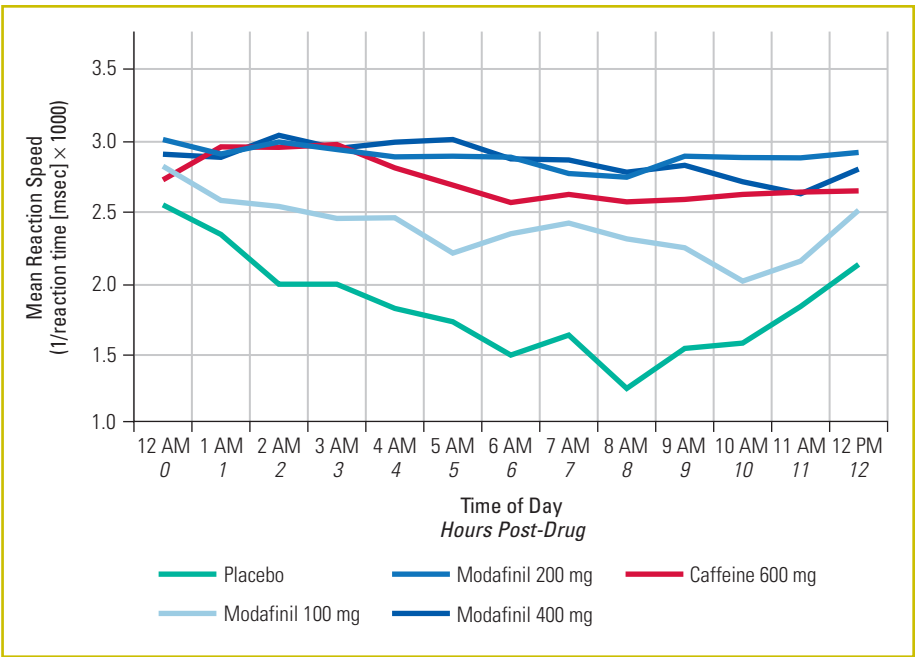
How does insomnia differ from nonsomnia, and what negative consequences occur when people deliberately resist their sleep drive?

Sleep deprivation can be dangerous
The most reliable consequence of sleep deprivation is increased sleepiness, which, in a situation like this one, can be deadly.



Adam Gault / Digital Vision / Getty

FIGURE 6.15 | Decline and rise in vigilance during sleep deprivation In this experiment, university students were kept awake for 42 hours before the start of testing. Just before testing, some received a strong dose of caffeine, others received one of three doses of another stimulant drug, modafinil, and still others received a placebo. In the test of vigilance, conducted every hour, the subjects watched a time display on a computer screen and were instructed to press a button as quickly as possible each time it began to increment. The measure was response speed (the reciprocal of response time). As shown in the graph, in the placebo group vigilance declined steadily from midnight to 8 a.m. and then began to increase. This is consistent with other evidence that the circadian rhythm of sleepiness and wakefulness continues even after long periods of sleep deprivation. There was no such decline in the groups receiving caffeine or in those receiving high doses of modafinil. (From Wesensten & others, 2002.)



People find it much harder to stay awake during the late night and early morning hours than they do during the rest of the 24-hour day, even after several days of sleep deprivation. In such experiments, scores on tests of vigilance, judgment, and creative thinking also wax and wane in a 24-hour cycle, keeping pace with sleepiness (Horne, 1979, 1988; Jennings & others, 2003). Scores on such tests decline when sleepiness rises, apparently because sleepy people have difficulty attending to the task and because their performance is often interrupted by brief moments of falling asleep, from which they arouse themselves. In general, stimulants such as caffeine, which counteract sleepiness, also remove the negative effects of sleep deprivation on the performance of such tasks. For an example of data showing the effects both of sleepiness and of stimulants on a test of vigilance, see Figure 6.15 (Wesensten & others, 2002).

In the real world outside of the laboratory, sleepiness, with its accompanying decline in attention and judgment, is dangerous. Many accidents in the workplace result from it, and sleepiness rivals drunkenness as a leading cause of traffic fatalities (Horne & Reyner, 1995, 2001).

Brain Mechanisms Controlling Sleep

In the early years of sleep research, some researchers believed that sleep is the natural state that the brain slips into when not aroused by external stimulation, so they saw no need to posit the existence of special sleep-inducing mechanisms. But such a view is inconsistent with the observation that sometimes sleepiness overwhelms us even when external stimulation is high, while at other times we can't sleep no matter how quiet, dark, and unstimulating the environment may be. We now know that sleepiness, like other drives, is actively promoted by neural mechanisms located in the hypothalamus and in brain areas closely connected to the hypothalamus.

Rhythm-Generating Neurons in the Hypothalamus Control the Daily Cycle of Sleepiness

32

What is some evidence that the sleep drive is affected by an internal clock, located in the hypothalamus, that can operate even without external time cues?

<<< In all animals, as I noted earlier, the sleep drive waxes and wanes in a cyclic manner over the 24-hour day. This cycle of sleepiness and wakefulness continues, in laboratory animals and human volunteers, even after many days in an artificial time-free environment—an environment where there are no regular changes in

lighting or other cues that could indicate the time of day. In such an environment, the cycle is typically a few minutes longer or shorter than 24 hours, and it varies from individual to individual, but it is remarkably constant within a given individual (Lavie, 2001; Takahashi & Zatz, 1982).

The technical term for any repetitive biological change that continues at close to a 24-hour cycle in the absence of external cues is **circadian** [*sir-kade-ee-un*] **rhythm**. The term comes from the Latin words *circa*, meaning “about,” and *dies*, meaning “day.” The clock that controls the circadian rhythm of sleep in all mammals is located in a specific nucleus of the hypothalamus called the *suprachiasmatic* [*soup-ra-kai-az-mat-ick*] *nucleus*. If this nucleus is damaged, animals lose their regular sleep-wake rhythms and sleep at rather random times over the 24-hour day, and the same is true of human patients (Cohen & Albers, 1991). They also lose their daily rhythms of change in body temperature and secretion of certain hormones, such as cortisol, because the suprachiasmatic nucleus controls those rhythms, too. This nucleus contains rhythm-generating neurons, which gradually increase and decrease their rate of action potentials over a cycle of approximately 24 hours, even when surgically isolated from other parts of the brain (Cermakian & Boivin, 2003).

Input from the Eyes Synchronizes the Hypothalamic Clock to the Light-Dark Cycle

Under normal conditions, the circadian clock is reset each day by the waxing and waning of daylight, so rhythms occur in periods of exactly rather than approximately 24 hours. Experiments with animals show that the cycle can be lengthened or shortened, by as much as a couple of hours either way, by artificially changing the period of light and dark. And experiments with humans as well as other animals show that the cycle can be reset through exposure to bright fluorescent lights.

Charles Czeisler and his colleagues (1989) found that just a few hours of bright fluorescent lighting at night, coupled with avoidance of bright light during the daytime, can, within a few days, reverse a person's circadian clock so that he or she becomes sleepy during the day and alert at night. Very bright light is most effective at delaying the sleep phase of the cycle, but even levels comparable to those already occurring in well-lit offices, factories, and homes have a significant effect (Boivin & others, 1996). Such knowledge can be applied to help night workers adapt their bodily rhythms to their work hours. In one experiment, people simulating night work for a week were more alert at work, slept better during the day, and showed a more complete shift in their body temperature cycle if their work environment was very brightly illuminated and their daytime sleep room was completely darkened than they did under more typical lighting conditions (Czeisler & others, 1990). The fact that light delays the onset of the sleep phase of the cycle may explain why most people living in modern societies, with electric lights on in the evening, have their peak periods of sleepiness in the wee hours of the morning rather than earlier at night (Lavie, 2001).

Brain researchers have found that changes in lighting influence the rhythm-generating neurons by way of a neural tract that runs from the retinas of the eyes to the suprachiasmatic nucleus. These neurons are different from those that are involved in vision, and they derive at least partly from light receptors in the retina that are different from the rods and cones (discussed in Chapter 7) that are involved in vision (Van Gelder, 2003).

>>>

33

What is some evidence that the internal clock is continuously reset by daily changes in light? Through what pathway does that resetting occur?

Resetting the circadian clock

Steven Lockley and his colleagues have found that blue light works better than white light in resetting the circadian clock that times the onset of sleepiness. Timed exposures to such light can help shift workers and world travelers adapt their sleep schedules to the requirements of their work or travel.



Kris Snibbe / Harvard University News Office

Other Neurons in the Hypothalamus and Pons Promote Sleepiness or Wakefulness

34 How are sleepiness and arousal controlled by a sleep center in the preoptic area of the hypothalamus and by arousal centers in the posterior hypothalamus and pons? How are these centers influenced by the circadian clock?

<<< Researchers have identified a number of other nuclei in the brain that are crucially involved in the generation of sleepiness and wakefulness. One such nucleus is the *ventrolateral preoptic nucleus*, which lies in the hypothalamus in front of the suprachiasmatic nucleus (Saper & others, 2001). Electrical stimulation in this area can cause a previously alert animal to fall asleep, and lesions there can result in permanent sleeplessness, eventuating in death in rats. In humans, a viral infection that destroys neurons in this nucleus greatly reduces sleepiness and average daily sleep time (Saper & others, 2001). This nucleus receives neural connections from the suprachiasmatic nucleus, and it also receives input from other brain areas that are involved in the control of sleep and wakefulness. Neurons from this nucleus connect to many brain areas and release the inhibitory neurotransmitter GABA (Saper & others, 2001). The resulting dampening of neural activity throughout the brain is believed to bring on the state of sleepiness and, subsequently, sleep itself.

Other brain areas, including a nucleus in the posterior (rear) portion of the hypothalamus and in the pons, are crucially involved in arousing the brain and maintaining a state of alertness. These nuclei receive input from a variety of sources, including the suprachiasmatic nucleus, and axons from these neurons release excitatory neurotransmitter molecules onto post-synaptic neurons in many parts of the brain (Aston-Jones & others, 2001; Siegel, 2004). An individual's level of sleepiness or arousal at any given moment is believed to derive from a balance between the excitatory action coming from these neural centers and the inhibitory action coming from the ventrolateral preoptic nucleus (Saper & others, 2001).

SECTION REVIEW

There is a drive for sleep, a state involving relative unresponsiveness to the environment.		
Basic Characteristics • Researchers use EEG records to divide sleep into stages, with stage 4 being deepest. • A sleeper progresses through the sleep stages, from light to deep and rapidly back to light, in approximately 90-minute cycles. • REM (rapid-eye-movement) sleep occurs at the transition between one sleep cycle and the next. • True dreams occur in REM sleep, and sleep thought occurs in non-REM sleep.	Functions of Sleep • The preservation and protection theory asserts that sleep is needed to conserve energy and increase safety. It is supported by cross-species comparisons of amount and timing of sleep. • The body-restoration theory contends that the body needs sleep to recover from wear and tear. It is supported by sleep-deprivation studies in animals. • REM sleep may function to maintain brain circuits and consolidate new learning. • Different people need different amounts of sleep. Sleep deprivation reduces performance, especially on tests of vigilance.	Brain Mechanisms • The suprachiasmatic nucleus in the hypothalamus acts as an internal clock for sleepiness and wakefulness. • Light synchronizes the internal clock with the light-dark cycle. • A hypothalamic nucleus that receives input from the suprachiasmatic nucleus induces sleep using an inhibitory neurotransmitter. • Other nuclei release excitatory neurotransmitters to waken the brain and keep it alert.

Foundations for Understanding Emotions

Enough sleep; wake up and face the challenges of the day. Midterm exams are just around the corner, your family is after you to get your life in order, your lover has just left you for another, the surgeon says your nose and left ear will have to go, and a hungry tiger is crouched behind you. Are you awake? All these events have something in common: All (if you believe them) have the potential to elicit strong emotions.

Emotion, like *motivation*, is a concept that applies to almost all of psychology. Much, if not all, of our thought and behavior is tinged with emotion. My goal here, in this section, is to introduce to you some ideas about the underlying processes and mechanisms of emotion. You will find much more information about particular emotions in other chapters.

The Nature and Value of Emotions

Emotion is a concept that has been difficult for psychologists to pin down and define in an agreed-upon way. According to one estimate, psychologists have generated at least 90 different definitions of emotion (Plutchik, 2001). My own preferred definition is one that closely resembles the way the term is used in everyday life: An **emotion** is a subjective feeling that is mentally directed toward some object. This brief definition requires some elaboration, however.

Emotions Are Feelings Directed Toward Objects

An emotion, as just defined, has two components: feeling and object. The *object* of an emotion may be a person, thing, or event, real or imagined, but that object is always something that is in some way important to the one who experiences the emotion. The feeling and the object are inextricably entwined in the emotional experience; the object is perceived as the cause of the emotion. The person *loves* John because John is so *loveable*, *hates* war because war is *detestable*, *fears* snakes because snakes are *frightening*, and *grieves* the loss of a friend because that loss is *grievous*. In these examples, John, war, snakes, and loss of friend are the objects of emotions. In some emotions—specifically, pride, shame, guilt, and embarrassment (all discussed in Chapter 14)—one’s self or one’s own behavior is the object. These are called *self-conscious emotions*.

The *feeling* associated with emotion, independent of the object, is referred to by some psychologists as *affect* [**aph**-ect]. Such feelings can vary along two dimensions. One dimension has to do with the degree of pleasantness or unpleasantness of the feeling, and the other has to do with the degree of mental and physical arousal. Figure 6.16 shows one way of describing and depicting various affects. Notice that by moving upward or downward on the circle, one goes toward increasing or decreasing arousal; by moving rightward or leftward, one goes toward increasing degrees of pleasure or displeasure. Most of the terms in the figure describe not emotions, but feelings devoid of objects. An emotion depends on the object as well as the feeling. Thus a feeling of pleasure may be experienced as the emotion of *pride* when the object is oneself and as the emotion of *love* when the object is someone else.

35
According to the definitions used here, how does *emotion* differ from *affect* and from *mood*?

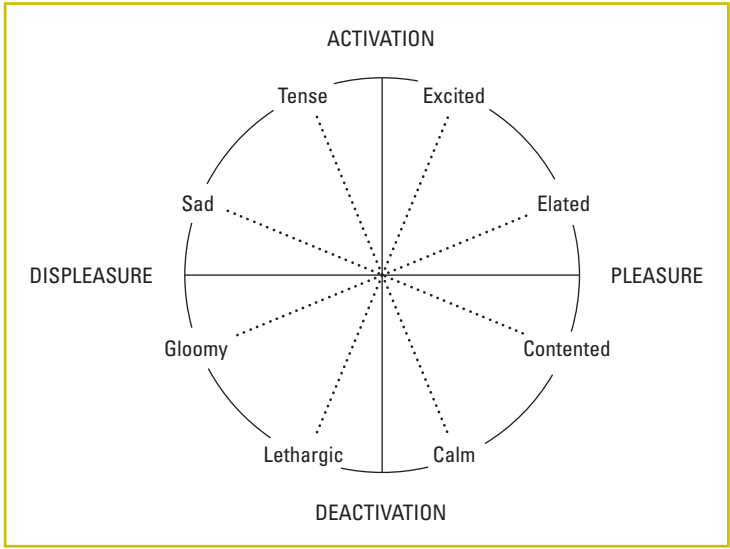


FIGURE 6.16 | Affects Basic affects are shown here arranged in a circle in which the vertical dimension represents degree of perceived arousal and the horizontal dimension represents degree of pleasure or displeasure. (Adapted from Russell, 2003.)

Emotional feelings are not always attached to objects. Sometimes an emotional feeling is experienced as free-floating rather than directed at a particular object, and in that case it is commonly referred to as a **mood**. Moods can last for hours, days, or even longer and can color all aspects of one's thought and behavior. In some cases everyday language provides different terms for moods and emotions that have similar underlying feelings. A feeling of being tense, jittery, and unhappy may be labeled as *anxiety* when it is a free-floating mood but as *fear* when it is an emotion associated with an object such as a snake or an upcoming examination. A feeling of being sad and upset may be labeled *depression* when it is free-floating and as *grief* when it is associated with a specific loss.

There Is No Agreed-upon Way to Classify Emotions

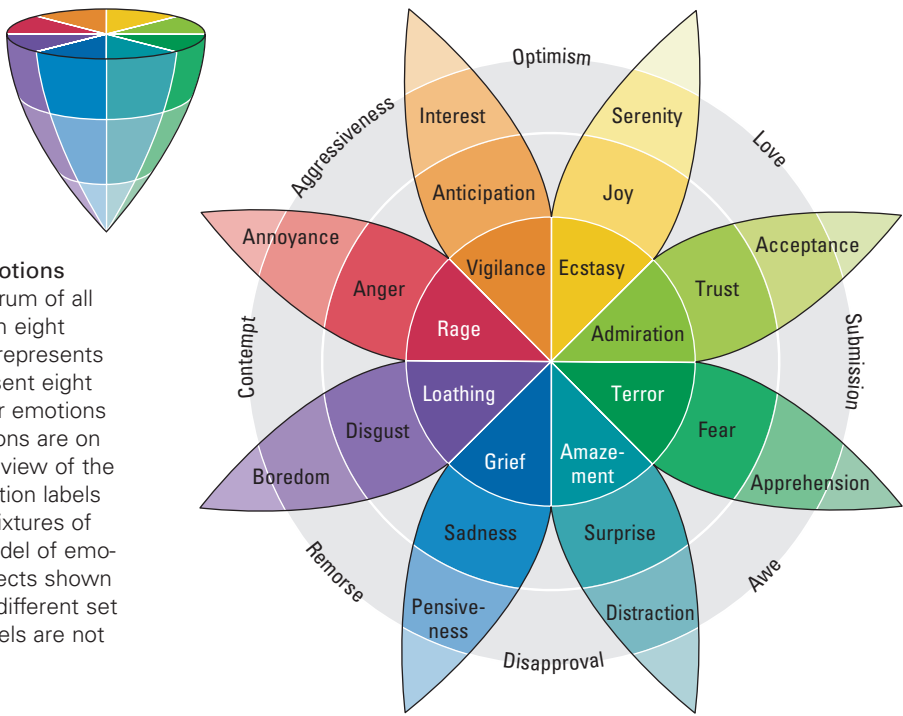
How many different emotions are there? The answer to that is quite arbitrary; it depends on how finely graded a taxonomy (system of classification) we wish to create. *Fear*, for example, might be understood as one emotion or as a cluster of emotions that all have some things in common. One's fear of snakes, fear of death, and fear of others' judgments, for example, may be experienced as qualitatively different from one another.

36 Through what strategy did Plutchik arrive at his model of eight primary emotions? Why is this model not universally accepted?

<<< Some psychologists have attempted to identify a set of primary emotions by analyzing the emotion labels that are common to a language or to several languages. In English several hundred different words are used as labels for emotions, but many of them are synonyms or near synonyms. By asking people to rate pairs of common emotion labels for the degree of similarity between the emotions they describe, psychologists have found that the labels can be clustered into a relatively small number of groups. Using this method, Robert Plutchik (2003) identified what he considers to be eight primary emotions, which can be arranged as four pairs of opposites: *joy* versus *sorrow*, *anger* versus *fear*, *acceptance* versus *disgust*, and *surprise* versus *expectancy*. According to Plutchik's model (depicted and described more fully in Figure 6.17), these primary emotions can mix with one another in any number of ways to produce an essentially infinite variety of different emotional experiences, just as a small set of primary colors can mix to create virtually all the colors the human eye can see.

Plutchik's model is elegant and useful, but by no means universally accepted. Other psychologists, using similar means, have generated different sets of basic

FIGURE 6.17 | A Model of primary emotions Robert Plutchik has proposed that the spectrum of all emotions can be represented by a cone with eight sectors. The vertical dimension of the cone represents emotional intensity. The eight sectors represent eight primary emotions, arranged such that similar emotions are next to one another and opposite emotions are on opposite sides of the cone. In the exploded view of the cone, with the sectors laid out flat, the emotion labels in the spaces between sectors represent mixtures of the adjacent two primary emotions. This model of emotions may appear similar to the model of affects shown in Figure 6.16, but it was developed from a different set of observations and premises; the two models are not directly comparable. (From Plutchik, 2001.)



emotions. Still others have argued that emotional experiences are so subjective and variable, and emotional descriptions are so much influenced by culture and language, that any attempt to arrive at a universally agreed-upon list of basic emotions is futile (Russell, 2003).

Emotions Serve Adaptive Functions

Our capacity for emotions must have come about through natural selection because of their adaptive value. This line of thinking draws attention to the motivating qualities of emotions. Implicit in the positive-negative dimension of emotional feelings is the idea that emotions motivate us to approach objects that can help us, and to avoid or repel objects that hinder us, in our efforts toward survival and reproduction. When passionately in love, we behave in all sorts of ways—some of which would otherwise seem foolish—to get close to our beloved, a potential partner in reproduction. When fearful, we run away, or freeze, or try to look inconspicuous, or in other ways attempt to avoid the dangerous predator or enemy. When angry, we puff ourselves up, threaten, and sometimes even fight, in order to repulse or change the behavior of one who stands in the way of what we want or need. Strong emotions of all sorts focus our attention narrowly on the object of the emotion and lead us to ignore, for a time, other events and drives that would distract us from that object.

Emotions also promote our survival and reproduction through their capacity to communicate our intentions and needs to others. Our expressions of love may signal to our beloved that we will be faithful mates. Our expressions of fear to human attackers may signal submission and thereby prevent a fight, or they may draw friends to our aid. Mere hints of anger in our demeanors may, all by themselves, convince the objects of our anger to change their ways in desired directions. As discussed in Chapter 3, many emotions are automatically and unconsciously expressed on the face in ways that are universal, and such expressions most likely came about through natural selection because of their communicative value.

Effects of Bodily Responses on Emotional Feelings

Most emotional states are accompanied by peripheral changes in the body. By *peripheral* changes, I mean all changes in the body outside of the brain itself. These include changes in heart rate, blood pressure, diversion of blood from one set of tissues to another, activation of certain glands, tension in particular muscles, and facial expression of the emotion. The changes, overall, are adaptive because of their communicative function or their role in helping prepare the body for possible action.

Common sense and everyday language tell us that these peripheral changes are caused by our emotions. We say: “My heart pounds and I tremble because I am afraid;” “My face is flushed and my teeth are clenched because I am angry”; “Tears well up in my eyes and a lump forms in my throat because I feel grief.” Over a hundred years ago, in his classic textbook, *The Principles of Psychology*, William James (1890/1950) turned common sense upside down and suggested that bodily reactions precede the emotions and cause them, rather than the reverse.

James's Peripheral Feedback Theory of Emotion

James's evidence for his theory of emotion came not from experiments but from introspection. Looking inward at his own emotions, James concluded that his emotional feelings were really sensations stemming from bodily changes. Thus, his feeling of fear was his feeling of a quickened heart, shallow breathing, goose-bumpy flesh, and trembling limbs. Similarly, his feeling of anger was his feeling of

>>> **37**
How can emotions promote adaptive ends through their motivating and communicative effects?



© ThinkStock LLC / Index Stock

Anger serves valuable functions
Sometimes the person we are most angry at is also the person we care most about. This woman's anger is a clear signal to her partner that something in their relationship is out of balance and needs correction.

>>> **38**
What is James's theory of emotion, what evidence did James supply for the theory, and what modern evidence is consistent with the theory?

a flushed face, dilated nostrils, and clenched teeth. James believed that he could identify a different constellation of bodily changes for each emotion and that if he could not feel these changes, he would not feel the emotion.

The essence of James's theory is that the bodily reaction to an emotion-provoking stimulus is automatic, occurring without conscious thought or feeling, and that the assessment of one's emotional state comes later and is based on the perception of the bodily state. The brain, at some unconscious level, judges the bear to be dangerous and precipitates a bodily change that helps prepare the person for flight. There is no time, in an emergency, for conscious reflection. The body reacts immediately. Then, later on, when the danger is over or at least reduced, the person may sense his beating heart and trembling knees and conclude that he is frightened. The contrast between James's theory and the common sense theory that he was arguing against is illustrated in the top two portions of Figure 6.18.

A considerable amount of evidence today tends to support James's theory. People throughout the world describe their emotions in terms of bodily changes and are quite consistent in the kinds of changes they associate with each emotion (Cacioppo & others, 1992; Rime & others, 1990). Researchers have also found that people who are particularly good at detecting changes in their own internal condition, such as changes in their heart rate, are more likely than others to detect and report emotional states in themselves (Critchely & others, 2004; Weins & others, 2000). Moreover, recent brain imaging studies have shown that a certain portion of

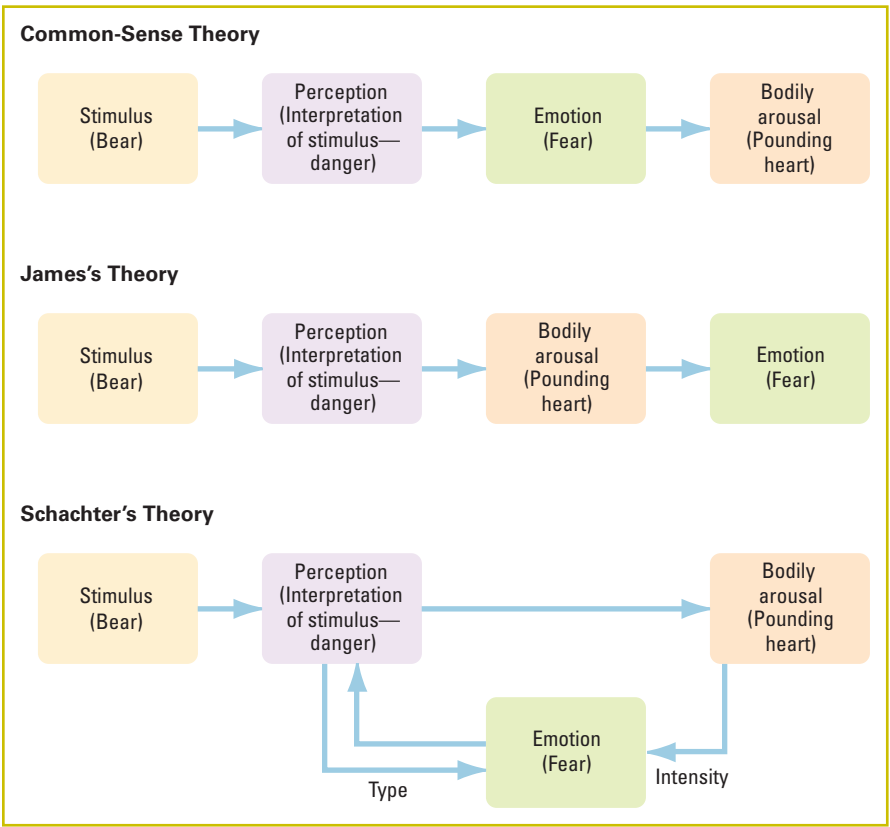


FIGURE 6.18 Three theories of emotion Each theory proposes a different set of causal relationships among perception of the stimulus, bodily arousal, and emotional feeling. According to the common-sense theory (so labeled by James), the emotional feeling precedes and causes the bodily arousal. James's theory reverses that relationship, and Schachter's theory holds that the intensity of the emotional feeling depends on the bodily response but the type of emotion experienced (such as fear, anger, or love) depends on the person's cognitive assessment of the external stimulus or situation.

the somatosensory area of the cerebral cortex, which becomes active when a person is sensing his or her own bodily state, also becomes active when a person is consciously assessing his or her own emotional state (Critchely & others, 2004; Damasio, 2001).

Schachter's Cognition-Plus-Feedback Theory

In the 1960s, Stanley Schachter developed a theory of emotion that can be understood as a variation of James's theory. According to Schachter, the feeling of an emotion depends not just on sensory feedback pertaining to the body's response but also on one's perceptions and thoughts (cognitions) about the environmental event that presumably evoked that response. More specifically, he proposed that perception and thought about the environment influence the *type* of emotion felt, and that sensory feedback about the degree of bodily arousal influences the *intensity* of the emotion felt (see the bottom portion of Figure 6.18). Thus, if you see a bear, your perception that it is dangerous leads you to interpret your emotion as fear, and your perception of the degree to which your heart is pounding and your knees are trembling determines the amount of fear you experience. Schachter also proposed that the intensity of the emotional feeling influences the interpretation of the stimulus. Thus, if your bodily arousal was already high, perhaps from drinking too much coffee, that arousal would contribute to your emotional intensity and might lead you to perceive the bear as more dangerous than you would otherwise perceive it to be.

In experiments testing his theory, Schachter (1971) injected people with either epinephrine (a substance also known as adrenaline, which raises heart rate and produces other effects associated with high arousal) or a placebo (an inactive substance) and then exposed them to various emotion-eliciting conditions. He found that epinephrine by itself did not produce any particular emotion (the subjects just said they felt jumpy), but when epinephrine was combined with an emotion-inducing situation, such as a horror film, it increased the intensity of the subject's emotion. As predicted by his theory, the kind of emotion subjects felt depended on the external situation, but the intensity was heightened by epinephrine. The epinephrine-injected subjects manifested and reported more anger when insulted, more fear when watching a frightening film, and more hilarity when watching a slapstick comedy than did placebo-injected subjects. This emotion-enhancing effect occurred only if the subjects had not previously been informed of the physiological effects of epinephrine. Thus, according to Schachter, high physiological arousal increases emotion only when people believe that the arousal is caused by the external situation.

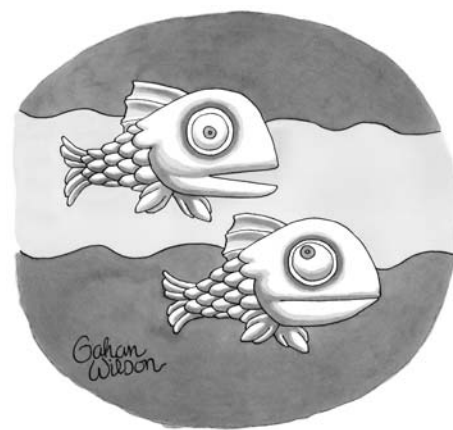
Schachter's theory fits well with the modern idea that emotions are defined not just by feelings but also by the perceived objects of those feelings.

Influence of Facial Feedback on Emotional Experience

Some years ago, Paul Ekman (1984) proposed a theory of emotions that is similar to James's peripheral feedback theory but focuses particularly on the role of the face. As discussed in Chapter 3, Ekman and his colleagues found that different basic emotions are associated with different facial expressions. Those expressions are produced rapidly and automatically (though they can be inhibited). According to Ekman, sensory feedback from facial expressions contributes both to emotional feelings and to the production of the full-body reactions that accompany emotions.

If you form your face into a smile, will you feel happier? A Pollyannaish suggestion, perhaps, but research suggests that there may be some truth to it. In one experiment, for example, some subjects were each asked to hold a pencil tightly between their teeth, which forced their faces into smiling expressions, and others were each asked to hold a pencil between their lips, in a manner that prevented smiling (see Figure 6.19), as they watched films of happy or funny scenes

>>> **39**
How does Schachter's theory differ from James's? How did Schachter support his theory with experiments?



© The New Yorker Collection 2004. Gahan Wilson from Cartoonbank.com. All rights reserved.

"Just once I would like to laugh or to frown."

>>> **40**
What is some evidence supporting Ekman's theory that feedback from the face influences one's felt emotion and produces whole-body reactions that are consistent with the facial expression?

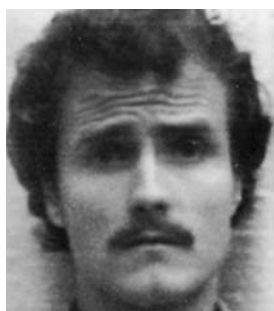


Courtesy of Robert Soussignan

FIGURE 6.19 Forcing a smile In Soussignan's (2002) experiment testing the facial feedback theory of emotion, some subjects were asked to hold a pencil between their teeth, in a manner that led to their lips being pulled back and their cheeks raised, similar to a full-faced smile, and others were asked to hold a pencil between their lips, which prevented smiling.

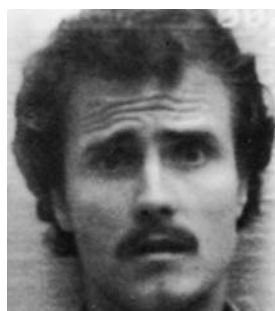
(Soussignan, 2002). The result was that the former reported more enjoyment of the films than did the latter. In another experiment, subjects were asked to contract certain facial muscles in ways that were (unbeknownst to the subject) designed to mold the face into a smile or a frown. While they were maintaining these facial postures, they were asked to examine certain emotion-arousing pictures. Then, as if it were incidental to the experiment, they were asked to fill out a questionnaire designed to assess their emotional feelings. The result was that the subjects whose faces had formed a frown reported more anger and less happiness than did those whose faces had formed a smile (Laird, 1974).

Ekman and his colleagues (1983) found that induced facial expressions not only can alter self-reports of emotion but also can produce physiological responses throughout the body that are consistent with the induced expression. In one experiment, these researchers asked subjects to move specific facial muscles in ways designed to mimic each of six basic emotional expressions (see Figure 6.20). For comparison, they asked other subjects to experience each emotion by mentally reliving an event in which that emotion had been strong. As the subjects held the facial expressions or imagined the emotional events, various indices of their physiological arousal were recorded. The main finding was that different patterns of arousal accompanied different emotions, but the pattern for a given emotion was the same whether the person had been asked to relive that emotion or simply to move certain facial muscles (see Figure 6.21). For instance, anger, whether it was relived or mimicked by facial molding, was accompanied by increases in skin

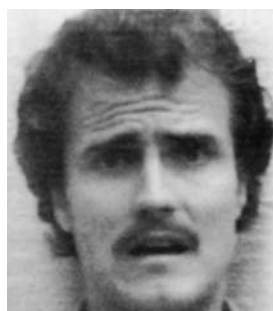


Paul Ekman

(a)



(b)



(c)

FIGURE 6.20 Inducing an expression of fear Shown here are frames from a videotape of a man following the instructions used by Ekman and his colleagues (1983) to induce an expression of fear: (a) "Raise your brows and pull them together," (b) "now raise your upper eyelids," and (c) "now stretch your lips horizontally, back toward your ears." Other instructions were used to induce other emotional expressions, producing the results shown in Figure 6.21.

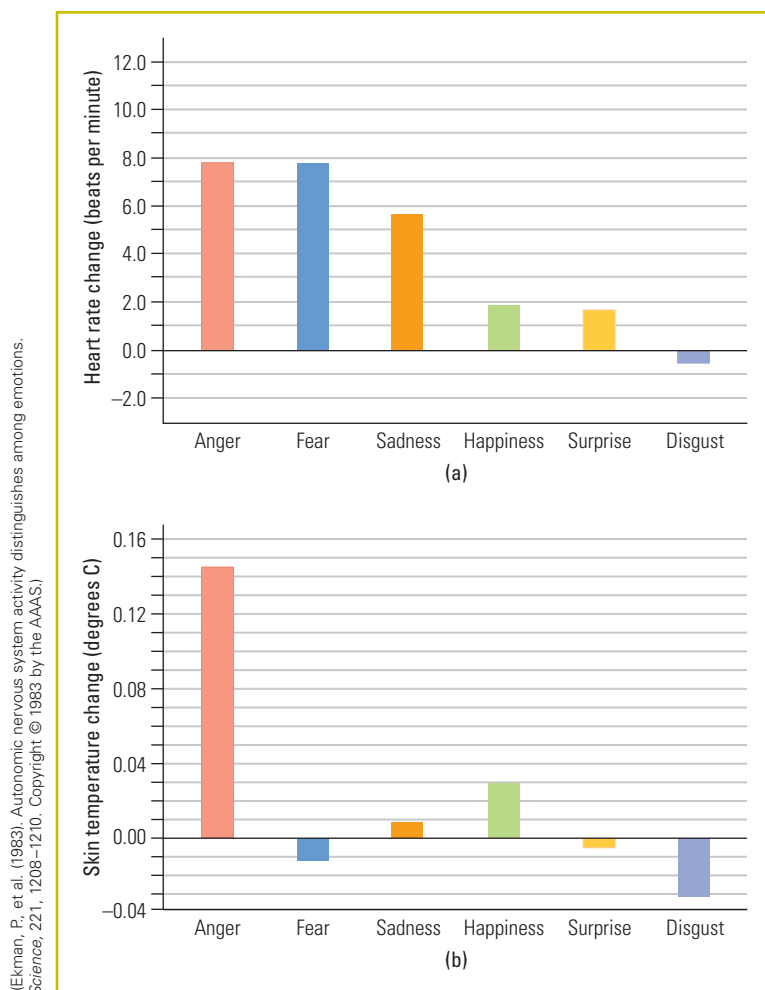


FIGURE 6.21 Effect of induced emotional expressions on heart rate and skin temperature Ekman and his colleagues (1983) found that (a) heart rate increased most when the induced facial expression was of anger, fear, or sadness, and (b) skin temperature increased most when the induced expression was of anger. Although not shown here, the same pattern of effects also occurred when subjects were asked to relive specific emotional experiences through imagination.

temperature that did not occur for the other emotions (consistent with evidence that blood tends to flow into the skin during anger). As another example, both anger and fear—in both the mimicking and the reliving conditions—increased the subjects' heart rates more than did any of the other emotions. Researchers subsequently replicated these findings with a wide variety of people, including people in non-Western cultures (Levenson, 1992; Levenson & others, 1992).

Brain Mechanisms of Emotion

Thus far I have focused on the role of peripheral bodily changes in emotion and have said little about the brain. But of course the brain is the center both for producing the bodily changes and for experiencing emotions. Research on the brain's emotional systems has focused particularly on two structures: the amygdala and the prefrontal portion of the cerebral cortex.

The Amygdala Assesses the Emotional Significance of Stimuli

In James's classic example (diagrammed in Figure 6.18), a person sees a bear and reacts in a manner that the person subsequently interprets as fear. What causes the initial reaction? James assumed that somehow the brain can quickly and unconsciously assess the significance of stimuli to the person and can generate appropriate bodily reactions. We now know that a crucial structure for this capacity of the brain is the amygdala.

41

What is some evidence that the amygdala initiates emotional reactions to stimuli and that this effect can occur even without conscious awareness of the emotion-eliciting stimuli?

<<< The **amygdala** [uh-mig-duh-luh] is part of the limbic system (discussed in Chapter 5). It is a cluster of nuclei buried deep, underneath the cerebral cortex, in the temporal lobe. This structure appears to be, among other things, the brain's early warning system. It receives stimulus input from all of the body's sensory systems; performs continuous, rapid assessments of that input; and alerts the brain and body when those assessments suggest that some sort of reaction is called for.

The amygdala receives sensory input by way of two routes: a very rapid *subcortical* route and a somewhat slower *cortical* route (diagrammed in Figure 6.22 for the sense of vision). Through the former, it analyzes incoming information even before that information has been processed by sensory areas of the cerebral cortex. Through the latter, it analyzes, in more detail, information that has been processed by the cerebral cortex. The amygdala sends its output to many other brain structures. Through these connections it alerts the rest of the brain to pay attention to the stimulus of concern and it generates such bodily reactions as increased heart rate and muscle tension (Davis, 1992).

In a set of classic experiments with monkeys, removal of the amygdala along with nearby portions of the temporal lobe of the cerebral cortex on both sides of the brain produced a dramatic set of changes in behavior described as *psychic blindness* (Klüver & Bucy, 1937; Weiskrantz, 1956). The monkeys could still see objects and could move in a coordinated fashion, but they seemed indifferent to the psychological significance of objects. They no longer responded fearfully to objects that had previously frightened them or aggressively to objects that had previously angered them. They also failed to distinguish in the usual ways between foods and non-foods or between appropriate and inappropriate objects for sexual attention.

People who have suffered partial damage to the amygdala likewise exhibit striking losses in fear and anger (Allman & Brothers, 1994; Berridge, 2003). For instance, one woman with such damage failed to react emotionally, or even to show much concern, when she and her husband were mugged (Bruce & Young, 1998). In

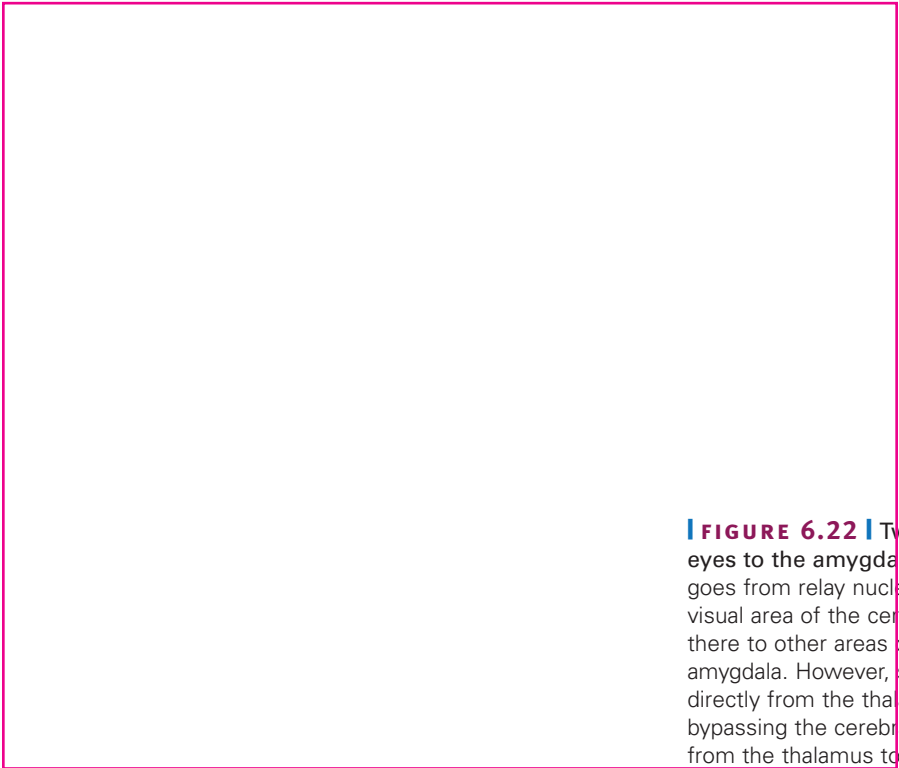


FIGURE 6.22 Two routes from the eyes to the amygdala. Most visual input goes from relay nuclei in the thalamus to the visual area of the cerebral cortex, and from there to other areas of the brain, including the amygdala. However, some visual input goes directly from the thalamus to the amygdala, bypassing the cerebral cortex. The direct input from the thalamus to the amygdala may be responsible for fast, unconscious emotional responses to visual stimuli.

experiments, people with damage to the amygdala generally fail to respond emotionally to stimuli that regularly elicit emotional responses and feelings in brain-intact people (Berridge, 2003; Helmuth, 2003). Moreover, in brain imaging studies with brain-intact people, increased neural activity in the amygdala correlates strongly with increases in fear, anger, or disgust (Hamann & others, 2002; Whalen, 1998). The amygdala may also be activated, but less so, by stimuli that induce positive emotions. In one brain imaging study, with men, the left amygdala (but not the right) became active in response to a variety of appealing pictures, including depictions of attractive semi-nude women and of appetizing foods (Hamann & others, 2002).

Researchers have found that people react emotionally even to stimuli—such as angry words or faces—that are flashed on a screen too rapidly for conscious perception (Katkin & others, 2001; Öhmann, 1999). This effect most likely occurs by way of the subcortical sensory input to the amygdala. Joseph LeDoux and his colleagues found that rats with lesions destroying the entire visual and auditory cortex, but not the amygdala, continued to respond emotionally to sights and sounds that had previously been paired with electric shock, but rats with lesions destroying the amygdala did not (LeDoux, 1996; LeDoux & others, 1989). Similarly, people who have damage to the visual cortex have been shown to respond emotionally to visual stimuli that they could not consciously see (Anders & others, 2004). Sensory areas of the cortex are essential for conscious perception of stimuli, but not for unconscious emotional responses to them. The fact that emotional responses can be generated by subcortical pathways to the amygdala helps explain why our emotions are often irrational and why they are so difficult to control through conscious reasoning.

The Prefrontal Cortex Is Essential for Conscious Experience of Emotion

While the amygdala is essential for unconscious emotional responses, the **prefrontal cortex**—that is, the foremost portion of the frontal lobe of the cerebral cortex (see Figure 6.22)—is essential for the full conscious experience of emotions and the ability to act in deliberate, planned ways based on those feelings. One line of evidence for this comes from observations many years ago of people who were subjected to *prefrontal lobotomy*—an operation that disconnects the prefrontal area from the rest of the brain and was, before the development of drug treatments, a relatively common treatment for severe mental disorders. The operation usually relieved people of their crippling emotional feelings, but it also left them unable to plan and organize their lives effectively (Valenstein, 1986). The prefrontal cortex receives input from the amygdala and from the somatosensory cortex, and such input provides it with information about the amygdala's assessment of the stimulus and the body's state of arousal.

The prefrontal cortex of the right hemisphere appears to be most involved in negative emotions, and that of the left hemisphere appears to be most involved in positive emotions (Haller & others, 1998; Davidson & others, 2003). Some of the evidence for this comes from experiments in which people were presented with emotion-eliciting stimuli while the overall rate of neural activity in each frontal lobe was recorded by means of an electroencephalograph. In one such experiment, for example, the right prefrontal cortex became most active in response to disturbing medical films of third-degree burns or limb amputations and the left prefrontal cortex became most active in response to happy films of a puppy playing with flowers or a gorilla taking a bath (Davidson & others, 1990).

Such research on the brain and emotions illustrates nicely the increasing conjoining of psychological theories with knowledge of physiological mechanisms. Conscious and unconscious emotional assessments, and positive and negative emotions, differ not just in the realm of subjective experience but also, apparently, in the brain pathways that underlie them.

What is some evidence that the prefrontal cortex is involved in the conscious feeling of emotions and that the right and left prefrontal cortices are involved with different classes of emotions?

42

SECTION REVIEW

Emotions are subjective feelings directed toward specific objects.		
Nature of Emotion • An <i>emotion</i> is a feeling that is tied subjectively to an object of that feeling, as exemplified by anger at someone who insulted you. • The feeling aspect of emotion, called <i>affect</i>, can vary in degree of arousal and degree of pleasantness of unpleasantness. • Psychologists have theories, but no consensus, about how to classify emotions. • Emotions have adaptive value, motivating us and communicating our intentions and needs to others.	Peripheral Feedback • James argued that peripheral bodily reactions precede and cause emotion. • Schachter proposed that peripheral feedback affects emotional intensity, but the type of emotion experienced depends on thoughts and perceptions. • Ekman suggested that feedback from facial muscles can induce feelings and bodily reactions indicative of emotion. • Evidence was presented supporting all three of these ideas.	Brain Mechanisms • The amygdala alerts the brain and body when emotion may be called for. It rapidly evaluates sensory information for its significance to survival or well-being and triggers bodily responses. • The prefrontal cortex is crucial for conscious emotional experience and deliberate action based on it. • The right prefrontal cortex is relatively specialized for negative emotions and the left for positive emotions.

Concluding Thoughts

As you review the mechanisms of reward, hunger, sex, sleep, and emotions described in this chapter, you might find the following two points useful in organizing some of your thoughts.

1. The attempt to relate behavioral states to physiological states Hunger, sexual drive, sleepiness, and emotions are examples of *behavioral states*. They all involve somewhat sustained but reversible changes in the way a person or other animal behaves in a given environment. In this chapter you read many examples of research aimed at finding relationships between behavioral states and *physiological states*. The goal of such work is to identify the somewhat sustained but reversible changes in the brain that provide the foundations for behavioral states and to identify the various neural and chemical inputs to the brain that bring on or terminate a given behavioral state. This goal is both fascinating and difficult to achieve because the machine being studied is so extraordinarily complicated. As you think about each type of behavioral state discussed in this chapter, ask yourself: What changes in the brain correspond to this state, and how are these changes regulated through means that tend to promote the individual's survival?

2. Two categories of methods in the physiological study of states Like all the other parts of the book, this chapter is about methods and evidence, not just about findings. As you review the specific methods described in the chapter, you will find that they fall into two broad categories. One category involves *intervention* in ongoing physiological processes to see what happens to the behaviorally measured state. What happens to hunger, or the sex drive, or the sleep drive, or an emotion if a particular

part of the brain is destroyed or stimulated or if the activity of a particular hormone or neurotransmitter is increased or decreased in some way? As you review the chapter, notice how often such methods were mentioned. Intervention is a powerful way to identify causal relationships between physiology and behavior. But most intervention procedures are harmful or at least risky to the subject, so they are used primarily in studies of animals other than humans. The intervention approach is approximated, however, in studies of people whose natural physiology has been disrupted through accident or disease. Several studies of that sort were described in the chapter.

The other category involves the *measurement* of physiological processes and the *correlation* of the measures with changes in behavioral state. What natural changes in brain activity (measured by EEG or by brain imaging methods) or in hormonal secretions, heart rate, or other physiological variables accompany specific changes in behavioral state? Most measurement procedures are safe and can be used with humans. Notice, as you review the chapter, how often this correlational method was used in the human studies described. This method helps identify reliable relationships but does not, by itself, tell us about cause and effect. The observation that brain waves slow down during sleep, or that skin temperature goes up during anger, or that the amygdala is active during fear tells us about correlations but not necessarily about cause and effect. To test cause-effect hypotheses, researchers seek to manipulate the physiological system and measure the effect on behavior. That is why intervention experiments with non-human animals often go hand-in-hand with correlational studies in humans to get at the full story.

Further Reading

ALEXANDRA W. LOGUE (2004). *The psychology of eating and drinking*, 3rd ed. New York: Brunner-Routledge.

This is an interesting introduction to the variables, both inside and outside the body, that affect hunger, thirst, and the ways they are expressed in behavior. The book also discusses the major disorders of eating and drinking.

JOANN RODGERS (2001). *Sex: A natural history*. New York: Holt.

This engaging 500-page book, written for nonspecialists, deals with almost every aspect of sex that one can imagine: meiosis, gametes, the physiology and psychology of orgasms, differences between males and females, determinants of sexual orientation, brain mechanisms, species differences, cultural influences, alternative sex, physical and psychological underpinnings of attraction, and

foundations for long-term mating attachments. Rodgers is a science writer, not a scientist, but she prepared well by reading much of the research literature and interviewing many of the leading researchers concerned with various aspects of sexual drive and behavior.

DAVID N. NEUBAUER (2003). *Understanding sleeplessness: Perspectives on insomnia*. Baltimore: The Johns Hopkins University Press.

This brief book is written primarily for health-care providers but is easily readable by nonprofessionals. Neubauer begins with a clear discussion of the biological bases of normal sleep and of the meaning of sleep in people's lives. He then goes on to use that discussion as a foundation for understanding and treating common sleep disorders.

ROBERT PLUTCHIK (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. Washington, DC: American Psychological Association.

This is a well-written, up-to-date introductory textbook on emotion by a long-time leading researcher into that topic. Plutchik contends that emotions are part and parcel of nearly all of the behavior of human and non-human animals and that they help to guide behavior in adaptive ways. From this perspective, he discusses the major theories of emotions, his own model of primary emotions, the role of emotions in human thought, the evolution of emotions, the measurement of emotions, the development of emotions in children, brain mechanisms of emotions, and emotional disorders.

FPO



SENSATION AND PERCEPTION

All that we learn we learn through our senses. All the images that come to mind when we remember or think are derived from our sensory experiences. Our senses are the conduits that connect our brains and minds to the rest of the world. Nothing is more fundamental to psychology than an understanding of the senses. Chapter 7 is about smell, taste, pain, and hearing, and Chapter 8 is about vision. In both chapters, the main question is this: How does our nervous system respond to and make sense of the patterns of energy in the physical world?

part 4

Smell, Taste, Pain, Hearing, and Psychophysics

OVERVIEW OF SENSORY PROCESSES

SMELL

- Anatomy and Physiology of Smell
- Age and Sex Differences in Olfactory Sensitivity
- Discriminating Among Individuals by Smell
- Smell as a Mode of Communication: Do Humans Produce Pheromones?

TASTE

- Anatomy and Physiology of Taste
- An Evolutionary Account of Taste Quality

PAIN

- Neural Pathways for Pain
- The Modulation of Pain

HEARING

- Sound and Its Transduction by the Ear
- Pitch Perception
- Making Sense of Sounds

PSYCHOPHYSICS

What would your mental life be like if you had no senses? What if, from birth, you could not see, hear, touch, taste, smell, or in any other way sense the world around you? You would not be able to react to anything, because reaction requires sensory input. You would not be able to learn anything, because learning begins with sensory input. Would you be able to think? What could you think about with no knowledge gained from the senses? Philosophers, from Aristotle on, have pondered these questions and have usually concluded that without sensation there would be no mental life. It is no wonder that the study of the senses has always been a fundamental part of the science of psychology.

Sensory systems have evolved in all animals for the purpose of guiding their behavior. To survive and reproduce, animals must respond to objects and events in the world in which they live. They must move toward food and mates, for example, and away from predators and other dangers. Sensory systems evolved not to provide full, objective accounts of the world's physical properties but, rather, to provide the specific kinds of information that the animal needs in order to survive and reproduce. To understand an animal's sensory systems is to understand its way of life. For instance, frogs' eyes contain "bug detectors," neurons that respond only to small, moving dark spots and that trigger tongue movements in the spot's direction; they also contain color detectors that are ideally tuned to distinguish the hue of the pond from that of the grass and lily pads (Muntz, 1964). Many species of migrating birds have a magnetic sense, sensitive to the earth's magnetic field, that allows the birds to fly in the correct direction even on a cloudy night, when visual cues are not available (Hughes, 1999). Bats, which fly at night, have tiny, almost useless eyes but huge ears. They emit ultrasonic beeps and hear the echoes, which they use to detect objects (such as edible insects) and barriers as they navigate in complete darkness (Griffin, 1986).

chapter 7

This chapter and the next one are about sensation and perception. Roughly speaking, **sensation** refers to the basic processes by which sensory organs and the nervous system respond to stimuli in the environment and to the elementary psychological experiences that result from those processes (such as our experience of the bitterness of a taste, loudness of a sound, or redness of a sight). **Perception**, in contrast, refers to the more complex organizing of sensory information within the brain and to the meaningful interpretations extracted from it (such as “This is strong coffee,” “My alarm clock is ringing,” or “That object is an apple”). Thus, the study of perception is more closely tied to the study of thought and memory than is the study of sensation. The distinction is fuzzy, however, because the organizing of stimulus information in ways useful for extracting meaning actually begins during the earliest steps of taking that information in.

This chapter begins with a brief overview of basic processes involved in sensation. It continues with sections on smell, taste, pain, and hearing, and it concludes with a discussion of psychophysics, which is an approach to finding and describing reliable relationships between physical stimuli and sensory experiences of those stimuli. The next chapter deals exclusively with vision, which is by far the most thoroughly studied of the senses. Whereas research on smell, taste, and pain provides us with an opportunity to link the senses to basic drives and emotions (discussed in the preceding chapter), research on vision links more strongly to cognition—to issues of memory and thought that are discussed in the chapters that follow the next one.

Overview of Sensory Processes

1

How can the process of sensation be described as a chain of three different kinds of events?

<<< Most broadly, the process of sensation can be diagrammed as follows:

physical stimulus → physiological response → sensory experience

We have here three classes of events, each of which is entirely different from the others: (1) The *physical stimulus* is the matter or energy of the physical world that impinges on sense organs; (2) the *physiological response* is the pattern of chemical and electrical activity that occurs in sense organs, nerves, and the brain as a result of the stimulus; and (3) the *sensory experience* is the subjective, psychological sensation or perception—the taste, sound, sight, or whatever—experienced by the individual whose sense organs have been stimulated. The sensory experience generally tells us something about the physical stimulus, but it is a very different thing from the physical stimulus. We receive molecules of caffeine on our tongue and we experience a bitter taste. The bitterness is not a chemical property of the caffeine molecules; it lies only in our sensory experience triggered by the molecules. We receive electromagnetic energy of a certain wavelength and we experience the color red. The redness is not a property of the electromagnetic energy that impinges on our eye but exists only in our sensory experience.

Sensory psychologists are interested in identifying lawful relationships among the three just-described classes of events. Here, before discussing particular senses separately, let's consider some general principles that apply to all sensory systems.

Each Sensory System Has Distinct Receptors and Neural Pathways

Ever since Aristotle, people have spoken of the *five senses*, identifying them as smell, taste, touch, hearing, and vision. Actually, humans have more than five senses, and any attempt to tally them up to an exact number is arbitrary, because what one person thinks of as one sense may be thought of as two or more by another. For example, our skin is sensitive not just to touch but also to temperature and pain, neither of which is included in Aristotle's five. Other senses omitted by Aristotle have to do with body position and the body's internal environment. We

TABLE 7.1 |
Stimuli, receptors, and the pathways to the brain for various senses

Sense	Stimulus	Receptors	Pathway to the brain
Smell	Molecules dissolved in fluid on mucous membranes in the nose	Sensitive ends of olfactory neurons in the olfactory epithelium in the nose	Olfactory nerve (1st cranial nerve)
Taste	Molecules dissolved in fluid on the tongue	Taste cells in taste buds on the tongue	Portions of facial, glossopharyngeal, and vagus nerves (7th, 9th, and 10th cranial nerves)
Touch	Pressure on the skin	Sensitive ends of touch neurons in skin	Trigeminal nerve (5th cranial nerve) for touch above the neck; spinal nerves for touch elsewhere
Pain	Wide variety of potentially harmful stimuli	Sensitive ends of pain neurons in skin and other tissues	Trigeminal nerve (5th cranial nerve) for pain above the neck; spinal nerves for pain elsewhere
Hearing	Sound waves	Pressure-sensitive hair cells in cochlea of inner ear	Auditory nerve (8th cranial nerve)
Vision	Light waves	Light-sensitive rods and cones in retina of eye	Optic nerve (2nd cranial nerve)

have a sense of balance, mediated by a mechanism in the inner ear, and a sense of limb position and movement, mediated by receptors in muscles and joints. Each sense has distinct sensory receptors and neural pathways to and in the brain.

Sensory receptors are specialized structures that respond to physical stimuli by producing electrical changes that can initiate neural impulses in sensory neurons. **Sensory neurons** (described and illustrated in Chapter 5) are specialized neurons that carry information from sensory receptors into the central nervous system. For some senses the receptors are simply the sensitive ends of sensory neurons, and for others they are separate cells that form synapses upon sensory neurons. For some senses the receptors all exist in a specific, localized sensory organ, such as the ear, eye, or nose, and for others they exist in a wide variety of locations. Pain receptors, for example, exist not just in the skin but also in muscles, tendons, joints, and many other places. The stimuli, receptors, and peripheral nerves involved in the most thoroughly studied senses are listed in Table 7.1

Regardless of whether they come from one location or many, the neurons for any given sense lead to sensory-specific pathways in the central nervous system. These pathways send their messages to many different parts of the brain, including specific **sensory areas** of the cerebral cortex (see Figure 7.1). Although brain structures

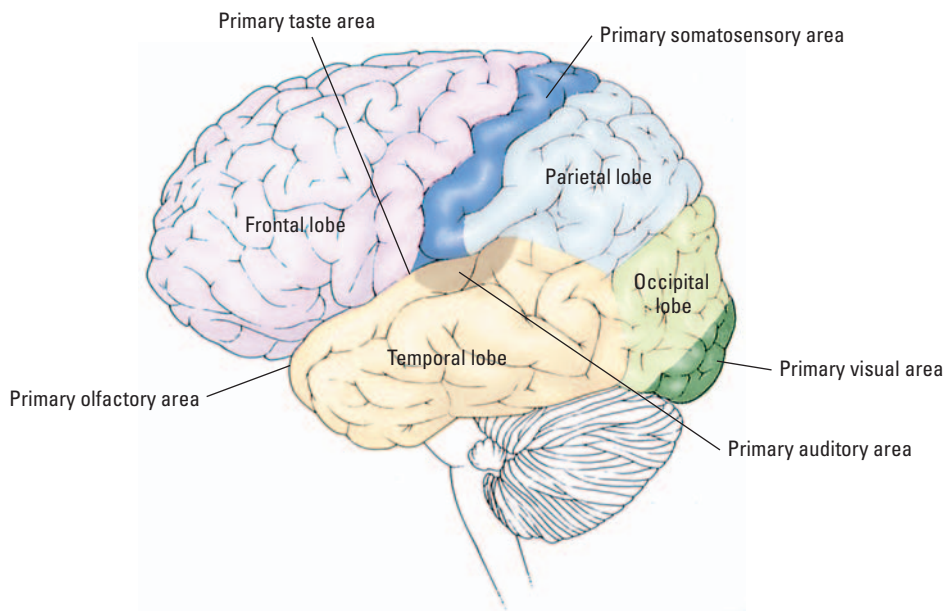


FIGURE 7.1 | Primary sensory areas of the cerebral cortex Shown here are the locations of the primary cortical areas for vision, hearing, somatosensation (which includes touch, temperature sensitivity, and pain), taste, and smell. The primary taste area overlaps with the primary somatosensory area for the tongue and mouth, and it lies buried in the fold between the parietal and temporal lobes. The primary olfactory area lies in a portion of the temporal lobe that wraps underneath the brain. Secondary sensory processing areas generally lie near the primary areas.

below the cortex can organize unconscious behavioral reactions to sensory stimuli, conscious sensory experiences depend on activity within the cerebral cortex.

Every sensation that you experience consciously is a product, ultimately, of some pattern of activity within sensory areas of your cerebral cortex. You see light because light receptors in your eyes are connected to visual areas of your cortex, and you hear sound because sound receptors in your ears are connected to auditory areas of your cortex. If we could somehow rewire those connections, sending your optic nerves to your auditory brain areas and your auditory nerves to your visual brain areas, you would hear light and see sound. When you bump the back of your head, you may “see stars,” because the bump artificially activates neurons in visual areas of your brain.

Sensory Receptors Generate Action Potentials Through a Process of Transduction

2

In general, how do physical stimuli produce action potentials in sensory neurons?

<<< The process by which a receptor cell produces an electrical change in response to physical stimulation is called **transduction**. It is the process by which receptors in the eye respond to light, receptors in the ear respond to sound, receptors on the tongue respond to chemicals dissolved there, and so on. Although the details of transduction are different for each sense, basic similarities exist across the senses. In every case, the membrane of the receptor cell becomes more permeable to certain electrically charged particles, such as sodium or potassium ions, when the appropriate type of stimulus energy acts on the receptor cell. These charged particles then flow through the membrane, either from outside the cell to inside or vice versa, and change the electrical charge across the membrane. This electrical change is called the **receptor potential**, and it is analogous to the post-synaptic potential produced on neurons by the action of synaptic transmitters (described in Chapter 5). Receptor potentials in turn trigger events that lead to the production of action potentials (also described in Chapter 5) in the axons of sensory neurons.

Sensory Systems Preserve Information About Stimulus Quantity and Quality

For senses to be useful, they must preserve—in the patterns of neural activity they produce—relevant information about the physical stimuli to which they are responding. That preservation of information is referred to as **sensory coding**.

Every form of energy can vary along at least two dimensions—a quantitative dimension and a qualitative one. The *quantitative variation* has to do with the amount or intensity of energy. A sound or light can be weak or strong; molecules stimulating taste or smell can be dilute or highly concentrated. The *qualitative variation* has to do with the precise kind of energy. Lights of different wavelengths (which we perceive as different colors) are considered to be qualitatively different, as are sounds of different frequencies (which we perceive as different pitches), as are different chemicals (which we perceive as different smells or tastes). For each of our senses, transduction occurs in such a way that information about the quantity and quality of the stimulus is preserved in the pattern of action potentials sent to the brain.

3

In general, how do sensory systems code information about the amount and kind of stimulus energy?

<<< Coding of stimulus *quantity* results from the fact that stronger stimuli produce larger receptor potentials, which in turn produce faster rates of action potentials in sensory neurons. The brain interprets a fast rate of action potentials as a strong stimulus and a slow rate as a weak stimulus. Coding of stimulus *quality* occurs because different receptors within any given sensory tissue are tuned to respond best to somewhat different forms of energy. In the eye, for example, three different kinds of receptor cells, each most sensitive to a different range of wavelengths of light, provide the basis for color vision. In the ear, different receptors are most sensitive to different sound frequencies. And in the nose and mouth, different receptors are most sensitive to different kinds of molecules. Thus, in general, qualitative variations are coded as different *ratios* of activity in sensory neurons coming from

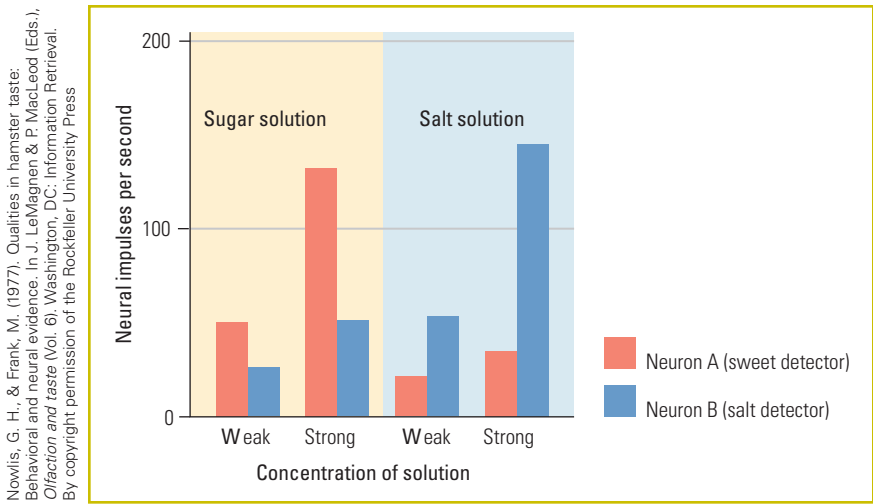


FIGURE 7.2 Quantitative and qualitative coding of taste Shown here are the rates of action potentials in two different taste sensory neurons when a weak or strong solution of sugar or salt is placed on the tongue. Each neuron responds at a faster rate to a strong solution of a given substance than to a weak one (quantitative coding); but neuron A always responds at a faster rate than neuron B when the stimulus is sugar, and the reverse is true when the stimulus is salt (qualitative coding). This illustrates the general principle that sensory quantity is coded in the overall rate of action potentials in sensory neurons and sensory quality is coded in the ratio of activity across different sets of neurons.

different sets of receptors. For an illustration of how quantitative and qualitative information about a stimulus can be independently coded, for the sense of taste, see Figure 7.2. The same principle applies to other senses, as you will discover later in this chapter.

Sensory Systems Respond to Changes More Than to Steady States

Our senses are designed to alert us to *changes* in our sensory environment and to be relatively oblivious to steady states. When you first put on your wristwatch, you feel the pressure on your skin, but later you don't. When you first enter a well-lit room from the dark, the room seems very bright, but later not so bright. When you first waded into a lake, the water may seem terribly cold, but later only slightly cool. When you first enter a chemistry lab, the odor may seem overwhelming, but later you hardly notice it. The change in sensitivity that occurs when a sensory system is either stimulated or not stimulated for a length of time is called **sensory adaptation**. In general, when the amount of stimulation of a sensory system increases for a period of time, the sensory system adapts by becoming less sensitive than it was before; and when the amount of stimulation decreases, the sensory system adapts by becoming more sensitive than it was before.

In many cases, sensory adaptation is mediated by the receptor cells themselves. If a stimulus remains for a period of time, the receptor potential and rate of action potentials are at first great, but over time they become much reduced, resulting in reduced sensation. In other cases, however, adaptation is mediated at least partly by changes farther inward, in the central nervous system. You can prove this yourself for the sense of smell (Matlin & Foley, 1997). If you place an odorous substance (such as an open bottle of nail-polish remover or cologne) on a desk in front of you, with one nostril plugged, you will adapt to it within about 5 minutes (it won't smell as strong). Then, if you unplug that nostril and quickly plug the other, you will find that you are still partly (but not completely) adapted to the odor, even though it is now acting on receptors in the other nostril, different from those that it was acting on before. Thus, adaptation for smell must result in part from changes in neurons in the brain that receive input from both nostrils.

>>> **4**
What is the value of sensory adaptation? How can you demonstrate that adaptation can occur in neurons in the brain, not just in receptors?

It's a nice place to work, but I wouldn't want to visit Can you imagine the pungent smell you would experience if you walked into this cave full of Roquefort cheese? Yet, as suggested by this worker's equable expression, one adapts after a while. In general, our senses are designed to register changes in the environment, not steady states.



SECTION REVIEW

Sensory experiences result from physiological responses to physical stimuli.		
Receptors and Pathways • Sensory receptors respond to physical stimuli with electrical changes, a process called transduction. • Those changes, called receptor potentials, can trigger neural impulses in sensory neurons. • Activity in a specific sensory area of the cortex (e.g., visual) produces a corresponding type of sensory experience (e.g., sight).	Sensory Coding • Coding preserves information about physical stimuli in the patterns of neural activity they give rise to. • Stronger stimuli produce larger receptor potentials and faster rates of firing in sensory neurons. • Coding of qualitative information depends on differential tuning of specific receptor types.	Sensory Adaptation • In general, continuous stimulation reduces the responsiveness of a sensory system and lack of stimulation increases its responsiveness. • The result of this is that senses are much more responsive to changing than to unchanging stimulation.

Smell

Smell and taste are called *chemical senses*, because the stimuli for them are chemical molecules. The chemical senses are first and foremost systems for warning and attracting. They play on our drives and emotions more than on our intellects. Think of the effects produced by a valentine gift of chocolates, perfume, or fresh roses; by the aroma and taste of your favorite meal; or by the stench of feces or rotting meat.

The human sense of smell is underappreciated. People often think of it as highly developed in dogs and other animals, but as an almost useless evolutionary vestige in people. In fact, although our sense of smell is much less sensitive than that of many other animals, it is still remarkably sensitive and useful. We can smell smoke at concentrations well below that needed to trigger even the most sensitive of household smoke detectors. We can distinguish among roughly 10,000 different chemicals by their smell (Axel, 1995). Blind people regularly identify individuals by their unique odors, and sighted people can do that too when they try. And smell contributes greatly to what we call “flavor” in foods.

Anatomy and Physiology of Smell

Great progress has been made within the past 10 years or so in understanding how the sense of smell works.

Transduction and Coding for the Sense of Smell

5 How do transduction, qualitative coding, and quantitative coding occur for the sense of smell?

<<< The basic layout of the olfactory (smell) system is illustrated in Figure 7.3. The stimuli for smell are molecules that evaporate into the air, are taken with air into the nasal cavity, and then become dissolved in the mucous fluid covering the *olfactory epithelium*, the sensory tissue for smell, which lines the top of the nasal cavity. The olfactory epithelium contains the sensitive terminals of roughly six million olfactory sensory neurons (Doty, 2001). Each terminal contains many *olfactory receptor sites*, which are large protein molecules woven into the cell membrane that are capable of binding molecules of specific odorants (odorous substances). The binding of a molecule to a receptor site changes the structure of the cell membrane, which results in an electrical change that tends to trigger action potentials in the neuron's axon. The greater the number of binding sites activated by odorous molecules, the greater is the rate of action potentials triggered in the axon.

The olfactory nerve contains roughly 400 different types of sensory neurons, each of which is characterized by a distinct type of binding site on its terminals within the olfactory epithelium (Dennis, 2004). Any given type of binding site can

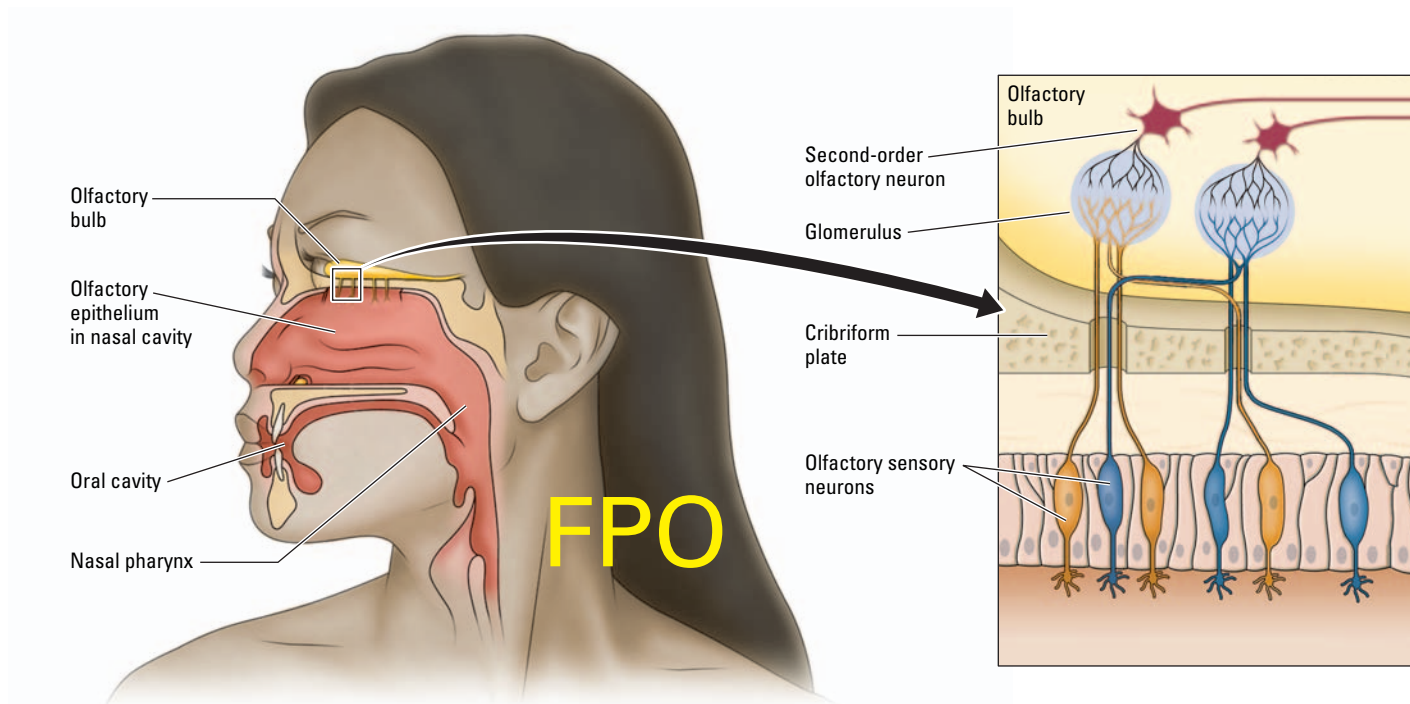


FIGURE 7.3 | The anatomy of smell Molecules of odorants enter the nose through the nostrils, become dissolved in the mucous fluid covering the olfactory epithelium, and bind to receptor sites on the sensitive tips of olfactory sensory neurons, where they initiate action potentials. The sensory neurons send their axons through the cribriform plate to form synapses on second-order olfactory neurons in the glomeruli of the olfactory bulb, directly above the nasal cavity. As illustrated in the right-hand diagram by the use of color, each glomerulus receives input from only one type of olfactory sensory neuron (defined by its type of receptor sites). Only two types of such neurons are shown here (depicted as yellow and blue), of the roughly 400 types that exist in the human olfactory system.

bind more than one odorant, but any given odorant binds more readily to some types than to others. Thus, each type of olfactory neuron differs from the other types in its degree of sensitivity to particular odorants.

The axons of the olfactory sensory neurons pass through a thin, porous bone (called the cribriform plate) into the *olfactory bulb* of the brain, where they form synapses upon other neurons in structures called *glomeruli* [glom-**air**-you-lee] (see Figure 7.3). The pattern of these connections is remarkably orderly. Each glomerulus in the olfactory bulb receives input from several thousand olfactory sensory neurons, but all these neurons are of the same type. For each of the 400 different types of olfactory sensory neurons, there is a different receiving glomerulus (or a set of two or three such glomeruli) in the olfactory bulb.

From this work, researchers have inferred the process by which qualitative and quantitative coding occurs for the sense of smell. Each odorant that we can distinguish is apparently characterized by its ability to produce a unique pattern of activity across the 400 different types of olfactory neurons and the 400 types of glomeruli of the olfactory bulb. Thus, odorant A might trigger a lot of activity in one set of glomeruli, a moderate amount in another set, and very little in others. The greater the amount of odorant A, the greater would be the total amount of activity triggered in each of the glomeruli that it affects, but the *ratio* of activity across glomeruli would remain relatively constant. Thus, the ratio indicates the type of odorant (quality of the smell), while the total amount of activity indicates the amount of odorant (quantity, or intensity of the smell).

Olfactory Brain Areas Beyond the Olfactory Bulb

The glomeruli, in the olfactory bulb, send output to various other parts of the brain. Most of this output goes to structures in the limbic system and the hypothalamus, which (as discussed in Chapters 5 and 6) are involved in basic drives and emotions. These connections, presumably, help account for the strong and often unconscious effects that smell can have on our motivational and emotional states. The connections from the olfactory bulb to the limbic system are so strong, in fact, that the limbic system was at one time often referred to as the *rhinencephalon*, which literally means “nose brain.” Output from the olfactory bulbs also goes to various portions of the cerebral cortex. The primary olfactory cortex is located in

the underside of the temporal lobe (where the temporal lobe wraps down underneath the rest of the brain). This area in turn sends output to a secondary olfactory area in the *orbitofrontal cortex*, located on the underside of the frontal lobe (Rolls, 2004). These cortical areas are crucial for the ability to experience odors consciously and to identify consciously the differences among different odors (Rolls, 2004; Buck, 2000b).

Smell as a Component of Flavor: The Mouth-to-Nose Connection

6

How do we smell foods that are already in our mouths, and what evidence indicates that smell contributes greatly to flavor?



An excellent bouquet This professional wine taster samples the wine's scent through both his nose and his mouth. Through the mouth, odorant molecules reach the nasal cavity by way of a connection called the nasal pharynx. Much of what we think of as taste is actually smell.

<<< Odorants can reach the olfactory epithelium through two different routes. The route that everyone knows about is through the nostrils. That's the route that allows you to smell smoke, roses, skunks, and other odor sources that are outside your mouth. The other route allows you to smell substances that have already entered your mouth. An opening (the nasal pharynx, shown in Figure 7.3) connects the back of the mouth cavity with the nasal cavity. The acts of chewing and swallowing push air from the mouth up into the nose—air that carries volatile molecules of whatever you are eating. What most people call taste—and what is properly called *flavor*—consists not just of true taste (from taste receptors in the mouth) but also of smell that has been triggered through this mouth-to-nose, back-door route. Remarkably, you *experience* this sensation as coming from the mouth, where the food exists, and as indistinguishable from taste, even though it actually comes from the olfactory epithelium (Bartoshuk & Beauchamp, 1994).

If you pinch your nostrils shut, you cut off both routes to the olfactory epithelium. If air can't flow out through the nostrils, it can't stream into the nasal cavity from the mouth. Experiments have shown repeatedly that people's ability to identify foods and drinks by flavor declines markedly when their nostrils are shut. You can easily demonstrate the role of smell in your own ability to identify flavors, using jellybeans as stimuli (suggested by Schiffman, 1996). With eyes closed, chew one jellybean at a time and try to identify its flavor. You will probably find this task to be relatively easy as long as you keep your nostrils open. Then try the same task with your nostrils pinched shut. You will most likely find that now all the jellybeans taste the same; all you can taste is the sugar. The differences that allow you to distinguish among cherry, grape, orange, and licorice depend on smell. Smell and taste inputs converge in a certain portion of the orbitofrontal cortex, and this area appears to be critical for the psychological experience of flavor (Rolls, 2004).

Age and Sex Differences in Olfactory Sensitivity

7

What age and sex differences in olfactory ability have been found?

<<< As we get older, our sense of smell declines. By age 65 roughly 11 percent of women and 25 percent of men have serious olfactory impairment, and by age 85 those numbers are 60 percent and 70 percent (Murphy & others, 2002). Many elderly people complain of loss in ability to taste foods, but tests typically show that their real loss is not taste but smell (Bartoshuk & Beauchamp, 1994). The most dangerous effect of loss of smell is the inability to smell smoke or toxic gases in the air. A high proportion of people who die from asphyxiation are elderly people who have lost much or all of their olfactory ability.

Researchers have long known that women are generally more sensitive to odors than are men (Doty, 2001). Recently, Pamela Dalton and her colleagues (2002) have shown that women, but not men, become increasingly sensitive to particular odors as a result of repeated exposure to those odors. These researchers tested women and men repeatedly, with certain odorous chemicals, to determine their absolute thresholds (the minimal concentrations of the chemical that they could smell). As shown in Figure 7.4, the women, but not the men, exhibited dramatic increases in

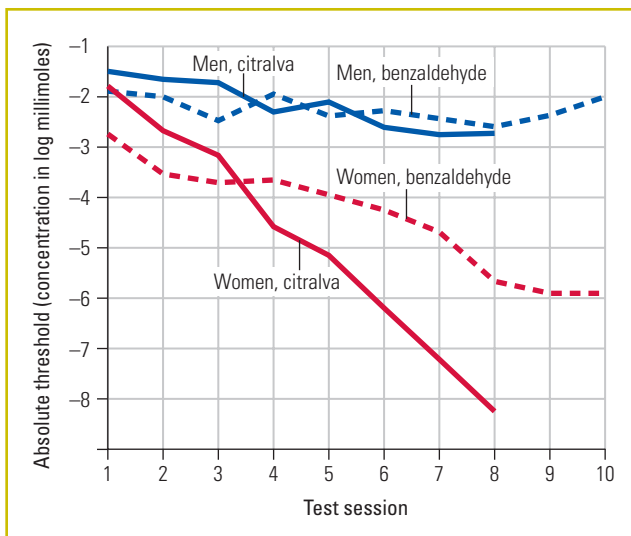


FIGURE 7.4 | A sex difference in olfactory ability When women and men were tested in repeated sessions for their ability to smell weak concentrations of benzaldehyde (an almond-cherry smell) and citralva (an orange-lemon smell), the women but not the men showed dramatic improvement. By the eighth session, the women's absolute threshold for citralva was more than five log units less than that of the men. In other words, the women were smelling this substance at concentrations less than one hundred-thousandth of the minimal concentration that the men could smell. (Data from Dalton & others, 2002.)

sensitivity (declines in absolute thresholds) with repeated testing. In further experiments, the researchers found the increased sensitivity only for women who were in their reproductive years; it did not occur for prepubescent girls or post-menopausal women. Such findings are consistent with theories that olfaction serves one or more special functions related to reproduction in women, such as choosing mates, avoiding toxins during pregnancy, or bonding with infants.

Discriminating Among Individuals by Smell

As dog owners well know, dogs greet and recognize others of their kind (and sometimes of our kind too) by smell. We humans living in a somewhat odor-phobic culture may not often admit it, but we too can identify individuals of our species by smell. In a typical experiment, one set of subjects wear initially clean T-shirts for a day without washing or using deodorants or perfumes, and then other subjects are asked to identify by smell alone which shirt had been worn by whom. Such experiments have revealed that parents can tell which of their children wore the shirt, children can tell which of their siblings wore it, and people generally can distinguish between the odors of two strangers (Porter, 1991; Wallace, 1977).

Role of Smell in Mother-Infant Bonding

Among some mammals, notably goats and sheep, odor recognition is a crucial part of the bond between mother and young (Kendrick & others, 1992). Might smell also play a role in human mother-infant bonding?

In one study conducted in a hospital maternity ward, 90 percent of mothers who had been exposed to their newborn babies for just 10 to 60 minutes after birth were able to identify by smell alone which of several undershirts had been worn by their own babies (Kaitz & others, 1987). In another study, breast-fed babies as young as 6 days old turned their heads reliably toward cotton pads that their own mothers had worn against their breasts, when given a choice between that and identical-looking pads that had been worn by other lactating women (Macfarlane, 1975). In still another study, babies who were exposed to a particular unusual odor (not that of their mothers) within the first hour after birth turned reliably toward that odor when given a choice between it and another unusual odor 6 days later (Varendi & others, 2002). All such evidence suggests that odor figures into the complex of stimuli that are involved in the attachment between human infants and their mothers. Unlike the case for goats and sheep, however, odor is certainly not essential for such attachment.

What is the evidence (a) that people can identify other individuals by smell; (b) that mothers can identify the scents of their infants very soon after birth; and (c) that infants quickly learn to identify their mother's scent?

A sweet aroma Every human being has a unique, identifiable odor.



B. Daemrich / The Image Bank

Possible Role of Smell in Choosing a Genetically Compatible Mate

9

From an evolutionary perspective, why might mice prefer to mate with others that smell most different from themselves? What evidence suggests that the same might be true of humans?

<<< In mice, odor has been shown to play a role in mating choices. Mice, like dogs and humans, can identify other individuals of their species by smell, and, remarkably, they prefer to mate with opposite-sex mice whose odor is *most different* from their own (Potts & others, 1991; Yamazaki & others, 1988). Why this preference? Researchers have found that the individual differences in odor that determine these mating preferences result from a set of about 50 highly variable genes (genes with many different alleles) referred to collectively as the *major histocompatibility complex (MHC)* (Yamazaki & others, 1994). These same genes also determine the precise nature of the cells used by the immune system to reject foreign substances and kill disease-producing bacteria and viruses. Thus, by choosing mates that smell most different from themselves, mice choose mates that (a) are not likely to be close relatives of themselves and (b) will add much new genetic variation to the mix of disease-fighting cells that develop in the offspring. As explained in Chapter 3, such genetic variety increases the chance that the offspring will survive.

The MHC also exists in human beings and contributes greatly to individual differences in odor (Penn, 2002). The advantages of mating with someone who has a very different MHC presumably exist in humans as much as in mice. Do humans, to any degree at all, prefer sexual mates who smell most different from themselves and differ most in MHC? At present, the answer to that is not known, but some research suggests that it might be yes (Penn, 2002). In one series of experiments, Claus Wedekind and his colleagues (1995, 1997) asked young men and women to rate the “pleasantness” (and in one study the “sexiness”) of the odors of T-shirts that had been worn by young adults of the opposite sex. All the subjects were assessed biochemically for differences in their MHCs. The result was that any given donor’s odor was, on average, rated as more pleasant by raters who differed from that person in MHC than by raters who were similar to that person in MHC. It remains to be seen whether this odor preference affects people’s actual mating choices, either in our culture (where we tend to hide our odors) or in any other. Research on this question has produced mixed results (Penn, 2002).

Smell as a Mode of Communication: Do Humans Produce Pheromones?

A *pheromone* [fair-uh-moan] is a chemical substance that is released by an animal and that acts on other members of the species to promote some specific behavioral or physiological response. The most dramatic examples occur in insects. For instance, sexually receptive female cabbage moths secrete a pheromone that attracts male cabbage moths from as far as several miles away (Lerner & others, 1990). Most species of mammals also produce pheromones, which serve such functions as sexual attraction, territorial marking, and regulation of hormone production (Hughes, 1999). Most species of mammals have in their nasal cavities a structure called the *vomeronasal* [voh-mair-oh-nay-zul] *organ*, which contains receptor cells specialized for responding to pheromones. Whereas the main olfactory epithelium is designed to distinguish somewhat imprecisely among many thousands of different odorants, the vomeronasal organ appears to be designed for very precise recognition of, and exquisite sensitivity to, a small number of specific substances, the species’ pheromones (Buck, 2000a).

10

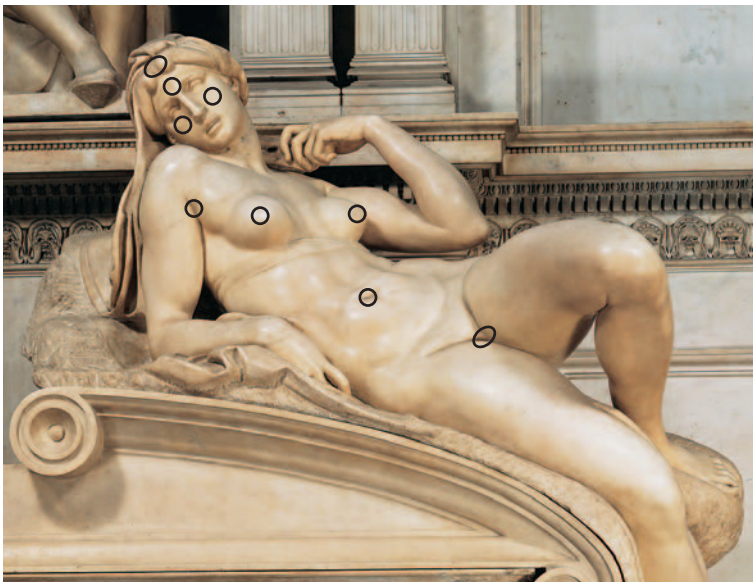
What anatomical characteristics of our species are consistent with the possibility that we produce and respond to pheromones? What observations and reasoning suggest that we do not produce sex-attractant pheromones?

<<< Do humans communicate by pheromones? We do have the structures that would make such communication possible. Like other mammals, we have specialized glands in the skin that secrete odorous substances. Such glands are especially concentrated in areas of the body where our species has retained hair—such as in the axillary region (armpits) and genital region (see Figure 7.5). One theory is that the function of hair in these locations is to hold on to the secretions and to provide a large surface area from which they can evaporate, so as to increase their effec-

tiveness as odorants (Stoddart, 1990). Some of the substances secreted by these glands are steroid molecules that resemble substances known to serve as pheromones in other mammals. We also have at least a rudimentary vomeronasal organ, but the evidence to date is inconclusive as to whether it functions in our species or is entirely vestigial (Hays, 2003).

Motivated partly by the perfume and cologne industry, most human pheromone research has centered on whether or not we produce sex-attractant pheromones. In many experiments, men and women have been exposed to various secretions taken from the other sex and have rated the attractiveness of the odor or changes in their own mood. To date, despite the often lurid claims in ads on the Internet, such experiments have failed to yield convincing evidence that such pheromones exist (Hays, 2003; McClintock, 2000). Certainly some people find some of the odorous substances secreted by other people to be pleasant, but individual differences are great and no specific human secretion has been found to be consistently attractive to members of the opposite sex. Perhaps that should not be surprising. Sex-attractant pheromones are valuable for animals that mate only at certain times of the year or only when the female is ovulating, as a means of synchronizing the sex drives of males and females to maximize the chance of conception. As discussed in Chapter 6, humans have taken a different evolutionary route, such that sexual drive and behavior are not tied to a season, cycle, or variable physiological state. For that reason, perhaps, there is little or no need for us to advertise by scent our readiness to mate.

Other research has suggested that women who live together tend to have synchronized menstrual cycles and that this effect results from pheromonal communication (McClintock, 1971, 2000). However, attempts to replicate these results have met with mixed success and critics have argued that the positive findings may derive from faulty methods of collecting and analyzing the menstrual-cycle reports (Hays, 2003; Schank, 2001). In sum, although researchers have long been searching for human pheromones, the evidence on whether or not they exist is mixed.



Scala / Art Resource

FIGURE 7.5 | Locations of maximal scent production by humans In humans, specialized scent-producing glands (apocrine glands) are concentrated most highly in the axillary region (underarms) and also exist in high concentrations in the genital area, the area around the nipples, the navel area, on the top of the head, and on the forehead and cheeks (Stoddart, 1990), as shown by the added circles. (The statue here is Michelangelo's *Aurora*, from the tomb of Lorenzo de Medici, in Florence.)

SECTION REVIEW

Smell, one of two human chemical senses, serves functions of warning and attraction.

Basic Facts of Smell • Roughly 400 different types of sensory neurons exist in olfactory nerves. The receptive ends of each type are most responsive to particular types of odorant molecules that reach the olfactory epithelium in the nose. • These different types of neurons are connected in an orderly way to corresponding parts of the olfactory bulb. Their differential responsiveness provides the basis for qualitative coding (ability to tell one scent from another). • Olfactory responses to odorants coming from the mouth contribute greatly to the flavor of foods, experienced as taste. • Olfactory sensitivity is commonly greater in women than in men, and, in both sexes, it declines with age.	Telling People Apart by Their Smell • Humans can identify individuals by smell alone. • Mothers and their infants quickly learn to recognize the other's smell, a fact that may contribute to bonding. • Odor allows mice (and perhaps humans) to choose mates that will enhance genetic variety in their offspring.	Communicating via Smell • For many animal species, including most mammals, pheromones serve as chemical messengers between individuals. • Though the evidence is mixed regarding the possibility of human pheromones, we do not need or appear to have sex-attractant pheromones.
--	--	---

Taste

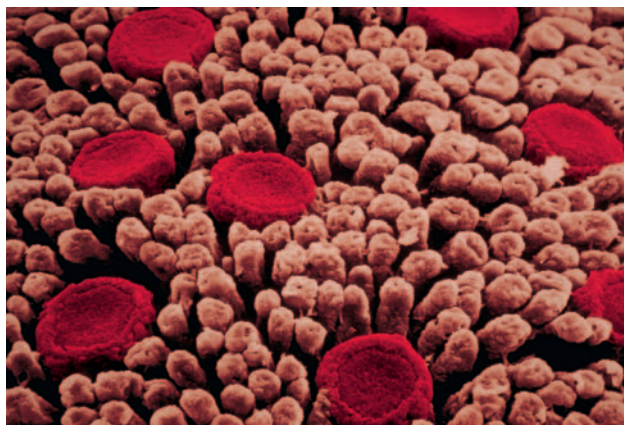
Some insects have taste receptors on their feet, which allow them to taste what they are walking on. Fish have taste receptors not just in their mouths but all over their bodies (Hara, 1994). They can taste their prey before they see it, and they use taste to help them track it down. For us and other mammals, taste has a more limited but still valuable function. Our taste receptors exist only in our mouths, and their function is to help us decide whether a particular substance is good or bad to eat.

Anatomy and Physiology of Taste

11 How, in general, does transduction occur in taste?

<<< The receptors for taste are found on specialized *taste receptor cells*, not directly on the sensory neurons (unlike the case for smell). These cells exist in spherical structures called *taste buds*. Each bud contains between 50 and 150 receptor cells, arranged something like segments in an orange (see Figure 7.6). Most people have between 2000 and 10,000 taste buds, about two-thirds of which are on the tongue and the rest of which are on the roof of the mouth and in the opening of the throat (Herness & Gilbertson, 1999). People who have more taste buds are typically more sensitive to tastes—especially to bitter tastes—than are people who have fewer taste buds (Bartoshuk & Beachamp, 1994).

To be tasted, a chemical substance must first dissolve in saliva and come into contact with the sensitive ends of appropriate taste receptor cells, where it triggers electrical changes that result in action potentials first in the taste receptor cells and then, by synaptic transmission, in sensory neurons that run to the brain (see Figure 7.6). The specifics of the transduction mechanism vary from one type of taste receptor cell to another.



Surface of the tongue The red disklike structures are fungiform papillae. Most taste buds are located within the walls of these papillae with their sensitive ends oriented toward the surrounding saliva-filled trough. The magnification is about 100X.

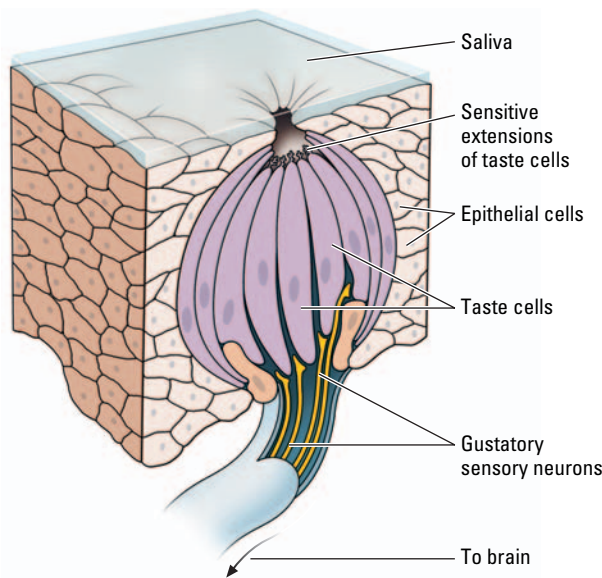


FIGURE 7.6 | A taste bud Taste buds are found in the surface tissue (epithelium) of the tongue and elsewhere in the mouth. Each one contains 50 or more taste receptor cells. From the tip of each cell, hairlike extensions make contact with the fluid lining the epithelial tissue. These extensions contain binding sites and channels where substances to be tasted exert their effects. The receptor cells produce action potentials in response to such input and, in turn, induce action potentials in taste sensory neurons through synaptic transmission (Herness & Gilbertson, 1999).

Five Primary Tastes, Five Types of Receptor Cells

For many years Western scientists believed that in humans (and other mammals) taste receptor cells were of just four types—*sweet*, *salt*, *sour*, and *bitter*—each named after the taste sensation that results when that type is activated. Every taste, it was believed, could be understood as a combination of those four primary tastes. Japanese scientists, in contrast, generally spoke of *five* primary tastes—the four just mentioned plus *umami*, which, loosely translated, means “savory” or “delicious” (Kurihara & Kashiwayanagi, 1998). Umami, they held, is not experienced subjectively as similar to any combination of the other primaries and is a major contributor to the taste of many natural foods, especially those that are high in protein, such as meats, fish, and cheese. Umami is also the taste produced by monosodium glutamate (MSG), a substance frequently used as a flavor enhancer in Asian cuisine. Taste researchers have recently identified distinct receptor cells and brain areas responsive to MSG, and now even Western taste specialists generally write of five rather than four primary tastes and types of taste receptor cells (Chaudhari & others, 2000; Rolls, 2004).



12 What are the five primary tastes, and how, in general, does transduction occur in taste receptor cells?

Taste Areas in the Brain

Taste sensory neurons have strong connections (by way of several way stations) to the limbic system and cerebral cortex. The connections to the primary taste area of the cortex (located in the rear part of the underside of the frontal lobe) are arranged in such a way that different sets of neurons are selectively responsive to each of the five basic categories of taste stimuli (Rolls, 2004). People with extensive damage to this area lose their conscious experience of taste (Pritchard, 1991), and artificial stimulation here produces experiences of taste in the absence of any stimuli on the tongue (Penfield & Faulk, 1955). The primary taste area in turn sends connections to several other areas of the cortex, including the orbitofrontal cortex, where, as mentioned in the section on smell, neural connections for taste and smell intermingle and enable us to experience the mixed taste-smell sensation called flavor.

Hmm, umami The savory umami taste of Chinese food is often enhanced by monosodium glutamate. Umami is believed to be one of the five primary tastes.

An Evolutionary Account of Taste Quality

The purpose of taste is to motivate us to eat some substances and avoid eating others. Generally speaking, salty, sweet, and umami are pleasant tastes, which, in the course of evolution, became attached to substances that are good for us (or at least were good for our evolutionary ancestors). A certain amount of salt intake is needed to maintain a proper salt balance in bodily fluids. Sugars, in fruits and other natural foodstuffs, were a valuable source of energy to our evolutionary ancestors. Protein (a source of umami flavor) is essential for building and restoring tissues. Our ability to taste and enjoy salt, sugars, and proteins is part of the mechanism that helps ensure that we consume adequate amounts of these nutrients.

Still speaking generally, sour and bitter are unpleasant experiences, which natural selection has attached to certain substances that are bad for us. Bacterial decay produces acidic compounds. Since decaying substances can cause disease when eaten, natural selection produced a taste system that experiences most acids as unpleasant (sour). Many plants and some animals, as part of their own evolution, have concentrated toxic substances into their tissues—substances that can harm or kill animals that eat them. As a consequence, our ancestors evolved a taste system that experiences most of those toxic substances as unpleasant (bitter).

13 From an evolutionary perspective, (a) what is the function of each of the primary tastes, (b) why do so many chemically diverse substances taste bitter, and (c) why does bitter sensation increase in women at pregnancy?



Michael Newman / PhotoEdit

Yuck, bitter This response may protect children from eating poisonous substances. The mammalian taste system evolved to sense many poisons as bitter. Among humans, young children are especially sensitive to, and rejecting of, bitter taste.

Evolution of the Ability to Taste Toxic Substances as Bitter

To pursue this evolutionary line of thinking, let's consider further the relationship between bitter taste and poisons. A wide variety of chemical substances taste bitter to us. They taste bitter because each of them is able to bind to one or another of the 50 to 80 different types of receptor sites located on bitter receptor cells. The receptor sites are specialized protein molecules that bind certain molecules that come into contact with them. Each bitter receptor cell apparently contains all of these types of receptor sites and is therefore capable of responding to the whole set of substances that we taste as bitter (Adler & others, 2000; Dulac, 2000). The binding of any substance to these sites triggers a chemical change in the cell. This chemical change results in action potentials in the sensory neurons and, ultimately, activity in areas of the brain that produce the bitter sensation.

From an evolutionary perspective it is no accident that there are so many different types of receptor sites on bitter receptor cells and that most of the chemicals that they bind are either poisonous to us or chemically similar to poisons. Imagine an early species of animal that ate plants, some of which contain substance A, a poison that kills by interfering with the animal's respiratory system. Any mutations affecting the taste system of an individual of this species in such a way as to make it dislike the taste of A would promote that individual's survival and reproduction, so those new A-rejecting genes would multiply from generation to generation.

Individuals having the new genes would experience substance A as having a negative, bitter taste and would avoid it; but they would continue to eat plants that contain another poison, B, which kills animals by paralyzing the muscular system. Because B is an entirely different compound from A, it does not attach to the A receptors on the taste cells that trigger bitterness, so the animals have no sensory grounds for avoiding it. Now imagine a single mutation, which results in a new variety of receptor protein on the surface of bitter taste cells, a protein that can bind molecules of B. The animal with this mutation would experience the same bitter sensation upon eating leaves containing B as it experiences when eating leaves containing A, and it would avoid both of them. It would be more likely to survive and send progeny into the next generation than would other members of the species, and many of those progeny would inherit the new gene.

Extending this scenario, you can imagine how natural selection might have produced modern mammals that have bitter taste cells containing 50 to 80 different types of receptor proteins, able collectively to bind molecules of most of the toxic substances found in plants and other potential foods. We can't go back and prove with certainty that this is how the present state of affairs relating to bitterness came about, but this account helps make sense of the facts of bitter taste.

Possible Explanation of Sex and Age Differences in Bitter Sensitivity

Although avoidance of bitter-tasting foods is generally adaptive, too much avoidance of them is not. Plants protect themselves from being eaten not just by concentrating truly poisonous substances but also by concentrating nonpoisonous substances that are chemically similar to poisons and that therefore bind to receptors on bitter taste cells and produce a bitter taste. Individuals that avoided all bitter tastes would lose the nutritional value of plants that taste bitter but are safe to eat. Through observation and experience, people and plant-eating animals can learn to eat and enjoy bitter foods that have no toxins or low levels of toxins.

Among humans, women are generally more sensitive to bitter taste than are men, and women become still more sensitive to it during the first 3 months of pregnancy (Duffy & Bartoshuk, 1996). A possible reason is found in evidence that the human fetus is highly subject to damage by poisons, especially during the first 3 months of its development (Profet, 1992). With the stakes being survival of offspring, the value of avoiding even mild toxins increases. Also, in general, young children are more sensitive to bitter tastes than are adults, which helps explain

why it is so difficult to get children to eat nutritious but bitter-tasting greens, such as spinach or Brussels sprouts. Children's extra sensitivity to bitter taste may help protect them from eating poisonous foods during their early development, before they have learned which materials are safe to eat and which are not.

SECTION REVIEW

Taste is a chemical sense that helps us decide what is good or bad to eat.

Basic Facts of Taste

- Human taste receptors are specialized cells contained in taste buds on the tongue and other parts of the mouth and throat.
- Five types of taste receptors have been identified, which correspond to five primary tastes—sweet, salty, sour, bitter, and umami.
- Taste receptors can trigger neural impulses in taste sensory neurons, which send input to the primary taste area in the frontal lobe and to other parts of the brain.

Evolution and Taste Quality

- Through evolution, we have generally come to experience beneficial nutrients as pleasant in taste (sweet, salty, or umami) and harmful substances as unpleasant (sour or bitter).
- Bitter receptor cells evolved to respond to many chemically varied substances that are poisonous or chemically similar to poisons.
- Natural selection may explain why children and women, especially pregnant women, are particularly sensitive to bitter tastes.

Pain

Pain is one of the somatosenses (introduced in Chapter 5). That is, like touch, temperature sensitivity, and proprioception (the sense of body position), pain is a sense that can originate from multiple places throughout the body rather than just from specialized sensory organs in the head. (Recall that *soma* means “body.”) Pain receptors exist over the whole surface of the skin and in many other bodily tissues.

Pain is a “body” sense in another way, too. When you see, hear, smell, taste, or touch something, you experience the sensation as coming from outside yourself (from the thing you are seeing, hearing, smelling, tasting, or touching); but when you feel pain, you experience it as coming from your own body. If you cut yourself with a knife, your feeling of pain is a sense not of the knife (that comes from your vision and touch) but of your own injured bodily state. Pain is not only a sense but also an emotion and a drive (Craig, 2003). As an emotion, strong pain overwhelms a person’s conscious mind, making it hard to think about anything except the pain; and, like many other emotions, pain has its own well-recognized facial expression (see Figure 7.7) (Williams, 2002). As a drive, pain motivates a person both to reduce the pain and to avoid future behaviors like the one that produced it (such as careless handling of knives). In psychology, pain is by far the most thoroughly studied of the somatosenses (Craig & Rollman, 1999). That, no doubt, is largely due to the dramatic ways that pain can affect, and be affected by, a person’s other psychological experiences.

The evolutionary value of pain—its role in promoting survival—is dramatically illustrated in those rare, unlucky people who are born with a genetic disorder that makes them insensitive to it (Brand & Yancey, 1993). They can experience all other sensations, including touch and temperature, and they can report increasing intensities of those feelings, but pain, with its warning and motivating qualities, is missing. Children with this disorder are not motivated to remove their hands from a hot stove, or to refrain from chewing on their tongues as they eat, or to change their body positions (as most of us do from minute to minute) to relieve the strain on muscles and joints. Even if they are constantly watched throughout childhood, and even if they learn intellectually to avoid certain activities, people with this disorder usually die young from the tissue deterioration and infections that result from their wounds.

>>> 14 In what ways is pain a “body” sense, an emotion, and a drive? How does observation of people born without pain sensitivity illustrate pain’s value?

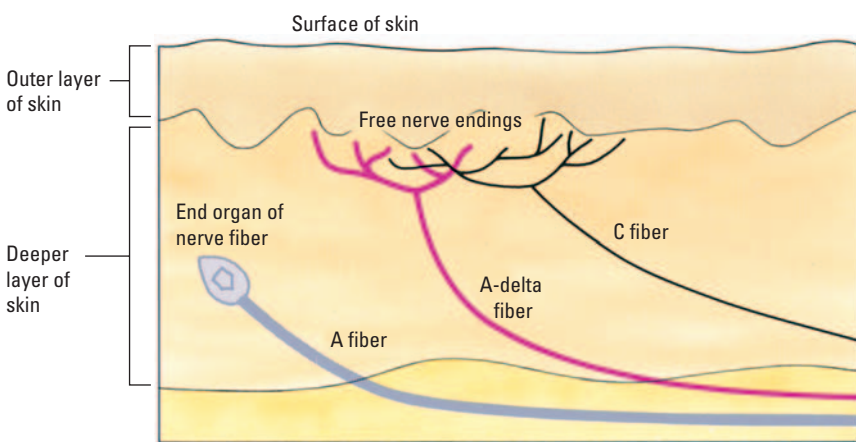


FIGURE 7.7 | The face of pain People everywhere, including newborn infants, express pain in a similar way. The full expression includes a lowered brow, closed or partly closed eyes, tightened eyelids, raised cheeks, wrinkled nose, and raised upper lip (Williams, 2002).

Neural Pathways for Pain

Anatomically, pain is closely related to the other somatosenses, such as touch and temperature sensitivity. For all these senses, the receptor cells are the sensory neurons themselves. These neurons have receptive endings in the skin and long axons that enter the central nervous system. Pain neurons are thinner than other neurons from the skin, and their sensitive terminals, called **free nerve endings**, are not encased in special capsules, or end organs, as are the endings of touch and temperature receptors (see Figure 7.8). Free nerve endings can be found in all body tissues from which pain is sensed (Lynn & Perl, 1996)—not just the skin but also the pulp of the teeth (from which comes the dreaded toothache), muscles (giving us the pain of cramps and muscle aches), membranes around bones and joints (from which we experience arthritis), and various visceral organs (giving us stomachaches and other inner pains).

FIGURE 7.8 | Pain receptors in the skin The pain receptors are the sensitive endings of sensory neurons, called free nerve endings. The slower second wave of pain is carried by the very thin C fibers, and the faster first wave is carried by the thicker A-delta fibers. The sense of touch is carried by still thicker (and faster) A fibers, whose endings are not “free” but, rather, are surrounded by a capsule, or end organ, that modifies the pressure stimulus.



Sensory Neurons for Two Waves of Pain

15 What is the anatomical basis for a distinction between first and second pain?

<<< Pain sensory neurons are of two general types—very thin, unmyelinated, slow-conducting neurons called *C fibers* and slightly thicker, myelinated, faster-conducting neurons called *A-delta fibers* (again see Figure 7.8). Some A-delta fibers are specialized to respond to strong pressure (such as from a pinprick), other A-delta fibers are specialized to respond to extremes of temperature (hot or cold), and C fibers respond to all sorts of stimuli that produce pain, including strong pressure, intense heat or cold, and chemicals that produce pain when applied to the skin (Basbaum & Jessell, 2000). When your skin is pricked or burned, you feel two separate waves of pain: a sharp, highly localized *first pain*, followed (1 or 2 seconds later) by a dull, burning, more diffuse, longer-lasting *second pain*. The first is mediated by A-delta fibers, and the second by the slower C fibers (Basbaum & Jessell, 2000). The C fibers also respond in a more prolonged way to a variety of chemicals that are released by damaged or infected cells, accounting for the persistent pain that accompanies burns, wounds, and infections.

Pain neurons enter the spinal cord (via a spinal nerve) or the brainstem (via the fifth cranial nerve) and terminate there on interneurons. Some of these interneurons promote reflexive responses—such as the automatic withdrawal of the hand from a hot stove—independent of conscious experience. Others send their axons to the thalamus, in the center of the brain, which, in turn, sends output to portions of the brain that are involved in the conscious experience of pain.

Brain Areas for Three Components of Pain Experience

16 What are three different components of pain experience, and what evidence links these to three different portions of the brain?

<<< Pain as a psychological experience can be divided meaningfully into three different components, each of which depends most critically on a different portion of the brain (see Figure 7.9):

1. The *sensory* component of pain depends largely on the *somatosensory cortex*, the area of the parietal lobe that receives input for touch and temperature as well as pain (for its location, look back at Figure 7.1, page 227). This area appears to be crucial for the ability to perceive pain as a sensation, to describe its intensity and qualities (as sharp or dull, for example), and to locate it in a particular portion of the body.
2. The immediate, *primary emotional and motivational* component of pain depends on portions of the limbic system referred to as the *cingulate cortex* and the *insular cortex*, which lie buried in the folds of the brain's frontal lobe. People with damage to these brain regions experience a condition called *asymbolia for pain* (Foltz & White, 1968; Price, 2000). They can perceive a painful stimulus and describe it as such, identify the location of the pain, describe its qualities, and rate its intensity; but they do not feel a normal desire to escape the pain. The pain doesn't bother them.
3. A third component of pain, in humans, is the more cognitively based *secondary emotional and motivational* component—the suffering that derives from the person's worrying about the future or about the meaning of the pain. The brain area that is crucial for this troubled state is the *prefrontal lobe*, the front-most portion of the neocortex, which is involved in all aspects of planning and concern for the future (as noted in Chapter 5). People with damage here feel and respond to the immediate threat and unpleasantness of pain, but they do not worry about it, just as they do not worry about or make plans based on other experiences (Price, 2000).

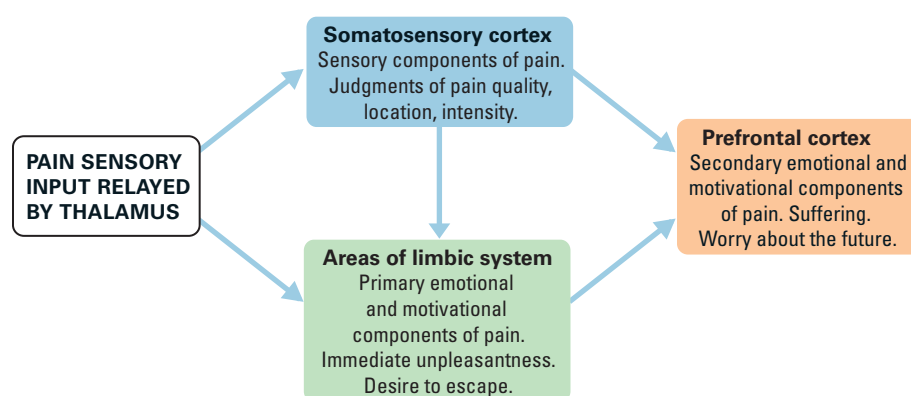


FIGURE 7.9 Brain areas involved in three components of pain experience. Pain input from tracts through the spinal cord and midbrain is relayed by one portion of the thalamus to the somatosensory cortex and by another portion to the insular cortex and certain portions of the limbic system. These areas, in turn, send output to the prefrontal cortex. These different areas account for three relatively distinct components of pain experience, described in the figure. (Based on information reviewed by Price, 2000.)

The experience of pain, with all three of the above components, does not always originate from stimulation of pain receptors. This fact is all too well known by many people who have had a limb amputated. Such people often feel as if the missing limb were still present and full of pain. Such *phantom-limb pain* can persist even if all the nerves from the limb's stump have been destroyed and even if the pain pathways entering the brain from the spinal cord have been destroyed (Melzack, 1992). Apparently, in such cases the brain's mechanism for experiencing pain and assigning that experience to a particular body location can be activated without sensory input from that part of the body. In fact, the *lack* of sensory input might trigger phantom-limb pain by removing a source of inhibition to the pain mechanisms of the brain.

The Modulation of Pain

The experience of pain depends not just on the physical stimulus or its damaging effects but also on other conditions that exist at the time the stimulus or damage occurs. The same degree of wound may at one time feel excruciatingly painful and

at another time be barely detected. Over 40 years ago, Ronald Melzack and Patrick Wall (1965, 1996) proposed a **gate-control theory** of pain, aimed at explaining such variability. In essence, the theory holds that the experience of pain depends on the degree to which input from pain sensory neurons can pass through a neural “gate” and reach higher pain centers in the brain. Conditions can increase or decrease pain by opening or closing the gate. Much research has confirmed this general theory and has added many details to it.

The major gate, where pain input is most strongly enhanced or inhibited, is at the first way station in the central nervous system. Pain sensory neurons enter the spinal cord and brainstem and terminate there on second-order pain neurons that send their axons upward. These second-order neurons are highly variable in their responsiveness. Their degree of responsiveness is controlled in part by pain-enhancing and pain-inhibiting neurons coming down from the brain. Pain-enhancing neurons tend to open the gate for pain, and pain-inhibiting neurons tend to close the gate, as illustrated in Figure 7.10.

Mechanisms of Pain Enhancement

17

How does illness produce a general increase in pain sensitivity, and how does injury produce a localized increase in pain sensitivity?

<<< As your own experience may have shown you, a general enhancement of pain sensitivity all over the body accompanies illness, especially if high fever is involved. This effect may have evolved to motivate ill individuals to rest rather than move around, in order to conserve energy needed to fight the disease (Kluger, 1991). Although the details are not known, this illness-induced effect is believed to occur through an action of the immune system on pain-enhancing neurons in the brain (again see Figure 7.10) (Watkins & Maier, 2000).

Pain sensitivity can also be increased in specific locations of the body as a result of injury at those locations. This heightened sensitivity occurs partly because of changes in the free nerve endings of C fibers and A-delta fibers that are induced by chemicals released from damaged cells (Basbaum & Jessell, 2000). The sensitized sensory neurons respond to much weaker stimuli than they would have before the injury. In addition, second-order pain neurons in the spinal cord and brainstem become sensitized by intense activation, such that they become more responsive to subsequent input for periods ranging from minutes to weeks (Ji & others, 2003; Woolf & Salter, 2000). The result is that even light touch to a recently burned or wounded area of skin can be intensely painful. Such pain-enhancing systems presumably evolved as a means of motivating individuals to protect damaged areas of their bodies.

Neural and Chemical Mechanisms of Pain Reduction

18

How can pain input be inhibited at its entry into the central nervous system, and how might endorphins be involved in this process?

<<< A major neural center for pain inhibition exists in a portion of the midbrain called the *periaqueductal gray* (abbreviated *PAG*). Neurons in this area send their axons down into the lower brainstem and spinal cord to inhibit pain input there, as illustrated in Figure 7.10. Electrical stimulation of the PAG has a powerful analgesic (pain-reducing) effect—so powerful, in fact, that abdominal surgery can be performed without drugs in animals that are receiving such stimulation (Reynolds, 1969; Mason, 2001). Electrical stimulation of this area has also, in humans, successfully reduced or abolished chronic pain that could not be relieved by other means (Hosobuchi & others, 1979).

Morphine and other *opiate drugs* (derivatives of opium) exert their well-known analgesic effects partly through direct action in the PAG. Morphine that passes into the brain is taken up at special binding sites on neurons in the PAG, where it increases neural activity and thereby reduces pain (Basbaum & Fields, 1984). Morphine binding sites are also found on neurons in the spinal cord, and injection of a tiny amount of morphine directly into the spinal cord can greatly reduce or eliminate pain in the part of the body that sends its sensory neurons into that spinal cord area (Basbaum & Jessell, 2000).

Of course, the pain-inhibiting system did not evolve to respond specifically to morphine or other foreign substances. Its basic function is to mediate the body's

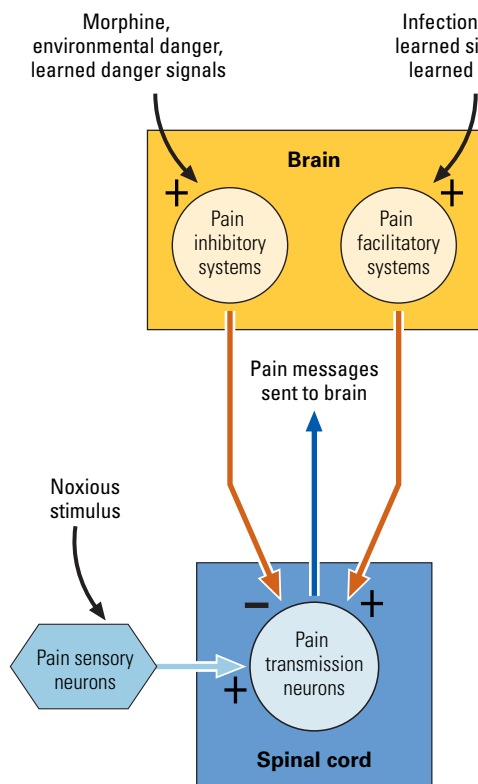


FIGURE 7.10 | Gate control of pain Pain transmission neurons in the spinal cord, which receive input from pain sensory neurons coming from the body, can be made more or less excitable by descending connections from the brain. In the figure, + indicates excitatory connections and – indicates inhibitory connections. Although not shown in the drawing, the pain inhibitory and excitatory systems can also produce their effects by inhibiting or facilitating the release of neurotransmitter from the sensory neurons. (Modified from Watkins & Maier, 2000, p. 31.)

own capacity to reduce pain. The body produces a variety of chemicals that act like morphine and are collectively referred to as **endorphins**. As was noted in Chapter 6, the term *endorphin* is short for *endogenous morphine-like substance* (*endogenous* means “created within the body”). Some endorphins are produced in the brain or spinal cord and serve as neurotransmitters or neuromodulators to alter the activity or excitability of neurons to which they bind. Others are secreted from the pituitary and adrenal glands, as hormones, into the blood and have a variety of effects both peripherally and in the central nervous system (Henry, 1986). Endorphins are believed to inhibit pain by acting both in the PAG and at the places (illustrated in Figure 7.10) where pain-carrying neurons enter the spinal cord and lower brainstem.

Stress-Induced Analgesia

During his search for the source of the Nile River, the famous explorer David Livingston was attacked by a lion. He survived the incident and later wrote that although the lion shook him “as a terrier does a rat” and crushed his shoulder, he had felt “no sense of pain nor feeling of terror, though quite conscious of all that was happening” (Livingston, 1857). Other people have had similar experiences. For example, soldiers severely wounded in battle often do not notice their wounds until the battle is over. We are apparently endowed with a mechanism that prevents us from feeling pain at times when, for survival purposes, it is best to ignore our wounds. A person or non-human animal faced with a predator or similar threat cannot afford to nurse a wound or favor it by limping; all the body’s resources must be used to fight or flee. The decreased pain sensitivity that accompanies highly stressful situations is referred to as **stress-induced analgesia**.

Many experiments have shown that stress-induced analgesia depends partly, if not entirely, on the release of endorphins. In one experiment, rats that were subjected to a series of electric shocks to their feet (the source of stress) became relatively insensitive to pain for several minutes afterward, as indicated by their lack of response to normally painful heat applied to their tails (Lewis & others, 1980). Rats treated with a drug that blocks the action of endorphins did not show this

19 What is some evidence that stress-induced analgesia is at least partly mediated by endorphins?

stress-induced analgesia, indicating that the effect must have been mediated by endorphins. In similar experiments, the mere presence of a cat (Lichtman & Fanselow, 1990) produced analgesia in rats; the presence of biting flies produced analgesia in mice (Kavaliers & others, 2001); a stressful math test produced analgesia in students (Bandura & others, 1988); and films depicting combat produced analgesia in veterans who had experienced the trauma of war (Pitman & others, 1990). In all of these cases injection of a drug that blocks the actions of endorphins prevented the analgesic effect from occurring.

Endorphins are also secreted during periods of prolonged, strenuous physical exertion, such as long-distance running, and may account for the pain reduction and euphoric “runner’s high” that many people enjoy during and after such exertion. In one experiment both the reduced pain and the sense of euphoria failed to occur in runners who had been treated with an endorphin-blocking drug (Janal & others, 1984).

Belief-Induced Analgesia

20 What is some evidence that pain can be reduced by belief?

<<< In humans, dramatic reduction in pain can also, at times, be produced by the power of *belief* or faith. Some religious groups engage in practices that most of us would regard as torture, yet the participants appear to feel no pain. One group in India, for example, practices a hook-hanging ritual. A man who has been chosen to represent the power of the gods is secured to a rope by two steel hooks that pierce the skin and muscles on his back. He hangs from this rope, swinging back and forth, while he blesses the children and the crops of the village. He feels honored to have been chosen and apparently feels little or no pain (Melzack & Wall, 1996).

A less dramatic example, in cultures where faith is more often placed in science and medicine, is the *placebo effect* on pain. In many cases a pill or injection that contains no active substance (the placebo) can reduce pain in a person who believes that the drug is a painkiller. A number of experiments have shown that placebo-induced pain reduction depends partly or entirely on the secretion of endorphins (ter Riet & others, 1998). In one of the first such experiments, people who had undergone a tooth extraction reported less pain if given a placebo than if not, and this placebo effect did not occur in subjects who were treated with an endorphin-blocking drug (Levine & others, 1979). Other experiments have shown that various cognitive techniques for relieving pain, such as meditating on the idea that the pain is disconnected from the rest of the body, also work at least partly through endorphins (Bandura & others, 1987). Might the man hanging from hooks in India also be secreting large quantities of endorphins? Nobody knows for sure, but most pain researchers would bet that he is.

SECTION REVIEW

Pain, which is not only a somatosense but also an emotion and a drive, has survival value.

Basic Facts of Pain

- Free nerve endings of pain sensory neurons are the receptors for pain and are located in many parts of the body.
- C fibers and A-delta fibers, two types of pain sensory neurons, mediate two different waves of pain.
- The experience of pain has three identifiable components—sensory, primary emotional and motivational, and secondary emotional and motivational—each relying on different areas of the brain.

Pain Modulation

- The gate-control theory of pain maintains that the degree of pain felt depends on how much input from pain neurons passes through a neural “gate” to higher areas of the brain.
- The enhanced pain sensitivity that accompanies illness or injury helps protect the body from further harm.
- The PAG and endorphins provide the body with a means of pain inhibition.
- Endorphins play a part in stress-induced and belief-induced analgesia.

Hearing

Among mammals, the greatest listeners are the bats. In complete darkness, and even if blinded, bats can flit around obstacles such as tree branches and capture insects that themselves are flying in erratic patterns to escape. Bats navigate and hunt by *sonar*, that is, by reflected sound waves. They send out high-pitched chirps, above the frequency range that we can hear, and analyze the echoes in a way that allows them to *hear* the scene in front of them. From echoes, they can hear such characteristics as the size, shape, position, direction of movement, and texture of a target insect (Feng & Ratnam, 2000).

Hearing may not be as well developed or as essential for us as it is for bats, but it is still enormously effective and useful. It allows us to detect and respond to potentially life-threatening or life-enhancing events that occur in the dark, or behind our backs, or behind occluding objects. We use it to identify animals and events such as upcoming storms. And, perhaps most important, it is the primary sensory modality of human language. We learn from each other through our ears.

Sound and Its Transduction by the Ear

If a tree falls in the forest and no one is there to hear it, does it make a sound? This old riddle plays on the fact that the term *sound* refers both to a type of physical stimulus and to the sensation produced by that stimulus.

Sound as a Physical Stimulus

As a physical stimulus, sound is the vibration of air or some other medium produced by an object such as a tuning fork, one's vocal cords, or a falling tree. The vibration moves outward from the sound source in a manner that can be described as a wave (see Figure 7.11). The height of the wave indicates the total pressure exerted by the molecules of air (or another medium) as they move back and forth, which is referred to as the sound's **amplitude**, or intensity, and corresponds to what we hear as the sound's **loudness**. Sound amplitude is usually measured in logarithmic units



John Mitchell / Photo Researchers

An auditory animal Bats, which navigate and hunt by sonar, have large, mobile outer ears. This is especially true of some species, such as *Plecotus auritus*, pictured here.

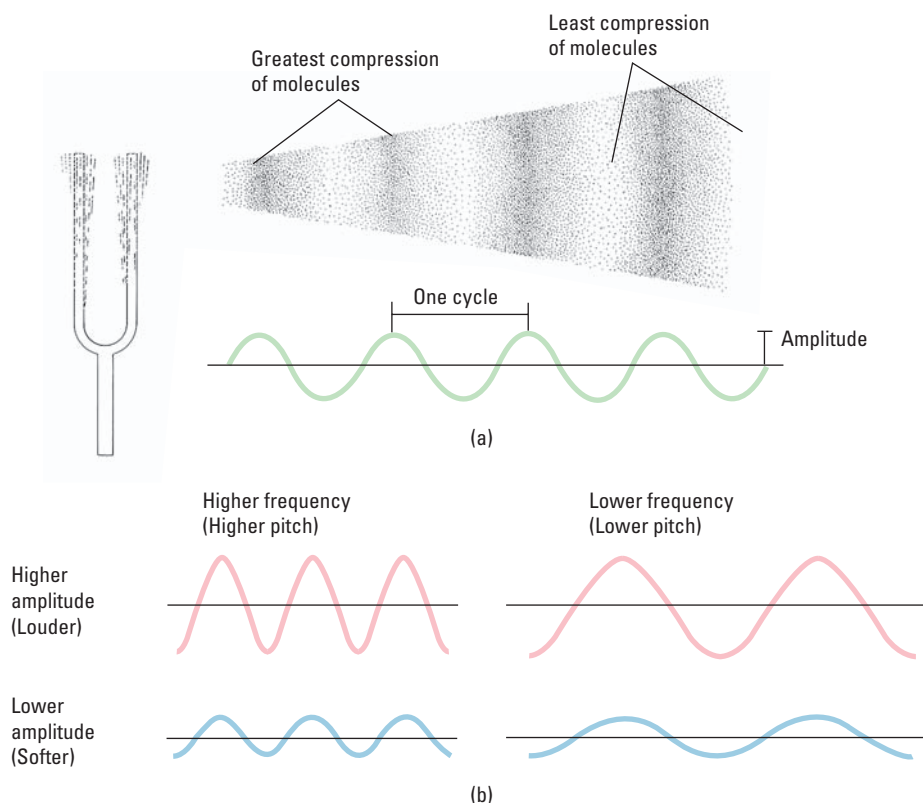


FIGURE 7.11 | Characteristics of sound The oscillating tuning fork (a) causes air molecules to vibrate in a manner that can be represented as a wave of pressure. Each wave contains an area in which the air molecules are more compressed (the dark regions in the upper diagram) and an area in which they are less compressed (the light regions) than normal. The peak pressure (the highest compression) of each wave defines the amplitude of the sound, and the number of waves that pass a given point per second defines the frequency. The higher the amplitude, the louder the sound; and the higher the frequency, the higher the pitch (b). All the wave drawings in this figure are sine waves, indicative of pure tones.
(Klinke, R. (1986). Physiology of hearing. In R. F. Schmidt (Ed.), *Fundamentals of sensory physiology*. New York: Springer-Verlag)

TABLE 7.2 |
Sound-pressure amplitudes of various sounds, and conversion to decibels*

Example	<i>P</i> (in sound-pressure units)	Log <i>P</i>	Decibels
Softest detectable sound	1	0	0
Soft whisper	10	1	20
Quiet neighborhood	100	2	40
Average conversation	1,000	3	60
Loud music from a radio	10,000	4	80
Heavy automobile traffic	100,000	5	100
Very loud thunder	1,000,000	6	120
Jet airplane taking off	10,000,000	7	140
Loudest rock band on record	100,000,000	8	160
Spacecraft launch (from 150 ft.)	1,000,000,000	9	180

*One sound-pressure unit (*P*) is defined as 2×10^{-5} newtons/square meter (Klinke, 1986). When measured in sound-pressure units, the amplitude range of human hearing is enormous. A reason for converting to logarithmic units is to produce a smaller range of numbers. The logarithm (log) of a number is the power to which 10 must be raised to produce that number. For example, the log of 10,000 is 4, because $10^4 = 10,000$. A decibel (dB) is defined as $20 \log P$. Thus, $4 \log \text{units} = 80 \text{ dB}$.
Sources: Lindsay, P. H. & Norman, D. A. (1977) *Human information processing*, 2nd ed. (p.161). New York: Academic Press. Matlin, M. W. & Foley, H. J. (1997). *Sensation and perception*, 4th ed. (p. 281). Boston: Allyn and Bacon.

of pressure called *decibels* (abbreviated *dB*). (See Table 7.2, which further defines decibels and contains the decibel ratings for a number of common sounds.)

In addition to varying in amplitude, sound waves vary in **frequency**, which we hear as the sound's **pitch**. The frequency of a sound is the rate at which the molecules of air or another medium move back and forth. Frequency is measured in *hertz* (abbreviated *Hz*), which is the number of complete waves (or cycles) per second generated by the sound source. Sounds that are audible to humans have frequencies ranging from about 20 to 20,000 Hz. To give you an idea of the relationship between frequency and pitch, the dominant (most audible) frequency of the lowest note on a piano is about 27 Hz, that of middle C is about 262 Hz, and that of the highest piano note is about 4186 Hz (Matlin & Foley, 1997). The simplest kind of sound is a pure tone, which is a constant-frequency sound wave that can be described mathematically as a sine wave (depicted in Figure 7.11). Pure tones, which are useful in auditory experiments, can be produced in the laboratory, but they rarely if ever occur in other contexts. Natural sources of sound, including even musical instruments and tuning forks, vibrate at several frequencies at once and thus produce more complex waves than that shown in Figure 7.11. Natural sounds can be thought of as consisting of many different tones at once.

How the Ear Works

21 What are the functions of the outer ear, middle ear, and inner ear?

<<< Hearing originated, in the course of evolution, from the sense of touch. Touch is sensitivity to pressure on the skin, and hearing is sensitivity to pressure on a special sensory tissue in the ear. In some animals, such as moths, sound is sensed through modified touch receptors located on flexible patches of skin that vibrate in response to sound waves. In humans and other mammals the special patches of skin for hearing have migrated to a location inside the head, and special organs, the ears, have developed to magnify the pressure exerted by sound waves as they are transported inward. A diagram of the human ear is shown in Figure 7.12. To review its structures and their functions, we will begin from the outside and work inward.

The **outer ear** consists of the *pinna*, which is the flap of skin and cartilage forming the visible portion of the ear, and the *auditory canal*, which is the opening into the head that ends at the eardrum (or *tympanic membrane*). The whole outer ear can be thought of as a funnel for receiving sound waves and transporting them

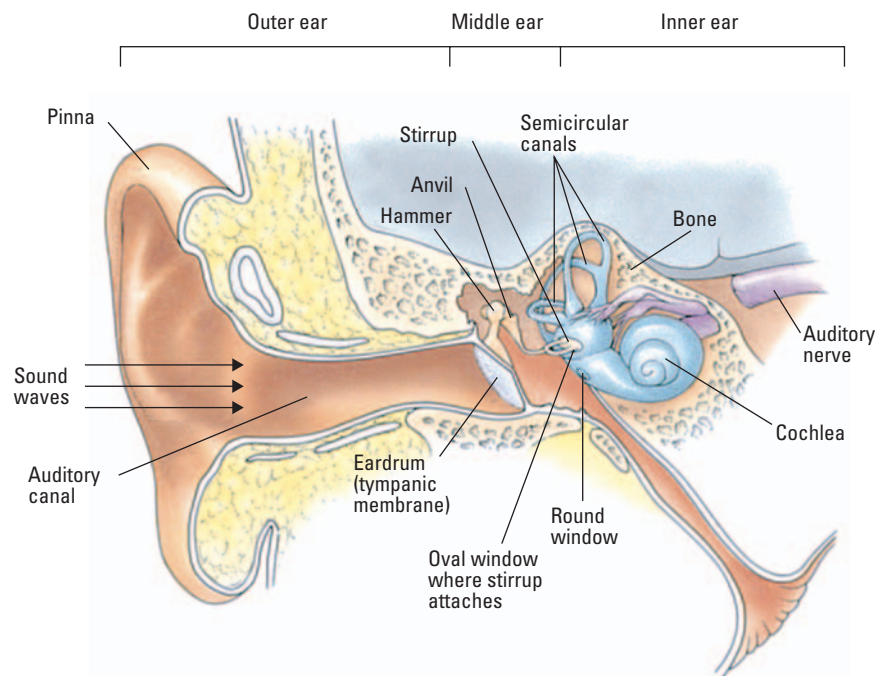


FIGURE 7.12 | Parts of the human ear
Sound waves (vibrations of air) that enter the auditory canal cause the eardrum to vibrate, which causes the ossicles (the hammer, anvil, and stirrup bones) to vibrate, which causes the oval window to vibrate, setting up waves of motion in the fluid inside the cochlea. The semicircular canals are involved in the sense of balance, not hearing.

inward. The vibration of air outside the head (the physical sound) causes air in the auditory canal to vibrate, which, in turn, causes the eardrum to vibrate.

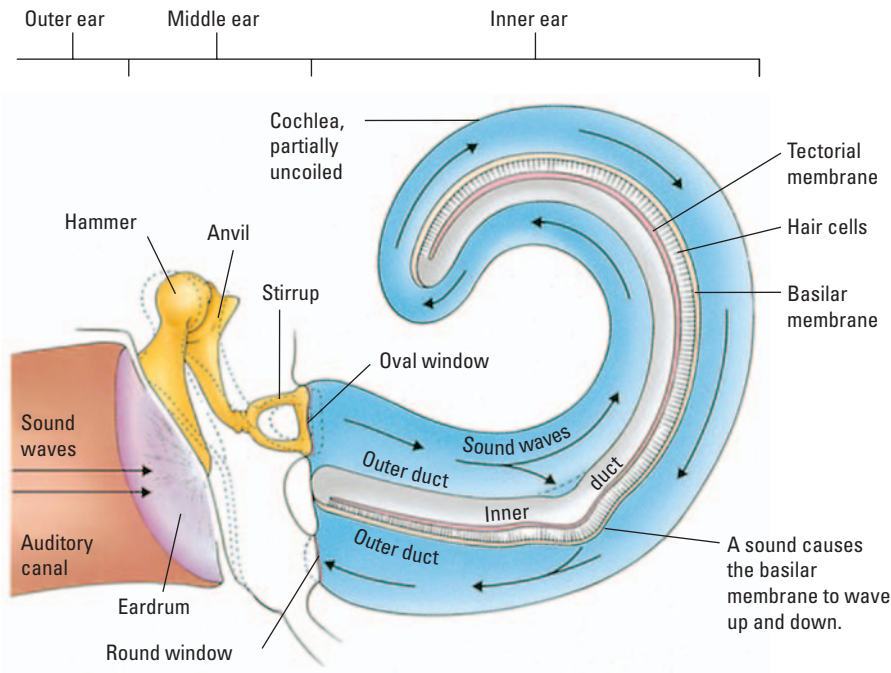
The **middle ear** is an air-filled cavity, separated from the outer ear by the eardrum. Its main structures are three tiny bones collectively called *ossicles* (and individually called the *hammer*, *anvil*, and *stirrup*, because of their respective shapes), which are linked to the eardrum at one end and to the *oval window*, a membrane of the cochlea, at the other end. When sound causes the eardrum to vibrate, the ossicles vibrate and push against the oval window. Because the oval window has only about one-thirtieth the area of the tympanic membrane, the pressure (force per unit area) that is funneled to it by the ossicles is about 30 times greater than the pressure on the eardrum. Thus, the main function of the middle ear is to increase the amount of pressure that sound waves exert upon the inner ear so that transduction can occur.

The coiled **cochlea**, in the **inner ear**, is where transduction finally takes place. The cochlea contains a fluid-filled *outer duct*. As depicted in the uncoiled view in Figure 7.13, this duct begins at the oval window, runs to the tip of the cochlea, and then runs back again to end at another membrane, the *round window*. Sandwiched between the outgoing and incoming portions of the outer duct is another fluid-filled tube, the *inner duct*. Forming the floor of the inner duct is the **basilar membrane**, on which are located the receptor cells for hearing, called **hair cells**. There are four rows of hair cells (three outer rows and one inner row), each row running the length of the basilar membrane. Tiny hairs (called *cilia*) protrude from each hair cell into the inner duct and abut against another membrane, called the *tectorial membrane*. At the other end from its hairs, each hair cell forms synapses with several *auditory neurons*, whose axons form the *auditory nerve* (eighth cranial nerve), which runs to the brain.

The process of transduction in the cochlea can be summarized as follows: The sound-induced vibration of the ossicles against the oval window initiates vibration in the fluid in the outer duct of the cochlea, which produces an up-and-down waving motion of the basilar membrane, which is very flexible. The tectorial membrane above the basilar membrane is less flexible and does not move when the basilar membrane moves, so the hairs protruding from the hair cells in the basilar membrane bend against the tectorial membrane each time the basilar membrane moves upward. This bending causes a physical change in the hair cell's membrane, which leads to a change in the electrical charge across the membrane (the receptor

>>> 22
How does transduction occur in the inner ear?

FIGURE 7.13 | Transduction mechanism in the inner ear This diagram depicts a longitudinal section of the cochlea (partially uncoiled), showing the outer and inner ducts. Sound waves in the fluid of the outer duct cause the basilar membrane to wave up and down. When the basilar membrane moves upward, its hairs bend against the tectorial membrane, initiating receptor potentials in the hair cells.



potential). This in turn causes each hair cell to release neurotransmitter molecules at its synapses upon auditory neurons, thereby increasing the rate of action potentials in those neurons (Hudspeth, 2000b).

Deafness and Hearing Aids

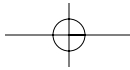
23 How do two kinds of deafness differ from each other in their physiological bases and in possible treatment?

<<< With this knowledge of the ear it is possible to understand the physiological bases for two varieties of deafness. **Conduction deafness** occurs when the ossicles of the middle ear become rigid and cannot carry sounds inward from the tympanic membrane to the cochlea. People with conduction deafness can hear vibrations that reach the cochlea by routes other than the middle ear. A conventional hearing aid is helpful for such people, because it magnifies the sound pressure sufficiently for vibrations to be conducted by other bones of the face into the cochlea.

Sensorineural deafness occurs from damage to the cochlea, the hair cells, or (less often) the auditory neurons. People with complete sensorineural deafness are not helped by a conventional hearing aid but can in many cases be helped by a newer form of hearing aid called a **cochlear implant**. This device performs the transduction task normally done by the ear's hair cells (though it does not do it nearly as well). It transforms sounds into electrical impulses and sends the impulses through thin wires permanently implanted into the cochlea, where they stimulate the terminals of auditory neurons directly. It can be used when deafness has re-

External portions of a cochlear implant A microphone hooked around the girl's ear picks up sound waves and sends signals down, through a cable connection, to an auditory processor attached to a belt around her waist. The auditory processor converts the signals into electrical pulses, suitable for stimulating the cochlea, and sends those pulses up the cable to a transmitter fastened to the girl's head behind her ear. The transmitter sends the pulses through the skin to a receiver implanted in her skull, and from there the pulses are sent by thin wire electrodes to stimulate appropriate sets of auditory neurons in the cochlea of the inner ear.





sulted from the destruction of hair cells, but not when the auditory nerve has been destroyed. The best cochlear implants today permit sufficient sound resolution to enable at least some deaf children to acquire vocal language at a nearly normal rate (Svirsky & others, 2000). They also allow adults who became deaf after learning language to regain their ability to understand speech (Dorman & Wilson, 2004).

Pitch Perception

How do receptors on the basilar membrane code the differences in sound frequency that permit pitch detection? That is one of the first questions that scientists asked about the workings of the ear. In the mid-nineteenth century, the great German physiologist Hermann von Helmholtz suggested a possible analogy between the ear's basilar membrane and the strings of a harp. A harp's strings vibrate not only when plucked but also when sound waves strike them, and different strings vibrate best to different frequencies. Helmholtz proposed that the basilar membrane contains separate fibers that vibrate (resonate) in response to different sound frequencies, like a harp's strings, and that activate different neurons in the auditory nerve (Zwislocki, 1981). We now know that Helmholtz was correct in theorizing that different portions of the basilar membrane are most sensitive to different sound frequencies, but he was wrong about the details.

The Traveling Wave as a Basis for Frequency Coding

Considerably after Helmholtz, in work begun in the 1920s that eventuated in a Nobel prize, Georg von Békésy developed a way to observe directly the action of the basilar membrane, and he discovered that it behaves not like a harp with separate strings but, rather, like a bedsheet when someone shakes it at one end. Sound waves entering the cochlea set up traveling waves on the basilar membrane, which move from the proximal end (closest to the oval window) toward the distal end (farthest away from the oval window). As each wave moves, it gradually increases in amplitude up to a certain maximum and then rapidly dissipates. Of most importance, Békésy found that the position on the membrane at which the waves reach their peak depends on the frequency of the tone. High frequencies produce waves that travel only a short distance, peaking near the proximal end, and low frequencies produce waves that travel farther, peaking near the distal end. From this observation, Békésy hypothesized that (a) rapid firing in neurons that come from the proximal end of the membrane, accompanied by little or no firing in neurons coming from more distal parts, is interpreted by the brain as a high-pitched sound and (b) rapid firing in neurons coming from a more distal portion is interpreted as a lower-pitched sound.

Subsequent research has confirmed the general validity of Békésy's hypothesis and has shown that the waves on the intact, living basilar membrane are in fact much more sharply defined than those that Békésy had observed. There is now strong evidence that the primary receptor cells for hearing are the inner row of hair cells and that the outer three rows serve mostly a different function. They have the capacity to stiffen when activated, and they do so in a manner that amplifies and sharpens the traveling wave (Géléoc & Holt, 2003). For an illustration of the effects that different tones and complex sounds have on the basilar membrane, see Figure 7.14.

Two Sensory Consequences of the Traveling-Wave Mechanism

The manner by which the basilar membrane responds to differing frequencies helps us make sense of a number of auditory phenomena. One such phenomenon is *asymmetry in auditory masking*, which is especially noticeable in music production. **Auditory masking** refers to the ability of one sound to mask (prevent the hearing of) another sound. Auditory masking is asymmetrical in that low-frequency sounds mask high-frequency sounds much more effectively than the reverse (Scharf, 1964). A bassoon can easily drown out a piccolo, but a piccolo cannot easily drown out a

>>> 24

How does the basilar membrane of the inner ear operate to ensure that different neurons are maximally stimulated by sounds of different frequencies?

>>> 25

How does the traveling-wave theory explain (a) an asymmetry in auditory masking and (b) the pattern of hearing loss that occurs as we get older?

bassoon. To see how this can be explained in terms of the waves the two instruments produce on the basilar membrane, look at Figure 7.15. The wave produced by a low-frequency bassoon note encompasses the entire portion of the basilar membrane that is encompassed by the piccolo note (and more). Thus, if the bassoon note is of sufficient amplitude, it can interfere with the effect of the piccolo note; but the piccolo note, even at great amplitude, cannot interfere with the effect that the bassoon note has on the more distal part of the membrane, because the wave produced by the piccolo note never travels that far down the membrane.

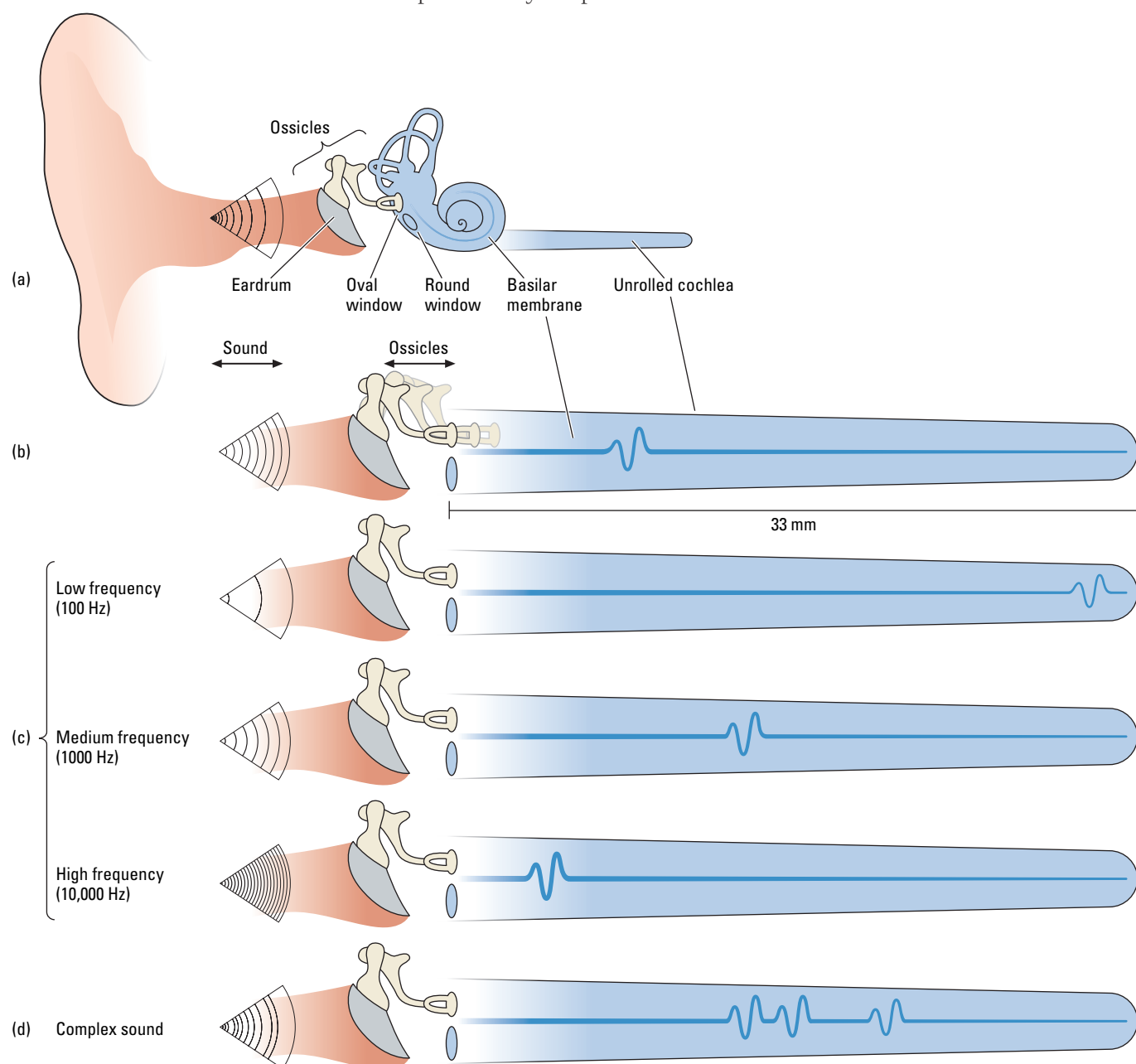


FIGURE 7.14 | Waves on the basilar membrane In this highly schematic diagram, the cochlea is uncoiled and stretched out and the basilar membrane is depicted as a line running down the center of the cochlea. Part (b) illustrates the back-and-forth movement of the ossicles in response to sound and the traveling wave that that motion sets up on the basilar membrane. Part (c) illustrates the relationship between the frequency of a tone and the place on the basilar membrane where the traveling wave reaches

its peak amplitude. Part (d) shows that a complex sound, made up of more than one frequency, produces wave peaks simultaneously at more than one location on the basilar membrane. The sizes of the wave peaks are greatly exaggerated in the diagram. The waves are shown only at the places where they peak. In fact, each wave must travel the length of the basilar membrane to the place where it reaches its peak, and then it dies out shortly after that. (Adapted from Hudspeth, 2000a, p. 595.)

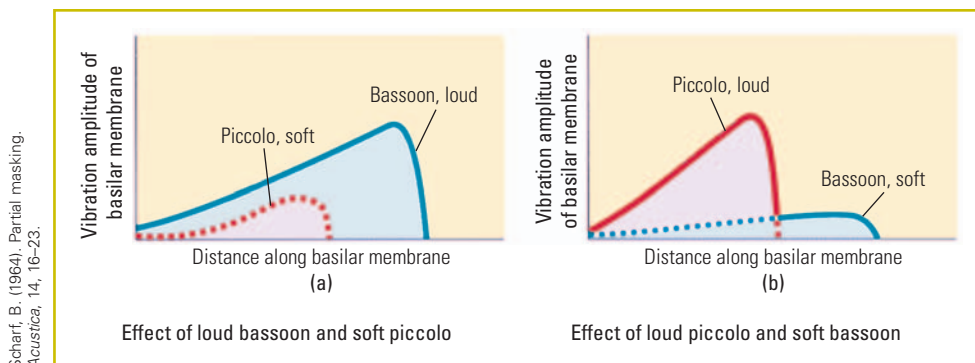


FIGURE 7.15 | Why a low-frequency tone masks a high-frequency tone better than the reverse (a) When a low-frequency tone (such as that from a bassoon) and a high-frequency tone (such as that from a piccolo) are played simultaneously, the bassoon can mask the piccolo because the bassoon's action on the basilar membrane encompasses the entire portion on which the piccolo acts. (b) But the piccolo cannot mask the bassoon because, even when played at high amplitude, the piccolo does not affect the distal part of the membrane to which the bassoon's waves extend.

Another effect of the traveling-wave mechanism concerns the pattern of hearing loss that occurs as we get older. We lose our sensitivity to high frequencies to a much greater degree than to low frequencies. Thus, young children can hear frequencies as high as 30,000 Hz, and young adults can hear frequencies as high as 20,000 Hz, but a typical 60-year-old cannot hear frequencies above about 15,000 Hz. This decline is greatest for people who live or work in noisy environments (see Figure 7.16) and is caused by the wearing out of hair cells with repeated use (Kryter, 1985). But why should cells responsible for coding high frequencies wear out faster than those for coding low frequencies? The answer may lie in the fact that the former are acted upon by all sounds (as shown in Figure 7.15), while the latter are stimulated only by low-frequency sounds.

Another Code for Frequency

Although the traveling-wave theory of frequency coding has been well validated, it is not the whole story. For frequencies below about 4000 Hz (which include most of the frequencies in human speech), perceived pitch depends not just on which part of the basilar membrane is maximally active but also on the timing of that activity (Moore, 1997). The electrical activity triggered in sets of auditory neurons tends to be locked in phase with sound waves, such that a separate burst of action potentials occurs each time a sound wave peaks. The frequency at which such bursts occur contributes to the perception of pitch.

Consistent with what is known about normal auditory coding, modern cochlear implants use both place and timing to produce pitch perception (Dorman & Wilson,

>>> 26
How does the timing of action potentials code sound frequency? How do cochlear implants produce perception of pitch?

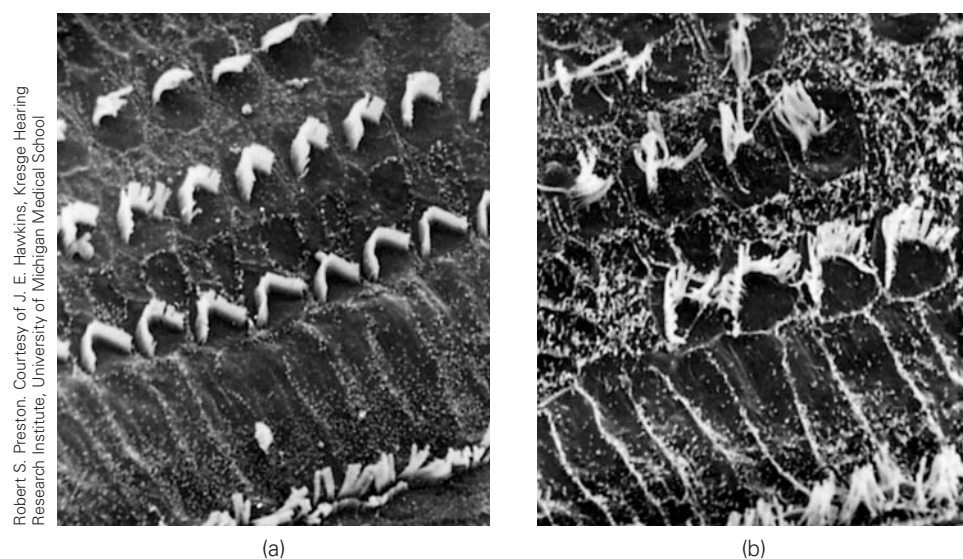
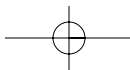


FIGURE 7.16 | Warning: Noise can be dangerous These electron micrographs show hair cells on the basilar membrane of a guinea pig (a) before and (b) after exposure to 24 hours of sound at an intensity comparable to a rock concert. The tiny hairs are disarranged on some cells and completely destroyed on others. Notice also that the hair cells exist in four rows—three outer rows and one inner row.



2004). These devices break a sound signal into several (typically six) different frequency ranges and send electrical pulses from each frequency range through a thin wire to a different portion of the basilar membrane. Pitch perception is best when the electrical signal sent to a given locus of the membrane is pulsed at a frequency similar to that of the sound wave that would normally act at that location.

Tonotopic Arrangement of the Primary Auditory Cortex

Auditory sensory neurons send their output to nuclei in the brainstem, which in turn send axons upward, ultimately to the primary auditory area of the cerebral cortex, located in each temporal lobe (look back at Figure 7.12). Neurons in the primary auditory cortex are *tonotopically* organized. That is, each neuron is maximally responsive to sounds of a particular frequency, and the neurons are systematically arranged such that high-frequency tones activate neurons at one end of this cortical area and low-frequency tones activate neurons at the other end. Ultimately, the pitch or set of pitches we hear depends largely on which neurons in the auditory cortex are most active.

As is true of other sensory areas in the cortex (discussed in Chapter 5), the response characteristics of neurons in the primary auditory cortex are very much influenced by experience. When experimental animals are trained to use a particular tone frequency as a cue guiding their behavior, the number of auditory cortical neurons that respond to that frequency greatly increases (Bakin & others, 1996). Heredity determines the general form of the tonotopic map, but experience determines the specific amount of cortex that is devoted to any particular range of frequencies.

Making Sense of Sounds

27

How does the difference between the two ears in their distance from a sound source contribute our ability to locate that source?

<<< Think of the subtlety and complexity of our auditory perception. With no cues but sound, we can locate a sound source within about 5 to 10 degrees of its true direction (Hudspeth, 2000a). At a party we can distinguish and listen to one person's voice in a noisy environment that includes many other voices and a band in the background. To comprehend speech, we hear the tiny difference between *plot* and *blot*, while ignoring the much larger differences between two different voices that speak either of those words. All sounds set up patterns of waves on our basilar membranes, and from those seemingly chaotic patterns our nervous system extracts all the information needed for auditory perception.

Locating Sounds

The ability to tell the direction of a sound source contributes greatly to the usefulness of hearing. When startled by an unexpected rustling, we reflexively turn toward it to see what might be causing the disturbance. Even newborn infants do this (Muir & Field, 1979), indicating that the ability to localize a sound does not require learning. Such localization is also a key component of our ability to keep one sound distinct from others in a noisy environment. People can attend to one voice and ignore another much more easily if the two voices come from different locations in the room than if they come from the same location (Feng & Ratnam, 2000).

Sound localization depends largely on the timing at which each sound wave reaches one ear compared to the other. A wave of sound coming from straight ahead reaches the two ears simultaneously, but a wave from the right or the left reaches one ear slightly before the other (see Figure 7.17). A sound wave that is just slightly to the left of straight ahead reaches the left ear a few microseconds (millionths of a second) before it reaches the right ear, and a sound wave that is 90 degrees to the left reaches the left ear about 700 microseconds before it reaches the right ear. Neurons in at least one way station in the auditory system receive input from both ears. Some of these neurons respond most to waves that reach both ears at once, and others respond most to waves that reach one ear some microseconds—

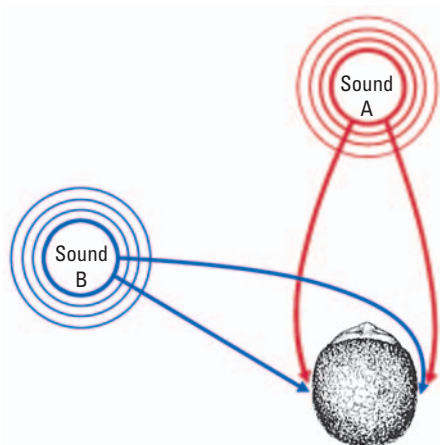


FIGURE 7.17 | Locating a sound
Sound waves coming from the left (sound B) reach the left ear sooner than the right ear. The converse would be true of sound waves coming from the right. Neurons receiving input from the two ears are sensitive to this difference, which provides a code for sound localization. Without moving the head, it is difficult to distinguish a sound in front of the head from one that is at the same angle in back of the head, but a head rotation quickly resolves that problem.

ranging from just a few on up to about 700—before, or after, reaching the other ear. These neurons, presumably, are part of the brain's mechanism for perceiving the direction from which a sound is coming (Hudspeth, 2000a; Pollak, 2002).

Analyzing Patterns of Auditory Input

In the real world (outside of the laboratory), sounds never come in the form of pure tones. The sounds we hear, and from which we extract meaning, consist of highly complex patterns of waveforms. How is it that you identify the sweet word *psychology* regardless of whether it is spoken in a high-pitched or low-pitched voice, or spoken softly or loudly? You identify the word not from the absolute amplitude or frequency of the sound waves, but from certain patterns of change in these that occur over time as the word is spoken.

Beyond the primary auditory area are cortical areas for analyzing such patterns (Poremba & others, 2003). For example, some neurons in areas near the primary auditory area respond only to certain combinations of frequencies, others only to rising or falling pitches, others only to brief clicks or bursts of sound, and still others only to sound sources that are moving in a particular direction (Baumgart & others, 1999; Phillips, 1989). In macaque monkeys, some cortical neurons respond selectively to particular macaque calls (Rauschecker & others, 1995). In the end, activity in some combination of neurons in your cerebral cortex must provide the basis for each of your auditory experiences, but researchers are only beginning to learn how that is accomplished.

Phonemic Restoration: An Auditory Illusion

In certain situations our auditory system provides us with the perception of sounds that are not really present as physical stimuli. A well-documented example is the sensory illusion of *phonemic restoration*. **Phonemes** are the individual vowel and consonant sounds that make up words (discussed further in Chapter 11), and phonemic restoration is an illusion in which people hear phonemes that have been deleted from words or sentences as if they were still there. The perceptual experience is that of really *hearing* the missing sound, not that of inferring what sound must be there.

Richard Warren (1970) first demonstrated this illusion in an experiment in which he removed an s sound and spliced in a coughing sound of equal duration in the following tape-recorded sentence at the place marked by an asterisk: *The state governors met with their respective legi*latures convening in the capital city*. People listening to the doctored tape could hear the cough, but it did not seem to coincide with any specific portion of the sentence or block out any sound in the sentence. Even when they listened repeatedly, with instructions to determine what sound was missing, subjects were unable to detect that any sound was missing. After they

>>> 28
Under what conditions does our auditory system fill in a missing sound? What might be the value of such an illusion?

were told which sound was missing, they still claimed to hear that sound each time they listened to the tape. In other experiments, thousands of practice trials failed to improve people's ability to judge which phoneme was missing (Samuel, 1991).

Which sound is heard in phonemic restoration depends on the arrangement of the surrounding phonemes and the meaningful words and phrases they produce. The restored sound is always one that turns a partial word into a whole word that is consistent in meaning with the rest of the sentence. Most remarkably, even words that occur *after* the missing phoneme can influence which phoneme is heard. For example, people heard the stimulus sound **eel* (again, the asterisk represents a cough-like sound) as *peel*, *heel*, or *wheel*, depending on whether it occurred in the phrase *The *eel was on the orange*, *The *eel was on the shoe*, or *The *eel was on the axle* (Warren, 1984). The illusory phoneme was heard as occurring at its proper place, in the second word of the sentence, even though it depended on words coming later.

One way to make sense of this phenomenon is to assume that much of our perceptual experience of hearing derives from a brief auditory sensory memory, which lasts for a matter of seconds and is modifiable. The later words generate a momentary false memory of hearing, earlier, a phoneme that wasn't actually present, and that memory is indistinguishable from the memory of a phoneme that actually did occur. Such brief, compelling memories and their role in perceptual experiences are discussed more fully in Chapter 9. Illusory restoration has also been demonstrated in music perception. People hear a missing note, in a familiar tune, as if it were present (DeWitt & Samuel, 1990).

A limiting factor in these illusions is that the gap in the sentence or tune must be filled with noise; it can't be a silent gap. In everyday life the sounds we listen to are often masked by bits of noise, never by bits of silence, so perhaps illusory sound restorations are an evolutionary adaptation by which our auditory system allows us to hear meaningful sound sequences in a relatively uninterrupted stream. When a burst of noise masks a phoneme or a note, our auditory system automatically, after the fact, in auditory memory, replaces that burst with the auditory experience that, according to our previous experience, belongs there. Auditory restoration nicely exemplifies the general principle that our perceptual systems often modify sensory input in ways that help to make sense of that input. You will find more examples and much more discussion of that principle in the next chapter, which deals with vision.

SECTION REVIEW

Hearing allows us to glean information from patterns of vibrations carried by air.

Basic Facts of Hearing - Physically, sound is the vibration of air or another medium caused by a vibrating object. A sound wave's amplitude is related to its perceived loudness and its frequency is related to its perceived pitch. - The outer ear funnels sound inward, the middle ear amplifies it, and the inner ear transduces and codes it. - Conduction deafness is due to middle ear rigidity; sensorineural deafness is due to inner ear or auditory nerve damage.	Pitch Perception - A sound's frequency determines the length of the traveling wave produced in the basilar membrane, thus allowing frequency to be coded. - The traveling-wave theory helps to explain the asymmetry of auditor masking and the typical pattern of age-related hearing loss. - For frequencies below 4000 Hz, the timing of action potentials also codes sound frequency. - The primary auditory cortex is tonotopically organized. The pitch we hear is determined by which cortical neurons are most active.	Making Sense of Sounds - Brain neurons that compare the arrival time of sound waves at our two ears enable us locate a sound's source. - Most sounds are complex waveforms requiring analysis in cortical areas beyond the primary auditory area. - The phonemic restoration effect illustrates the idea that context and meaning influence our sensory experience.
---	--	---

Psychophysics

Psychophysics is the study of relationships between physical characteristics of stimuli and the sensory experiences produced by those stimuli. (You can remember the term by recalling that *psychophysics* relates *psychological* sensory experiences to *physical* characteristics of stimuli.) Sensory experience is typically assessed by asking subjects to make some judgment about each stimulus, such as whether it is present or absent or whether it is the same as or different from another stimulus. Many experiments already alluded to in this chapter were psychophysical, including those assessing people's abilities to detect or identify particular substances by taste or smell and those relating the frequency of a tone to its perceived pitch.

This section describes some psychophysical methods and findings related to (a) the detection of weak stimuli, (b) the detection of small changes in stimuli, and (c) the attempt to develop a general law relating the physical intensity of a stimulus to the intensity of the sensory experience it produces.

As you will see, psychophysics is more mathematical than most other areas of psychology. That is one of the reasons why some psychologists find it exciting. Psychophysics is just the right cup of tea for those psychologists who like a degree of precision in their science, are drawn by the elegance of mathematics, and are fascinated by the idea that certain psychological phenomena can be described meaningfully with algebraic equations. Historically, research in psychophysics by such scientists as Ernst Weber and Gustav Fechner, in the nineteenth century, played a vital role in the founding of psychology as a science, and many of the specific questions they raised are still topics of research and debate today.

The Absolute Threshold Is a Measure of Sensitivity

Psychophysicists refer to the faintest detectable stimulus, of any given type, as the **absolute threshold** for that type of stimulus. For example, the weakest intensity of a particular tone (say, middle C on the piano) that you can hear is your absolute threshold for that tone, and the weakest solution of sodium chloride (table salt) that you can taste is your absolute threshold for sodium chloride. Absolute thresholds vary from person to person and are used clinically as measures of a person's sensitivity to specific types of stimuli. You may well have undergone a test of your absolute threshold for various sound frequencies, as a clinical test of your hearing.



Phanie / Photo Researchers

Finding the absolute threshold In a hearing test, faint sounds of varying frequencies and amplitudes are presented to one ear at a time. To find the absolute threshold for a given frequency, for a given ear, the sound is presented at decreasing amplitudes until the person cannot detect it, and at increasing amplitudes until the person can detect it.

In general, for most senses, absolute thresholds are higher (meaning that sensitivity is lower) for older adults than for younger adults. Some research applications of absolute thresholds were mentioned earlier in this chapter. For example, Figure 7.4 (page 233) shows differences between men and women in absolute thresholds for detecting certain odors.

The Difference Threshold Depends on the Magnitude of the Original Stimulus

Another measure of sensitivity is the **difference threshold**, defined as the minimal difference in magnitude (or intensity) between two stimuli that is required for the person to detect them as different. Another name for the difference threshold is the **just-noticeable difference**, abbreviated **jnd** (by convention, small letters are used for this abbreviation).

In the early nineteenth century, the German physicist Ernst Weber (1834) conducted a systematic study of just-noticeable differences. Weber asked: What is the relationship between the jnd and the magnitude of the original stimulus? Is the jnd the same for a given type of stimulus regardless of the magnitude of the original stimulus, or does it vary in some systematic way with that magnitude?

29

How did Weber derive a law from data on just-noticeable differences? How can Weber's law be used to predict the degree to which two stimuli must differ for a person to tell them apart?

<<< In one series of experiments, Weber applied this question to people's abilities to judge differences between weights. On each trial he asked the subjects to pick up each of two weights (the "original" weight and a comparison weight), one at a time, and to judge which was heavier. Weber found that the jnd varied in direct proportion to the original weight. Specifically, he found that for any original weight that he used (within a certain range), the jnd was approximately 1/30 of that weight (Gescheider, 1976). Thus, a typical subject could just barely detect the difference between a 15-gram and a 15.5-gram weight or between a 90-gram and a 93-gram weight. In the first case the jnd was 0.5 gram, and in the second it was 3 grams, but in both cases it was 1/30 of the original weight (which was always the lower of the two weights being compared).

In other experiments, Weber studied people's ability to discriminate between the lengths of two lines, one presented after the other, and again he found a constant proportionality between the original stimulus and the difference threshold. For this task, however, the constant fraction was 1/100 rather than 1/30. Thus, a typical subject could just barely detect the difference between a 100- and a 101-millimeter line or between a 1000- and a 1010-millimeter line.

On the basis of these and similar experiments, Weber formulated a general law, now called **Weber's law**, stating that *the jnd for stimulus magnitude is a constant proportion of the magnitude of the original stimulus*. The law can be abbreviated as

$$\text{jnd} = kM$$

in which *M* is the magnitude or intensity of the stimulus used as the original stimulus and *k* is a proportionality constant referred to as the *Weber fraction*, which is different for different sensory tasks (1/30 for weight judgment and 1/100 for length judgment in the examples just cited). Since Weber's time, researchers have confirmed Weber's law for many different types of stimuli, including sounds, sights, tastes, and smells. The law holds up rather well over a wide portion of the possible range of intensities or magnitudes for most types of stimuli, but not at the very low (near the absolute threshold) and very high ends of the range.

Sensory Magnitude Is Lawfully Related to Stimulus Magnitude

When a physical stimulus increases, our sensory experience of it also increases. When a sodium chloride solution becomes more concentrated, we taste its saltiness as stronger; when a sound increases in amplitude, we hear it as louder; when a light becomes more intense, we see it as brighter; and so on. Is it possible to specify in a mathematical equation the relationship between the magnitude of a stimu-

lus and the magnitude of the sensory experience it produces? Such an equation was proposed by Gustav Fechner in the middle of the nineteenth century and was tested experimentally and modified by S. S. Stevens a century later.

Fechner's Logarithmic Law Fechner (1860/1966), in Germany, used Weber's law to derive a mathematical relationship between stimulus magnitude and sensory magnitude. The jnd is measured in physical units (such as grams or sound-pressure units), yet it reflects a sensory phenomenon, the just-noticeable difference between two sensations. Therefore, Fechner reasoned, the jnd could serve as a unit for relating physical and sensory magnitudes. His crucial assumptions were (a) that every jnd along a sensory dimension is equivalent to every other jnd along that dimension in the amount it adds to the sensory magnitude and (b) that jnd's can be added together. In other words, he assumed that a sound that is 100 jnd's above threshold would sound twice as loud as one that is 50 jnd's above threshold, or one-tenth as loud as one that is 1000 jnd's above threshold.

Fechner assumed that jnd's are subjectively equal, but he knew from Weber's work that they are not physically equal. As you just learned, the jnd is directly proportional to the magnitude of the original stimulus. Thus, Fechner assumed that the amount of physical change needed to create a constant sensory change is directly proportional to the magnitude of the stimulus, and he showed mathematically that this can be expressed as a logarithmic relationship (if you wish to see how he could prove this, turn to page A-9 of the Statistical Appendix at the back of the book). Thus, Fechner derived what he believed to be a general psychophysical law, now called **Fechner's law**, stating that *the magnitude of the sensory experience of a stimulus is directly proportional to the logarithm of the physical magnitude of the stimulus*. This law can be abbreviated as

$$S = c \log M$$

where S is the magnitude of the sensory experience, c is a proportionality constant, and M is the magnitude of the physical stimulus.

To gain a sense of the meaning of Fechner's law (if you are a bit foggy on logarithms), look at Table 7.3, which shows hypothetical data that are consistent with the law. Imagine a person judging the brightness of a room's illumination with one light bulb on, then two bulbs, then three bulbs, and so on. Although each successive bulb adds a constant amount (100 units) to the physical intensity (M) of the room's illumination, each bulb adds a progressively smaller amount to the perceived brightness (S). Thus, in the example, the second bulb adds 0.30 unit to the brightness, the third adds another 0.18 unit, the fourth adds 0.12, and so on. Notice too that every time the physical intensity (M) doubles, the sensed brightness (S) increases by a constant amount (0.30 unit in the example). Stated differently, as the physical scale increases geometrically, the sensory scale increases arithmetically. That is the essence of a logarithmic relationship. Consequently, a huge intensity range on the physical scale is condensed to a much smaller range on the psychological scale.

Stevens's Power Law Although Fechner believed his law was valid on theoretical grounds, he did not believe it could be tested experimentally. He wrote (in 1860), "A real measure of sensation would demand that we be able to call a given sensation twice, thrice, or so-and-so as many times as intense as another—but who would say such a thing?" (quoted in Stevens, 1975). The belief that people could not report the magnitudes of their sensations in a consistent way went relatively unchallenged until the early 1950s, when S. S. Stevens, a Harvard psychologist, began a series of experiments in which he asked people to do exactly that.

What reasoning led Fechner to propose that sensory magnitude is proportional to the logarithm of the stimulus magnitude? If a physical stimulus keeps doubling in magnitude, how will the sensory magnitude change, according to Fechner?

TABLE 7.3
Illustration of the relation between physical magnitude (M) and sensory magnitude (S) according to Fechner's law

Number of light bulbs	M	S^* ($1 \log M$)	
1	100	2.00	
2	200	2.30	← +.30
3	300	2.48	
4	400	2.60	← +.30
5	500	2.70	
6	600	2.78	
7	700	2.84	
8	800	2.90	← +.30

*In this hypothetical example, c in the formula $S = c \log M$ is 1. A change in c would change the values of S but would not change the fact that each doubling of M adds a constant increment to S , which is the point that the table is designed to illustrate.

31 How did Stevens test Fechner’s law, and what did he find? How does Stevens’s law differ from Fechner’s?

<<< Stevens’s technique, called the *method of magnitude estimation*, involved asking subjects to assign numbers to the magnitudes of their sensations. For example, he would present a standard stimulus and call that a “10,” and then he would present a comparison stimulus and ask the subject to give it a number that best approximated its sensory magnitude compared with that of the standard. Thus, a sensation that appeared to the subject to be twice that of the standard would be called “20,” one that seemed half that of the standard would be called “5,” and so on. Stevens found that people had little difficulty carrying out these instructions and that their responses were remarkably consistent for any given set of stimuli.

If Fechner’s law were correct, Stevens should have found that his subjects’ magnitude estimates were directly proportional to the logarithms of the stimulus intensities he used. He found, however, that the logarithmic relationship was only roughly accurate for most senses and quite inaccurate for some senses. For every sense, according to Stevens, the results could be described better by a different mathematical relationship—a power relationship. On this basis, Stevens proposed an alternative to Fechner’s logarithmic law. According to ***Stevens’s power law***, the magnitude of a sensation is directly proportional to the magnitude of the physical stimulus raised by a constant power. The law can be abbreviated as

$$S = cM^p$$

where S is the reported magnitude of the sensory experience, M is the physical magnitude of the stimulus, p is the power (or exponent) to which M must be raised (which differs from one sensory dimension to another), and c is a constant whose value depends on the size of the measurement units used.

Stevens and his colleagues performed dozens of experiments, involving magnitude estimates for many different kinds of stimuli, and they almost always found that the results could be quite well represented by a power equation. For each kind of stimulus, they could determine a unique exponent (p) to which the physical magnitude had to be raised to approximate the experienced magnitude. Table 7.4 shows the exponents that they compiled for several different kinds of stimuli. For most tasks the exponent is less than 1, but for one task shown in the table (estimating the length of a line) it is exactly 1, and for another (estimating the pain of an electric shock) it is greater than 1.

In cases where p is less than 1, equal physical changes produce smaller sensory changes at the high end of the scale than at the low end, as was also true with Fechner’s logarithmic law. When p is greater than 1, however, the opposite relationship holds. Thus, adding a certain amount of electric shock to a relatively strong shock produces a greater increase in pain than does adding the same amount to a relatively weak shock. Finally, when p is equal to 1, equal physical changes produce the same amount of sensory change regardless of whether one is starting with a strong or a weak stimulus. All these relationships are graphically portrayed in Figure 7.18.

TABLE 7.4 |
Power-law exponents for various stimuli

Type of stimulus	Measured exponent (p)*
Brightness of a spot of light in the dark	0.33
Loudness of a 3000-cps tone	0.67
Smell of heptane	0.60
Taste of saccharine	0.80
Length of a line	1.00
Pain of an electric shock on the fingers	3.50

*The exponent (p) is the power to which the stimulus magnitude must be raised to approximate the sensory magnitude.
Source: From Stevens, S. S. (1975) *Psychophysics: Introduction to its perceptual, neural, and social prospects* (p.13). New York: Wiley.

Why a Power Law?

32 Why is it advantageous that our senses operate according to a power law rather than a logarithmic law?

<<< Why do our senses obey a power law for so many different kinds of stimuli? Is it just coincidence, or is there some advantage that would lead each sense, in the course of evolution, to operate in accordance with a power law? Stevens (1975) thought a good deal about that question, and the answer he suggested was essentially the following:

Our world of stimuli is constantly changing. As we move closer to or farther from the sources of stimulation, or as day turns to dusk, the overall intensity of the energy reaching us from specific objects in the environment changes greatly. If we are to recognize the same scenes, sounds, and smells under such varying conditions, then we must extract those features of each stimulus constellation that re-

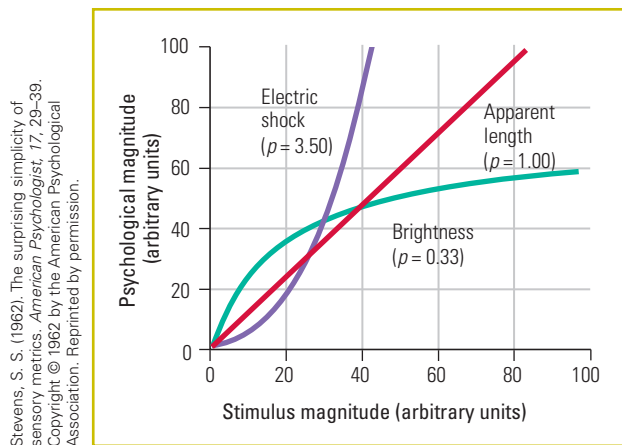


FIGURE 7.18 | Power law illustrated for three sensory dimensions This shows how subjects' estimates of the sensory magnitude that they experienced increased as the stimulus magnitude increased, separately for the pain of an electric shock, the length of a line, and the brightness of a spot of light. Notice that the curvature is upward or downward depending on whether the exponent, p , is greater or less than 1.

main constant. One such constancy is the ratio of the magnitudes of the stimulus elements with respect to each other.

As you move toward or away from a sound source, the ratio of the amplitudes of the various tones in the sound remains relatively constant, even though the overall amplitude increases or decreases greatly. Similarly, the ratio of light reflected from a darker compared with a lighter portion of a visual scene remains nearly constant as the overall intensity fades at dusk. A power law, and only a power law, preserves constant sensory ratios as the overall intensity waxes or wanes. (You can find a proof of this statement on page A-10 of the Statistical Appendix at the back of the book.) For example, in the case of the power function for brightness, with an exponent of 0.33, every eightfold change in light intensity results in a twofold change in apparent brightness, no matter where on the intensity continuum we start. Thus, if the light illuminating a visual scene decreases in physical intensity to one-eighth of what it was before, each part of the scene will appear half as bright as before, and the ratios of brightnesses among the parts will remain what they were before. The elegant feature of the power law, with p less than 1, is that it compresses large physical changes down to smaller sensory changes, as does the logarithmic law; but, unlike the logarithmic law, it does this while preserving the constancy of stimulus ratios.



Todd Gipstein / National Geographic / Getty

Value of the power law Because of the power law relating the physical intensity of light to its perceived brightness, the ratios in brightness among different elements of this scene remain constant as the overall illumination changes, from daylight to dusk.

SECTION REVIEW

Psychophysics relates sensations to physical characteristics of stimuli.

Sensory Thresholds

- A person’s absolute threshold for a specific type of stimulus is the lowest detectable intensity of the stimulus.
- The smallest difference in intensity of a given type of stimulus that a person can detect is a difference threshold, or jnd.
- According to Weber’s law, the jnd is a constant proportion of the magnitude of the original stimulus.

Sensory Magnitude Related to Physical Magnitude

- Fechner’s law, derived from Weber’s, holds that the magnitude of a sensory experience is directly proportional to the logarithm of the physical magnitude of the stimulus triggering that experience. For example, a sound’s perceived loudness is proportional to the logarithm of the sound’s physical intensity.
- Using the method of magnitude estimation, Stevens found evidence for an alternative to Fechner’s law: The magnitude of a sensation is directly proportional to the magnitude of the physical stimulus raised to a constant power.
- A power relationship between sensory and physical magnitudes may have evolved because it maintains constant ratios among the sensory magnitudes produced by different stimuli in the face of changes that affect both (or all) of the stimuli equally. For example, because of the power law, the ratio of the perceived loudness of sound A to sound B remains constant as one moves away from or closer to the two sound sources.

Concluding Thoughts

To review this chapter, you would do well to focus on major themes. Even details are easier to remember if tied to larger themes, or arguments, than if seen as isolated facts. Here are two themes that ran through the chapter and may help you to organize your review:

1. The mechanisms of transduction and coding All sensory systems respond to physical stimuli by producing action potentials (the process of transduction), and all sensory systems do this in such a way as to preserve useful information about the stimulus (coding). For each sense discussed in this chapter—smell, taste, pain, and hearing—you might think about each of the following questions pertaining to transduction and coding: (a) To what type of physical stimulus does this sense respond, and what is the range of stimuli to which it responds? (b) How is the sensory organ designed for receiving (and possibly concentrating or amplifying) the stimulus? (c) What are the receptors for the stimulus, and how do they respond in such a way as to generate action potentials in sensory neurons? (d) How does the transduction process code the different qualities of stimuli to which the sensory system responds? (e) How do neural mechanisms in the central nervous system alter or reorganize the input and for what purposes? The chapter does not answer all these questions (especially not the last) completely for each sense, but the questions provide a good framework for organizing and thinking about the information that is provided.

2. The survival functions of sensory processes Our sensory systems, like all the basic mechanisms underlying our behavior, evolved through natural selection because they promoted

our ancestors’ survival and reproduction. They are not unbiased recorders of physical energies but biological tools designed to pick out from the sea of energy around us the information that is potentially most useful. We are sensitive to some kinds of energies and not others, and, within the kinds to which we are sensitive, our senses extract and enhance some relationships and not others.

Here are some examples, described in the chapter, of how sensory processes can be understood in terms of their survival advantages: (a) Sensory adaptation (the decline in sensitivity to prolonged, constant stimuli) helps us to ignore stimuli that remain unchanged and to notice changes. (b) Smell and taste work together to produce flavors that are experienced as pleasant or unpleasant, in ways that are generally consistent with what is good for us or bad for us (or for our evolutionary ancestors) to eat. The association of bitter taste with poisons is an example. (c) Pain is a sensory system for warning us when our actions are damaging our tissues and for motivating us to avoid such actions. Evolved mechanisms increase pain sensitivity at times of illness, when it is best to rest, and decrease pain sensitivity at times of threat, when strenuous action without favoring one’s wounds may be necessary. (d) The phonemic-restoration illusion helps us to hear speech in a continuous, meaningful flow and to ignore extraneous noises and interruptions. (e) The fact that our senses obey a power law in converting stimulus magnitude to sensory magnitude may have come about because the power law preserves the constancy of ratios, helping us recognize a pattern in sound or light as the same pattern even when its overall intensity increases or decreases.

Further Reading

HOWARD C. HUGHES (1999). *Sensory exotica: A world beyond human experience*. Cambridge, MA: MIT Press.

The fit between an animal's senses and its habitat and behavior becomes most apparent when we look at animals whose sensory world is very different from our own. This fascinating, well-researched book is about sonar in bats and dolphins, magnetic-field sensitivity in migrating birds, sun compasses in bees and ants, electroreception in fishes, and pheromone communication in insects and mammals.

GARY K. BEAUCHAMP & LINDA BARTOSHUK (Eds.) (1997). *Tasting and smelling*. San Diego: Academic Press.

This brief, scholarly book is clearly written and quite readable by the serious beginning student. It consists of six chapters, each by a different expert or pair of experts, which deal with the transduction mechanisms of taste, coding in taste, psychophysics of taste, psychophysics of smell, clinical disorders of taste and smell, and the development of flavor perception in infants.

SCOTT FISHMAN, WITH LISA BERGER (2000). *The war on pain*. Dublin, Ireland: Newleaf.

In this relatively brief, nontechnical book, Fishman, a physician specializing in the treatment of pain, shares poignant stories of people's suffering and of his efforts (sometimes successful, sometimes not) to relieve their suffering. Along the way, he tells us about the many causes of pain and the many weapons in the arsenal against it, ranging from acupuncture and behavior therapy to drugs and surgery. The epilogue is about making peace with pain in cases where it can't be defeated.

TIFFANY FIELD (2003). *Touch*. Cambridge, MA: MIT Press.

In this fascinating book, based on her own research and that of others, Field describes the psychological functions of the sense of touch. Of greatest interest is research on the medical benefits of touch and massage, including growth-promoting effects of touch in infants.

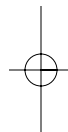
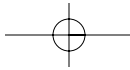
BRIAN C. J. MOORE (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego: Academic Press.

Clear and up-to-date, this paperback textbook discusses all aspects of the physiology and psychology of hearing. Separate chapters are devoted to the auditory perception of loudness, pitch, space, and temporal qualities (such as rhythm). The final two chapters deal with speech perception and with practical applications in such realms as development of sound reproduction systems (such as compact disc players) and aids for the hard of hearing.

S. S. STEVENS (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

A brilliant description of the research leading to the power law (which relates sensory magnitude to stimulus magnitude) and the implications of that law, this slender classic is quite readable by the beginning student who is not intimidated by exponents or algebraic equations. Stevens also shows how the power law applies to magnitude judgments other than those about sensations, such as judgments about the seriousness of crimes.

FPO



The Psychology of Vision

HOW THE EYE WORKS

Functional Organization of the Eye
Differences Between Cone Vision and Rod Vision

SEEING COLORS

How Color Varies with the Physical Stimulus
Two Classic Theories of Color Vision

SEEING FORMS AND PATTERNS

Enhancement of Contours
The Detection and Integration of Stimulus Features
Gestalt Principles of Perceptual Grouping
Evidence That Wholes Can Affect the Perception of Parts

RECOGNIZING OBJECTS

Beyond the Primary Visual Cortex:
Two Streams of Visual Processing
Clues for Object Recognition

SEEING IN THREE DIMENSIONS

Cues for Depth Perception
The Role of Depth Cues in Size Perception

We are visual creatures. Our eyes are our primary gateway for perceiving and understanding the physical world in which we survive. We say “I see,” to mean “I understand,” and when we doubt some claim, we say “I’d have to see it to believe it.” Our visual system provides us with such rich, clear, solid-looking, and generally useful perceptions of the physical world that it is easy for us to forget that the physical world and our sight of it are not one and the same thing. In reality, our visual perceptions are subjective, psychological experiences, which our brains create almost instantly, and continuously as long as our eyes are open, from clues that lie in the patterns of light reflected off objects. The machinery that underlies our ability to produce such perceptions is incredibly complex. Brain scientists estimate that somewhere between 25 and 40 percent of the human brain is devoted exclusively or primarily to the analysis of input from the eyes (Gross, 1998; Sereno & others, 1995). It is no wonder that vision is the sense to which psychologists have paid by far the greatest attention.

Vision begins with activity in the eyes, and that is where this chapter begins. From there we shall go on to examine our abilities to perceive colors, patterns, objects, and depths in three-dimensional space.

How the Eye Works

Life on earth evolved in a world illuminated by the sun during the day and by starlight and the moon’s reflected sunlight at night. Most forms of earthly life are sensitive in one way or another to that light (Land & Fernald, 1992). Even single-celled organisms contain chemicals that respond to light and alter the organism’s activity in survival-promoting ways. In many species of multicellular animals, specialized light-detecting cells called **photoreceptors** evolved and became connected to the animal’s nervous system. Earthworms, for example, have photoreceptors distributed throughout their skin. Stimulation of these cells by light causes the worm to wriggle back down into the earth, where it finds safety, moisture, and food.

1 Through what steps might sophisticated eyes like ours have evolved from primitive beginnings?

<<< Cross-species comparisons, based on homologies (discussed in Chapter 3), suggest that eyes such as ours evolved through the following steps (Gregory, 1996): In some early organisms photoreceptors became concentrated into groups, forming primitive light-detecting organs, or *eye spots*, on the animal’s skin. With further evolution these eye spots became recessed into depressions, or pits. This step reduced glare and increased the animal’s chance of detecting moving boundaries of light and dark that could signal an approaching predator. Subsequent

evolution led to the development of a transparent membrane over such pits, which protected the pits from becoming filled by sand or other debris that could block incoming light. Mutations that thickened the center of this membrane converted it into a crude lens, which at first may have served merely to magnify the light reaching the photoreceptors. With further evolutionary refinement, the lens became capable of projecting an image onto the lining of photoreceptors. Through such gradual steps, coupled with appropriate changes in the nervous system, the primitive ability to detect light and shadows evolved into the more advanced ability to see the shapes of things and eventually into the marvelously complex and precise visual ability that this chapter is about.

Functional Organization of the Eye

The main parts of the human eye are shown in Figure 8.1. The photoreceptors lie in the **retina**, a membrane lining the rear interior of the fluid-filled eyeball. The rest of the eye is a device for focusing light reflected from objects in such a way as to form images on the retina.

Structures at the Front of the Eye Focus Images on the Retina

2 How do the cornea, iris, and lens help to form images on the retina?

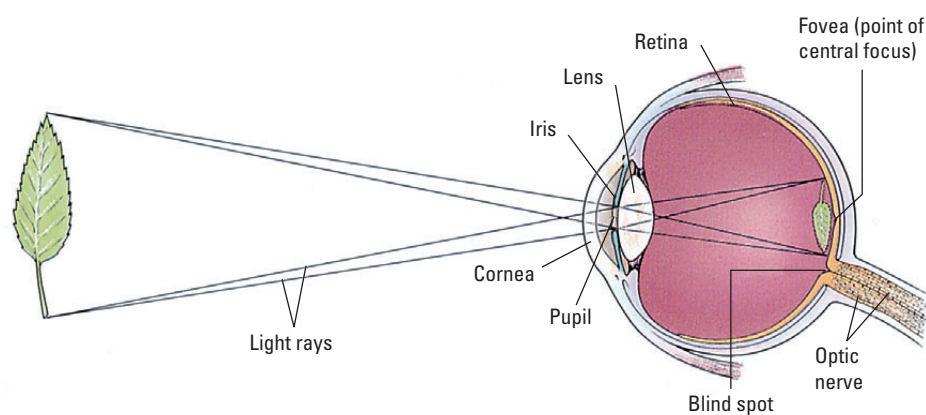
<<< The front of the eyeball is covered by the **cornea**, a transparent tissue that, because of its convex (outward) curvature, helps focus the light that passes through it. Immediately behind the cornea is the pigmented, doughnut-shaped **iris**, which provides the color (usually brown or blue) of the eye. The iris is opaque, so the only light that can enter the interior of the eye is that which passes through the **pupil**, which is simply a hole in the center of the iris. Muscle fibers in the iris enable it to increase or decrease the diameter of the pupil to allow more or less light to enter. Behind the iris is the **lens**, which adds to the focusing process begun by the cornea. Unlike the cornea, the lens is adjustable—it becomes more spherical when focusing on objects close to the eye and flatter when focusing on those farther away. Light rays diverge as they move toward the eye from any given point on a visual object. The focusing properties of the cornea and lens bring the light rays back together at a particular point on the retina, thereby forming an image of the object on the retina. The image is upside down (as shown in Figure 8.1), but that does not matter, since its only purpose is to trigger patterns of activity in neurons running to the brain. The brain is wired to the retina in such a way that input from lower on the retina is interpreted as up, and input from higher on the retina is interpreted as down.

Transduction Occurs in the Retina

3 How are cones and rods distributed on the retina, and how do they respond to light?

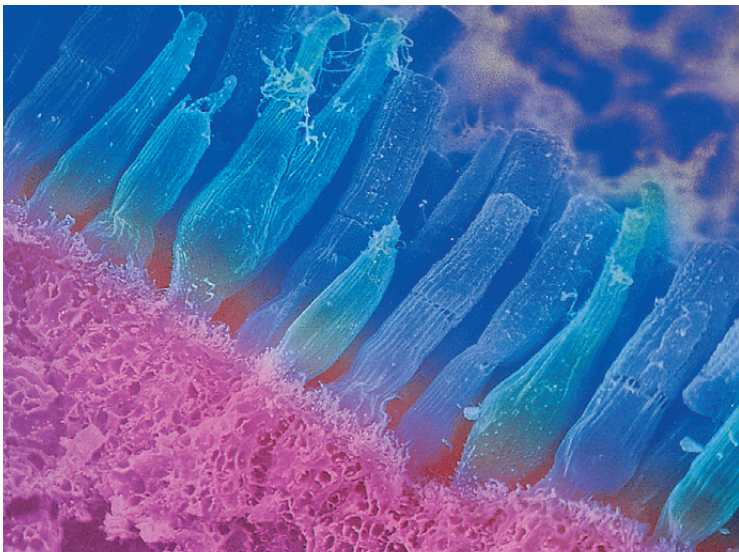
<<< As noted in Chapter 7, the process by which a stimulus from the environment generates electrical changes in neurons is referred to as **transduction**. In vision, transduction is the function of the photoreceptor cells. In each eye millions of

FIGURE 8.1 Cross section of the eye, depicting the retinal image The light rays that diverge from any given point on the surface of an object are brought back together (focused) at a distinct point on the retina to create an image of the object on the retina. This drawing shows light rays diverging and being brought together for just two points on the leaf, but the same process is occurring for light rays coming from every point.



photoreceptor cells are arranged, mosaic-like, in one thin layer of the multilayered retina. These cells are of two types: **cones**, which permit sharply focused color vision in bright light, and **rods**, which permit vision in dim light. They are so named because of their shapes (see Figure 8.2). Cones are most concentrated in the **fovea**, the pinhead-size area of the retina that is in the most direct line of sight (look again at Figure 8.1), which is specialized for high visual *acuity* (the ability to distinguish tiny details). The concentration of cones decreases sharply with increasing distance from the fovea. Rods, in contrast, exist everywhere in the retina except in the fovea and are most concentrated in a ring about 20 degrees away from the fovea (see Figure 8.3). Each human retina contains about 6 million cones and 120 million rods (Wade & Swanston, 1991).

The outer segment of each photoreceptor contains a *photochemical*—a chemical that reacts to light. The rods' photochemical is called **rhodopsin** [roh-dop-sin]. When hit by light, rhodopsin molecules undergo a structural change that triggers a series of chemical reactions in the rod's membrane, which in turn causes an electrical change across the membrane (Schnapf & Baylor, 1987). The transduction process for cones is similar to that for rods, but (for reasons discussed in the section on color vision) three varieties of cones exist and each contains a different photochemical. The electrical changes in rods and cones cause electrical responses in other cells in the retina, which lead to the production of action potentials (neural impulses) in neurons that form the **optic nerve**, which runs from the back of the eye to the brain. At the place on the retina where the axons of these neurons converge to form the optic nerve there is a **blind spot**, due to the absence of receptor cells (shown in Figure 8.3). We normally do not notice the blind spot, but you can demonstrate its existence by following the instructions in Figure 8.4.



Omikron / Photo Researchers

FIGURE 8.2 | Rods and cones
This electron micrograph shows that the photoreceptors are aptly named. Rods are responsible for vision in dim light, and cones for vision in bright light.

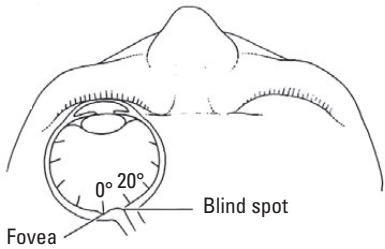
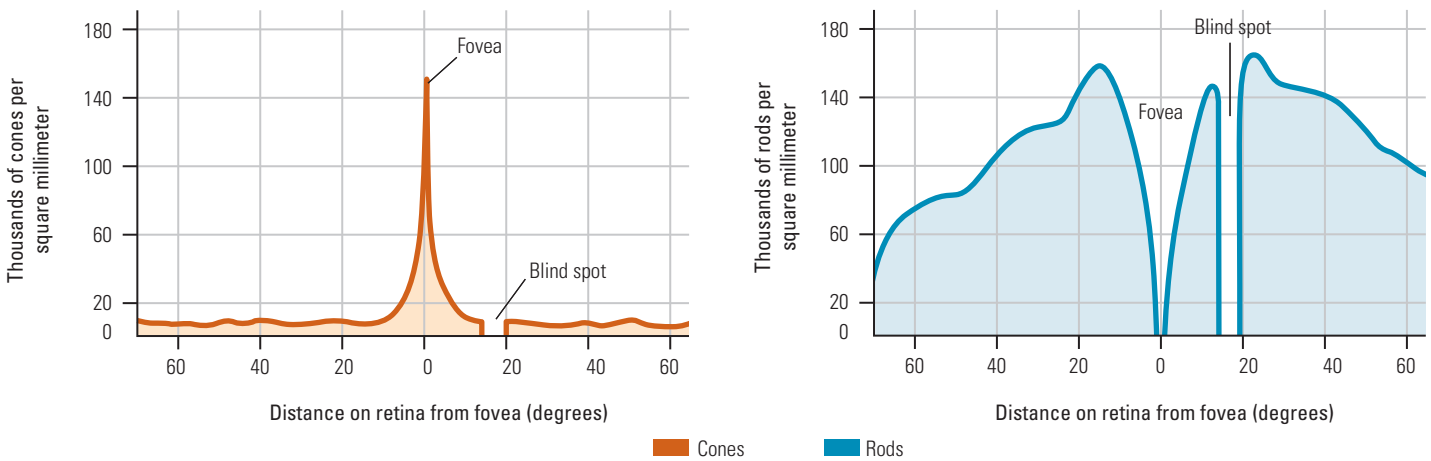


FIGURE 8.3 | Distribution of cones and rods in the retina
Cones are most concentrated in the fovea. Rods are absent from the fovea and most concentrated in a ring 20 degrees away from it. No receptors at all exist in the blind spot.



Lindsay, P. H., & Norman, D. A. (1977). *Human information processing* (2nd ed.). New York: Harcourt Brace Jovanovich

FIGURE 8.4 | Demonstration of the blind spot Close your left eye and focus on the X with your right eye. Start with the page a little more than a foot from your eye and move it gradually closer, still focusing on the X. At about 10 inches, the bird will disappear. At that location, the image of the bird falls on the blind spot of the retina, shown in Figure 8.3. Yet you will probably still see the bars of the cage running across the area where the bird was located, and the background color may fill the space previously occupied by the bird. The bars and color are perceptually filled in by your visual system.

X



Differences Between Cone Vision and Rod Vision

4 How do cone vision and rod vision differ?

<<< Cones and rods provide the starting points for what can be thought of as two separate but interacting visual systems within the human eye. **Cone vision**, also called *photopic vision* or *bright-light vision*, is specialized for high acuity (the ability to see fine detail) and for color perception. **Rod vision**, also called *scotopic vision* or *dim-light vision*, is specialized for sensitivity (the ability to see in very dim light). Rod vision lacks acuity and the capacity to distinguish colors, but, according to calculations from laboratory studies, it is sensitive enough to allow a person on a clear night to detect a single candle flame from 30 miles away if no other lights are present (Galanter, 1962). In terms of evolutionary function, cone vision came about to allow us to see fine details during daylight, and rod vision came about to allow us to see at least the vague outlines of objects at night.

In very dim light, too dim to activate cones, you see only with rod vision; you can make out the general shapes of objects but not their details and colors. In such light you can see dim objects best when you don't look directly at them—because, as noted before, the fovea (the part of the retina in the direct line of sight) contains no rods. If you find yourself in the country on a starry but moonless night, away from all artificial lights, try to stare at the dimmest star that you can see. It will disappear when you look straight at it, but reappear when you look just a little off to one side. You can see it best when looking 20 degrees away, because this angle allows the light from the star to strike the part of the retina where rods are most concentrated.

Roles of Rods and Cones in Dark Adaptation and Light Adaptation

One of the many problems that our visual system has to solve is that of adjusting to the enormous range of light intensities that occur over the course of a day. A white object in sunlight at noon reflects roughly 100 million times as much light as the same object reflects on a starlit but moonless night (Riggs, 1965), yet we can see the object in either condition.

5 What is the chemical basis for dark adaptation and light adaptation? Why do we see mostly with cones in bright light and with rods in dim light?

<<< As you know from experience, it takes time for your vision to adapt to sudden large changes in illumination. The gradual increase in sensitivity that occurs after you enter a darkened room or turn off the lights is called **dark adaptation**, and the more rapid decrease in sensitivity that occurs after you turn on a bright lamp or step out into sunlight is called **light adaptation**. The iris contributes to these adaptive processes by dilating (widening) the pupil in dim light and constricting it in bright light. A fully dilated pupil allows in about 16 times as much light as a fully constricted pupil (Matlin & Foley, 1997). Temporary changes in the sensitivity of visual neurons that receive input from the receptor cells also contribute (Hood,

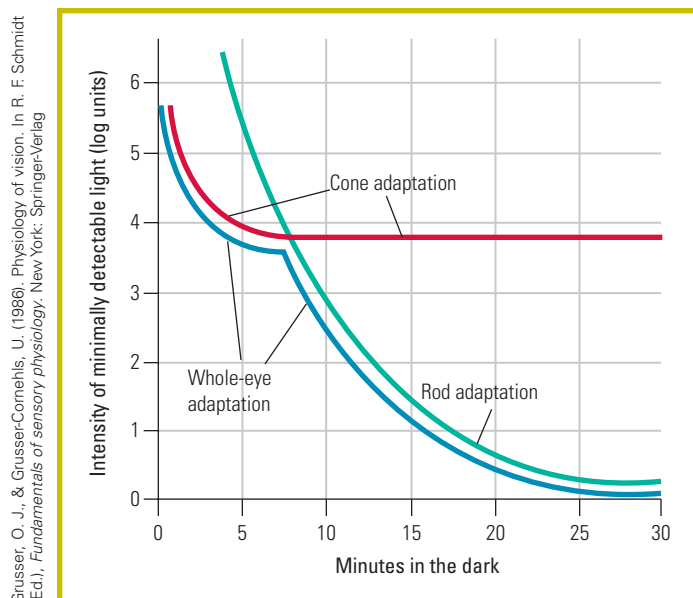


FIGURE 8.5 Dark-adaptation curves The blue curve depicts the minimal intensity of a spot of light that a person with normal vision can see after varying periods in the dark after having been in bright light. The lower the curve, the greater the sensitivity to light. For the first 7 minutes or so, the cones are more sensitive than the rods, but after that the rods are more sensitive. The two-part nature of the curve can be understood by comparing it to the dark-adaptation curve obtained for a person who has only rods (the green curve) and to that obtained if the light is presented in such a way that it strikes only the fovea, where only cones exist (the red curve).

1998). However, the major contribution to dark and light adaptation comes from the different sensitivities of rods and cones.

Rhodopsin, the rod photochemical, is much more sensitive to light than are the cone photochemicals (Kefalov & others, 2003). Bright light causes rhodopsin to break down into two inactive substances, making the rods nonfunctional. In sunlight, therefore, you see entirely with your cones, and even in a typical artificially lit room you see mostly with your cones. When you first step from bright light into a dark room, you may see nothing, because your rods are still inactive and there isn't enough light to activate your cones. With time in the dark, however, rhodopsin regenerates, your rods regain their sensitivity, and you gradually see more and more. It takes about 25 minutes for rhodopsin to regenerate fully. If you then step out of the dark room into sunlight, you are at first blinded by the dazzling light, because your highly sensitive rods are responding maximally and indiscriminately. Within about 5 minutes, however, the rhodopsin breaks down and you begin to see normally, with your less sensitive cones. The cone photochemicals also break down somewhat in bright light and regenerate in dim light and thereby contribute to light and dark adaptation, but this change is smaller than that which occurs in rods (see Figure 8.5).

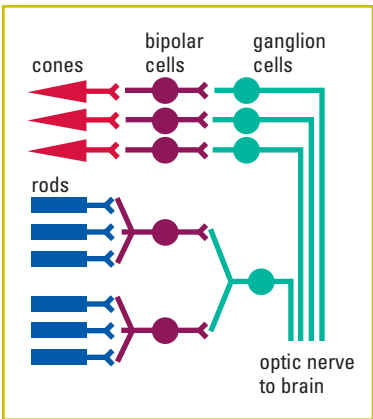
Neural Convergence as a Basis for Differences Between Rod and Cone Vision

The superiority of rods over cones in detecting dim light stems not just from the greater sensitivity of rhodopsin but also from a difference in the neural wiring that connects these receptor cells to the brain. This same difference in wiring also helps explain cone vision's superior acuity.

Both rods and cones form synapses on short neurons called **bipolar cells**, which in turn form synapses on longer neurons called **ganglion cells**, whose axons leave the eye at the blind spot to form the optic nerve. As depicted schematically in Figure 8.6, there is a great deal of *convergence* in the connections from rods to bipolar cells to ganglion cells. Many rods synapse on each bipolar cell, and many bipolar cells synapse on each ganglion cell. Such convergence increases the sensitivity of rod vision in dim light by adding up the output of many rods to generate activity in the ganglion cells. The principle here is similar to that of a funnel, which can collect raindrops from over a relatively wide area and channel them into a single steady stream.

>>> 6 How does the pattern of neural connections within the retina help account for the greater sensitivity and reduced acuity of rod vision compared with cone vision?

FIGURE 8.6 | Difference between rods and cones in neural convergence
This schematic diagram illustrates the fact that there is much more neural convergence between rods and ganglion cells than between cones and ganglion cells in the retina. In the actual human retina, the convergence for rods is much greater than shown here. On average, about 1000 rods converge onto any given ganglion cell in the periphery of the retina (Masland, 2001).



While it increases sensitivity, neural convergence necessarily decreases acuity. Because of convergence, two nearby spots of light striking a peripheral portion of the retina, where rods predominate, will activate the same ganglion cell, with the result that they will be seen as one blurry spot rather than two distinct spots. To be seen as distinct, two spots must activate different ganglion cells, whose axons carry separate messages to the brain. Again, the analogy to a funnel is helpful. Just as two raindrops lose their distinctiveness when they are channeled together by the funnel, two spots in a visual stimulus striking the retina lose their distinctiveness when their neural effects are channeled together by convergence.

In the fovea, where all the receptor cells are cones, there is little or no convergence (Masland, 2001). The result is reduced sensitivity to dim light, because there is no funneling of activity, but high acuity, because each small spot of visual stimulation activates a separate pathway to the brain (again see Figure 8.6).

SECTION REVIEW

The sophisticated human eye focuses light to form images on photoreceptors in the retina.

Basic Structures and Functions in the Eye

- The cornea and the adjustable lens focus incoming light onto the retina at the back of the eye's interior.
- Light energy is then transduced by two types of photoreceptor cells in the retina—rods and cones—which contain photochemicals that react to light with structural changes.
- These changes can lead to further chemical and electrical changes, ultimately triggering action potentials in the cells of the optic nerve.

Cone and Rod Vision

- While cones provide color vision, high visual acuity, and our ability to see in bright illumination, rods provide the sensitivity that allows vision in dim illumination.
- The rate and degree of dark adaptation and light adaptation are different for rods and cones. A result of these differences is that we see only with rods in very dim light and only with cones in bright light.
- Neural convergence helps to explain the greater sensitivity and lower acuity of rod vision compared to cone vision.

Seeing Colors

Many animals—including humans, other primates, birds, bees, and most fishes and reptiles—that depend greatly on vision to identify objects have evolved color vision, which helps to make objects stand out vividly from their backgrounds. The colors that we and other animals see in objects depend on the *wavelengths* of the light that is reflected from those objects.

How Color Varies with the Physical Stimulus

7 How can light be described physically, and what is the relationship of its wavelength to its perceived color?

<<< Light is a form of electromagnetic energy, and like all such energy it can be described as both particles and waves. The particles, or individual packets of light, are called *photons*, and as they travel through space they pulsate in a wavelike way. Light travels at a constant speed (186,000 miles per second), and the distance that photons travel between the beginning of one pulse and the beginning of the next defines the wavelength of the light. The wavelengths of light visible to humans range from about 400 to 700 nm (1 nm, or nanometer, is a billionth of a meter). Shorter waves of electromagnetic energy, below our visible range, include ultra-

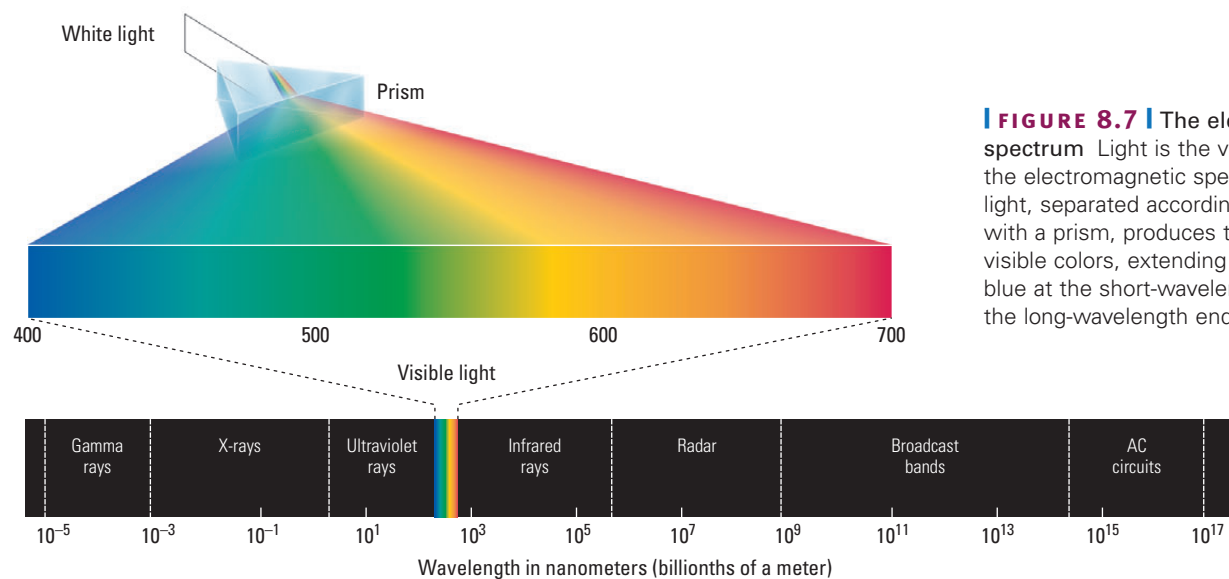


FIGURE 8.7 The electromagnetic spectrum Light is the visible portion of the electromagnetic spectrum. White light, separated according to wavelength with a prism, produces the spectrum of visible colors, extending from violet and blue at the short-wavelength end to red at the long-wavelength end.

violet rays, x-rays, and gamma rays; longer waves, above our visible range, include infrared rays, radar rays, and radio waves.

White light, such as that from the sun, contains all visible wavelengths combined (plus some of the non-visible wavelengths). When the wavelengths in white light are separated—by shining the light through a prism, for example—the visual effect is an array of colors like that of a rainbow, because different wavelengths of visible light are seen as different colors. Progressing from the shortest to the longest wavelengths of visible light, we see the familiar rainbow spectrum, from violet through shades of blue, blue-green, green, green-yellow, yellow, orange, and red, in that order (see Figure 8.7).

Objects vary in the wavelengths of light that they reflect because they have on their surfaces different *pigments*, chemicals that absorb some wavelengths and thereby prevent them from being reflected. A pigment that absorbs short and medium waves, for example, appears red, because only long (red-appearing) waves are reflected. Similarly, a pigment that allows only short waves to be reflected appears violet or blue, and one that allows only medium waves to be reflected appears green or yellow. A surface that reflects all wavelengths about equally appears white, gray, or black, depending on whether the relative amount of each wavelength reflected is, respectively, high (none absorbed), moderate (some absorbed), or low (all absorbed).

Subtractive Color Mixing

Because pigments create the perception of color by *subtracting* (absorbing) some of the light waves that would otherwise be reflected to the eye, the mixing of pigments is called **subtractive color mixing**. As illustrated in Figure 8.8 (on page 270), if a blue pigment, which absorbs long waves, is mixed with a yellow pigment, which absorbs short waves, only medium-length waves will be reflected, and the mixture will be seen as green. When you were a child playing with watercolors, you probably proved the basic facts of subtractive color mixing many times. You may remember being disappointed when you mixed all the paints together and produced something pretty close to black rather than the brilliant reddish-yellowish-greenish-blue that you had hoped for. In that experiment you subtracted all the wavelengths by combining all the pigments.

Two Psychological Laws of Additive Color Mixing

The opposite of subtractive color mixing is **additive color mixing**, which occurs when colored lights rather than pigments are mixed. Additive color mixing can be

8 How do pigments affect the perceived color of an object in white light? How does the mixing of pigments affect color by subtracting from the light that is reflected to the eye?

9 How does additive color mixing differ from subtractive color mixing? What are the two laws of additive color mixing, and how is each illustrated in the standard chromaticity diagram?

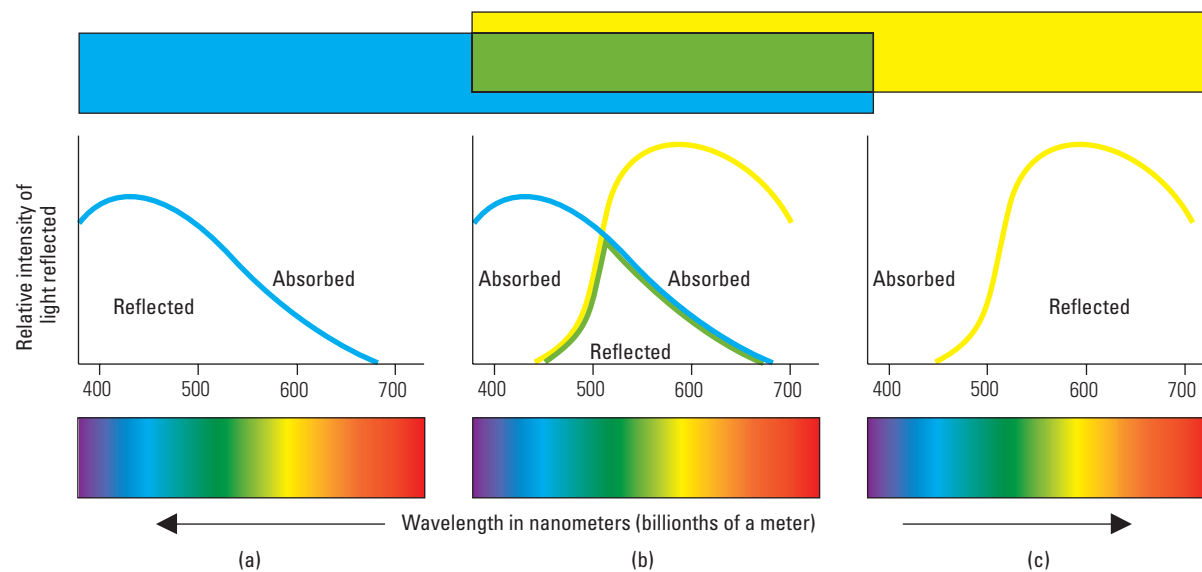


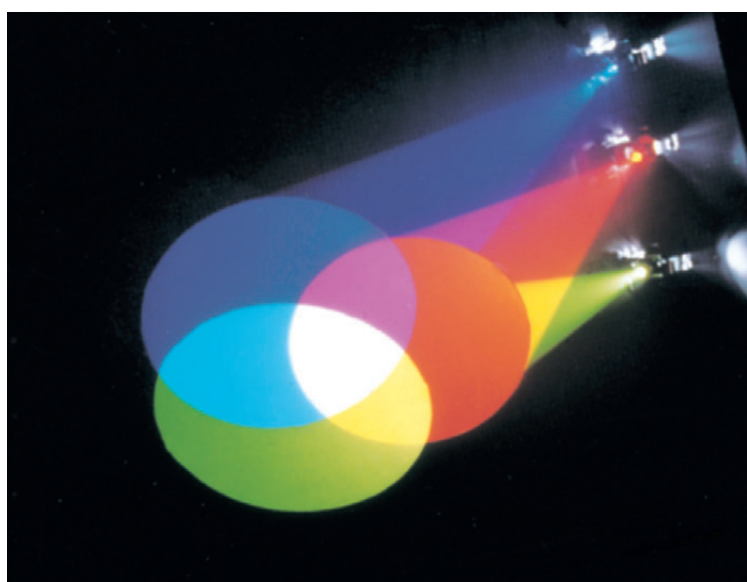
FIGURE 8.8 | Subtractive color mixing In this example, the blue pigment (a) absorbs most of the light that has wavelengths above 550 nm, and the yellow pigment (c) absorbs most of the light that has wavelengths below 500 nm. When the two pigments are mixed (b), the only light that is not strongly absorbed is that with wavelengths lying between 500 and 550 nm. This is the light that will be reflected, causing the mixture to appear green.

demonstrated by shining two or more beams of light of different wavelengths at the same spot on a white screen; the screen then reflects them back mixed together (see Figure 8.9). By the early eighteenth century, scientists had discovered that additive color mixing follows two quite surprising laws—the three-primaries law and the law of complementarity—neither of which could be predicted from anything that one could know about the physics of light. These early researchers realized that these laws of color mixing are laws of psychology, not physics.

According to the **three-primaries law**, three different wavelengths of light (called *primaries*) can be used to match any color that the eye can see if they are mixed in the appropriate proportions. The primaries can be any three wavelengths as long as one is taken from the long-wave end of the spectrum (red), one from the short-wave end (blue or violet), and one from the middle (green or green-yellow). According to the **law of complementarity**, pairs of wavelengths can be found that, when added together, produce the visual sensation of white. The wavelengths of light in such a pair are referred to as *complements* of each other.

All the facts associated with the two laws of additive color mixing are taken into account in the *standard chromaticity diagram*, shown in Figure 8.10. The colors along the curved edge of the diagram are produced by single wavelengths and are

FIGURE 8.9 | Additive color mixing Additive color mixing occurs when lights of different wavelengths are mixed by shining them together on a surface that reflects all wavelengths. By varying the intensity of the three lights shown here, it would be possible to produce all the colors that the eye can see.



Fritz Goro, LIFE Magazine © 1944 Time Warner, Inc.

called *saturated* colors. As you move from any point near the edge toward the white center, the color comes more and more to resemble white; that is, it becomes increasingly *unsaturated* (white is regarded as fully unsaturated). For example, as you go along the line from the 620-nm point on the edge to the center, you go from red (produced by light with a wavelength of 620 nm) to progressively whiter shades of pink (unsaturated red) to white. (Because of the limitations of color printing, Figure 8.10 does not show the gradualness of this change in saturation, but you can imagine it.) The figure's caption describes how the diagram can be used to determine (a) the color that will result from mixing the three standard primaries in any given proportion and (b) which pairs of wavelengths are complements of each other.

It is important and rather exciting to realize that the facts of color mixing portrayed by the chromaticity diagram are facts of *psychology*, not physics. The wavelengths of the three primaries do *not* become physically blended into one wavelength when added together to match the color produced by a fourth wavelength. A machine that detects wavelengths has no difficulty distinguishing, say, a 550-nm light from the mixture of three primaries that would exactly match its greenish-yellow color. Similarly, when two complementary wavelengths are mixed to produce the sensation of white, they do not physically produce white light (which contains all the wavelengths). Instead, such color matches, in which physically distinct stimuli look identical, occur because of processes in the eye and the brain. Indeed, the matches just described provided the insight that led to two theories of how the eye and brain produce color vision.

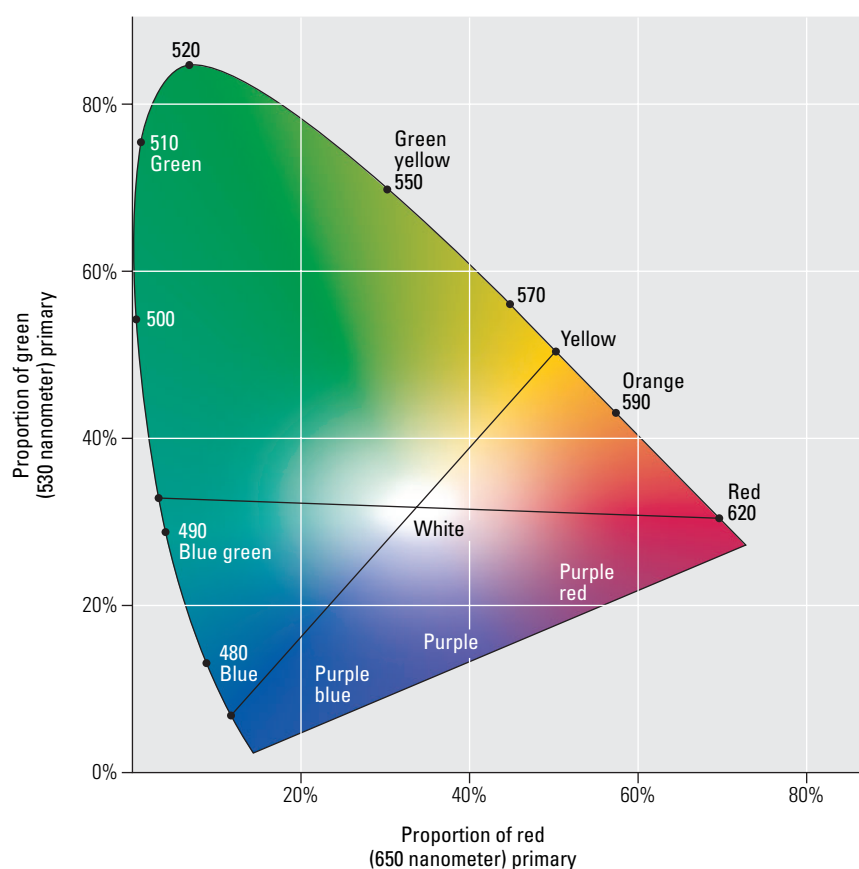


FIGURE 8.10 The standard chromaticity diagram All the facts of additive color mixing—related to both the three-primaries law and the law of complementarity—are summarized in this diagram. The three primaries here are lights of 460 nm (blue), 530 nm (green), and 650 nm (red). The proportions of red and green primaries that must be added to the blue primary to match any given color on the diagram are shown, respectively, on the horizontal and vertical axes. The proportion of blue can be calculated by subtracting the other two proportions from 100 percent. For example, if you wish to match the blue-green produced by a 490-nm light, the figure indicates that your mixture must contain about 5 percent red primary, 30 percent green primary, and 65 percent blue primary ($100\% - 5\% - 30\% = 65\%$). As another example, the figure shows that the best white is produced by approximately equal (33.3 percent) proportions of the three primaries.

The chromaticity diagram can also be used to find all possible pairs of complementary colors. Two colors are complementary if their additive mixture produces white. In the diagram, the possible colors produced by mixing any two wavelengths lie along the straight line connecting the points representing the two wavelengths on the diagram. Thus, the two ends of any straight line passing through the white center of the diagram represent complementary colors. Two such lines are drawn on the figure for purposes of illustration. Notice that wavelengths in the red to orange part of the spectrum have their complements in the green to blue-green part and that wavelengths in the yellow part of the spectrum have their complements in the blue part.

Two Classic Theories of Color Vision

The two theories of color vision, both developed in the nineteenth century, are the trichromatic theory and the opponent-process theory. For many years these two theories were thought to be contradictory to one another, but we now know that both are true.

The Trichromatic Theory

10 How does the trichromatic theory explain the three-primaries law? How was the theory validated by the discovery of three cone types?

According to the *trichromatic theory*, color vision emerges from the combined activity of three different types of receptors, each most sensitive to a different range of wavelengths. This idea was proposed first (in 1802) by Thomas Young and later by Hermann von Helmholtz (1852) as an attempt to explain the three-primaries law of additive color mixing. Young and Helmholtz reasoned that if every color we see is the result of a unique proportion, or ratio, of activity among three types of receptors, then the three-primaries law would be an inevitable result: It would be possible to match any visible color by varying the relative intensities of three primary lights, each of which acts maximally on a different type of receptor. Young and Helmholtz developed their theory purely from behavioral data on perceptual effects of color mixing, at a time when nothing was known about photoreceptors in the retina. We now know from physiological studies that their theory was correct. Three types of cones indeed exist in the human retina, each with a different photochemical that makes it most sensitive to the light within a particular band of wavelengths.

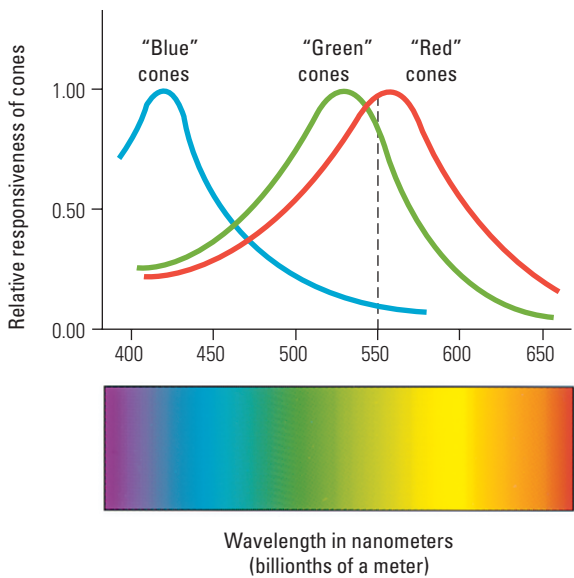
In Figure 8.11 you can see an approximation of the actual sensitivity curves for each type of cone. The cones are labeled “blue,” “green,” and “red,” after the color that is experienced when that type of cone is much more active than the other types. Notice that any given wavelength of light produces a unique ratio of activity in the three types of cones. For example, a 550-nm light, which is seen as greenish-yellow, produces a slightly larger response in “red” cones than in “green” cones and a very low response in “blue” cones. That same ratio of response in the three cone types could be produced by shining into the eye a mixture of red, green, and blue primaries, with the first two much more intense than the last. The result would be a perceptual experience of greenish-yellow indistinguishable from that produced by the 550-nm light.

Exceptions to Trichromatic Vision

11 Why does vision in some people obey a two-primaries law rather than the three-primaries law, and why are these people not good at picking cherries? How does the color vision of most non-primate mammals, and that of most birds, differ from that of most humans?

Some people, referred to as *dichromats*, have only two, not three, types of cone photochemicals. These people, as you might expect, obey a two-primaries law of color mixing rather than the three-primaries law. For them, any color that they can

FIGURE 8.11 How the three types of cones respond to different wavelengths of light Any given wavelength produces a unique ratio of activity in the three cone types, and that ratio provides the initial code that permits us to see different wavelengths as different colors. For example, a 550-nm light, which is seen as greenish-yellow, produces a slightly larger response in “red” cones than in “green” cones and a very little response in “blue” cones. Any combination of lights that would produce that same ratio of responses would be seen as greenish-yellow. (For precise data, see Bowmaker & Dartnall, 1980, or Merbs & Nathans, 1992.)



see can be matched by varying the proportion of just two different wavelengths of light. The most common forms of dichromia involve the absence of the normal photochemical for either the “red” or the “green” cones (usually the latter) due to a defect in the gene that produces that photochemical (Neitz & others, 1996). Because the defective gene is recessive and the genes for both the “red” and the “green” photochemicals are located on the X chromosome, this trait appears much more often in men than in women. Men (as discussed in Chapter 3) have only one X chromosome, inherited from the mother, so a single defective gene on that chromosome can produce color blindness. Women have two X chromosomes, one inherited from each parent, and are color-blind only if the defective gene exists on both of them.

Approximately 5 percent of men and 0.25 percent of women lack either the “red” or “green” cone photochemical and are *red-green color-blind*, meaning that they have difficulty distinguishing colors ranging from green to the red end of the spectrum (Masland, 2001). If you look again at Figure 8.11, you will see why this is so. The normal ability to distinguish colors in this range (from about 500 to 700 nm) is mediated almost entirely by differential activity in the “red” and “green” cones, because “blue” cones are quite inactive in this range. If either the “red” or the “green” cone photochemical is missing, then the person has no or little physiological basis for distinguishing one wavelength from another in this range. Many people who have red-green color blindness don't know it and may wonder why certain perceptual tasks that are hard for them are easy for others. One man's red-green color blindness was not discovered until he told his family how much he admired the perceptual skill of cherry pickers: “After all,” he said, “the only thing that tells 'em it's a cherry is . . . that it's round and the leaves aren't. I just don't see how they find 'em in those trees!” (Coren & Ward, 1989).

Most non-primate mammals, including dogs and cats, have just two types of cones, corresponding to our “blue” and “green” cones, and have the same difficulty discriminating among colors at the long-wavelength end of the spectrum as do people who are red-green color blind (Kolb, 2003). Most birds, in contrast, have four types of cones (Dalton, 2004). Birds rely heavily on vision to guide their flight and to locate food, and they have evolved color vision that is better than ours. In addition to “red,” “green,” and “blue” cones, they have a fourth set of cones that are maximally sensitive to wavelengths in the ultraviolet range (shorter than 400 nm), which we cannot see (see Figure 8.12).

The Opponent-Process Theory

The trichromatic theory explains the three-primaries law and certain types of color blindness well, but it does not explain the law of complementarity—how certain pairs of wavelengths produce the experience of white. To explain that, Ewald Hering, another nineteenth-century scientist, developed the *opponent-process theory*. Hering was most impressed by the observation that complementary



David Frazier / The Image Works

The value of color vision Our sensitivity to the wavelength of reflected light helps us to distinguish objects of interest from their background.

FIGURE 8.12 | Ultraviolet mouths These starling nestlings' gaping beaks and mouths would appear black to us, but, as shown in this UV photograph, they reflect ultraviolet light intensely. They are ideally designed to attract the attention of the mother starling without attracting that of mammalian predators who cannot see ultraviolet light. (From Dalton, 2004, p. 596.)



Bengt Lundberg / Nature Picture Library

12

How does the opponent-process theory explain (a) the law of complementarity in color mixing and (b) the complementarity of afterimages?

colors (blue and yellow, or green and red) seem to swallow each other up, erasing each other's color, when added together. For example, if you begin with blue light and gradually add more of its complement (yellow), the result is not “bluish-yellow” but an ever-paler (more unsaturated) blue, which finally becomes white. To explain such observations, Hering (1878/1964) proposed that color perception is mediated by physiological units (which we now call neurons) that can be either excited or inhibited, depending on the wavelength of light, and that complementary wavelengths have opposite effects (that is, they activate “opposing processes”) on these opponent-process units.

More specifically, Hering's proposal (in modern language) was that the ability to see blues and yellows is mediated by blue-yellow opponent neurons, which are excited by wavelengths in the blue part of the spectrum and inhibited by those in the yellow part, or vice versa. Similarly, he proposed that the ability to see greens and reds is mediated by green-red opponent neurons, which are excited by wavelengths in the green part of the spectrum and inhibited by those in the red part, or vice versa. In addition, he proposed that the ability to distinguish bright from dim light, independent of wavelength, is mediated by a third set of neurons (brightness detectors), which are excited by lights of any wavelength. This theory nicely accounts for the facts of complementary colors. A mixture of wavelengths from the blue and yellow parts of the spectrum, or from the green and red parts, appears white (colorless but bright) because the two sets of wavelengths cancel each other out in their effects on color detectors but act in concert to excite brightness detectors.

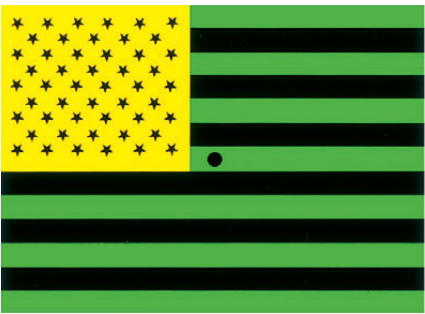
Color Afterimages Explained by the Opponent-Process Theory

The opponent-process theory also accounts wonderfully for another psychological phenomenon, the *complementarity of afterimages*. For a demonstration of this phenomenon, follow the instructions in the first paragraph of the caption of Figure 8.13. You will see that the colors in the afterimage are the complements of those in the original: What was green becomes red; what was yellow becomes blue; and what was black becomes white. How does the opponent-process theory explain this phenomenon?

Consider the example of green becoming red in the afterimage. The neurons in the retina that respond most strongly to the green-appearing (middle-wavelength) light as you stare at the picture become fatigued. Therefore, when you shift your eyes to the white paper (which reflects all wavelengths), those neurons don't respond as strongly as they normally would, but other neurons, including those that respond to red-appearing (long-wavelength) light, do respond strongly. Thus, opponent-process neurons that are normally excited by red-appearing light and inhibited by green-appearing light in that part of the retina become excited, resulting in the perception of red. To convince yourself that this adaptive process occurs early in the visual pathway, before the inputs from the two eyes converge, follow the instructions in the second paragraph of the caption of Figure 8.13.

FIGURE 8.13 | Complementarity of afterimages Stare at the dot in the middle of the flag for at least half a minute. Then look at the dot on the white space beside the flag. What do you see?

To demonstrate that this effect involves a change within the retina, or at least somewhere early in the visual system before the input from the two eyes converges, repeat the demonstration but this time keep one eye closed as you stare at the middle of the flag. Then look at the dot in the white space first with one eye and then with the other. What happens?



A Physiological Reconciliation of the Two Theories

For many years the trichromatic and opponent-process theories were thought to be contradictory, but in the 1950s and 1960s research showed that both theories are fundamentally correct (De Valois & others, 1966; Hurvich & Jameson, 1957). The retina indeed contains three types of cones, consistent with Young and Helmholtz's trichromatic theory. But the cones feed into ganglion cells (the neurons of the optic nerve) in a pattern that translates the trichromatic code into an opponent-process code, conforming to Hering's theory. Some ganglion cells behave in a red-green opponent manner: They are excited by input from "red" cones and inhibited by input from "green" cones, or vice versa. Others behave in a blue-yellow opponent manner: They are excited by input from "blue" cones and inhibited by input from both "green" and "red" cones (which, in combination, respond best to yellow light), or vice versa. To see how connections from the three cone types could produce such opponent-process cells, look at Figure 8.14. Neurons in the cerebral cortex that are most directly involved in color perception maintain these opponent-process characteristics (Dacey, 2000).

The research and theories on color vision just presented are a wonderful early illustration of the value of combining behavioral and physiological research. The trichromatic and opponent-process theories were developed, in the nineteenth century, from behavioral evidence having to do with the perceptual effects of additive color mixing, before anything was known about the physiology of receptors and neurons. Later, both theories were confirmed physiologically, and today neuroscientists and psychologists are working out the finer details of the theories.

>>> **13**
How has the opponent-process theory been validated in studies of the activity of neurons that receive input from cones?

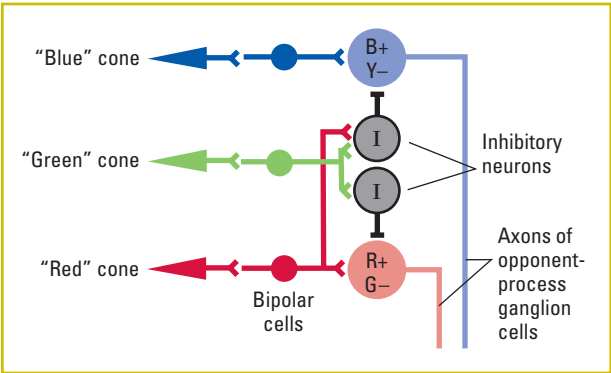


FIGURE 8.14 | Reconciliation of the trichromatic and opponent-process theories Shown here is a schematic diagram illustrating how connections from the three types of cones can result in opponent-process cells—cells that are excited by blue and inhibited by yellow (B+Y−) and cells that are excited by red and inhibited by green (R+G−). Excitatory neural connections are shown as Y-shaped axonal endings, and inhibitory connections are shown as straight lines perpendicular to the axon. The cells marked “I” are inhibitory intermediary neurons. As an exercise, try constructing diagrams that would produce B−Y+ and R−G+ opponent neurons.

SECTION REVIEW

The experience of color results from the visual system’s response to wavelengths of light.		
Wavelength and Color • We experience different wavelengths of light as different colors. • Objects appear colored because their pigments absorb some wavelengths from white light and reflect others. • Mixing pigments is called subtractive color mixing because each pigment removes the wavelengths it absorbs. • Additive color mixing involves mixing lights of different wavelengths. It is characterized by the three-primaries law and the law of complementarity.	Trichromatic Theory of Color Vision • This theory was derived from the three-primaries law. It states that we have three types of receptors, each most responsive to a particular range of wavelengths. • The theory was confirmed by the identification of three types of cones, each with a different sensitivity curve.	Opponent-Process Theory of Color Vision • This theory was derived from the law of complementarity. It states that there are physiological units that are affected in opposite ways (are excited or inhibited) by complementary wavelengths. • The theory explains the complementarity of afterimages. • The theory was confirmed by the discovery of visual neurons that behave just as the theory predicted.

Seeing Forms and Patterns

The purpose of human vision is not to detect the simple presence or absence of light or different wavelengths, but to identify meaningful objects and actions. As I look at the top of my desk, my retinas are struck by a continuous field of light, varying from point to point in intensity and wavelength. But I see the scene neither as a continuous field nor as a collection of points. Instead, I see objects: a computer, a pencil, a stapler, and a pile of books. Each object stands out sharply from its background. My visual system has sorted all the points and gradations that are present in the reflected light into useful renditions of the objects on my desk. It has provided me with all the information I need to reach out and touch, or pick up, whichever object I want to use next. How does my visual system accomplish this remarkable feat?

Vision researchers generally conceive of object perception as a type of unconscious problem solving, in which sensory information provides clues that are analyzed using information that is already stored in the person's head. In this section and the next, we shall be concerned with how our visual system organizes sensory information in ways that allow us to see forms, patterns, and objects.

Enhancement of Contours

14

What is the value of the visual system's ability to exaggerate contrast at contours, and how can that ability be demonstrated? How is contrast enhancement explained by lateral inhibition?

<<< Objects are defined principally by their *contours*, that is, by their edges or borders. Contours give objects their distinctive shapes. The contours of the pencil on my desk, for example, consist of two long, parallel edges, which are connected at one end by the curved edge of the pencil's eraser and at the other end by the two converging edges of the pencil's sharpened point. That outline, by itself, would allow me to distinguish the pencil from other objects on my desk. Visually, contours are sudden changes in brightness or color that separate objects from their backgrounds.

Exaggeration of Contrast at Contours

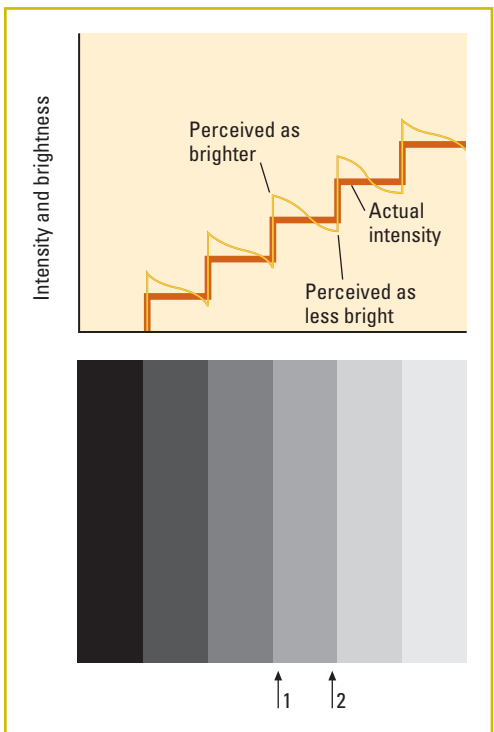
As an aid in detecting objects, our visual system has evolved in such a way that it exaggerates the amount of physical change that occurs across contours. To experience that effect, look at Figure 8.15. Each stripe is physically uniform in its *lightness* (that is, in the actual amount of light it reflects), but our perception of it is not uniform. Each stripe appears to be lighter near its boundary with a darker stripe, and darker near its boundary with a lighter stripe, than it is in the middle. Compare the perceived lightness just above arrow 1 with that just above arrow 2, for example. The graph above the stripes shows how the actual lightness and the perceived lightness change from stripe to stripe. The exaggeration of contrast that our visual system imposes on these stripes helps us to see each one as distinct from its neighbors.

Lateral Inhibition as a Mechanism for Enhancing Contours

Neuroscientists have discovered that enhancement of contrast, such as that illustrated in Figure 8.15, occurs through a process of *lateral inhibition*. Some neurons in the retina and in visual areas of the brain have inhibitory connections with their neighboring neurons. The result is that activity in any given neuron declines when its neighboring neurons are active, by an amount that is proportional to the degree of activity in those neighbors.

FIGURE 8.15 | Enhancement of contrast Each solid gray stripe appears lighter near its boundary with a darker stripe and darker near its boundary with a lighter stripe. This effect is believed to result from lateral inhibition among neighboring neurons in the visual system. (Adapted from Matlin & Foley, 1997.)

Matlin, M. W. (1988). *Sensation and perception* (2nd ed.). Boston: Allyn & Bacon. Copyright © 1988 by Allyn and Bacon. Adapted by permission



To understand how lateral inhibition can enhance contrast at edges, imagine a person looking at the two patches of gray shown in Figure 8.16. Neighboring portions of this stimulus activate neighboring neurons in the visual system, and neighboring neurons tend to inhibit one another. For simplicity, imagine four non-neighboring neurons—A, B, C, and D—which are excited, respectively, by light coming from the circular spots labeled A, B, C, and D in the figure. In this example, neuron A, which is excited by a spot in the middle of the lighter patch, will be less active than neuron B, which is excited by a spot in the lighter patch near its border with the darker patch. Both neurons receive the same amount of excitatory input, but neuron A is completely surrounded by very active neighbors, while neuron B is only partly surrounded by very active neighbors. Some of B's neighbors are receiving input from the darker patch and are therefore less active than A's neighbors. As a result, neuron B receives less inhibitory input than does neuron A and is more active.

Similar logic, applied to the neurons responding to light reflected from the darker patch, shows that neuron D, next to the contour, will be less active than neuron C. Both neurons receive the same (relatively small) amount of excitatory input, but neuron D receives more inhibitory input because some of its neighbors are stimulated by the lighter patch and are therefore very active. The result of all this, compounded over the thousands of neurons that receive input from these gray patches, is an exaggeration of the contrast at the border. The light patch appears lighter near its border with the dark patch than elsewhere, and the dark patch appears darker near its border with the light patch than elsewhere.

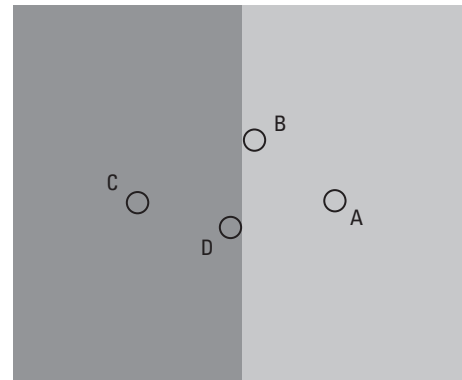


FIGURE 8.16 | Neural explanation of contrast enhancement In this figure, spots A, B, C, and D represent areas of the visual stimulus that activate neurons A, B, C, and D, respectively, in the visual system. As explained in the text, because of lateral inhibition, neuron B in this example will be more active than neuron A, and neuron D will be less active than neuron C. Such differences in neural activity, over all of the neurons that receive input from this stimulus, account for contrast enhancement at edges.

The Detection and Integration of Stimulus Features

Any object that we see can be thought of as consisting of a set of elementary stimulus features. The features may include the various straight and curved lines that form the object's contours, the brightness and color of the light that the object reflects, and the object's movement or lack of movement with respect to the background. Somehow our visual system registers all these features and brings them together to form one unified perception of the object. Over the past 50 years, researchers have performed many experiments aimed at understanding how the brain detects and integrates the elementary stimulus features of objects.

Feature Detectors in the Visual Cortex

One route to learning about how the brain registers stimulus features involves recording the activity of individual neurons in visual areas of the brain. As depicted in Figure 8.17, ganglion cells of the optic nerve run to the thalamus, in the middle of the brain, and synapse there on other neurons that carry their output to the **primary visual area** of the cerebral cortex (also introduced in Chapter 5, Figure 5.17). Within the primary visual area, millions of neurons are involved in analyzing the sensory input. Researchers have lowered thin electrodes into individual neurons in this area of the cortex, in animals, in order to record the neurons' rates of action potentials as various visual stimuli are presented to one or both of the animal's eyes. In a typical experiment the animal—usually a cat or a monkey—is anesthetized, so it is unconscious, but its eyes are kept open and the neural connections from the eyes to the relevant brain neurons are intact.

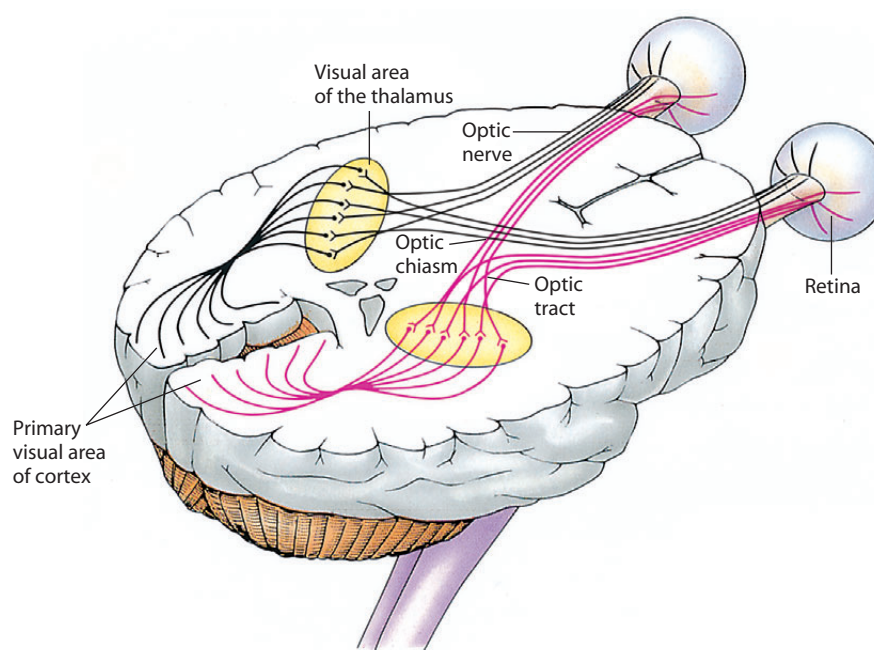


FIGURE 8.17 | Pathway from the eyes to the primary visual cortex Neurons in the optic nerves come together at the optic chiasm at the base of the brain and form the optic tracts, which run to nuclei in the thalamus, where they synapse on neurons that run to the primary visual area of the cerebral cortex.

15 What sorts of stimulus features influence the activity of neurons in the primary visual cortex?

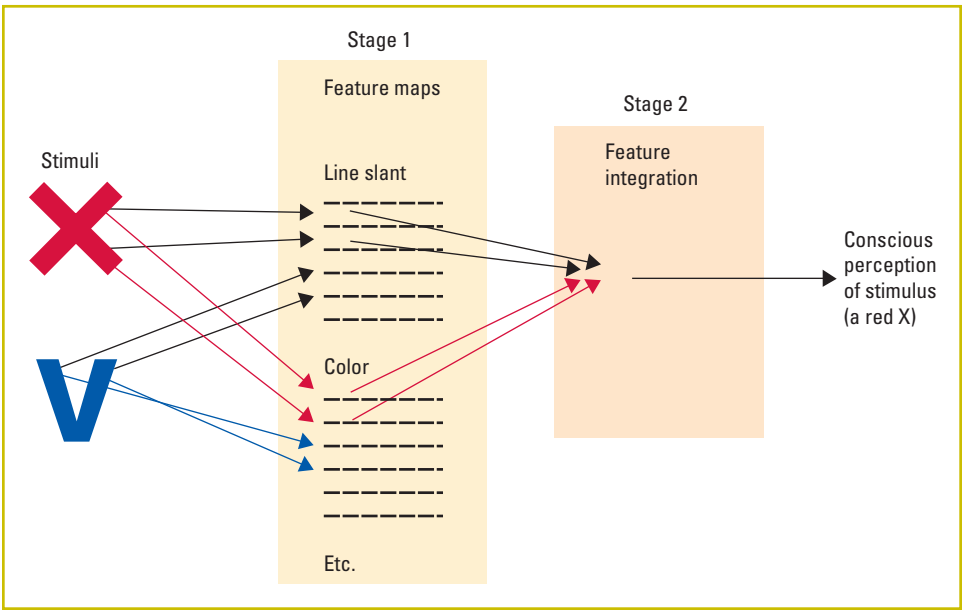
<<< In a classic series of experiments of this sort, David Hubel and Torsten Wiesel (1962, 1979) recorded the activity of individual neurons in the primary visual cortex of cats while the cats' eyes were stimulated by simple black-and-white patterns of various shapes, sizes, and orientations. They found that different neurons responded preferentially to different patterns. For example, some neurons responded best to stimuli that contained a straight contour separating a black patch from a white patch. Hubel and Wiesel referred to these neurons as edge detectors. Other neurons responded best to a narrow white bar against a black background, or to a narrow black bar against a white background. The researchers referred to these as bar detectors. Hubel and Wiesel found, moreover, that any given edge or bar detector responded best to a particular orientation of the edge or bar. Some responded best when it was oriented vertically, others responded best when it was oriented horizontally, and still others responded best when it was slanted at a specific angle.

Subsequent research showed that neurons in the primary visual cortex are sensitive not just to the orientation of visual stimuli, but also to other visual features, including color and rate of movement. Thus, one neuron might respond best to a yellow bar on a blue background, tilted 15 degrees clockwise and moving slowly from left to right. Recording from visual cortical areas just outside of the primary area, researchers have identified individual neurons that are most sensitive to contours that are more complex than those described for the primary visual area. For example, they have found neurons that respond most strongly to arcs or circles of a specific radius and other neurons that respond most strongly to angles of a specific number of degrees (Hegd  & Van Essen, 2003). Taken as a whole, the neurons of the primary visual cortex and nearby areas seem to keep track of all the bits and pieces of visual information that would be available in a scene. Because of their sensitivity to the elementary features of a scene, these neurons are often referred to as *feature detectors*.

16 What is the difference between parallel processing and serial processing? What role does each play in Treisman's feature-integration theory of perception?

<<< Some years ago, Anne Treisman (1986, 1998) developed a theory of visual perception that she called a *feature-integration theory*. She developed the theory partly from neurophysiological evidence concerning feature detectors, but mostly from behavioral evidence concerning the speed with which people can perceive various stimuli. The theory begins with the assertion that any perceived stimulus, even a simple one such as the X shown in Figure 8.18, consists of a number of distinct

FIGURE 8.18 Treisman's theory of feature detection and integration During stage 1 of visual processing, in Treisman's theory, the primitive features of all stimuli that reach the eyes are registered automatically and simultaneously—that is, in parallel. In this example, all the features of both the X and the V are registered simultaneously in the appropriate feature-mapping areas of the brain. Integration of features occurs at stage 2, during which information is processed serially, from one localized area of the visual field at a time. In this example, only the information from one stimulus, the X, is being processed. An instant later, stage 2 might operate on the V, but it cannot operate on the X and V at the same time.



primitive sensory features, such as its color and the slant of its individual lines. To perceive the stimulus as a unified entity, the perceptual system must detect these individual features and integrate them into a whole. The essence of Treisman's theory is that the detection and integration occur sequentially, in two fundamentally different steps or stages of information processing.

The first stage in Treisman's theory is the *detection of features*, which occurs instantaneously and involves **parallel processing**. Parallel processing means that this step operates simultaneously on all parts of the stimulus array. That is, according to Treisman, our visual system picks up at once all the primitive features of all the objects whose light rays strike our retinas. Even if we are paying attention just to the X in Figure 8.18, our visual system picks up at the same time the primitive features of the V in that figure and of all other stimuli in our field of view.

The second stage is the *integration of features*, which requires more time and leads eventually to our perception of whole, spatially organized patterns and objects. This step involves **serial processing**, which occurs sequentially, at one spatial location at a time, rather than simultaneously over the entire array. When looking at Figure 8.18, we can integrate the features of the X and then, an instant later, the features of the V, but we cannot integrate the two sets of features simultaneously.

Research Support for Treisman's Theory

To understand some of the evidence on which Treisman based her theory, look at the array of stimuli in Figure 8.19a. Notice that no effort is needed to find the single slanted line. You don't have to scan the whole array in serial fashion to find it; it just “pops out” at you. According to Treisman, this is because line slant is one of the primitive features that are processed automatically through parallel processing. Now look at Figure 8.19b and find the single slanted green line among the vertical green lines and slanted red lines. In this case the target does not pop out; you have to scan through the array in serial fashion to find it (though you can still find it quite quickly).

In controlled experiments, Treisman and Stephen Gormican (1988) measured the time it took for people to locate specific target stimuli in arrays like those of Figure 8.19 but with varying numbers of *distractors* (defined as the non-target stimuli). As long as the target differed from all the distractors in one or more of Treisman's list of primitive features—such as slant, curvature, color, brightness, and movement—subjects detected it equally quickly no matter how many distractors were present. This lack of effect of the number of distractors on detection time is indicative of parallel processing. In contrast, when the target did not differ from the distractors in any single visual feature, but differed only in its conjoining of two

17
How do pop-out phenomena and mistakes in joining features provide evidence for Treisman's theory?

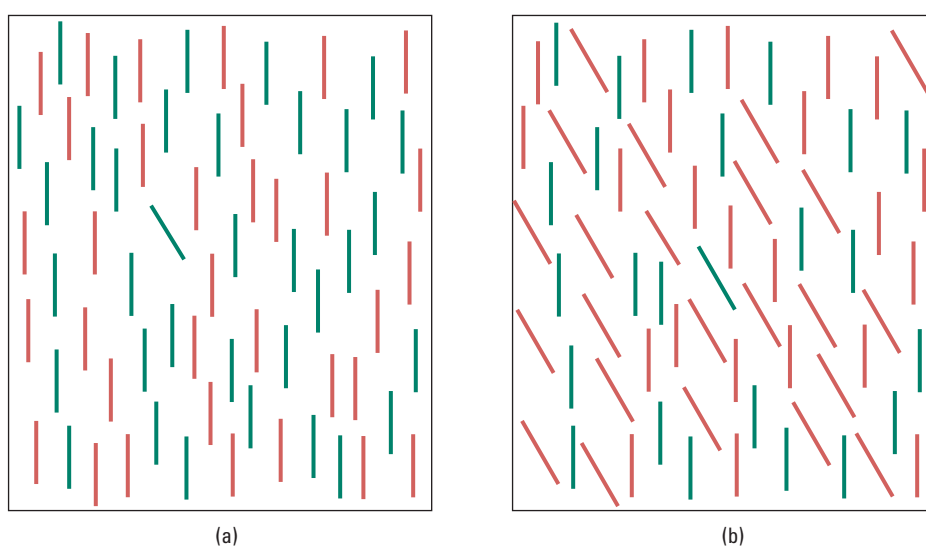


FIGURE 8.19 Stimuli that pop out or do not pop out These stimulus arrays are similar to those used by Treisman and Gormican (1988). In (a) the target stimulus (slanted green line) differs from the distractor stimuli in a primitive feature; it is the only slanted line in the array. Here the target pops out at you; you notice it immediately even if you aren't looking for it. In (b) the target stimulus (slanted green line) does not differ from the distractors in a primitive feature; its greenness is matched by some distractors and its slant is matched by other distractors. Rather, this target is distinct in the way its features are conjoined. It is the only line that is both green and slanted. This target does not pop out at you; you have to look for it to notice it.

or more features that were present in different sets of distractors, as in Figure 8.19b, the amount of time required to locate the target increased in direct proportion to the number of distractors. This increase in detection time is indicative of serial processing, the necessity to attend to each item (or to small groups of items) separately until the target is found. Identification of a single unique feature can be accomplished with parallel processing (stage 1 in Treisman's model), but identification of a unique conjoining of two or more features requires serial processing (stage 2 in Treisman's model).

Treisman also found that subjects who saw simple stimuli flashed briefly on a screen easily identified which primitive features were present but sometimes misperceived which features went together, a phenomenon called *illusory conjunctions*. For example, when shown a straight red line and a green curved one, all subjects knew that they had seen a straight line and a curved line, and a red color and a green color, but some were mistaken about which color belonged to which line. Such findings led Treisman to conclude that stage 1 (parallel processing) registers features independently of their spatial location and that different features that coincide in space (such as the color and curvature of a given line) are joined perceptually only at stage 2 (serial processing), which requires separate attention to each spatial location. In Treisman's theory, stage 2 operates by assigning primitive features to specific spatial locations and then weaving the features together in patterns that reflect their locations.

As additional evidence for her theory, Treisman and her colleagues reported on a man who had suffered damage in both hemispheres to the portion of the parietal cortex that is most critical for identifying the spatial location of visual stimuli (Friedman-Hill & others, 1995; Treisman, 1998). The man could see objects but often made mistakes when asked to point to them or to say where they were. As would be predicted by Treisman's theory, he was especially prone to illusory conjunctions. He had no difficulty registering the primitive features of briefly flashed stimuli but had great difficulty integrating those features, apparently because of the loss in his visual ability to represent spatial locations. He found that he could see objects better if he looked at them one at a time through a narrow viewing tube. By cutting off his view of other objects, he prevented the melding of their features into those of the object that he was inspecting. According to Treisman's analysis, this man's brain damage did not interfere with his capacity for stage 1 processing, but significantly reduced his capacity for stage 2 processing.

Not all research findings are consistent with Treisman's theory. Some researchers have argued, with evidence, that the registration and integration of visual features are not as separate as Treisman proposed (Quinlan, 2003). However, to date, no well-accepted newer theory has emerged to replace Treisman's.

Gestalt Principles of Perceptual Grouping

In the early twentieth century, long before Treisman had developed her model of feature integration, adherents to the school of thought known as ***Gestalt psychology*** had argued that we automatically perceive whole, organized patterns and objects. *Gestalt* is a German word that can be translated roughly as “organized shape” or “whole form.” The premise of the Gestaltists—including Max Wertheimer, Kurt Koffka, and Wolfgang Köhler—was that the mind must be understood in terms of organized wholes, not elementary parts. A favorite saying of theirs was, “The whole is different from the sum of its parts.” A melody is not the sum of the individual notes, a painting is not the sum of the individual dots of paint, an idea is not the sum of its elementary concepts. In each of these examples, something new emerges from the arrangement of the parts, just as meaning emerges when words are arranged to form a sentence. From the Gestalt point of view, the attempt by psychologists to understand perception by focusing on elementary features was like trying to explain the beauty of the *Mona Lisa* by carefully weighing the amount of paint used to produce each part of the masterpiece.

Most early Gestalt research was in the area of visual perception. In many laboratory demonstrations, Gestalt psychologists showed that in conscious experience whole objects and scenes take precedence over parts. For example, when looking at a chair, people perceive and recognize the chair as a whole before noticing its arms, legs, and other components. Treisman and other modern perceptual psychologists would not disagree with this point. In our conscious experience we do typically perceive wholes before we perceive parts; the building up of the wholes from the parts occurs through unconscious mental processes. They would also agree that the whole is different from the sum of its parts, because the whole is defined by the way the parts are *organized*, not just by the parts themselves.

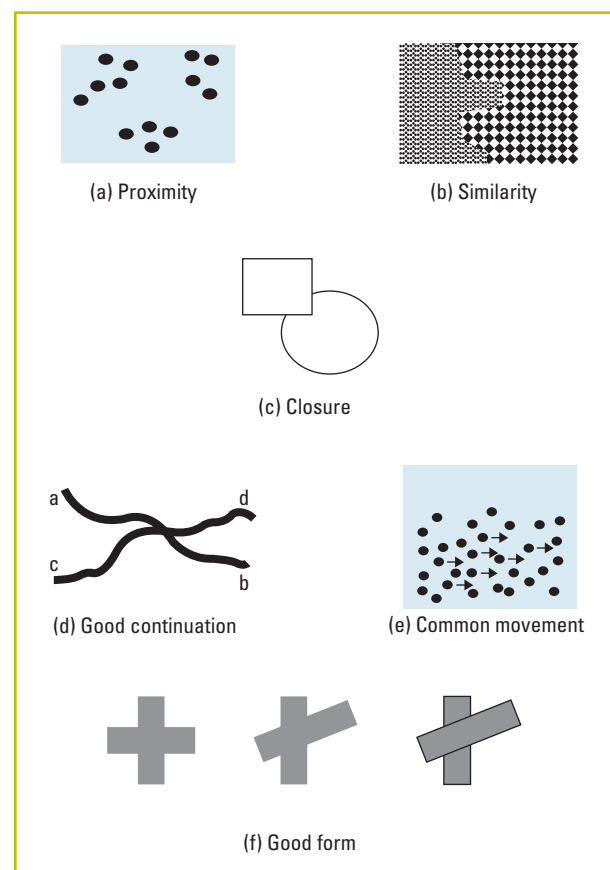
Built-in Rules for Organizing Stimulus Elements into Wholes

The Gestaltists proposed that the nervous system is innately predisposed to respond to patterns in the stimulus world according to certain rules or *principles of grouping*. These principles include the following (Koffka, 1935; Wertheimer, 1923/1938).

1. **Proximity** We tend to see stimulus elements that are near each other as parts of the same object and those that are separated as parts of different objects. This helps us organize a large set of elements into a smaller set of objects. In Figure 8.20a, because of proximity, we see 3 clusters of dots rather than 13 individual dots.
2. **Similarity** We tend to see stimulus elements that physically resemble each other as parts of the same object and those that do not resemble each other as parts of different objects. This helps us distinguish between two adjacent or overlapping objects on the basis of a change in their texture elements, as illustrated in Figure 8.20b. (Texture elements are repeated visual features or patterns that cover the surface of a given object.)
3. **Closure** We tend to see forms as completely enclosed by a border and to ignore gaps in the border. This helps us perceive complete objects even when they are partially occluded by other objects. For example, in Figure 8.20c we automatically assume that the boundary of the circle is complete, continuing behind the square.
4. **Good continuation** When lines intersect, we tend to group the line segments in such a way as to form continuous lines with minimal change in direction. This helps us decide which lines belong to which object when two or more objects overlap. In Figure 8.20d, for example, we see two smooth lines, *ab* and *cd*, rather than four shorter lines or two sharply bent lines such as *ac* or *bd*.
5. **Common movement** When stimulus elements move in the same direction and at the same rate, we tend to see them as part of a single object. This helps us distinguish a moving object (such as a camouflaged animal) from the background. If the dots marked by arrows in Figure 8.20e were all moving as a group, you would see them as a single object.
6. **Good form** The perceptual system strives to produce percepts that are elegant—simple, uncluttered, symmetrical, regular, and predictable (Chater, 1996; Koffka, 1935). This rather unspecific principle encompasses the other principles listed above but also includes other ways by which the perceptual system organizes stimuli into their simplest (most easily explained) arrangement. For example, in Figure 8.20f, the left-hand figure, because of its symmetry, is more likely than the middle figure to be seen as a single object. The middle figure, because of its lack of symmetry, is more likely to be seen as two objects, as shown to its right.

What are some principles of grouping proposed by Gestalt psychologists, and how does each help explain our ability to see whole objects?

FIGURE 8.20 Gestalt principles of grouping These drawings illustrate the six Gestalt principles that are described in the text.



Principle of similarity When we look at a scene such as this one, we automatically group together portions that contain similar stimulus elements. The similarity may be in the color, brightness, texture, or orientation of contours.

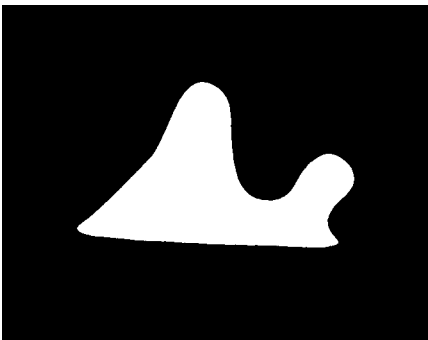


Stefano Scata / Tips Images

Figure and Ground

In addition to proposing the six principles of grouping just listed, the Gestaltists (particularly Rubin, 1915/1958) called attention to our automatic tendency to

divide any visual scene into **figure** (the object that attracts attention) and **ground** (the background). As an example, look at Figure 8.21. The illustration could be described as two unfamiliar figures, one white and one black, whose borders coincide, but you probably do not see it that way. Most people automatically see it as just one white figure against a black background. According to the Gestaltists, the division into figure and ground is not arbitrary, but is directed by certain stimulus characteristics.



In Figure 8.21, the most important characteristic is probably *circumscription*: Other things being equal, we tend to see the circumscribing form (the one that surrounds the other) as the ground and the circumscribed form as the figure.

<<< The figure-ground relationship is not always completely determined by characteristics of the stimulus. You can reverse your perception of figure and ground in Figure 8.21 by imagining that the illustration is a black square with an oddly shaped hole cut out of it, sitting on a white background. When cues in the scene are sparse or ambiguous, the mind may vacillate in its choice of which shape to see as figure and which as ground. This is illustrated by the **reversible figure** in Figure 8.22, where at any given moment you may see either a white vase against a dark ground or two dark profiles against a white ground. In line with the Gestalt figure-ground principle, the same part of the figure cannot simultaneously be both figure and ground, and thus at any instant you may see either the vase or the faces, but not both.



FIGURE 8.21 | Figure and ground Because the white form is completely surrounded by the black form, we tend to see the white form as the figure and the black form as the ground.

19 How do reversible figures illustrate the visual system's strong tendency to separate figure and ground, even in the absence of sufficient cues for deciding which is which?

FIGURE 8.22 | Reversible figure Because it lacks strong cues as to which is figure and which is ground, this image—developed by the Danish Gestalt psychologist Edgar Rubin—may be seen either as a white vase against a dark ground or as two dark profiles against a white ground. If you stare at it, your perception may alternate between the two.

Evidence That Wholes Can Affect the Perception of Parts

When you look at a visual scene, the elementary stimulus features certainly affect your perception of the whole, but the converse is also true: The whole affects your perception of the features. Without your conscious awareness, and at a speed measurable in milliseconds, your visual system uses the sensory input from a scene to draw inferences about what is actually present—a process referred to as *unconscious inference*. Once your visual system has hit upon a particular solution to the problem of what is there, it may actually create or distort features in ways that are consistent with that inference. Examples of such creations and distortions are illusory contours and illusory lightness differences.

Illusory Contours

Look at Figure 8.23. You probably see a solid white triangle sitting atop some other objects. The contour of the white triangle appears to continue across the white space between the other objects. This is not simply a misperception caused by a fleeting glance. The longer you look at the whole stimulus, the more convinced you may become that the contour (border) between the white triangle and the white background is really there; the triangle seems *whiter* than the background. But if you try to see the contour isolated from the rest of the stimulus, by covering the black portions with your fingers or pieces of paper, you will find that the contour isn't really there. The white triangle and its border are illusions.

The white triangle, with its illusory contour, apparently emerges from the brain's attempt to make sense of the sensory input (Parks & Rock, 1990). The most elegant interpretation of the figure—consistent with the Gestalt principle of good form and with expectations drawn from everyday experience—is to assume that it contains a white triangle lying atop a black triangular frame and three black disks. That is certainly simpler and more likely than the alternative possibility—three disks with wedges removed from them and three unconnected black angles, all oriented with their openings aimed at a common center. According to this unconscious-inference explanation, the perceptual system uses the initial stimulus input to infer that a white triangle must be present (because that makes the most sense), and then it *creates* the white triangle, by influencing contour-detection processes in the brain in such a way as to produce a border where one does not physically exist in the stimulus.

Illusory contours cannot be explained in simple stimulus terms, having to do, for example, with the amount of actual lined-up contour existing in the figure. Many experiments have shown that people are more likely to see illusory contours in cases where they are needed to make sense of the figure than in cases where they are not, even when the amount of actual dark-light border is constant (Hoffman, 1998; Gillam & Chan, 2002). For an example, look at Figure 8.24. Most people see a clearer illusory contour, outlining a white square, in pattern *b* more than in pattern *a*, even though the actual black-white borders at the corners of the

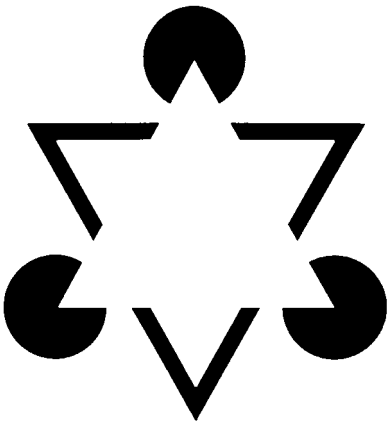


FIGURE 8.23 Illusory contour In response to this stimulus, the perceptual system creates a white triangle, the borders of which appear to continue across the white page, such that the triangle seems whiter than the white page.



FIGURE 8.24 Which pattern shows the clearer illusory contour? Most people see an illusory contour (outlining a white square) more clearly in pattern (b) than in pattern (a), a finding that is consistent with the unconscious-inference explanation of pattern perception. (Based on Hoffman, 1998, p 58.)

imagined white square are identical in the two figures. The unconscious-inference explanation of this is that the white square is more needed in *b* than in *a* to make sense of the stimulus input. The arrangement of four black angular objects in *a* is more likely to occur in everyday experience than is the arrangement of four disks with wedges cut out in *b*.

Illusory Lightness Differences

In Figure 8.25, most people see the small gray square on the left as lighter than the one on the right. In reality, the two squares are identical in lightness; they are exactly the same shade of gray. A common explanation of this illusion is similar to that discussed earlier for contour enhancement. The right-hand square may appear darker because the neurons that respond to it are inhibited by nearby neurons, which are strongly stimulated by the greater amount of light surrounding the square. The left-hand square may appear lighter because the surrounding neurons are not very active, because of the small amount of light reflected from the dark surround, and therefore provide relatively little inhibition.

21

What is some evidence that illusory lightness differences cannot be fully explained by lateral inhibition? How can such differences be explained in terms of unconscious inference?

<<< However, not everyone agrees that such lightness illusions can be explained by simple lateral inhibition. Dale Purves and his colleagues (2004) have argued that more complex processes, involving unconscious inference, must be involved. These researchers suggest that the illusion depicted in Figure 8.25 may derive from the brain's unconscious inference that the left-hand portion of the figure is in a shadow, or for some other reason is receiving less light than is the right-hand portion. In order to compensate for the reduced light, the brain may respond in a way that lightens the appearance of any object in the left-hand portion. This explanation is supported by the observation that the contrast effect is even stronger in realistic scenes and photographs where it is obvious that one object is in a shadow and another is not, as illustrated in Figure 8.26. Notice

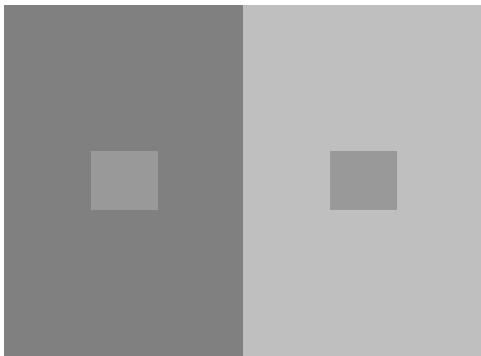


FIGURE 8.25 | A lightness illusion The small gray square on the left looks lighter than that on the right, but the two squares are actually identical in lightness. The lateral-inhibition theory of this illusion has been challenged by evidence such as that illustrated in Figures 8.26 and 8.27.



FIGURE 8.26 | Lightness contrast in a real scene In this photograph (from Purves & others, 2002) tile A looks white and tile B looks dark gray. In reality, the light reflected from the two tiles is an identical shade of gray. You can prove this to yourself by cutting two small holes in a sheet of paper, the appropriate distance apart, and positioning them over the tiles to compare them without seeing the rest of the picture. In this example the perception that tile A is lighter than tile B is gives us useful, real-world information. In reality, tile A is white and tile B is dark gray. They reflect the same amount of light only because tile A is in a shadow and tile B is in bright light. Our visual system takes the shadow and the bright light into account and adjusts the perceived lightness of the tiles appropriately.

that the “illusion” in Figure 8.26 gives us a more accurate understanding of the objects in question (tiles A and B) than we would obtain if our visual system recorded the physical properties of the reflected light without taking the whole scene into account.

The strongest support for unconscious-inference explanations of lightness illusions comes from examples that run directly counter to what would be expected from simple lateral inhibition. Look at Figure 8.27. The gray patches on the left appear to be lighter than those on the right, when, in reality, all of the gray patches are identical to one another. This illusion cannot be explained in any straightforward way by lateral inhibition. The patches on the left are actually surrounded by more white and less black than are those on the right, so the lateral-inhibition theory would predict that those on the left should appear darker, not lighter, than those on the right. The illusion must result from some more complex analysis of the scene. One possible explanation, in terms of unconscious inference, is that the visual system interprets the left-hand patches to be parts of a vertical stripe painted down the rails of a fence (or something like a fence) and interprets the right-hand patches to be visible portions of a post that stands behind the fence. Given this interpretation, the relevant background for the painted stripe is the set of black rails, and the relevant background for the post is the white space behind the fence. To make the stripe and the post stand out more clearly from their respective backgrounds, the visual system lightens the stripe and darkens the post.

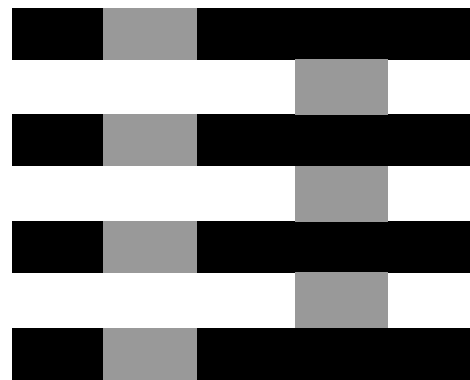


FIGURE 8.27 | Lightness contrast that can't be explained by lateral inhibition The gray patches on the left appear lighter than the physically identical patches on the right. This illusion is the opposite of what one would expect from lateral inhibition.

Unconscious Inference Involves Top-Down Control Within the Brain

When psychologists explain perceptual phenomena in terms of unconscious inference, they do not mean to imply that these phenomena result from anything other than neural processes in the brain. All reasoning, unconscious or conscious, is a product of neural activity. What is implied by unconscious-inference theories is that the phenomena in question result from neural activity in higher brain areas, which are able to bring together the pieces of sensory information and make complex calculations concerning them.

Neuroscientists have learned that the connections between the primary visual area and higher visual areas in the brain are not one-way. The higher areas receive essential input from the primary visual area, but they also feed back to that area and influence neural activity there. Thus, complex calculations made in perceptual portions of the temporal, parietal, and frontal lobes (discussed later) can feed back to the primary visual area in the occipital lobe and influence the activity of feature detectors there. Researchers have found, for example, that visual stimuli that produce illusory contours activate edge-detector neurons, in the primary visual cortex, that are receiving input from precisely that part of the stimulus where the illusory contour is seen (Lee, 2002). In this case, the activity of the edge detector is not the direct result of input from the optic nerve to the edge detector, but, rather, is the result of descending connections from higher visual centers that have determined that there *should* be an edge at that location.

Brain scientists and perceptual psychologists refer to control that comes from higher up in the brain as **top-down control**, and they refer to control that comes more directly from the sensory input as **bottom-up control**. Perception always involves interplay between bottom-up and top-down control in the brain. Bottom-up processes bring in the sensory information that is actually present in the stimulus. Top-down processes bring to bear the results of calculations based on that sensory information plus other information, such as that derived from previous experience and from the larger context in which the stimulus appears. Remember, all these calculations are being conducted by a calculating machine—the human brain—that is vastly more complex and powerful than any non-biological computer that humans have developed.

>>> 22
How is unconscious inference described as top-down control within the brain? What is the difference between top-down and bottom-up control?

SECTION REVIEW

A central purpose of the human visual system is to identify objects.			
Contour Enhancement • An object’s shape is defined by its contour, the edge between it and its background. • Lateral inhibition among sensory neurons heightens the perceived contrast in lightness at contours. This helps us see objects’ shapes.	Integration of Features • Neurons in the primary visual cortex and nearby areas are maximally responsive to specific visual features—such as particular line orientations, movements, and colors. • Behavioral evidence suggests that features are detected through rapid parallel processing and then are integrated spatially through serial processing.	Gestalt Principles • Gestalt psychologists asserted that whole objects are not merely the sums of their parts and that wholes take precedence in conscious perception. • The Gestalt principles of grouping describe rules by which we automatically organize stimulus elements into wholes. • We also automatically separate figure from ground.	Top-Down Processes • Wholes influence our perception of parts through unconscious inference, as illustrated by illusory contours and illusory lightness differences. • The effects of unconscious inference occur through top-down control mechanisms in the brain.

Recognizing Objects

To recognize an object is to categorize it. It’s a bird, or an airplane, or Superman. To recognize an object visually, you must form a visual perception of the object that is sufficient to allow you to match it with your stored definition, or understanding, of the appropriate object category. Some of the most interesting research on object recognition has been conducted with people who suffer from brain damage that leaves them able to see the parts of objects but unable to see and identify whole objects.

Beyond the Primary Visual Cortex: Two Streams of Visual Processing

23 What are the anatomical and functional distinctions between two different visual pathways in the cerebral cortex?

◀◀ The primary visual cortex, which occupies the rearmost part of the occipital lobe, sends its output to (and receives feedback from) many other visual-processing areas, which occupy the rest of the occipital lobe and extend forward into much of the temporal and parietal lobes. Some years ago, largely on the basis of evidence from single-cell recordings in monkeys and effects of brain damage in people, Leslie Ungerleider and Mortimer Mishkin (1982) proposed that the visual areas beyond the primary area exist in two relatively distinct cortical pathways, or “streams,” which serve different functions. As shown in Figure 8.28, one stream runs into the lower portion of the temporal lobe and the other runs upward into the parietal lobe.

Ungerleider and Mishkin proposed that the temporal stream is specialized for processing all the information needed to identify objects and that the parietal stream is specialized for maintaining a map of three-dimensional space and locating objects in that space. Subsequent research largely confirmed their proposal, and the two tracks came to be known, respectively, as the “*what*” and “*where*” pathways. More recently, however, research has shown convincingly that a primary function of the parietal pathway is to use visual input to guide movements. The pathway is concerned not just with *where* the object is located, but also with *how* the person must move in order to pick up the object, or move out of its way, or interact with it in some other way. For this reason, I use the label “*where-and-how*,” rather than just “*where*,” to refer to the parietal pathway.

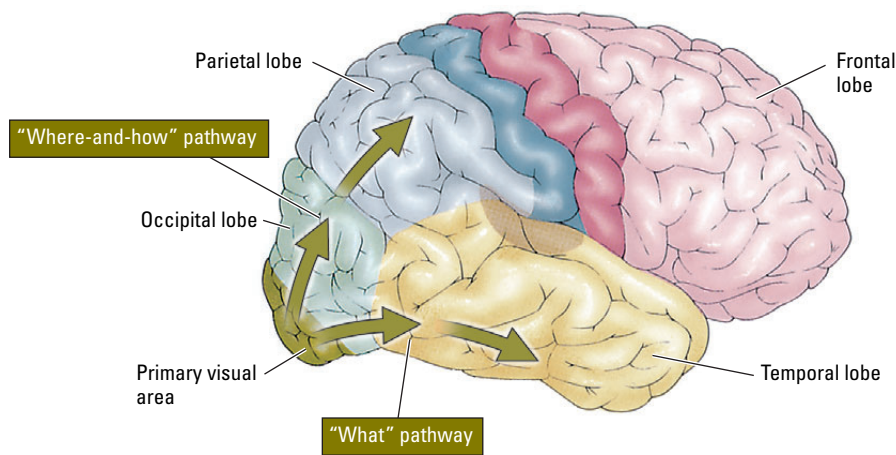


FIGURE 8.28 | The “what” and “where-and-how” visual pathways
Neurons in the primary visual area send output into two relatively distinct streams for further visual processing. The “what” pathway, into the lower temporal lobe, is specialized for perceiving shapes and identifying objects. The “where-and-how” pathway, into the parietal lobe, is specialized for perceiving spatial relationships and for guiding actions.

Effects of Damage in the “What” Pathway

Loss of Ability to Identify Forms, or Objects People who have suffered damage to specific portions of the “what” pathway on both sides of the brain can see but cannot make sense of what they see, a condition called **visual agnosia**. This term was coined in the late nineteenth century by Sigmund Freud, who at that time was a young, little-known neurologist (Goodale & Milner, 2004). It comes from the Greek words *a*, meaning not, and *gnosia*, meaning knowledge. People with visual agnosia can see, but they do not know what they are seeing.

Visual agnosias have been classified into a number of general types (Farah, 1989; Milner & Goodale, 1995). People with **visual form agnosia** can see that something is present and can identify some of its elements, such as its color and brightness, but cannot perceive its shape (Milner & Goodale, 1995). They are unable to describe or draw the outlines of objects or patterns that they are shown (see Figure 8.29). In contrast, people with **visual object agnosia** can identify and, with considerable effort, draw the shapes of objects they are shown but cannot identify the objects. For instance, shown an apple, such a person might draw a recognizable apple but would still be unable to say what he or she had just drawn. The problem is not one of language, since the person can name objects identified through other means, such as touch.

In a famous essay entitled “The Man Who Mistook His Wife for a Hat,” the neurologist Oliver Sacks (1970) described the plight of one of his patients, who, after a stroke, suffered from visual object agnosia. When Sacks showed the man a rose and asked him to identify it, the man’s response was, “About six inches in length. A convoluted red form with a linear green attachment. It lacks the simple symmetry of the Platonic solids, although it may have a higher symmetry of its own” After a period of such erudite reasoning about the object’s parts, the man finally guessed, uncertainly, that it might be some kind of flower. Then Sacks asked him to smell it. “Beautiful!” he exclaimed. “An early rose. What a heavenly smell!” He could tell a rose by smell but not by sight, even though he could see and describe each of its parts in terms of geometry and color.

The existence of these two types of agnosia is consistent with the theory that seeing an object involves a sequence of separable steps: The individual sensory elements are picked up and analyzed, then these are grouped by the brain into basic shapes, and finally the shapes are organized into a whole and recognized as a particular object. Visual form agnosia appears to result from destruction of the mechanisms needed for the second of these steps, and visual object agnosia appears to result from destruction of the mechanisms needed for the final step.

What perceptual disorders are observed in people with damage to the “what” pathway on both sides of the brain? How do these observations support the idea that object perception occurs in a sequence of steps?

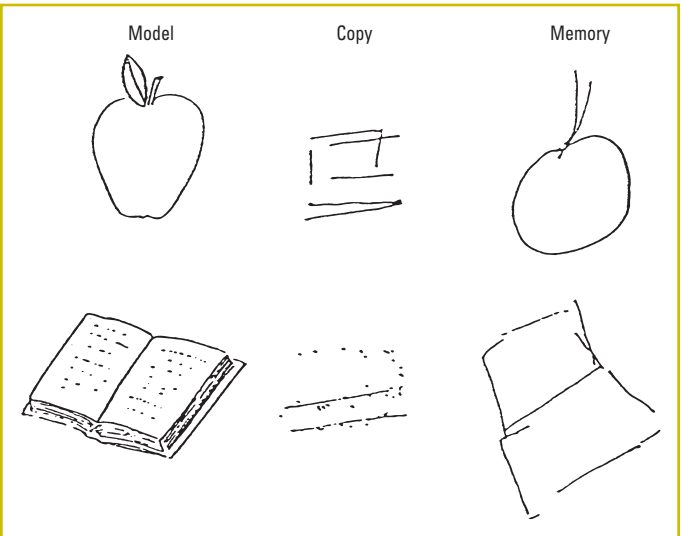


FIGURE 8.29 | Drawings made by a person with visual form agnosia
The patient D. F. was unable to recognize the drawings on the left. When she was asked to try to copy these models, she produced drawings (middle column) that bore no or very little resemblance to the originals. When asked to draw an apple and a book from memory, however, she succeeded (right-hand column). Later, when she was shown the drawings she had produced from memory, she had no idea what they represented. (From Milner & Goodale, 1995, p. 127.)

25

What abilities are preserved in people with damage to the “what” pathway but lost in people with damage to the “where-and-how” pathway?

**Preserved Ability to Act on Objects**

People with damage to the “what” pathway show deficits in their ability to see forms or objects but nevertheless retain the ability to reach for objects and act on them in coordinated ways, guided by vision. A dramatic example is the case of a woman known as D. F., the same person whose drawings are reproduced in Figure 8.29, who has been extensively studied by Melvyn Goodale and A. David Milner (2004). This woman suffers from very severe visual form agnosia, stemming from carbon monoxide poisoning that destroyed portions of the “what” pathway close to the primary visual area on both sides of her brain. Despite her complete inability to perceive consciously the shapes of objects, D. F. responds to objects in ways that take into account shape as well as size, position, and movement. When she walks, she moves around obstacles with ease. She is good at catching objects that are thrown to her, even though she can’t consciously see the object coming toward her. When asked to pick up a novel object placed in front of her, she moves her hand in just the correct way to grasp the object efficiently.

In one experiment, D. F. was shown an upright disk with a slot cut through it. She claimed to be unable to see the orientation of the slot and, indeed, when she was asked to hold a card at the same angle as the slot, her accuracy (over several trials with the slot at varying orientations) was no better than chance. But when she was asked to slip the card into the slot as if mailing a letter, she did so quickly and accurately on every trial, holding the card at just the right orientation before it reached the slot! She could not use conscious perception to guide her hand, but her hand moved correctly when she wasn’t trying to use conscious perception to guide it.

Effects of Damage in the “Where-and-How” Pathway

Damage in the “where-and-how” pathway—in the upper parts of the occipital and parietal lobes of the cortex—interferes most strongly with people’s ability to use vision to guide their actions. People with damage here have relatively little or no difficulty identifying objects that they see, and often they can describe verbally where the object is located, but they have great difficulty using visual input to coordinate their movements. They lose much of their ability to follow moving objects with their eyes or hands, to move around obstacles, or to reach out and pick objects up in an efficient manner (Goodale & Milner, 2004; Schindler & others, 2004). Even though they can consciously see and describe an object verbally, and report its general location, they reach for it gropingly, much as a blind person does. They frequently miss the object by a few inches and then move their hand around until they touch it (see Figure 8.30). Only when they have touched the object do they begin to close their fingers around it to pick it up.



Sight unseen: An exploration of conscious and unconscious vision (2003) by Goodale, M. A., & Milner, A. D. By permission of Oxford University Press.

FIGURE 8.30 Efficient and inefficient reaching for an object People with damage to the “what” pathway reach for an object in the efficient manner depicted in (a), even though they can’t consciously see the object’s shape. In contrast, people with damage to the “where-and-how” pathway reach in the inefficient, groping manner depicted in (b), even though they can consciously see the object’s shape.

Complementary Functions of the Two Visual Pathways in the Intact Brain

The two just-described visual pathways apparently evolved to serve different but complementary functions. The “what” pathway provides most of our conscious vision. It provides the input that allows us to see and identify objects consciously, to talk about those objects, and to make conscious plans concerning them. In contrast, the “where-and-how” pathway provides the input that is needed for the automatic, rapid, and largely unconscious visual control of our movements with respect to objects. This pathway is able to register the shape of an object to the degree that shape is necessary for effectively reaching for and picking up the object, but it does not register shape in a manner that enters consciousness.

Think for a moment about the unconscious roles that vision plays in your movements. As you follow a moving object with your eyes, your eye movements are constantly guided by visual input. As you walk through a world of objects, you automatically use visual input to stay on the path and to avoid bumping into things. You can demonstrate the unconscious operation of your “where-and-how” pathway with a simple exercise. If you focus your eyes at a specific point and someone places an object in your peripheral vision, far enough away from your direct line of sight that you can just barely see that something is there but can’t make out its size and shape, you can, without moving your eyes, reach out and grab that object, accurately and without fumbling (Goodale & Murphy, 1997). As you do this, you make hand movements that are well attuned to the object’s size and shape, even though you can’t consciously see that size and shape. The visual system that guides your hand has enough information about the size and shape of the object, as well as its location, to allow you to reach for it effectively, but that information does not enter your consciousness.

>>> 26 In sum, what are the distinct functions of the “what” and “where-and-how” visual pathways?

Clues for Object Recognition

Studying the visual abilities and disabilities of people with brain damage is one route for learning about the processes involved in object recognition. Another route is studying the minimal stimulus conditions that are necessary for normal, brain-intact people to recognize objects. What clues in the visual stimulus do observers use to identify an object at which they are looking? Here we shall look briefly at clues having to do with shape, context, and motion.

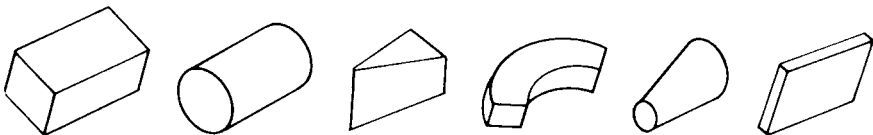
Recognizing Objects on the Basis of Components

Think of all the different visual patterns that you would recognize as, say, an airplane. An airplane produces a very different array of stimuli depending on whether it is viewed from the bottom, or the side, or the front, and different airplanes differ in many of their visual aspects, yet you can recognize all of them as airplanes. How do you do that? There are an infinite number of different specific arrangements of visual elements that you will call an airplane. It seems logical that you must somehow reduce that infinite set to something finite as part of the process of deciding what you are looking at.

Biederman’s Recognition-by-Components Theory In what he calls a *recognition-by-components theory*, Irving Biederman (1987, 1989) has proposed that, to recognize a three-dimensional object, the visual system first organizes the stimulus information into a perceived set of basic components, and then uses the arrangement of the components as a basis for recognizing the object. The components are elementary three-dimensional shapes, referred to as *geons*. On the basis of fundamental properties of three-dimensional geometry, Biederman has proposed that there are 36 different geons, some of which appear in Figure 8.31 (on page 290). By smoothing the edges and ignoring the details, according to Biederman, we can depict any object as a small subset of such geons organized in a certain way. You may

>>> 27 According to Biederman’s theory, how might a finite set of geometric forms (geons) provide the basis for perception of an infinite set of objects?

FIGURE 8.31 | Some geons From principles of geometry, Biederman developed a list of 36 simple, three-dimensional forms, which he labeled “geons,” six of which are shown here. He suggests that geons provide the basic perceptual components of more complex forms.



already be familiar with this general idea from experience with learn-to-draw books that recommend sketching any object as a set of three-dimensional geometric shapes before fleshing it out with details. To see an object, according to Biederman, your visual system first organizes the elementary stimulus features in such a way as to detect the geons, and then organizes the geons in such a way as to identify the object.

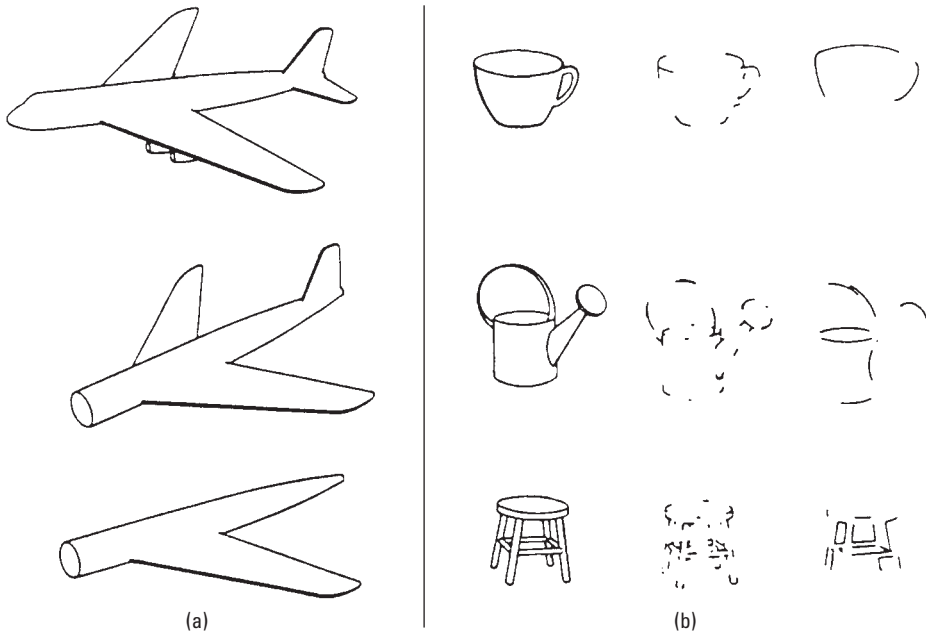
28

How is the recognition-by-components theory supported by experiments on object recognition in people with intact brains and by observations of people with brain damage?

<<< Some Evidence for Biederman's Theory Biederman's own evidence for the theory comes from studies showing that object recognition by normal, brain-intact people depends on their ability to detect at least some of the object's geons and their arrangement relative to one another. In a series of experiments, Biederman (1987) asked people to identify objects in pictures that were flashed briefly on a screen. He found that the speed and accuracy of recognizing the objects depended very much on the intactness and arrangement of an object's geons and very little on other aspects of the stimulus, such as the amount of detail occurring within the geons or the exterior outline of the object as a whole. Figure 8.32 shows some of the stimuli that Biederman used. The airplane in Figure 8.32a was recognized as an airplane even when most of its geons were removed, thereby changing its overall outline, as long as the geons still present were intact and properly arranged. The various objects in Figure 8.32b were recognized when the lines were degraded in a way that preserved the recognizability of individual geons and their connections to one another (middle column) but not when the same amount of line degradation occurred in a way that obscured the geons and their connections (right-hand column).

Biederman's theory can also be applied to the observations concerning people with visual agnosias, previously discussed. Using Biederman's term, we could say that people with visual form agnosia are unable to perceive the geons of an object,

FIGURE 8.32 | Support for Biederman's recognition-by-components theory Part (a) shows an airplane consisting of nine, four, or two components (geons). Even with just a few components present, it is recognizable. Part (b) shows a set of line drawings of objects degraded in two different ways. The degradation in the middle column preserves the connections between adjacent components, and that in the right-hand column does not. When subjects saw the degraded figures alone, they recognized those in the middle column but not those in the right-hand column.



and people with visual object agnosia are able to perceive the geons but unable to put them together to form a meaningful perception of the whole object. The man who could not identify a rose by sight could nevertheless see and describe its individual geons—“the convoluted red form” and “the linear green attachment.” Another patient with this disorder described a bicycle that he was shown as a pole and two wheels but could not identify it as a bicycle or guess its function (Hécaen & Albert, 1978).

Recognizing Objects on the Basis of Context

Our recognition of an object depends not just on cues from the object itself but also on the context in which the object appears. When you look at a tree, you can perceive leaves and twigs that, because of their angle or distance from you, would be unrecognizable if seen away from the tree. Likewise, you probably recognize your psychology professor more easily in the classroom than you would at a shopping mall and more easily at a shopping mall than at a rock concert. Numerous experiments have shown that people recognize objects more quickly and accurately in accustomed contexts than in unaccustomed contexts (Henderson & Hollingworth, 1999). To experience the effect of context, try to identify the three objects depicted in Figure 8.33, and then turn to Figure 8.44 on page 299 and identify the same three objects there. Expectations based on context provide part of the top-down contribution to object identification.

Recognizing Objects on the Basis of Motion

Perceptual psychologists most often study the perception of static images, typically using still pictures as the stimuli to be identified. But in the real world many of the things we observe (including people, animals, and vehicles) move, and their movement helps us identify them.

To study the role of movement cues independent of other indices of shape, Gunnar Johansson and his colleagues in Sweden attached a small light to each major joint of a person's body and then filmed the person in such a way that only the lights could be seen as the person engaged in various activities (see Figure 8.34). When observers saw a single still frame of the film, they had no idea what they were looking at; but as soon as the lights began to move, they recognized the object as a person performing a particular action. Just 1/10 second of viewing time, enough to go through two frames of the film, was sufficient for most viewers to see the lights as a human being walking (Johansson, 1976/1994). Other experiments have shown that viewers can even judge the sex of the person according to how the tiny lights move as the person walks or jumps (Troje, 2002). Our ability to perceive subtle differences in movement patterns may play a larger role than we normally realize in identifying people and other animate objects.

>>> 29
How does context contribute to the top-down processes in object recognition?



FIGURE 8.33 | Objects out of context Can you identify these objects? After making your guesses, turn to Figure 8.44 at the end of the chapter.

>>> 30
How have researchers shown that patterns of movement provide a foundation for object recognition?

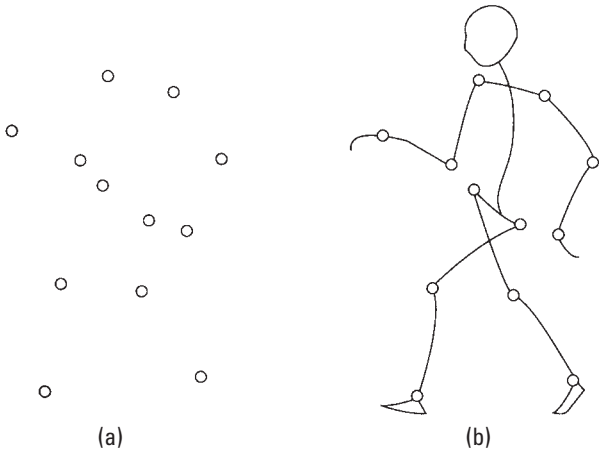


FIGURE 8.34 | Lights that are perceived as a person when the person moves Lights attached to a person's major joints (a) are not perceived as a recognizable object in the dark when the person stands still. However, when the person begins to move (b), the lights are perceived immediately as a human form.

SECTION REVIEW

To identify an object, we must perceive it well enough to match it to a stored representation.

Two Streams of Visual Processing

- Visual processing beyond the primary visual cortex in the occipital lobe takes place along two independent streams (pathways)—a “what” stream leading into the temporal lobe and a “where-and-how” stream leading into the parietal lobe.
- Bilateral damage to parts of the “what” stream can cause visual agnosias, disorders in which people can see features or geometric components of objects but fail to identify the objects.
- Damage to the “where-and-how” stream impairs one’s ability to use vision to direct physical actions with respect to objects, such as grasping or moving around an object.

Clues for Recognizing Objects

- The recognition-by-components theory suggests that we identify three-dimensional objects by perceiving their basic component shapes (geons), which provide a level of structure between features and wholes.
- Through top-down processing, an object’s context can help us to identify it.
- The unique way an object moves can provide powerful clues for recognizing it.

Seeing in Three Dimensions

One of the great early puzzles of vision, about which philosophers speculated for centuries, was the puzzle of depth perception. We automatically, effortlessly, see the world in three dimensions. Objects occupy and move in space that includes not only a vertical (up-down) and a horizontal (right-left) dimension but also a dimension of depth, or distance from our eyes. Our retinas and the images that are projected on them are flat, two-dimensional. It is relatively easy to understand how our retinas might record the vertical and horizontal dimensions of our visual world, but how do they record the third dimension, the dimension of depth?

31 How did Helmholtz describe perception as a problem-solving process?

<<< A major step toward answering this question was the publication of a treatise on vision, in the mid-nineteenth century, by Hermann von Helmholtz (1867/1962), the same German physiologist who developed the trichromatic theory of color vision. He argued that seeing is an active mental process. The light focused onto our retinas is not the scene we see but is simply a source of hints about the scene. Our brain infers the characteristics and positions of objects from cues in the reflected light, and those inferences are our perceptions. Helmholtz pointed out that the

steps in this inferential process can be expressed mathematically, in equations relating information in the reflected light to conclusions about the positions, sizes, and shapes of objects in the visual scene. We are not conscious of these calculations and inferences; our brain works them out quickly and automatically, without our conscious awareness.

Helmholtz was the first scientist to use the term *unconscious inference* to describe visual perception. Earlier in the chapter you read about unconscious inferences that affect the perception of an object’s contours and lightness. Now, in the more direct tradition of Helmholtz, let us examine how unconscious inferences may provide the foundation for our ability to perceive the depth and size of objects.

Hermann von Helmholtz
Considered by many to be the greatest of all nineteenth-century physiologists, Helmholtz was also a pioneer of what we now call cognitive psychology. His unconscious-inference theory of perception posits that the mind constructs, through unconscious calculations, meaningful perceptions from cues picked up by the senses.



The Granger Collection

Cues for Depth Perception

Depth perception works best when you use both eyes. You can prove that with a simple demonstration. Pick up two pencils and hold them in front of you, one in each hand, with their points toward each other. Close one eye and move the pencils toward each other to make the points touch at their tips. Chances are, you will miss by a little bit on your first attempt, and your subsequent adjustments will have something of a trial-and-error quality. Now repeat the task with both eyes open. Is it easier? Do you find that now you can see more clearly which adjustments to make to bring the points together and that you no longer need trial and error? You can see depth reasonably well with one eye alone, but you can see it better with both eyes together.

Binocular Cues for Depth

Two types of cues contribute to the binocular (two-eye) advantage. The less important is *eye convergence*, the inward turning of the eyes that occurs when you look at an object that is close to you. The closer an object is, the more the two eyes must converge in order to focus on it. To experience eye convergence consciously, try focusing both eyes on your finger as you hold it a few inches in front of your nose. In theory, the perceptual system could judge the distance of an object from the degree to which the eyes converge when looking at it. In practice, however, researchers have found that eye convergence is a poor distance cue even for objects close to the eyes and a useless one for objects more than a few feet away (Arditi, 1986; Hochberg, 1971).

The other, far more important cue requiring two eyes is **binocular disparity**, which refers to the slightly different (disparate) views that the two eyes have of the same object or scene. Because the eyes are a few centimeters apart, they view any given object from slightly different angles. To see how the degree of binocular disparity varies depending on an object's distance from your eyes, hold your index finger about a foot in front of your eyes with a wall in the background. Look at your finger with just your right eye open, then with just your left eye open, alternately back and forth. As you alternate between the two eyes, your finger appears to jump back and forth with respect to the background wall. That is because each eye views the finger from a different angle. Now move your finger farther away (out to full arm's length), and notice that your view of it jumps a smaller distance with respect to the wall as you again alternate between right-eye and left-eye views. The farther your finger is from your eyes, the smaller is the difference in the angle between each eye and the finger; the two lines of sight become increasingly parallel.

Thus, the degree of disparity between the two eyes' views can serve as a cue to judge an object's distance from the eyes: The less the disparity, the greater the distance. In normal, binocular vision your brain fuses the two eyes' views to give a perception of depth. Helmholtz (1867/1962) showed mathematically how the difference in two objects' distance from the eyes can be calculated from differences in the degree of binocular disparity. More recently, researchers have found that neurons in an area of the visual cortex close to the primary visual area respond best to stimuli that are presented to both eyes at slightly disparate locations on the retina (Thomas & others, 2002). These neurons appear to be ideally designed to permit depth perception. For another demonstration of binocular disparity, see Figure 8.35 (on page 294).

Illusions of Depth Created by Binocular Disparity

The ability to see depth from binocular disparity—an ability called stereopsis—was first demonstrated in the early nineteenth century by Charles Wheatstone (described by Helmholtz, 1867/1962). Wheatstone wondered what would happen if he drew two slightly different pictures of the same object or scene, one as seen by the left eye and one as seen by the right, and then viewed them simultaneously, each with the appropriate eye. To permit such viewing, he invented a device called a

32
How can binocular disparity serve as a cue for depth?

33
How do stereoscopes provide an illusion of depth?

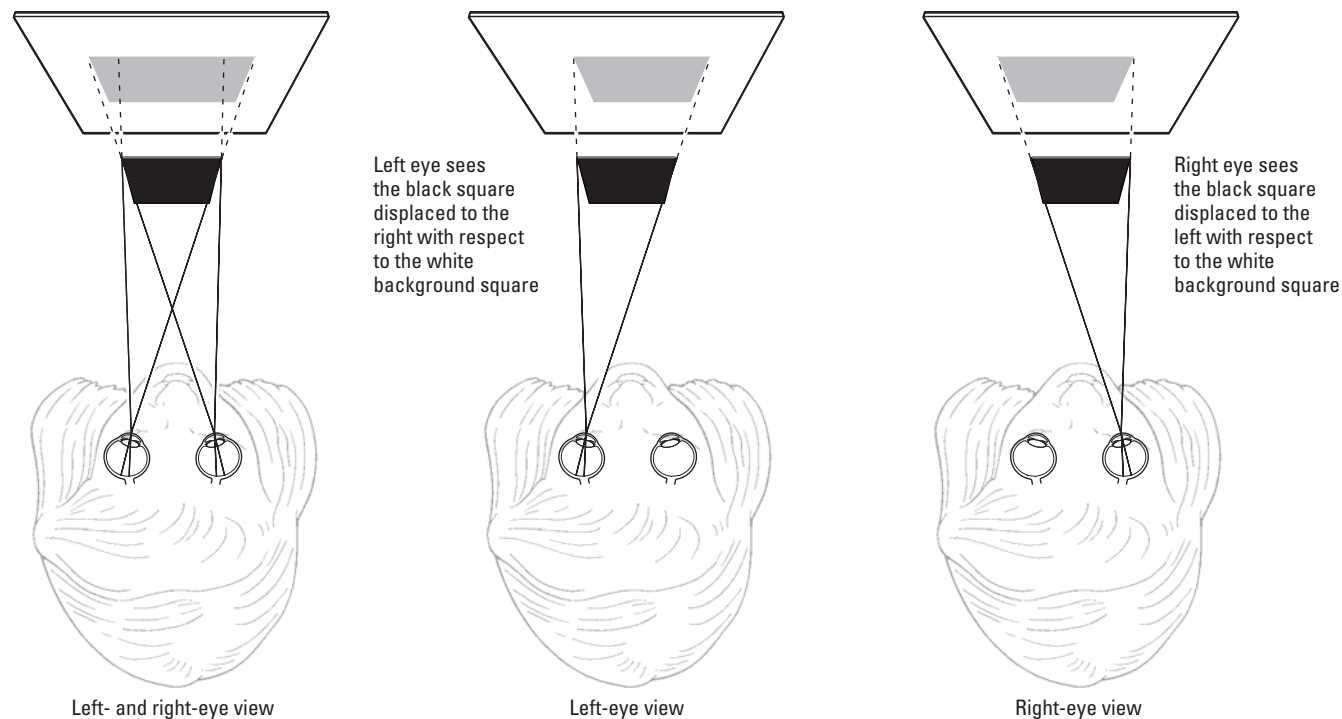


FIGURE 8.35 | Demonstration of binocular disparity The two eyes see somewhat different views of the relationship between the closer figure and the more distant figure. The disparity (degree of difference) between the two views is proportional to the distance between the two objects, and that information is used by the perceptual system to perceive the depth between them.

stereoscope. The effect was dramatic. When viewed through the stereoscope, the two pictures were fused perceptually into a single image containing depth. Stereoscopes became a great fad in the late nineteenth century. People could see scenes such as Buckingham Palace or the Grand Canyon in full depth by placing cards that contained two photographs of the same scene, shot simultaneously from slightly different angles, into their stereoscope. (The Viewmaster, a child's toy, is a type of stereoscope in common use today.) Three-dimensional motion pictures and comic books employ the same general principle. In the simplest versions, each frame of the film or comic strip contains an overlapping pair of similar images, each in a different color, one slightly displaced from the other, and the viewer wears colored glasses that allow only one image to enter each eye. You can demonstrate stereopsis without any special viewer by looking at the two patterns in Figure 8.36 in the manner described in the caption.

Monocular Cues for Depth

Although depth perception is most vivid with two eyes, it is by no means absent with one. People who have just one functioning eye can drive cars, shoot basketballs, and reach out and pick objects up without fumbling around.

Entertainment in depth An early method to produce three-dimensional movies, popular in the 1950s, was to project overlapping dual images on the screen, which differed in color and, slightly, in position. When viewed through differently colored lenses, which filtered a different image to each eye, the result was binocular disparity and a dramatic three-dimensional illusion. The viewer might see a bull charging directly at him, for example.



© Benjamin Randel / Corbis



FIGURE 8.36 A depth illusion created by binocular disparity. The two patterns are constructed to appear as they would to the left and right eye, respectively, if the dark square were actually a certain distance in front of the white square (like that shown in Figure 8.35). In order to experience the three-dimensional effect, hold the book about a foot in front of your eyes and let your eyes drift in an unfocused way until you see double images of everything. You will

see four renditions of the white frame with a darker square center—two renditions of (a) and two of (b). When all four of these images are clear, converge or diverge your eyes a little in order to get the right-hand image of (a) to sit right atop the left-hand image of (b). You have fused your left eye's view of (a) and your right eye's view of (b) into a single image, which now appears to be three-dimensional: The dark square seems to float in space in front of the white square.

Motion Parallax

Perhaps the most valuable monocular (one-eye) cue is *motion parallax*, which refers to the changed view one has of a scene or object when one's head moves sideways. To demonstrate motion parallax, hold your finger up in front of your face and view it with one eye as you rock your head back and forth. As your head moves, you gain different views of the finger, and you see it being displaced back and forth with respect to the wall in the background. If you now move your finger farther away from your eye, the same head movement produces a less changed view. Thus, the degree of change in either eye's view at one moment compared with the next, as the head moves in space, can serve as a cue for assessing the object's distance from the eyes: The smaller the change, the greater the distance.

As you can infer from this demonstration, motion parallax is very similar to binocular disparity. In fact, binocular disparity is sometimes called *binocular parallax*. The word *parallax* refers to the apparent change in an object or scene that occurs when it is viewed from a new vantage point. In motion parallax the changed vantage point comes from the movement of the head, and in binocular parallax (or disparity) it comes from the separation of the two eyes.

>>> **34**
How can motion parallax serve as a cue for depth, and how is it similar to binocular disparity?

Pictorial Cues Motion parallax depends on the geometry of true three-dimensionality and cannot be used to depict depth in two-dimensional pictures. All the remaining monocular depth cues, however, can provide a sense of depth in pictures as well as in the real three-dimensional world, and thus they are called *pictorial cues for depth*. You can identify some of these by examining Figure 8.37 (on page 296) and thinking of all the reasons why you see some objects in the scene as standing in the foreground and others as more distant. The pictorial cues include the following.

>>> **35**
What are some cues for depth that exist in pictures as well as in the actual, three-dimensional world?

- 1. Occlusion** The trees occlude (cut off from view) part of the mountains, which indicates that the trees are closer to us than are the mountains. Near objects occlude more distant ones.
- 2. Relative image size for familiar objects** The image of the woman (both in the picture and on the viewer's retina) is taller than that of the mountains. Because we know that people are not taller than mountain's, we take the woman's larger image as a sign that she must be closer to us than are the mountains.
- 3. Linear perspective** The rows of plants converge (come closer together) as they go from the bottom toward the mountains, indicating that objects toward the mountains are farther away. Parallel lines appear to converge as they become more distant.
- 4. Texture gradient** Texture elements in the picture—specifically, the individual dots of color representing the flowers—are smaller and more densely packed

FIGURE 8.37 | Pictorial cues for depth Depth cues in this picture include occlusion, relative image size for familiar objects, linear perspective, texture gradient, position relative to the horizon, and differential lighting of surfaces.



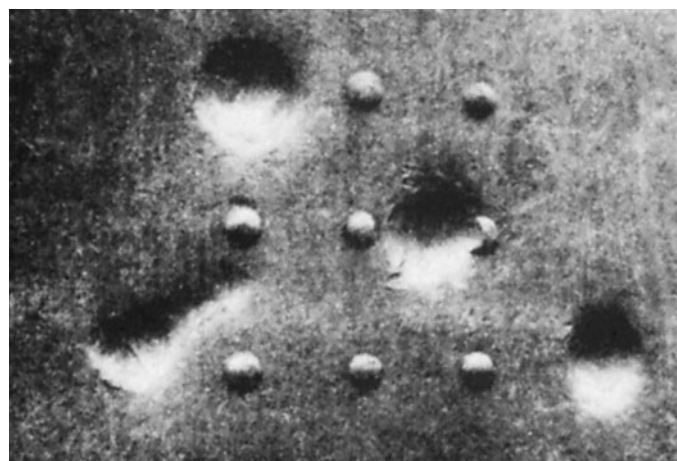
The Image Bank / Getty

near the trees and mountains than they are at the bottom of the picture. In general, a gradual decrease in the size and spacing of texture elements indicates depth.

5. Position relative to the horizon The trees are closer to the horizon than is the woman, indicating that they are farther away. In outdoor scenes, objects nearer the horizon are usually farther away than those that are displaced from the horizon in either direction (either below it or above it). If there were clouds in this picture, those seen just above the edge where the earth and sky meet (close to the horizon) would be seen as farther away than those seen farther up in the picture (farther from the horizon).

6. Differential lighting of surfaces In real three-dimensional scenes the amount of light reflected from different surfaces varies as a function of their orientation with respect to the sun or other source of light. The fact that the sides of the rows of lavender are darker than the tops leads us to see the rows as three-dimensional rather than flat. We see the brightest parts of the plants as their tops, closest to us (as we look down on the plants); and we see the darker parts as their sides, shaded by the tops, and farther from us. For an even more dramatic demonstration of an effect of lighting, see Figure 8.38.

FIGURE 8.38 | Depth perception created by light and shade Because we automatically assume that the light is coming from above, we see the smaller disruptions on the surface here as bumps and the larger ones as pits. Turn the picture upside down and see what happens. (The bumps and pits reverse.)



Kevin Berbaum

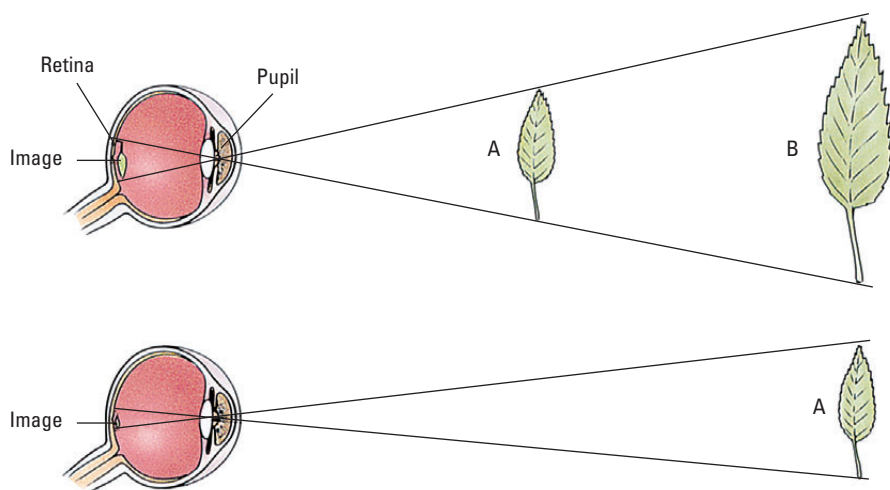


FIGURE 8.39 | Relationship of retinal-image size to object size and distance If, as in the upper sketch, object B is twice as tall and wide as object A and also twice as far away from the eye, the retinal images that the two objects produce will be the same size. If, as in the lower sketch, object A is moved twice its former distance from the eye, the retinal image produced will be half its former height and width.

The Role of Depth Cues in Size Perception

The ability to judge the size of an object is intimately tied to the ability to judge its distance. As Figure 8.39 illustrates, the size of the retinal image of an object is inversely proportional to the object's distance from the retina. Thus, if an object is moved twice as far away, it produces a retinal image half the height and width of the one it produced before. But you don't see the object as smaller, just farther away. The ability to see an object as unchanged in size, despite change in the image size as it moves farther away or closer, is called **size constancy**. For familiar objects, such as a pencil or a car, previous knowledge of the object's size may contribute to size constancy. But size constancy also occurs for unfamiliar objects if cues for distance are available, and even familiar objects can appear to be drastically altered in size if misleading distance cues are present (for an example, see Figure 8.40).

>>> **36**
In theory, why does size perception depend on distance perception?

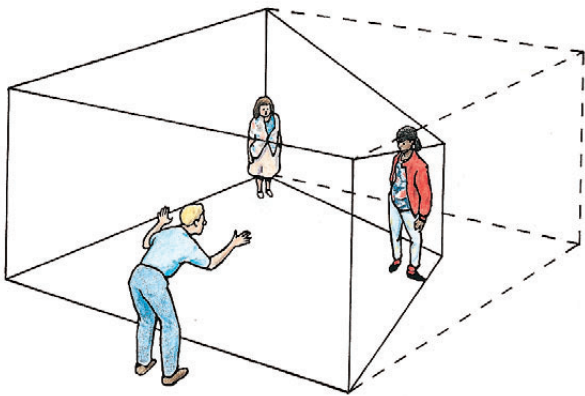


FIGURE 8.40 | A size-distance illusion We know that these young women must be approximately the same size, so what explains this illusion? It's the room they're standing in that's distorted. The back wall and both windows are actually trapezoidal in shape, and the wall is slanted so that its left-hand edge is actually twice as tall and twice as far away from the viewer as its right-hand edge (see drawing at right). When we view this scene through a peephole (or the camera's eye), we automatically assume that the walls and window are normal, that the occupants are the same distance away, and therefore that their size is different. This distorted room is called an *Ames room*, after Adelbert Ames, who built the first one.

37

How might the unconscious assessment of depth provide a basis for the Ponzo, Müller-Lyer, and moon illusions?

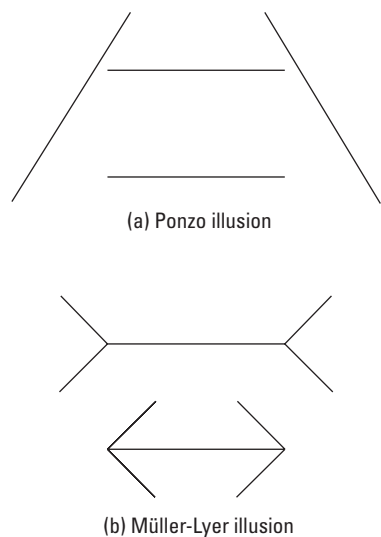


FIGURE 8.41 | The Ponzo and Müller-Lyer illusions In both (a) and (b), the top horizontal bar looks longer than the bottom one, although they are actually the same length.

Unconscious Depth Processing as a Basis for Size Illusions

<<< It is not difficult to produce drawings in which two identical lines or objects appear to be different in size. Two classic examples are the **Ponzo illusion** (first described by Mario Ponzo in 1913) and the **Müller-Lyer illusion** (first described by F. C. Müller-Lyer in the mid-nineteenth century), both of which are illustrated in Figure 8.41. In each illusion two horizontal bars appear to be different in length; but if you measure them, you will discover that they are in fact identical. Some years ago, Richard Gregory (1968) offered a *depth-processing theory* to account for these and various other size illusions. This theory—consistent with everything said so far about the relation between size and distance—maintains that one object in each illusion appears larger than the other because of distance cues that, at some early stage of perceptual processing, lead it to be judged as farther away. If one object is judged to be farther away than the other but the two produce the same-size retinal image, then the object judged as farther away will be judged as larger. This theory applies most readily to the Ponzo illusion, in which the two converging lines provide the depth cue of linear perspective, causing (according to the theory) the upper bar to be judged as farther away and hence larger than the lower one. The photograph in Figure 8.42 makes this point clear.

The application of the depth-processing theory to the Müller-Lyer illusion is a bit more subtle. The assumption here is that people register the figures as three-dimensional objects, something like sawhorses viewed from above. The object with wings extending outward (top drawing in Figure 8.43) resembles an upside-down sawhorse, with legs toward the viewer, and the one with inward wings (bottom drawing) resembles a right-side-up sawhorse, with its horizontal bar closer to the observer. If real sawhorses were viewed this way, the horizontal bar of the upside-down one would be farther from the observer than that of the right-side-up one, and if it produced the same-size retinal image, it would in fact be longer (see Figure 8.43).

The depth-processing theory seems to offer an elegant explanation of the Ponzo and Müller-Lyer illusions, but perhaps you have already thought of an objection to it. Most people claim to see the Müller-Lyer figures not as three-dimensional objects but as flat, arrowlike objects, yet they experience the illusion. Even with the Ponzo illusion, many people do not notice depth in the picture; they do not see the top bar as being farther away than the bottom one, yet they see the illusion. Researchers who support the depth-processing theory of these illusions are well



FIGURE 8.42 | Depth-processing explanation of the Ponzo illusion If this were a real, three-dimensional scene, not a photograph, and the red bars really existed as shown, the one in the background would not only look larger but would be larger than the one in the foreground. (Adapted from Gregory, 1968.)

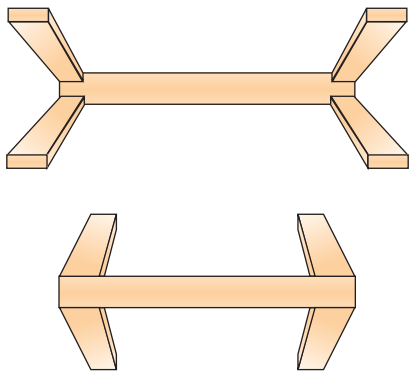


FIGURE 8.43 | Depth-processing explanation of the Müller-Lyer illusion Compare these sawhorses with the Müller-Lyer drawings in Figure 8.41. If these were real sawhorses, viewed from above in the three-dimensional world, the upside-down one would be longer than the right-side-up one.

aware of this objection, but it does not lead them to abandon the theory. They remind the dissenters, as did Helmholtz, that the mental processes leading to perceptions are unconscious. The cues to depth in the Müller-Lyer and Ponzo figures may be too weak, or too ambiguous, to lead to a conscious perception of depth, yet they may still lead to an *unconscious* assessment of depth. That unconscious assessment may enter into the mental calculation that leads to a conscious perception of size differences.

The Moon Illusion

The **moon illusion** has provoked debate since ancient Greek and Roman times. You have probably noticed that the moon looks huge when it is near the earth's horizon, just above the trees or buildings in the distance, but looks much smaller when it is closer to the zenith (directly overhead). This difference is truly an illusion. Objectively, the moon is the same size, and the same distance from us, whether it is at the horizon or the zenith. If you view the horizon moon through a peephole, so that you see it in isolation from other objects such as trees and buildings, the illusion disappears and the moon looks no larger than it does at the zenith.

A depth-processing account of this illusion was first proposed by the Greek astronomer Ptolemy in the second century, was revived by Helmholtz (1867/1962) in the nineteenth century, and has been supported by modern research conducted by Lloyd Kaufman and his colleagues (Kaufman & Kaufman, 2000; Kaufman & Rock, 1962). The account can be summarized as follows: Our visual system did not evolve to judge such huge distances as that from the earth to the moon, so we automatically assess its distance in relation to more familiar earthly objects. Most objects that we see near the earth's horizon are farther away than objects that we see farther from the horizon (as noted earlier, in the discussion of pictorial cues for depth). For example, birds or clouds seen near the horizon are usually farther away than are those seen higher up, closer to the zenith. Thus, our perceptual system assumes that the moon is farther away at the horizon than at the zenith, even though in reality it is the same distance away from us in either position. As in the case of the Ponzo and Müller-Lyer illusions, when two objects produce the same-size retinal image and are judged to be different distances away, the one that is judged to be farther away is seen as larger than the other.

Even today, the main objection to this explanation of the moon illusion is that people do not consciously see the horizon moon as farther away than the zenith moon (Hershenson, 1989, 2003). When people see the large-appearing horizon moon and are asked whether it seems farther away or closer than usual, they usually say closer. Again, however, as with the Ponzo and Müller-Lyer illusions, Kaufman and Irving Rock (1989) contend that we must distinguish between unconscious and conscious assessments. From their perspective, the sequence of perceptual assessments about the horizon moon might be described as follows: (1) Unconscious processes judge that the moon is farther away than usual (because objects near the horizon are usually farthest away). (2) Unconscious processes judge that the moon is larger than usual (because if it is farther away but produces the same-size retinal image, it must be larger), and this judgment enters consciousness. (3) If asked to judge distance, most people say that the horizon moon looks closer (because they know that the moon doesn't really change size, so its large apparent size must be due to closeness). This explanation has been referred to as the *farther-larger-nearer theory* (Ross & Plug, 2002).

Not all perceptual psychologists agree with the depth-processing account of the moon illusion, or with that of the Ponzo or Müller-Lyer illusions, but that account



The moon illusion The moon at the horizon looks very large. Seen in isolation, with no cues indicating that it is at the horizon, the moon would appear to be the same size at the horizon as at the zenith.



FIGURE 8.44 | Objects in context The schematic ear, mouth, and eye are easy to recognize in the context of the face, but they may not have been recognizable in Figure 8.33 (on page 291), where they were shown out of context.

seems to be supported by more evidence and logic than any other explanations that have been offered to date (Hershenson, 1989; Kaufman & Kaufman, 2000; Ross & Plug, 2002). It is perhaps fitting that we end this chapter, now, on a tentative note. Although perceptual psychologists have made great strides since the days of Ptolemy, we are still a long way from a full account of the calculations that our brains make to infer the sizes, distances, and shapes of all the objects in our field of view.

SECTION REVIEW

We see three-dimensionally—that is, with depth—through the retina is two-dimensional.

Depth Perception Cues

- Various cues are used by our visual system to infer the depth (distance) of objects or parts of objects.
- Binocular disparity is a major depth cue that derives from the fact that the two eyes, because of their different spatial positions, receive somewhat different images of an object.
- Another depth cue, motion parallax, is similar to binocular disparity but can occur with just one eye. It makes use of the different images of an object that either eye receives as the head moves right or left.
- Pictorial depth cues, such as linear perspective, do not depend on actual three-dimensionality. They allow us to infer depth even in two-dimensional pictures.

Size Perception

- The size of an object’s retinal image is inversely proportional to its distance from the viewer, so size perception depends on depth perception.
- Size constancy is the ability to perceive an object as the same size when its retinal image size varies as a result of changes in its distance.
- The Ponzo, Müller-Lyer, and moon size illusions may derive from unconscious inferences about depth. If two objects create identical retinal images, the one that is unconsciously judged as farther away will be seen as larger.

Concluding Thoughts

As you review this chapter, you may find it useful to think about the following three themes.

1. The survival functions of vision Our visual system, like all the basic mechanisms underlying behavior, evolved through natural selection to promote our ancestors’ survival and reproduction. Our visual system is not an unbiased recorder of physical patterns of light. Rather, it is a biological tool designed to pick out the information in patterns of light that is potentially most useful. As you review, think about the survival advantage of each of the following processes, phenomena, or mechanisms: (a) the processes of light and dark adaptation, (b) the distinctions between cone vision and rod vision, (c) color vision, (d) the enhancement of contrast at borders, (e) feature detection by neurons, (f) Treisman’s postulated two stages of feature pickup and integration, (g) the complementary roles of parallel and serial processing, (h) the Gestalt laws of perceptual grouping and of figure and ground, (i) illusory contours, (j) illusory lightness differences, (k) the distinction between the “what” and “where-and-how” visual pathways, (l) Biederman’s idea of geons serving as a basic alphabet of perception, (m) the distinctive-feature theory of perceptual learning, (n) the roles of context and motion in object identification, (o) the multiple cues that allow us to assess depth, and (p) the role of depth cues in perception of size. In most cases the “illusions” that our visual system produces are useful in helping us to recognize and interact effectively with objects in the real world in which we live.

2. The interaction of bottom-up and top-down processes in vision The neural processes that bring sensory information into the brain and move it along to higher analyzing centers are referred to as bottom-up processes, and those that bring information down from higher centers to contribute to the analysis occurring at lower centers are referred to as top-down processes. Bottom-up processes discussed in this chapter include (a) the coding of variations in light by rods and cones; (b) the organization of that information to register elementary sensory features such as color, lightness, contour orientations, and degree of binocular disparity; and (c), in Treisman’s theory, the integration of sensory features that occupy a given spatial location. Top-down processes are assumed to be involved whenever the perception of the whole, or expectations generated by the context, are brought to bear in the analysis of the sensory information. The phenomena discussed in this chapter that are assumed to involve top-down control include illusory contours, illusory lightness differences that cannot be explained by lateral inhibition, effects of context (such as being in a shopping mall) on object recognition, and size judgments that are based on unconscious assessments of depth.

3. The integration of behavioral and neurophysiological discoveries Herman Helmholtz realized long ago that all our perceptions result somehow from the interaction between physical properties of the external world and physiological processes occurring in our sensory organs and brain. His trichromatic the-

ory of color vision is an example of his attempt to explain perceptual phenomena—in this case, the three-primaries law of color mixing—in physiological terms. In recent years, great strides have been made in carrying out the general program of research pioneered by Helmholtz.

Examples of modern explanations of perceptual phenomena, discussed in this chapter, that are in the direct tradition of Helmholtz include (a) the explanation of differences between bright-light vision and dim-light vision in terms of differences between cones and rods in their photochemicals and their degree of neural convergence; (b) the explanations of the two laws of additive color mixing in terms of three varieties of cones and the pattern of neural wiring of those cones to opponent-process

neurons; (c) the explanation of contour enhancement in terms of lateral inhibition; (d) the explanation of rapid stimulus feature detection in terms of the properties of individual neurons in the primary visual cortex; (e) the explanation of distinctions between conscious visual perception and unconscious use of visual input to guide movement in terms of two separate pathways of visual processing; (f) the explanation of the role of components in object perception in terms of processes occurring in the “what” pathway of visual processing; and (f) the explanation of binocular disparity as a depth cue in terms of neurons that receive input from slightly different locations on the retinas of the two eyes.

Further Reading

RICHARD GREGORY (1997). *Eye and brain: The psychology of seeing* (5th ed.). Princeton, NJ: Princeton University Press.

This relatively brief book is an often-referred-to classic. With clear prose and illustrations, Gregory introduces the reader to the entire range of issues concerning visual psychology, beginning with a chapter on the nature of light and progressing through chapters on the eye, the visual brain, brightness perception, movement perception, color perception, the role of learning in vision, art, and illusions.

DALE PURVES & R. BEAU LOTTO (2003). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinauer.

This book is concerned with the relation between physical patterns of light and the percepts that our visual system generates in response to those patterns. It is filled with illusions—of lightness, color, depth, and motion—that are not readily explained in terms of bottom-up neural wiring or straightforward logical calcula-

tions. According to the authors' empirical theory, evolution and learning have shaped the visual system such that it can match any given pattern of reflected light, probabilistically, with some percept that, in the past, has been most useful when a similar pattern was present. The book is clearly written and beautifully illustrated, and it challenges much conventional thinking about vision.

MELVYN GOODALE & A. DAVID MILNER (2004). *Sight unseen: An exploration of conscious and unconscious vision*. Oxford: Oxford University Press.

In this brief, well-written, well-illustrated book, Goodale and Milner develop, for nonspecialists, their thesis that the brain contains two rather distinct visual-processing systems, one for conscious visual perception and one for immediate, unconscious guidance of action. Included is a full account of the visual abilities and deficits of D. F., the patient who suffers from visual form agnosia but can nevertheless act on objects in ways that take form into account.

HELEN E. ROSS & CORNELIS PLUG (2002). *The mystery of the moon illusion: Exploring size perception*. Oxford: Oxford University Press.

This book discusses the many theories and lines of evidence concerning the moon illusion that have been presented since ancient times. Considerable evidence emerges supporting the *farther-larger-nearer* theory described in this chapter, but the authors conclude that other factors probably contribute to the illusion as well. The book reveals some of the methodological problems associated with research on size perception.

ROGER SHEPARD (1990). *Mind sights*. New York: Freeman.

This is a book of playful drawings by a distinguished perceptual psychologist who is also a talented artist and humorist. The drawings use size illusions, figure-ground ambiguities, and other visual tricks to present jokes and puns. The author's commentary on his drawings explains what the various tricks tell us about perception.

FPO



THE HUMAN INTELLECT

The effectiveness of our behavior depends on knowledge we have stored as memory. It also depends on our ability to call up and combine the portions of that knowledge that are useful for the task at hand. How do we store and organize our memories? How do we recall memories when we need them? How do we manipulate knowledge in our minds in order to reason and solve problems? What causes individual differences in problem-solving ability? These big questions are what concern us in Chapter 9, on memory and consciousness, and Chapter 10, on intelligence and reasoning.

part 5

Memory and Consciousness

chapter 9

OVERVIEW: AN INFORMATION-PROCESSING MODEL OF THE MIND

ATTENTION: THE PORTAL TO CONSCIOUSNESS

- The Ability to Focus Attention and Ignore the Irrelevant
- The Ability to Shift Attention to Significant Stimuli
- Unconscious, Automatic Processing of Stimulus Input
- Brain Mechanisms of Attention and Preattentive Processing

WORKING MEMORY: THE ACTIVE, CONSCIOUS MIND

- Verbal Working Memory: The Phonological Loop
- Visual Working Memory: The Visuospatial Sketchpad
- Testing the Limits of Working Memory: Dual Task Performance
- Brain Areas Involved in Working Memory

ENCODING INFORMATION INTO LONG-TERM MEMORY

- Elaboration Promotes Encoding
- Organization Promotes Encoding
- Visualization Promotes Encoding
- Brain Mechanisms of Long-Term-Memory Encoding

RETRIEVING INFORMATION FROM LONG-TERM MEMORY

- Mental Associations and Memory Retrieval Cues
- Memory Construction as a Source of Distortion

MULTIPLE MEMORY SYSTEMS: BEYOND THE MODAL MODEL

- Distinctions Among Explicit- and Implicit-Memory Systems
- Neuropsychological Evidence for Separate Memory Systems

Repeatedly, while working on this book, I have lamented my seeming lack of memory. I can't remember who did that experiment. I forgot to bring home the articles I need for this section. I've looked up that word a thousand times but still don't remember how to spell it. Like digestion, memory is one of those abilities that we tend to take for granted except when it fails us. Usually we are more aware of forgetting than of remembering. But if we stop to think about it, we realize that our remembering is far more impressive than our forgetting. Every waking moment is full of memories. Every thought, every learned response, every act of recognition is based on memory. It can reasonably be argued that memory is the mind.

Memory, clearly, is intimately tied to learning. Memory is often thought of as the change within an individual, brought on by learning, that can influence the individual's future behavior: Learning → memory → effect on future behavior. Chapter 4 of this book examines basic learning processes that occur in other animals as well as humans. Consistent with its behavioral theme, the focus in that chapter is on the relationship between observable aspects of the learning experience (the training conditions) and subsequent behavior, with little concern for the inner change—memory—that mediates that relationship. The present chapter, in contrast, is primarily about that inner change, and it deals with types of learning and memory that may be unique to human beings. Our main focus here is on the conscious human mind.

Consciousness is one of those words that different philosophers and psychologists use in different ways and about which they often debate. For practical purposes, most psychologists define consciousness simply as the experiencing of one's own mental events in such a manner that one can report on them to others (Baars & Franklin, 2003). Defined this way, *consciousness* and *awareness* are synonyms. If you tell me, correctly, that a picture I show you has a bluebird in it, then I assume that you consciously see the bluebird. If, sometime later, you say, "a bluebird," when I ask you what was in the picture, then I assume you have consciously recalled the bluebird.

The chapter begins with a general model of memory that psychologists use as a framework for talking and thinking about the human mind. It then discusses issues of attention, mental

imagery, the formation and recall of long-term memories, and the idea that the human mind is composed of several different but interacting memory systems that have distinct neural bases.

Overview: An Information-Processing Model of the Mind

As was noted in Chapter 1, cognitive psychologists commonly look at the mind (or brain) as a processor of information, analogous to a computer. Information is brought into the mind by way of the sensory system, and then it can be manipulated in various ways, placed into long-term storage, and retrieved when needed to solve a problem. **Memory**, broadly defined from this perspective, refers to all of the information in a person's mind and to the mind's capacity to store and retrieve that information. Memory is too large and multifaceted a topic to be studied or talked about all in one piece. Progress has been made in understanding memory by breaking it into components that can be described and studied separately. But any such breakdown implies a theory about memory—a theory that memory consists of the proposed components, that these can be studied relatively separately from one another, and that an understanding of the components contributes to an understanding of the whole.

1 What are the main components of the so-called modal model of the mind? What is the purpose of such a model?

<<< In cognitive psychology, theories are commonly called *models* and are often summarized visually with diagrams that use boxes to represent the mind's components and arrows to represent the movement of information from one component to another. Such models attempt to describe at an abstract level the general ways by which the mind processes information—that is, the operations through which the mind modifies, stores, and uses the input it receives from the senses. Each component in the model is considered to be a center for a particular set of such operations. Cognitive psychologists test their models by using them to develop specific hypotheses about how people will perform on particular tasks. A model is considered to be successful if the predictions prove to be accurate and unsuccessful if the predictions prove to be inaccurate. In general, cognitive models do not attempt to explain how mental tasks are accomplished at the neural level. However, in recent years, many cognitive psychologists have attempted to match up the compartments of their hypothetical models with physical locations in the brain, as a step toward understanding the neural mechanisms.

The model that guides much of the discussion in this chapter (depicted in Figure 9.1) has been so influential, for so long, that it has come to be called the **modal model of the mind**, where *modal* means “standard.” Versions of this model were first proposed in the 1960s (by Waugh & Norman, 1965, and Atkinson & Shiffrin, 1968), and ever since it has served as a general framework for thinking and talking about the mind. As you use this model throughout this chapter, keep in mind that it is only a model. Like any general model in psychology, it is simply a

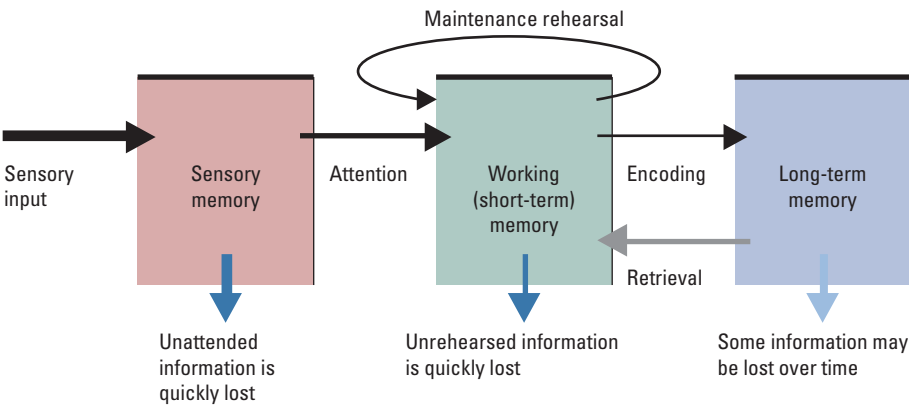


FIGURE 9.1 The modal model of the mind. This model has long served as a framework for thinking about the human mind, and we will use it for that purpose throughout the chapter.

way of trying to make sense of the data from many behavioral studies. And like any model, it can place blinders on thought and research if taken literally rather than as a metaphor.

The model portrays the mind as containing three types of **memory stores**—sensory memory, working (or short-term) memory, and long-term memory—conceived of as places where information is held and operated on. Each store type is characterized by its *function* (the role it plays in the overall workings of the mind), its *capacity* (the amount of information it can hold at any given instant), and its *duration* (the length of time it can hold an item of information). In addition to the stores, the model specifies a set of **control processes**, including *attention*, *rehearsal*, *encoding*, and *retrieval*, which govern the processing of information within stores and the movement of information from one store to another. As a prelude to the more detailed later discussions, here is a brief description of the three stores and the control processes.



Erich Hartmann / Magnum Photos

Sensory Memory: The Brief Prolongation of Sensory Experience

When lightning flashes on a dark night, you can still see the flash and the objects it illuminated for a split second beyond its actual duration. Somewhat similarly, when a companion says, “You’re not listening to me,” you can still hear those words, and a few words of the previous sentence, for a brief time after they are spoken. Thus, you can answer (falsely), “I was listening. You said . . .”—and then you can repeat your annoyed companion’s last few words even though, in truth, you weren’t listening when the words were uttered. These observations demonstrate that some trace of sensory input stays in your information-processing system for a brief period—less than 1 second for sights and up to several seconds for sounds—even when you are not paying attention to the input. This trace and the ability to hold it are called **sensory memory**.

A separate sensory-memory store is believed to exist for each sensory system (vision, hearing, touch, smell, and taste), but only those for vision and hearing have been studied extensively. Each sensory store is presumed to hold, very briefly, all the sensory input that enters that sensory system, whether or not the person is paying attention to that input. The function of the store, presumably, is to hold on to sensory information long enough for it to be analyzed by unconscious mental processes and for a decision to be made about whether or not to bring that information into the next store, that of working memory. Most of the information in our sensory stores does not enter into our consciousness. We become conscious only of those items that are transported, by the selective process of *attention*, into working memory.

Why don’t we see this way? If you’re wondering why multiple representations of the icon don’t distort your view of the visual world, the answer is that each new image from instant to instant blocks out the previous image.

>>> 2
What is the function of sensory memory?

Working Memory: Conscious Perception and Thought

According to the modal model, some of the information processed in the sensory store moves into the next compartment, which is called **working memory** (look again at Figure 9.1). This is conceived of as the major workplace of the mind. It is, among other things, the seat of conscious thought—the place where all conscious perceiving, feeling, comparing, computing, and reasoning take place. An older and still often-used alternative name for this memory compartment is **short-term memory**, a term that calls attention to the relatively fleeting nature of information in this store; each item fades quickly, and is lost within seconds, when it is no longer actively attended to or thought about.

As depicted by the arrows in Figure 9.1, information can enter working memory from both the sensory-memory store (representing the present environment) and the long-term-memory store (representing knowledge gained from previous expe-

>>> 3
What are the basic functions of working memory, and how is this memory store equated with consciousness? In what way is working memory like the central processing unit of a computer?



The passing moment The flow of thought through working memory is not unlike the flow of scenery past the window of a moving train.

riences). In this sense, working memory is analogous to the central processing unit of a computer. Information can be transmitted into the computer's central processing unit from a keyboard (comparable to input from the mind's sensory store), or it can be entered from a long-term storage device such as a CD or the computer's hard drive (comparable to input from the mind's long-term store). The real work of the computer—computation and manipulation of the information—occurs within its central processing unit.

The sensory store and long-term-memory store both contribute to the continuous flow of conscious thought that constitutes the content of working memory. *Flow* is an apt metaphor here. The momentary capacity of working memory is very small—only a few items of information can be perceived or thought about at once. Yet the total amount of information that moves through working memory over a period of minutes or hours can be enormous, just as a huge amount of water can flow through a narrow channel over time.

Long-Term Memory: The Mind's Library of Information

Once an item has passed from sensory memory into working memory, it may or may not then be encoded into **long-term memory** (again, see Figure 9.1). Long-term memory corresponds most closely to most people's everyday notion of memory. It is the stored representation of all that a person knows. As such, its capacity must be enormous. Long-term memory contains the information that enables us to recognize or recall the taste of an almond, the sound of a banjo, the face of a grade-school friend, the names of the foods eaten at supper last night, the words of a favorite sonnet, and the spelling of the word *sonnet*. Usually we are not aware of the millions of items of information in our long-term store. According to the model, the items lie dormant, or relatively so, like books on a library shelf or digital patterns on a computer disk until they are called into working memory and put to use.

4 In the modal model, what is the function of long-term memory, and how is this memory store different from working memory?

<<< In the modal model, long-term memory and working memory are sharply differentiated. Long-term memory is passive (a repository of information), and working memory is active (a place where information is thought about). Long-term memory is of long duration (some of its items last a lifetime), and working memory is of short duration (items disappear within seconds when no longer thought about). Long-term memory has essentially unlimited capacity (all your long-lasting knowledge is in it), and working memory has limited capacity (only those items of information that you are currently thinking about are in it).

Control Processes: The Mind's Information Transportation Systems

5 In the modal model, what are the functions of attention, encoding, and retrieval?

<<< According to the modal model, the movement of information from one memory store to another is regulated by the control processes of *attention*, *encoding*, and *retrieval*, all indicated in Figure 9.1 by arrows between the boxes.

Attention is the process that controls the flow of information from the sensory store into working memory. Because the capacity of sensory memory is large and that of working memory is small, attention must restrict the flow of information from the first into the second.

Encoding is the process that controls movement from working memory into the long-term store. When you deliberately memorize a poem or a list of names, you are consciously encoding it into long-term memory. Most encoding, however, is not deliberate; rather, it occurs incidentally, as a side effect of the special interest that you devote to certain items of information. If you become interested in, and think about, ideas in this book you will, as a side effect, encode many of those ideas and the new terms relating to them into long-term memory.

Retrieval is the process that controls the flow of information from the long-term store into working memory. Retrieval is what we commonly call *remembering* or *re-*



Control processes at play The man who best controls his control processes of attention, encoding, and retrieval is likely to win this game, unless, of course, he was dealt a lousy hand.

calling. Like both attention and encoding, retrieval can be either deliberate or automatic. Sometimes we actively search our long-term store for a particular piece of information. More often, however, information seems to flow automatically into the working store from the long-term store. One image or thought in working memory seems to call forth the next in a stream that is sometimes logical, sometimes fanciful.

Now, having previewed the memory stores and processes represented in the model, let's examine them in more detail, beginning with sensory memory and attention.

SECTION REVIEW

The modal model of the mind describes three memory stores and various control processes.			
Sensory Memory • A separate sensory store for each sensory system (vision, hearing, etc.) holds brief traces of all information registered by that system. • Unconscious processes may operate on these traces to decide which information to pass on to working memory.	Working Memory • This store is where conscious mental work takes place on information brought in from sensory memory and long-term memory. • It is also called short-term memory because information no longer attended to quickly disappears from it.	Long-Term Memory • This is the repository of all that a person knows. • Information here is dormant, being actively processed only when it is brought into working memory.	Control Processes • Attention brings information from sensory memory into working memory. • Encoding brings information from working memory into long-term memory. • Retrieval brings information from long-term memory into working memory.

Attention: The Portal to Consciousness

Imagine one of our prehistoric hunting-and-gathering ancestors foraging for certain nutritious roots. Most of her conscious attention is devoted to perceiving and analyzing visual stimuli in the soil that tell her where the roots might be. At the same time, however, at some level of her mind, she must monitor many other sensory stimuli. The slight crackling of a twig in the distance could signal an approaching tiger. The stirring of the infant on her back could indicate that the baby needs comforting before it begins to cry and attract predators. A subtle darkening of the sky could foretell a dangerous storm. On the positive side, a visual clue in the foliage, unrelated to stimuli she is focusing on, could indicate that an even more nutritious form of vegetation exists nearby.

6 What two competing needs are met by our attentional system? How do the concepts of preattentive processing and top-down control of the attentive gate pertain to these two needs?

<<< Natural selection endowed us with mechanisms of attention that can meet two competing needs. One need is to focus mental resources on the task at hand and not be distracted by irrelevant stimuli. The other, opposing need is to monitor stimuli that are irrelevant to the task at hand and to shift attention immediately to anything that signals some danger or benefit that outweighs that task. Cognitive psychologists have been much concerned with the question of how our mind manages these two competing needs.

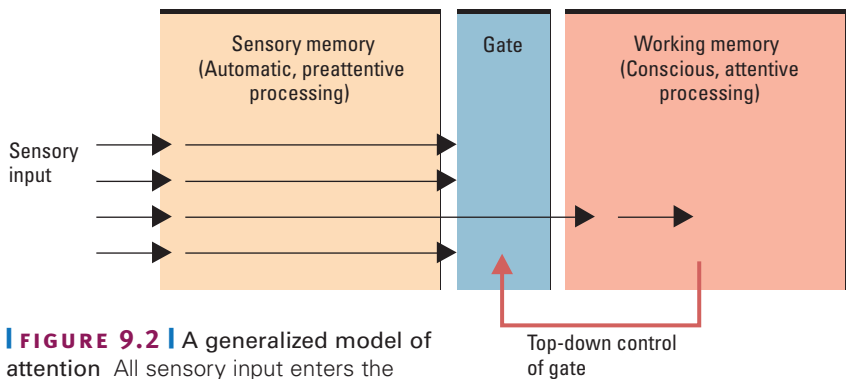


FIGURE 9.2 | A generalized model of attention All sensory input enters the sensory memory store, where it is processed preattentively. Then some of it is selected to pass through the gate into conscious, working memory. The arrow going from working memory to the gate indicates top-down control of the selection criteria.

Figure 9.2 depicts a very general model, in which attention is portrayed as a gate standing between sensory memory and working memory. According to this model, all information that is picked up by the senses enters briefly into sensory memory and is analyzed to determine its relevance to the ongoing task and its potential significance for the person's survival or well-being. That analysis occurs at an unconscious level and is referred to as **preattentive processing**. Logically, such processing must involve some comparison of the sensory input to information already stored in working memory and long-term

memory. Without such comparison, there would be no basis for distinguishing what is relevant or significant from what is not. Activity in higher parts of the brain that are involved in preattentive processing must somehow operate on the attention gate to help determine what items of information will be allowed to pass into the limited-capacity, conscious, working-memory compartment at any given moment. In Figure 9.2, that top-down control is depicted by the arrow running from the working-memory compartment to the gate. The degree and type of preattentive processing that occurs, and the nature of the top-down control of the gate, are matters of much speculation and debate (LaBerge; 1995; Pashler, 1998).

With this general model in mind, let's now consider some research findings that bear on the two competing problems that the attention system must solve: focusing attention narrowly on the task at hand and monitoring all stimuli for their potential significance.

The Ability to Focus Attention and Ignore the Irrelevant

Human beings are generally very good at attending to a relevant train of stimuli and ignoring stimuli that are irrelevant to the task they are performing. Here is some evidence for that.

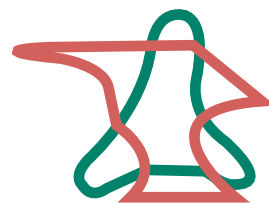
Selective Listening

7 How have researchers shown that people very effectively screen out irrelevant sounds and sights when focusing on difficult perceptual tasks?

<<< The pioneering research on attention, beginning in the 1940s and 1950s, centered on the so-called *cocktail-party phenomenon*, the ability to listen to and understand one person's voice while disregarding other, equally loud or even louder voices nearby. In the laboratory this ability is usually studied by playing recordings of two spoken messages at once and asking the subject to *shadow* one message—that is, to repeat immediately each of its words as they are heard—and ignore the other message. The experiments showed that people are very good at this as long as there is some physical difference between the two voices, such as in their general pitch levels or in the locations in the room from which they are coming (Hawkins & Presson, 1986). When asked immediately after the shadowing task about the unattended voice, the subjects could usually report whether it was a woman's voice or a man's, but were usually unaware of any of the message's meaning or even whether the speaker switched to a foreign language partway through the session (Cherry, 1953; Cherry & Taylor, 1954).



Selective listening Cocktail parties are not the only setting in which the ability to attend to one voice and ignore others is valuable.



Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 275-285. Copyright © American Psychological Association. Reprinted by permission.

FIGURE 9.3 Overlapping forms used in an experiment on attention To assess the degree to which vision can be selective, Rock and Guttman (1981) directed subjects to attend to either just the red or just the green shape in slides such as this and then tested their recognition of both shapes in each slide.

Selective Viewing

On the face of it, selective viewing seems to be a simpler task than selective listening; we can control what we see just by moving our eyes, whereas we have no easy control over what we hear. But we can also attend selectively to different, nearby parts of a visual scene without moving our eyes, as demonstrated in a classic experiment by Irvin Rock and Daniel Gutman (1981). These researchers presented, in rapid succession, a series of slides to viewers whose eyes were fixed on a spot at the center of the screen. Each slide contained two overlapping forms, one green and one red (see Figure 9.3), and subjects were given a task that required them to attend to just one color. Most of the forms were nonsense shapes, but some were shaped like a familiar object, such as a house or a tree. After viewing the sequence, subjects were tested for their ability to recognize which forms had been shown. The result was that they recognized most of the forms that had been presented in the attended color but performed only at chance level on those that had been presented in the unattended color, regardless of whether the form was nonsensical or familiar.

The most dramatic evidence of selective viewing comes from experiments in which subjects who are intent on a difficult visual task fail to see large, easily recognized objects directly in their line of sight. In one such experiment (Simons & Chabris, 1999), college students watched a 75-second video in which three black-shirted players tossed a basketball among themselves, and three white-shirted players tossed another basketball among themselves, while all six moved randomly around in the same small playing area. Each subject's task was to count the number of passes made by one of the two groups of players while ignoring the other group. Midway through the video, a woman dressed in a gorilla costume walked directly into the center of the two groups of players, faced the camera, thumped her chest, and then, several seconds later, continued walking across the screen and out of view (see Figure 9.4). Remarkably, when questioned immediately after the video, 50 percent of the subjects claimed they had not seen the gorilla. When these subjects were shown the video again, without having to count passes, they expressed amazement—with such exclamations as, “I missed *that*?!”—when the gorilla came on screen.

FIGURE 9.4 The unseen gorilla Subjects who were asked to count the number of times one group of players threw the basketball failed to see the gorilla. (From Simons & Chabris, 1999.)



Daniel J. Simons and Christopher F. Chabris

The Ability to Shift Attention to Significant Stimuli

If our hunter-gatherer ancestors frequently missed real gorillas, or lions or tigers, because of their absorption in other tasks, our species would have become extinct long ago. Similarly, neither you nor I would be alive today if we didn't shift our attention to unexpected dangers while crossing the street. We are good at screening out irrelevant stimuli when we need to focus intently on a task, but we are also good (although far from perfect) at shifting our attention to stimuli that signal danger or benefit or are otherwise significant to us. Our ability to shift attention appears to depend, in part, on our capacity to listen or look backward in time and “hear” or “see” stimuli that were recorded a moment earlier in sensory memory. A major function of sensory memory, apparently, is to hold on to fleeting, unattended stimuli long enough to allow us to turn our attention to them and bring them into consciousness if they prove to be significant (Lachter & others, 2004). Here is some evidence for that idea.

8

According to the theory described here, how does sensory memory permit us to hear or see, retroactively, what we were not paying attention to? How have experimenters measured the duration of auditory and visual sensory memory?

Attending to Information in Auditory Sensory Memory

<<< Auditory sensory memory is also called **echoic memory**, and the brief memory trace for a specific sound is called the *echo*. Researchers have found that the echo fades over a period of seconds and vanishes within at most 10 seconds (Cowan & others, 2000; Gomes & others, 1999).

In a typical experiment on echoic memory, subjects are asked to focus their attention on a particular task, such as reading a prose passage that they will be tested on, and to ignore spoken words that are presented as they work. Occasionally, however, their work on the task is interrupted by a signal, and when that occurs they are to repeat the spoken words (usually digit names, such as “seven, two, nine, four”) that were most recently presented. The typical result is that subjects can repeat accurately the last few words from the spoken list if the signal follows immediately after the last word, but performance drops off as the delay between the last word and the signal is increased, and it vanishes at about 8 to 10 seconds (Cowan & others, 2000). The subjective experience of people in such experiments is that they are only vaguely aware of the spoken words as they attend to their assigned task, but when the signal occurs, they can shift their attention back in time and “hear” the last few words as if the sounds were still physically present.

In experiments such as this, it is the experimenter's signal that alerts the subjects to shift attention. In other cases, however, the shift occurs because of the subject's own unconscious, preattentive processing of the unattended stimuli. In the original experiments on selective listening, in which people shadowed one voice and ignored another, subjects typically showed no conscious memory for any of the ideas or words spoken by the unattended voice. However, if their own names appeared in the unattended voice, they noticed and reported that about one-third of the time (Moray, 1959). How were they able to hear their own names when they had heard nothing earlier in that voice? Presumably all the words spoken by the unattended voice entered into sensory memory and were to some degree processed preattentively for meaning. Since one's own name has special meaning, preattentive processing alerted the attention mechanism, and the person could then “hear” his or her own name because it was still stored in sensory memory.

Attending to Information in Visual Sensory Memory

Visual sensory memory is also called **iconic memory**, and the brief memory trace for a specific visual stimulus is called the *icon*. The first psychologist to suggest the existence of iconic memory was George Sperling (1960). In experiments on visual perception, Sperling found that when slides containing rows of letters were flashed for one-twentieth of a second, people could read the letters, as if they were still physically present, for up to one-third of a second after the slide was turned off. For instance, if a slide was flashed containing three rows of letters and then a signal came on that told the subjects to read the third row, subjects could do that if the signal occurred within a third of a second after the slide was flashed. Such findings

led Sterling to propose that a memory store must hold visually presented information, in essentially its original sensory form, for about a third of a second beyond the termination of the physical stimulus.

A number of experiments have shown that people who are attending to one set of visual stimuli and successfully ignoring another set frequently notice their own name in the ignored set (Mack & others, 2002). A plausible explanation, again, is that preattentive processes decipher the name, recognize it as significant, and shift the person's attention to it while it is still available in sensory memory.

Effect of Practice on Attentional Capacity

The gate between sensory memory and working memory is narrow. Only a few items of information can cross through it at any given instant. If an array of visual stimuli—such as digits or simple shapes—are flashed quickly on a screen, and then a masking stimulus is flashed so as to erase the iconic memory for the original array, most people can identify only about three of the stimuli in the original array. Some psychologists have long argued that the capacity to attend to several items at once can be increased with practice (Neisser, 1976). Perhaps the best evidence that it can come from recent research on effects of video-game playing.

In an initial study, C. Shawn Green and Daphne Baveller (2003) tested the visual attentional capacity of two groups of young men—one group who regularly played action video games and another group who never or very rarely played such games. On every measure, the video-game players greatly outperformed the other group. In one test, for example, the video-game players correctly identified an average of 4.9 items in a briefly flashed array, compared to an average of 3.3 for the non-video-game players. In another test, in which target stimuli appeared in random places amid distracting stimuli on the video screen, video-game players correctly located twice as many target stimuli as did the non-players. Subsequently, the researchers performed an experiment in which people who had previously not played video games were assigned to play a video game for one hour per day for ten days. One group, randomly chosen, was assigned to play Medal of Honor, an action game that requires one to track known enemies while watching constantly for new enemies and other dangers. The other group was assigned to play Tetris, a game that requires players to focus on only one object at a time. The results were dramatic. The Medal of Honor group improved markedly on all the tests of attentional capacity, and the Tetris group did not improve at all.

>>> 9
What is the effect of action-video-game playing on attentional capacity?



Unconscious, Automatic Processing of Stimulus Input

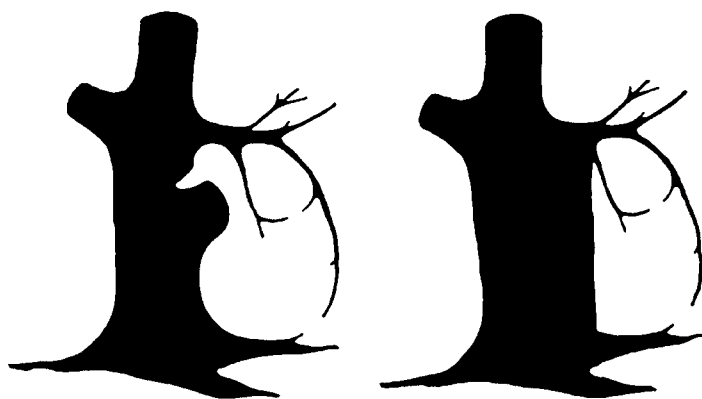
The modal model, as depicted in Figure 9.1 (p 000), is a useful beginning point for thinking about the mind, but it is far from complete. Perhaps its greatest deficiency is its failure to account for unconscious effects of sensory input. In the model, sensory information can influence behavior and consciousness only if it is attended to and enters the conscious working-memory compartment; otherwise, it is simply lost from the system. In fact, however, there is much evidence that sensory input can alter behavior and conscious thought without itself becoming conscious to the person. One means by which it can do this is called *priming*.

Unconscious Priming of Mental Concepts

Priming can be defined as the activation, by sensory input, of information that is already stored in long-term memory. The activated information then becomes more available to the person, altering the person's perception or chain of thought. The activation is not experienced consciously, yet it influences consciousness. Most relevant for our present discussion, there is good evidence that such activation can occur even when the priming stimulus is not consciously perceived.

Learning to split attention Action video games require players to notice target stimuli that can appear at any time while they are attending to other stimuli. Researchers have found that playing such games increases people's ability to attend to many stimuli at once.

FIGURE 9.5 | Tree with and without a duck Subjects who briefly saw the stimulus on the left did not consciously see the duck, but were subsequently more likely to draw a scene having to do with ducks than were subjects who saw the stimulus on the right. (From Eagle & others, 1966.)



Eagle, M., et al. (1966). Imagery: Effect of a concealed figure in a stimulus. *Science*, 151, 838. Copyright © 1966 by the AAAS

10 What is some evidence that concepts stored in long-term memory can be primed by stimuli that are not consciously perceived?

<<< One of the earliest demonstrations of unconscious priming was an experiment in which researchers showed students either of the two visual stimuli shown in Figure 9.5 (Eagle & others, 1966). The left-hand stimulus contains the outline of a duck, formed by the tree trunk and its branches. The researchers found that subjects who were shown the duck-containing stimulus, in three 1-second flashes on a screen, did not consciously notice the duck. Yet when all subjects were subsequently asked to draw a nature scene, those who had been shown the duck-containing stimulus were significantly more likely to draw a scene containing a duck or a duck-related object (such as a pond) than were the other subjects.

Another early example involves a selective-listening procedure (MacKay, 1973). Subjects wore headphones that presented different messages to each ear and were instructed to shadow (repeat) what they heard in one ear and ignore what was presented to the other ear. The shadowed message contained sentences, such as *They threw stones at the bank*, that contained words with two possible meanings. At the same time, the other ear was presented with a word that resolved the ambiguity (*river* or *money* in this example). After the shadowing task, the subjects were asked to choose from a pair of sentences the one that was most like the shadowed sentence. In the example just cited, the choice was between *They threw stones at the savings and loan association* and *They threw stones toward the side of the river*. Although the subjects could not report the non-shadowed word, they usually chose the sentence that was consistent with the meaning of that word. Thus, the unattended word influenced their interpretation of the shadowed message, even though they were unaware of having heard that word.

In everyday life, priming provides a means by which contextual information that we are not attending to can help us make sense of information that we are attending to. I might not consciously notice a slight frown on the face of a person I am listening to, yet that frown might prime my concept of sadness and cause me to experience more clearly the sadness in what he is saying. As you will discover later in the chapter, priming also helps us retrieve memories from our long-term store at times when those memories are most useful.

Automatic, Obligatory Processing of Stimuli

11 How is the concept of automatic, unconscious processing of stimuli used to help explain (a) people's ability to do more than one task at once, and (b) the Stroop interference effect?

<<< A wonderful adaptive characteristic of the mind is its capacity to perform routine tasks automatically, which frees its limited, effortful, conscious working memory for more creative purposes or for dealing with emergencies. Such automatization depends at least partly on the mind's ability to process relevant stimuli preattentively, unconsciously, and to use the results of that processing to guide behavior. When you were first learning to drive, for example, you probably had to devote most of your attention to such perceptual tasks as monitoring the car right ahead of you, watching for red traffic signals, and attending to sensory feedback from your steering wheel, brake, and accelerator. With time, however, these tasks became automatic, allowing you to devote ever more attention to other tasks, such as carrying on a conversation or looking for a particular street sign.

Another example of a skill at which most of us are experts is reading. When you look at a common printed word, you read it automatically, without any conscious effort. You unconsciously, preattentively, use the word's individual letters to make sense of the word. Because of that, when you read a book you can devote essentially all of your attention to the meaningful words and ideas in the book and essentially no attention to the task of recognizing the individual letters that make up the words. In fact, researchers have found that, in some conditions at least, reading is not only automatic but obligatory (incapable of being suppressed). In certain situations, people can't refrain from reading words that are in front of their open eyes, even if they try.

An often-cited demonstration of the obligatory nature of reading is the **Stroop interference effect**, named after J. Ridley Stroop (1935), who was the first to describe it. Stroop presented words or shapes printed in colored ink to subjects and asked them to name the ink color of each as rapidly as possible. In some cases each word was the name of the color in which it was printed (for example, the word *red* printed in red ink); in others it was the name of a different color (for example, the word *blue* printed in red ink); and in still others it was not a color name. Stroop found that subjects were slowest at naming the ink colors for words that named a color different from the ink color. To experience this effect yourself, follow the instructions in the caption of Figure 9.6.

	GREEN	RED	TRUCK
	RED	BLUE	TOP
	BLUE	GREEN	COUCH
	BLACK	BLACK	HAND
	BLUE	GREEN	COUCH
	RED	BLUE	TOP
	GREEN	BLACK	TRUCK
	BLACK	RED	TOP
	RED	BLUE	COUCH
	BLUE	GREEN	HAND
(a)	(b)	(c)	(d)

FIGURE 9.6 | The Stroop interference effect Time yourself (or a friend) on each of the following tasks: (1) Name the colors in each box in column (a). (2) Name the ink colors used to print the words in columns (b), (c), and (d), timing each column separately. Column (c) will take the longest, because the color words interfere with naming the ink colors—that is the Stroop interference effect. Column (d) may also take somewhat longer than either (a) or (b), because the noncolor words interfere somewhat with naming the ink colors. Column (b) should be quickest, because there the words facilitate naming the ink colors.

A variety of specific explanations of the Stroop effect have been offered (Luo, 1999; Melara & Algom, 2003), but all agree that it depends on people's inability to prevent themselves from reading the color words. Apparently, the preattentive processes involved in reading are so automatic that we cannot consciously stop them from occurring when we are looking at a word. We find it impossible *not* to read a color name that we are looking at, and that interferes with our ability to think of and say quickly the ink color name when the two are different. As you would expect, children who have not yet learned to read are not susceptible to the Stroop effect, and children who are just learning to read are only slightly susceptible to it (Gibson, 1971). In fact, the Stroop effect is sometimes used as an index of the extent to which reading has become automatic for a child.

Brain Mechanisms of Attention and Preattentive Processing

For many years psychologists were content to develop hypothetical models about mental compartments and processes, based on behavioral evidence, without much concern about how these are physically instantiated in the brain. In recent years,

with the advent of fMRI and other new methods for studying the intact brain, that has changed. Many studies have been conducted to see how the brain responds during preattentive processing of stimuli, during attentive processing of stimuli, and at moments when shifts in attention occur. Much is still to be learned, but so far three general conclusions have emerged from such research:

12 What three general conclusions have emerged from studies of brain mechanisms of preattentive processing and attention?

- <<<
1. *Stimuli that are not attended to nevertheless activate sensory and perceptual areas of the brain.* Specific stimuli activate specific sensory and perceptual areas of the cerebral cortex whether or not they are consciously noticed. Such activation is especially apparent in the primary sensory areas, but it also occurs, to some degree, in areas farther forward in the cortex that are involved in the analysis of stimuli for their meaning. For instance, in several fMRI studies, words that were flashed on a screen too quickly to be consciously seen activated neurons in portions of the occipital, parietal, and frontal cortex that are known, from other studies, to be involved in reading (Dehaene & others, 2001; 2004). Apparently, the unconscious, preattentive analysis of stimuli for meaning involves many of the same brain mechanisms that are involved in the analysis of consciously perceived stimuli.
 2. *Attention magnifies the activity that stimuli produce in sensory and perceptual areas of the brain.* Attention, at the neural level, seems to be a process that temporarily sensitizes the relevant neurons in sensory and perceptual areas of the brain, increasing their responsiveness to the stimuli that they are designed to analyze (Treue, 2001). Thus, spoken words activate certain areas of the left temporal lobe, which are involved in word comprehension, more if the person is consciously attending to those words than if the person is ignoring them.
 3. *Neural mechanisms in anterior (forward) portions of the cortex are responsible for shifts in attention.* Certain regions in the frontal lobe and in anterior portions of the temporal and parietal lobes become active at moments when shifts in attention occur. This and other evidence suggests that these anterior regions control shifts of attention and that they do so by acting, top-down, on sensory and perceptual cortical areas farther back (Hopfinger & others, 2000; Shipp, 2004).

Attention, whatever else it is, appears to be a state of the brain in which more neural resources are devoted to analyzing certain selected stimuli than to analyzing other stimuli that are picked up by the senses.

SECTION REVIEW

Attention is the means by which information enters consciousness.			
Focused Attention • Selective listening and viewing studies show that we can effectively focus attention, screening out irrelevant stimuli. • In general, we are aware of the physical qualities, but not of the meaning, of stimuli that we do not attend to.	Shifting Attention • We unconsciously monitor unattended stimuli, in sensory memory, so that we can shift our attention if something significant turns up. • Such monitoring includes preattentive processing for meaning. • Practice can enhance the capacity to attend to several items of information at once.	Preattentive Processing • Through preattentive processing, unattended sensory information can affect conscious thought and behavior. • For example, in unconscious priming, stimuli that are not consciously perceived can activate information in long-term memory, which can influence conscious thought. • Preattentive processing is automatic and in some cases obligatory, as exemplified by the Stroop interference effect.	Brain Mechanisms • Preattentive and conscious processing of stimuli for meaning involve many of the same brain areas. • However, attention causes greater activation of the relevant sensory and perceptual areas. • Areas of the cerebral cortex in front of the perceptual areas control shifts in attention.

Working Memory: The Active, Conscious Mind

As defined earlier in the chapter, working memory, or short-term memory, is the center of conscious perception and thought. This is the part of the mind that thinks, makes decisions, and controls such processes as attention and retrieval of information from long-term memory. Most cognitive psychologists today conceive of working memory as consisting of a number of separate but interacting components. According to one influential model, proposed by Alan Baddeley (1986, 2003), these include a **phonological loop**, responsible for holding verbal information; a **visuospatial sketchpad**, responsible for holding visual and spatial information; and a **central executive**, responsible for coordinating the mind's activities and for bringing new information into working memory from the sensory and long-term stores. We have already discussed one function of the central executive, attention, and will discuss some of its other functions later in this chapter and in Chapter 10. Here we'll focus on the phonological loop and the visuospatial sketchpad.

Verbal Working Memory: The Phonological Loop

As a test of one aspect of your own working memory, read the digits at the end of this sentence and then close your eyes and try to keep them in mind for a minute or so: 2 3 8 0 4 9 7. What did you do to keep them in mind? If you are like most people, you repeated the digit names over and over to yourself in the order you read them: *two three eight zero four nine seven*. Some IQ tests (discussed in Chapter 10) include a measure of *digit span*, the number of digits that the person can keep in mind for a brief period and report back accurately. Most people have a digit span of about seven digits. More generally, the number of pronounceable items—such as digits, other words, or nonsense syllables—that a person can keep in mind and report back accurately after a brief delay is called the **span of short-term memory**. According to Baddeley's model, it might better be called the *span of the phonological loop of working memory*. The phonological loop is the part of working memory that holds on to verbal information by subvocally repeating it.

Research has shown that the span of short-term memory, measured this way, depends on how rapidly the person can pronounce the items to be remembered (Baddeley, 1986). As a general rule, people can keep in working memory about as much verbal material as they can state aloud in 2 seconds (Baddeley & others, 1975). Unrehearsed items fade quickly; some of them begin to disappear within about 2 seconds or slightly longer. People who can speak rapidly have larger spans than people who cannot speak so rapidly. The span for single-syllable words is greater than that for multiple-syllable words. Try repeating from memory the following seven-word list, with eyes closed, immediately after reading it: *disentangle appropriation gossamer anti-intellectual preventative foreclosure documentation*. Was that list harder than the list of digits? Any intervention that interferes with a person's ability to articulate the words to be remembered interferes with verbal short-term memory (Baddeley, 2003). Try to hold seven digits in mind while repeating over and over, out loud, the word *the*. You probably can't do it; the act of saying *the* interferes with your ability to articulate to yourself the digit names.

Keeping a list of memory items in the phonological loop is a bit like a circus performer's keeping a set of plates spinning on the ends of sticks. As the number of plates increases, the performer must work more frantically to get back to each and renew its spinning before it falls. Performers who can move quickly can spin more plates than performers who move slowly. Larger plates take longer to set in motion than smaller ones, so the performer can't spin as many large plates as small ones. If the performer attempts to do at the same time another task that involves his or her arms and hands—such as building a tower of cups and saucers—the number of plates he or she can spin decreases. Of course, in everyday life we don't normally use our phonological loop to keep nonsensical lists in mind, any more than we use our hands to keep plates spinning. Rather, we use it for useful work. We say words

>>> **13**
What is some evidence that people keep information in the phonological loop through subvocal repetition?



Richard T. Nowitz / Corbis

silently to ourselves, and we bring ideas together in the form of words, as we reminisce about our experiences, solve problems, make plans, or in other ways engage in *verbal thought* (discussed in Chapter 10).

Visual Working Memory: The Visuospatial Sketchpad

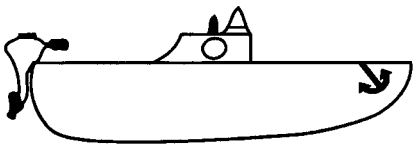
14

How do we use the visuospatial sketchpad to make judgments about spatial relationships?

<<< We think not only with words but also with mental pictures. We are especially likely to think in pictures when concerned with the spatial layout of things. I can never remember which way is west when I am facing south unless I picture myself standing on a map of North America looking toward South America. Then I can “see” immediately that the West Coast is on my right. Similarly (but at a slightly higher level of thought), Albert Einstein claimed that the concept of relativity came to him through visualizing what would happen if he chased a beam of light and caught up to it (Kosslyn & Koenig, 1992).

If you rotate the second letter of the alphabet 180 degrees clockwise, do you get a *p* or a *q*? Who had a bigger nose, your fifth-grade teacher or your sixth-grade teacher? Examine the boat in Figure 9.7 for a few seconds; then turn to page 000 and answer the question posed there in the margin near the top. If you are like most people, you probably answered each of these questions by creating and examining a visual image. According to Baddeley’s model, you held that image on the visuospatial sketchpad of your working memory, where you examined and possibly manipulated the image to answer the question.

FIGURE 9.7 | Sample stimulus used in a mental-imagery experiment Subjects saw line drawings such as this, and later, when answering questions based on their memory of a drawing, showed reaction times suggesting that they were scanning a mental image. (From Kosslyn, 1980.)



Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press

15

How did Kosslyn demonstrate a similarity between examining a visual image and looking?

<<< No cognitive psychologist would say that a visual image really is, physically, a picture in the mind; but examining such an image is in some ways like examining a picture. In a classic study demonstrating that, Stephen Kosslyn (1973) showed people drawings, including the one in Figure 9.7, and later asked them to visualize each drawing from memory. He asked them to focus, in their memory, on either the left or the right end of the drawing and then, as quickly as possible, to indicate (by pushing one of two buttons) whether or not a particular named component was present in the drawing. For example, referring to the boat drawing, he would say “motor” or “porthole” or “anchor” while the person was focusing at either the left or the right end of the boat in memory. He found that the farther away the named object was from the place of focus, the longer it took subjects to push the correct button. These time lags suggested that the subjects had to scan across the mental image to find the named component before they could respond, just as they would if they were looking at an actual picture. Kosslyn took the lags as evidence that mental images indeed are organized spatially, like pictures. Other studies have revealed that when people form a mental image of a picture that they had looked at previously, they physically move their eyes in much the same ways that they did when they were originally looking at the picture (Laeng & Teodorescu, 2002).

Testing the Limits of Working Memory: Dual-Task Performance

People are much better at doing two mental tasks at once if one task involves the phonological loop and the other involves the visuospatial sketchpad than they are if both tasks involve the same working-memory component. You can easily rotate a *b* in your mind, to see what it looks like upside down, while repeating a set of

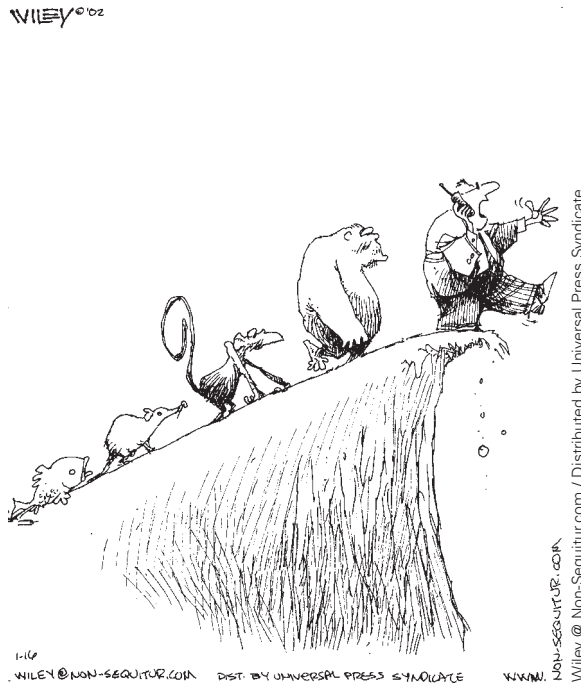
digit names; but you cannot easily do so while trying to visualize those same digits as mental pictures.

In a set of experiments, subjects were required to perform two challenging working-memory tasks at the same time (Cocchini & others, 2002). One task was to keep a series of digits in mind and repeat them later. The other was to keep a visual pattern of randomly arranged black-and-white checks in mind and reproduce it later. The result was that neither task interfered with the other. Subjects made about the same number of errors on each task when they performed the two simultaneously as they did when they performed each one separately. In contrast, the addition of a second verbal task to the digit-span task, or of a second visual task to the pattern-memory task, sharply increased the number of errors. From such findings, Baddeley (2003) has concluded that the phonological loop and the visuospatial sketchpad are two separate components of working memory rather than two aspects of the same component. As long as one task is verbal and the other is visual, we can operate, more or less, as if we have two minds working together in parallel.

In everyday life, however, few tasks are purely visual or purely verbal. Consider, for example, the dual tasks of conversing on a cell phone and driving a car. Driving is largely a visual task, but not entirely so. You need to read road signs (which is both visual and verbal), and you may need to think verbally about such issues as where you are supposed to turn and how to operate a particular piece of the car's equipment. Conversely, engaging in a phone conversation is largely a verbal task, but not entirely so. Any conversation is likely to call up visual images. For instance, if your conversational partner asks you how she can recognize a particular person, you are likely to form a visual image of that person in order to describe him. It is not surprising, therefore, that research indicates that phone conversations while driving contribute heavily to the rate of traffic accidents.

In a correlational study, involving only drivers who sometimes used their cell phones while driving, the accident rate during phone use was four times that for the same group when they did not use their phones (Redelmeier & Tibshirani, 1997). In a simulated-driving experiment in the laboratory, the requirement of conversing on a phone doubled the number of driving errors made (Strayer & Johnston, 2001). In both of these studies, the disruptive effect on driving was as great for hands-free phones as for hand-held phones. The interference is a mental one, involving competing uses of working memory, not a motor one, involving competing uses of the hands. Phone conversations appear to interfere with driving much more than do direct conversations with passengers, apparently because passengers, unlike phone partners, can see when the driving becomes difficult, so the conversation generally stops at those times (Strayer & Johnston, 2001).

How have dual-task experiments demonstrated that verbal and visual working memory can operate independently of each other? Why, even so, is it dangerous to talk on a cell phone while driving?



Brain Areas Involved in Working Memory

Brain research supports the idea that holding information in the phonological loop of working memory involves an internalized, mental process that is similar to actually speaking words and listening to spoken words. In neuroimaging studies, people who are asked to hold a list of words in working memory manifest extra neural activity in portions of the left hemisphere that are known to be involved in articulating words and in listening to words (Awh & others, 1996; Paulesu & others, 1993). Moreover, brain damage that interferes with either the ability to articulate words or the ability to understand words that are heard reduces a person's ability to hold words in working memory (Wilson, 2001).

Brain research also supports the idea that holding visual information in the visuospatial sketchpad involves mental processes akin to looking and seeing (Jonides

How has brain research supported Baddeley's model of the phonological loop and the visuospatial sketchpad?

& others, 2005; Wilson, 2001). As was discussed in Chapter 8, visual areas of the brain are divisible into two pathways. The “what” pathway is involved in recognizing patterns and objects. The “where-and-how” pathway is involved in perceiving spatial locations of objects and controlling a person’s movements in relation to those objects. People who have deficits in visual perception resulting from damage to either of these pathways have comparable deficits in visual working memory (Baddeley, 2003).

For example, after suffering damage in the “what” pathway of his cortex, one man lost much of his ability to see the shapes of objects even though he could still see the relative locations of objects in a scene (Farah, 1989). In memory tests for information that he would have learned before the accident, he performed poorly on such shape-dependent questions as “Does a bear have a long tail or a short tail?” and “What does a cocker spaniel’s ear look like?” Just as he was unable to see shapes in objects before his eyes, he was unable to “see” them in images recalled from memory. Yet he had no difficulty answering questions about the relative locations of objects. He could describe accurately, from memory, the layout of the furniture in his living room or the arrangement of buildings along a familiar street. Another man, with damage in the “where-and—how” pathway, showed just the opposite pattern of working-memory deficits: He could describe from memory the shapes of objects but not their locations (Farah, 1989).

18 What general roles are played by the prefrontal cortex in working memory?

<<< In addition to involving verbal or visual areas of the cortex, working-memory tasks also involve the *prefrontal cortex*, the anterior portion of each frontal lobe. In neuroimaging studies, increased activity in the prefrontal cortex occurs whenever a person is deliberately holding either verbal or visual information in mind (Baddeley, 2003). In one study, neural activity in the prefrontal cortex was significantly greater in trials in which the person successfully kept the information in mind than in unsuccessful trials (Sakai & others, 2002). The prefrontal cortex appears to be the neural hub of the central executive portion of working memory (Baddeley, 2003). It is the part of the brain that somehow organizes the efforts of the other portions of the brain and keeps them focused on the task. People with damage in the prefrontal cortex suffer from an inability to organize their behavior in meaningful, goal-directed ways, apparently because they cannot focus on a plan or call forth information in the right sequence to complete a task (Kimberg & others, 1997).

SECTION REVIEW

Working memory is the seat of conscious mental activity.			
The Phonological Loop • The phonological loop is a part of working memory that maintains verbal information through subvocal repetition. • Without active maintenance, the material is lost from working memory • This subcomponent of working memory permits verbal thought.	The Visuospatial Sketchpad • The visuospatial sketchpad is a part of working memory that permits us to hold and process mental images. • There are similarities between the way the mind deals with mental images and the way it deals with external visual stimuli. For example, spatial relationships such as distance are meaningful for both.	Dual Task Performance • We can perform two tasks simultaneously, without interference, if one involves only the phonological loop and the other only the visuospatial sketchpad. This shows that these two components of working memory function independently. • However, real-life tasks (e.g., driving or talking on a cell phone) usually involve both verbal and visual processing, so doing both at once leads to mental interference.	Brain Areas Involved • Neuroimaging shows that mental rehearsal of words involves activity in the same brain areas as speaking and listening to words. • The “what” and “where-and-how” pathways are involved in mental imagery in ways that correspond to their roles in visual perception. • The prefrontal cortex serves as the neural hub for the central executive subcomponent of working memory.

Encoding Information into Long-Term Memory

As you read a book, or attend to a conversation, or admire scenery, some of the sensory information that reaches your conscious mind enters your long-term-memory store, allowing you to recall it later. Why does some but not all of the information that reaches working memory get encoded into the long-term store?

As was discussed in the previous section, verbal information can be maintained in working memory simply by repeating it over and over. This is not, however, a good way to encode information into long-term memory. People who participate in a digit-span test, holding a list of digits in mind by reciting them over and over, rarely remember the digits even a minute after the test is over. Everyday life also has many examples of the failure of repetition to promote long-term memory. Long ago a psychologist named Edmund Sanford (1917/1982) illustrated this point by describing his own poor memory of four short prayers that he had read aloud thousands of times over a 25-year period as part of his daily religious practice. He found, when he tested himself, that he could recite, on average, no more than three or four successive words of each prayer from memory before having to look at the printed text for prompting. I myself have looked up a certain telephone number dozens of times and held it in working memory long enough to dial it, without ever encoding it into long-term memory. Yet I remember some other numbers that I have looked up only once or twice. The ones I remember are the ones I thought about in some way that helped me encode them into long-term memory.

Cognitive psychologists today distinguish between two kinds of rehearsal. **Maintenance rehearsal** is the process by which a person holds information in working memory for a period of time, and **encoding rehearsal** is the process by which a person encodes information into the long-term store. The activities that are effective for maintenance are not necessarily effective for encoding. Research suggests that some of the most effective activities for encoding involve elaboration, organization, and visualization.

Elaboration Promotes Encoding

Most of what we learn and remember in our everyday lives does not come from consciously trying to memorize. Rather, we remember things that capture our interest and stimulate our thought. The more deeply we think about something, the more likely we are to remember it later. To think deeply about an item is to do more than simply repeat it; it is to tie that item to a structure of information that already exists in long-term memory. Psychologists who study this process call it **elaboration**, or *elaborative rehearsal*. The immediate goal of elaboration is not to memorize but to understand, yet attempting to understand is perhaps the most effective of all ways to encode information into long-term memory.

Memory techniques centering on elaborative rehearsal capitalize on the human tendency to remember things that conform to some sort of logic, even if the logic is fictional. There is no intrinsic logic to the fact that stone formations hanging down in a cave are called *stalactites* while those pointing up are called *stalagmites*. But you can invent a logic: A stalactite has a *c* in it, so it grows from the ceiling; a stalagmite has a *g*, so it grows from the ground. Memory of a person's name can be improved by thinking of a logical relation between the name and some characteristic of the person. Thus, you might remember Mr. Longfellow's name by noting that he is tall and thin or, if he is actually short and stout, by recalling that he is definitely not a long fellow. I suspect that my students easily remember my name by relating it to the color of my hair.



Robert McElroy / Woodfin Camp & Associates

Thinking through the lines It is not exactly clear how performers commit long passages to memory, but, as suggested by the thoughtful expression on opera star Marilyn Horne's face, they apparently do much more than repeat the passages over and over.

19 What is some evidence, from the laboratory and from the classroom, that the more deeply a person thinks about an item of information, the more likely it is to be encoded into long-term memory?

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in long-term memory. *Journal of Experimental Psychology: General*, 104, 274

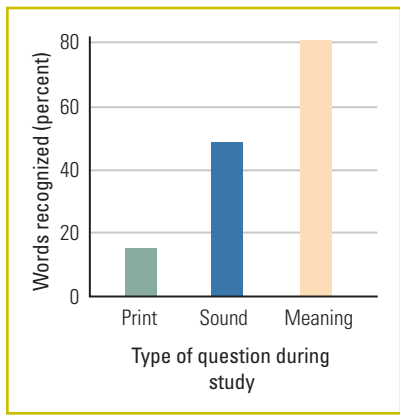


FIGURE 9.8 | The method of elaboration influences the likelihood of encoding into long-term memory. Subjects were shown a long sequence of words and were asked questions that required them to focus on the way the word was printed, how it sounded, or what it meant. The type of question dramatically affected the subjects' later ability to recognize the word as one that had appeared in the sequence. (Adapted from Craik & Tulving, 1975.)

Laboratory Evidence for the Value of Elaboration

<<< In a classic experiment demonstrating the value of elaboration, Fergus Craik and Endel Tulving (1975) showed subjects a long series of printed words, one at a time, and for each word asked a question that required a different form of thought about the word. In some cases, the question was simply about the printing of the word ("Is it in capital letters?"). In other cases, the question asked about the word's sound ("Does it rhyme with *train*?"). In still others, the question referred to the word's meaning ("Would it fit in the sentence, *The girl placed the _____ on the table?*"). As you can see by looking at Figure 9.8, subjects remembered many more words when they had been asked questions that focused on meaning than they did in the other conditions. Subsequent experiments showed that memory is even better when people are asked to relate each word's meaning to their own experiences or characteristics ("Does the word describe you?") (Rogers & others, 1977).

The Value of Elaboration for School Learning

In a study of fifth-graders, John Bransford and his colleagues (1982) found that students who received high marks in school were far more likely to use elaborative rehearsal than were those who received lower marks. The researchers gave the children written passages to study for a later test and asked them to explain what they did as they studied each passage. For example, one passage described two different kinds of boomerangs—a returning kind and a non-returning kind—each used for different purposes. Academically successful students often reported that they rehearsed the material by asking themselves questions about it. They might wonder what a non-returning boomerang looks like or why it is called a boomerang if it doesn't return, and this caused them to think deeply about what a boomerang really is and about the information in the passage. Less successful students, in contrast, usually studied the passages simply by rereading them.

Bransford's study was correlational in nature, so it does not prove that elaborative study caused better test performance; it shows only that the two tended to go together. But other research suggests that elaborative study can improve students' grades. In one college learning program, students are taught to write down questions about every textbook section that they read as they read it and about the lecture notes they take as they take them. The process of generating these questions and trying to answer them presumably deepens students' thought about the ideas and facts that they are reading about or hearing and thereby improves both understanding and memory. In a series of field experiments, students who were taught these techniques subsequently achieved higher grades in their college courses than did otherwise comparable students who received either subject-matter tutoring or no special help (Heiman, 1987).

Such findings are compatible with the following advice for studying this or any other textbook:

- Don't passively highlight or copy passages as you read, for later rereading. Focus on the ideas, not the author's exact words.
- Constantly ask yourself questions such as these: Do I understand the idea that the author is trying to convey here? Do I agree with it? Is it relevant to my own life experiences? Has the author given evidence supporting it? Does the evidence seem reasonable? How is this idea relevant to the larger issues of the chapter? (In this textbook, the numbered focus questions that appear in the margins give you a start on this task.)
- As you ask such questions, jot down notes in the margins that bear on your answers, such as "This idea seems similar to . . . ," or "I don't understand what he means by"

Through this active process, you will encode the material in a far richer and more lasting way than you could accomplish by simple rereading. You will also, in the process, generate questions that you might ask other students or your instructor.

Organization Promotes Encoding

As a memory strategy, organization is closely tied to elaboration. Organizing items to be remembered is itself a means of elaboration; you must think about the items, not just repeat them, in order to organize them. Moreover, organization can improve memory by revealing or creating links among items that would otherwise be perceived as separate.

Chunking

One way to increase memory efficiency is to group adjacent items that are at first perceived as separate, thus making them a single item. This procedure, known as **chunking**, decreases the number of items to be remembered and increases the amount of information in each item (Miller, 1956). As a simple illustration, if you had to memorize the series *M D P H D R S V P C E O I H O P* your task would be made easier if you saw it not as a string of 16 independent letters but as a set of five common abbreviations—M.D., Ph.D., RSVP, CEO, and IHOP. You could make your task still easier if you then chunked these five abbreviations into one sentence: “The M.D. and Ph.D. RSVPed to the CEO of IHOP” In developing such a story, you would not only be chunking but also elaborating—making the information more meaningful by adding some new information of your own.

Beginning music students find it easier to remember the notes on the lines of the treble clef as one sentence, “Every Good Boy Does Fine,” than as the senseless string of letters *E G B D F*. Similarly, physiology students can recall the seven physiological systems (skeletal, circulatory, respiratory, digestive, muscular, nervous, and reproductive) by matching their first letters to the consonants in SACRED MANOR. Both devices involve chunking. In the first example, the five notes are chunked into one meaningful sentence, and in the second the seven systems are chunked into two meaningful words. By reducing the number of separate items, and by attaching more meaning to each item, chunking provides an advantage both for maintaining information in working memory and for encoding it into long-term memory.

>>> **20**
How can chunking be used to increase the amount of information that can be maintained in short-term memory or encoded into long-term memory?

The Role of Chunking in Expert Memory

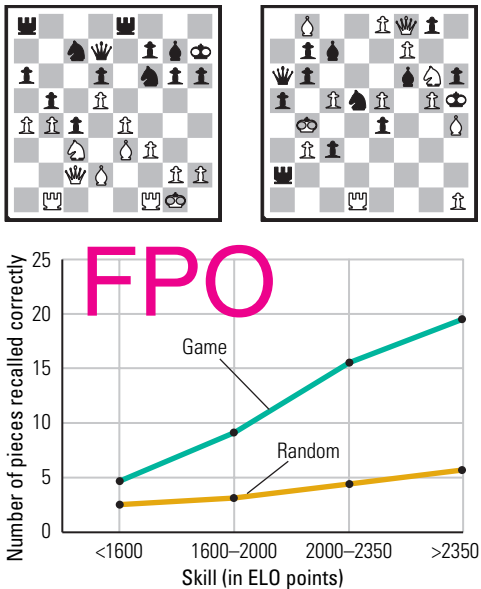
We are all much better at forming long-term memories for information that is within rather than outside our realm of expertise. For example, master chess players can look for just a few seconds at a chess game in progress and form a long-term memory of the locations of all the pieces on the board (de Groot, 1965). Similarly, football coaches have excellent memories for diagrams of football plays (Garland & Barry, 1991), architects have excellent memories for floor plans (Atkin, 1980), and physicians have excellent memories for information gained in diagnostic interviews of patients (Coughlin & Patel, 1987).

As a step toward explaining the expertise advantage in memory, K. Ander Ericsson and his colleagues have posited the existence of a special kind of long-term memory, called *long-term working memory* (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995). They conceive of this as memory for the interrelated set of items (such as a patient's case history or the pieces on a chess board) that is crucial for solving the problem or completing the task at hand. Such memories are encoded into long-term storage in a manner that makes the entire structure of information easily accessible to working memory, at least until the problem is solved or the task is finished. Such memories allow a physician to puzzle over a particular patient's symptoms as she drives home from work or a chess master to mull over the possibilities inherent in a particular set of chess positions while he is away from the game.

Chunking plays a major role in the formation of long-term working memories. In order to form such a memory of, say, a particular arrangement of pieces on a chess board, a person must already have in long-term storage a great deal of well-established information about possible and likely ways that such items might be

>>> **21**
How does chunking figure into experts' excellent memories for information that is within their realm of expertise?

FIGURE 9.9 | Chess memory The top drawings exemplify the types of chess layouts used in chess memory research. On the left is a board taken from a master's game and on the right is one with randomly arranged pieces. The graph shows the memory performance (averaged over 13 experiments) of chess players of various skill levels who studied either a game board or a random board for 10 seconds or less. The average number of pieces per board was 25. Skill level is measured in ELO points: 1600 is Class B, 2000 is Expert, and 2200 is Master. (Adapted from Gobet & others, 2001.)



arranged. This knowledge provides a foundation for the efficient chunking of new items of information. Chess games normally progress in certain logical ways, so logical relationships exist among the pieces, which experts can chunk together and remember as familiar formations rather than as separate pieces. If the chess pieces are arranged randomly rather than in ways that could occur in a real game, masters are no better, or little better, than novices at remembering their locations (Gobet & others, 2001; see Figure 9.9). Experts in other realms also lose their memory advantage when information is presented in a random manner rather than grouped in ways that make sense to them (Vicente & Wang, 1998; Weber & Brewer, 2003).

The Value of Hierarchical Organization

22 What is a hierarchical organization, and how can such an organization facilitate encoding into long-term memory?

◀◀ Perhaps the most useful format for organizing information is the hierarchy. In a hierarchy, related items are clustered together to form categories, related categories are clustered to form larger (higher-order) categories, and so on.

In an experiment demonstrating the advantage of hierarchical organization for long-term memory, Andrea Halpern (1986) gave subjects a chart listing 54 well-known song titles to be memorized. In some cases the chart was organized in a hierarchical manner, with songs arranged according to meaningful categories and subcategories. In other cases a similar chart was used but organized randomly, with no systematic relation among categories, subcategories, and song titles. When tested later for their memory of the song titles, subjects who had studied the organized chart recalled accurately many more titles than did those who had studied the disorganized chart. During the test the former group of subjects would first recall a category name and then the songs that had been listed under that name.

The information in this textbook (like that in nearly all textbooks) is hierarchically arranged: Each main heading refers to a set of related ideas, and each subheading refers to a smaller subset of ideas within the larger set. An efficient way to summarize the information in almost any textbook chapter is to sketch it out in a manner that preserves the author's hierarchical organization. As an illustration, a hierarchical sketch of the section you are now reading appears in Figure 9.10. Notice that the top node indicates in key words the theme of the section, the nodes subordinate to it indicate the main ideas pertaining to that theme, and the comments under them indicate the examples or evidence provided for each idea. You could summarize the whole chapter with six such sketches, one for each of the chapter's main sections. Such a summary would be a far more efficient aid in helping you commit the information to memory for a test than would a string of terms

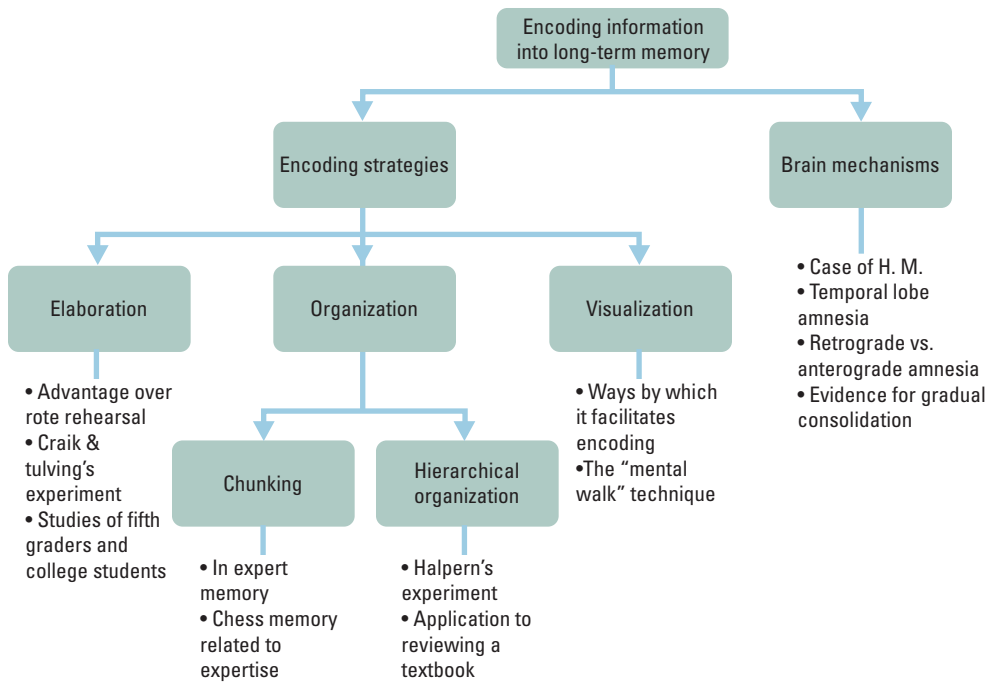


FIGURE 9.10 | Hierarchical summary of a textbook section
Summarized here are the ideas and evidence presented in this section (pp. 000–000) of the textbook. Such a chart is an aid to thought and memory because it reveals the logical connections among the items of information. Notice that most of the boxed items correspond to headings within the section.

and names that does not preserve the connections of ideas to each other or of ideas to evidence. The section reviews of this textbook also preserve the hierarchical organization of ideas and evidence within each section (as discussed in Chapter 1).

Visualization Promotes Encoding

Our discussion of encoding so far has centered mostly on memory for verbal information. But we can also encode pictures or visual scenes into long-term memory, apparently in a non-verbal form, which can be recalled later into the visuospatial sketchpad of working memory. Visual and verbal memories interact and supplement one another in our everyday experience. If you asked me to describe my living room, I would summon a pictorial memory of that room and then find the words to describe it.

A good deal of research suggests that people can improve their memory for verbally presented information if they encode that information visually as well as verbally. One study, for example, demonstrated that memory for information in news stories improves if relevant pictures accompany the stories (Prabu, 1998). Lacking pictures, people can improve their memory for verbally presented information by constructing visual images to represent that information (Paivio, 1986).

Visualization may improve memory through several different means. It may provide a distinct visual memory trace to supplement the verbal memory trace, thereby increasing the chance that the memory will be recalled in the future (Paivio, 1986). It may also provide an efficient way to chunk several verbally presented ideas together. For example, a verbal description of a person you haven't met may contain many separate items of information that you can combine into a single visual image. In addition, visual imagery may improve memory by linking newly learned information to information that is already well encoded in long-term memory. An example of that is found in a memory technique called the *mental walk*, used by nearly all contestants in the annual World Memory Championships when they are asked to memorize, in order, the names of a long list of objects (Maguire & others, 2003). As they hear the list of objects, they imagine themselves taking a walk along a familiar route and leaving each object next to a familiar landmark on the route. Then, during recall, they mentally walk the route again, inspect each landmark, and “see” each item that they had previously deposited there.

>>> **23**
Through what means might visualization help improve memory for verbally presented information?

Brain Mechanisms of Long-Term-Memory Encoding

Imagine what life would be like if, as a result of an operation, you became unable to form new explicit (conscious) long-term memories. At any given moment you would be fully aware of your environment and able to think about it, but you would have no idea how you arrived at that moment. You would live first in one moment, then in another, then in another, with no memory of any of the previous moments. If you made a plan for the future—even the future of just a few minutes ahead—you would forget the plan forever the instant you stopped thinking of it. Your life would be like that of a man known in the psychological literature as H. M., who indeed did lose his ability to form new explicit long-term memories.

The Case of H. M.: A Man Unable to Encode New Long-Term Memories

In 1953, at age 27, H. M. underwent surgery as treatment for severe epilepsy. A portion of the temporal lobe of the cortex and underlying parts of the limbic system on each side of his brain were removed. The surgery was effective against the epilepsy, but it left him unable to encode new explicit long-term memories. At the time of this writing, H. M. is still alive. Over the years he has participated in hundreds of memory experiments (Corkin, 2002).

24 How does the case of H. M. support the idea of a sharp distinction between working memory and long-term memory?

<<< H. M. can still remember events that occurred well before the operation. His long-term-memory store is full of knowledge acquired largely in the 1930s and 1940s. He can converse, read, solve problems, and keep new information in mind as long as his attention remains focused on it. But the minute his attention is distracted, he loses the information he had just been thinking about, and he cannot recall it later. This has been true for his entire life since the operation. To hold information in mind for a period of time, H. M. sometimes uses elaborate memory schemes. In one test, for example, he successfully kept the number 584 in mind for 15 minutes, and when asked how he did this, he replied: “It’s easy. You just remember 8. You see, 5, 8, 4 add to 17. You remember 8; subtract from 17 and it leaves 9. Divide 9 by half and you get 5 and 4, and there you are—584. Easy.” Yet a few minutes later, after his attention had shifted to something else, he could not remember the number or the memory scheme he had used, or even that he had been given a number to remember (Milner, 1970).

H. M.’s memory impairment has made it impossible for him to live independently. For many years now he has lived in a nursing home (Hilts, 1995). He must be accompanied wherever he goes and needs constant reminders of what he is doing. He is aware of his memory deficit and once described it in the following way (Milner, 1970): “Right now, I’m wondering, have I done or said anything amiss? You see, at this moment everything looks clear to me, but what happened just before? That’s what worries me. It’s like waking from a dream. I just don’t remember.”

Involvement of Temporal-Lobe Structures and Prefrontal Cortex in Encoding

25 What evidence indicates that the hippocampus and temporal-lobe structures near it are involved in encoding explicit long-term memories?

<<< There have been many other studies of people who have a memory loss like H. M.’s, though usually not as complete, after strokes or other sources of brain damage. The disorder is called **temporal-lobe amnesia**, and the areas of destruction most strongly correlated with it are the hippocampus (the limbic-system structure buried within the temporal lobe, depicted in Figure 9.11) and cortical and subcortical structures closely connected to the hippocampus in both halves of the brain (Squire, 1992).

Neuroimaging studies complement the evidence from brain-damage research. When people with intact brains are presented with new information to memorize, they manifest increased activity in the hippocampus and adjacent parts of the temporal lobe, and the degree of that increase correlates positively with the likelihood that they will recall the information successfully in a later test (Otten & others,

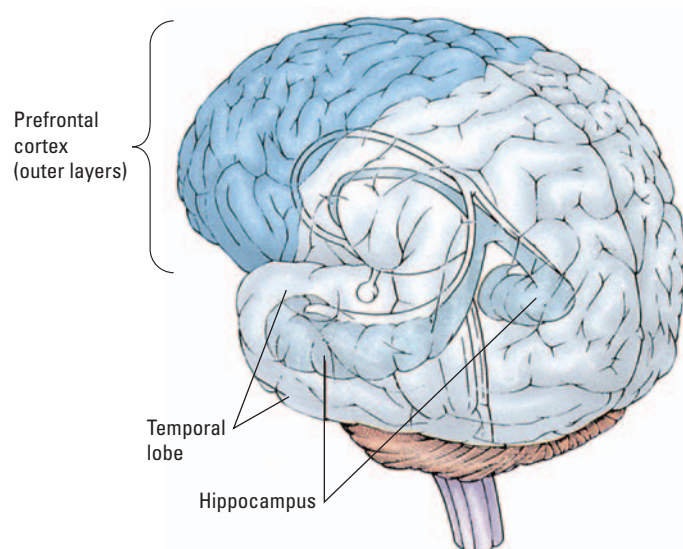


FIGURE 9.11 | Brain area involved in temporal-lobe amnesia The hippocampus is buried within the temporal lobe of the brain. Destruction of this structure in both halves of the brain produces a severe deficit in ability to encode new episodic memories. Destruction of this structure plus some of the surrounding areas of the temporal lobe produces a severe deficit in ability to encode new explicit memories of both types—semantic and episodic.

2001; Reber & others, 2002). Such studies also show that certain areas of the prefrontal cortex (also shown in Figure 9.11) become active during the encoding period. Activity in the prefrontal cortex appears to coincide with a conscious attempt to encode—that is, with the mental work involved in elaborative rehearsal—and activity in the temporal-lobe structures appears to coincide with success in encoding (Reber & others, 2002).

Retrograde Amnesia and Evidence for Gradual Consolidation of Long-Term Memories

The term *amnesia* refers to any loss of long-term memory, usually resulting from some sort of physical disruption or injury to the brain. The most dramatic type of amnesia observed in H. M. and other temporal-lobe patients is **anterograde** [an-tear-oh-grade] **amnesia**, the loss of capacity to form long-term memories of events that occur *after* the injury. However, these patients also show a considerable degree of **retrograde amnesia**, loss of memories of events that occurred *before* the injury.

Retrograde amnesia is generally time-graded; it is greatest for memories acquired just before the injury and least for those acquired long before (Wixted, 2004, 2005). H. M., for example, lost all his memories of events that occurred within several days of his surgery, some of his memories of events that occurred within a few years of his surgery, and essentially none of his memories of events that occurred in his childhood, many years before the surgery (Eichenbaum, 2001). Such time-graded retrograde amnesia is also seen in people who have suffered a severe blow to the head or have had a series of electroshock treatments—a procedure (discussed in Chapter 17) in which the brain is subjected to a jolt of electricity as a treatment for severe depression.

The time-graded nature of retrograde amnesia suggests that long-term memories are encoded in the brain in at least two forms—a labile, easily disrupted form and a solid, not easily disrupted form. Long-term memories appear to be encoded first in the labile form. Then, gradually over time, they apparently are either re-encoded in the solid form or are lost (forgotten). The process by which the labile memory form is converted into the solid form is referred to as **consolidation**.

A prominent theory today is that the labile form of long-term memory involves neural connections in the hippocampus and structures closely tied to the hippocampus, and the solid form does not involve these structures (Eichenbaum, 2001; Wittenberg & Tsien, 2002). This theory is supported by the time-graded retrograde amnesia observed after loss of the temporal-lobe structures, and it is also supported by neuroimaging research with people who have intact brains and normal

26 What evidence supports the theory that long-term memories first exist in a labile, temporal-lobe-dependent state and then, if not lost, are gradually consolidated into a more stable form that doesn't depend on temporal-lobe structures?



memories. When people recall memories that were acquired relatively recently, their temporal-lobe structures usually become active; but when they recall memories that were first acquired many years earlier, those structures usually do not become active (Haist & others, 2001).

Nobody knows just how memory consolidation occurs. It is also not clear why some memories appear to consolidate relatively quickly (within days) and others much more slowly (over years) or why many memories are forgotten within minutes, hours, or days and never become consolidated. There is evidence that memories that are recalled and used repeatedly, over relatively long time periods, are the ones most likely to be consolidated (Altmann & Gray, 2002). From a functional perspective, this makes sense. Much of the information that entered my long-term-memory store today—such as where I parked my car this morning and when I am expected home for dinner—is of use to me only today. Other information—such as when my publisher wants me to finish this chapter—is of use to me for a period of several days or weeks, but not a lifetime. There is no need to clutter up my permanent collection of memories with information such as that. It is better to reserve that collection for information that might always be useful.

SECTION REVIEW

Thinking about information increases the chances of encoding it into long-term

Elaboration	Organization	Visualization	Brain Mechanisms
- Elaboration involves actively extending one's understanding of something and meaningfully connecting it with information already in long-term memory. - Evidence that elaboration promotes encoding comes from both laboratory research and field research (e.g., of school learning).	- Mentally organizing information, which is a form of elaboration, greatly facilitates encoding. - Chunking is a form of organization in which separate items are grouped meaningfully into one higher-level unit (e.g., SACRED MANOR). - Hierarchies are highly effective formats for mental organization.	- Long-term memory for verbal information is strengthened if it is encoded visually as well as verbally. - Visualization may improve memory by creating an additional memory trace, by chunking separate items together, and by forming links to information already well-established in memory.	H. M. and people with temporal-lobe amnesia are unable to form new long-term memories. - Studies of such brain damage and neuroimaging studies suggest that temporal-lobe structures (e.g., the hippocampus) and the prefrontal cortex are critically involved in encoding. - The time-graded nature of retrograde amnesia suggest that long-term memories exist first in a labile form, becoming stable only through consolidation.

Retrieving Information from Long-Term Memory

Once we encode an item into long-term memory, we may or may not be able to find it there when we need it. Why are we sometimes unable to remember a name or fact that we “know” that we know? Why do we at other times remember a name or fact that we thought we had long since forgotten? Why, in some cases, do we remember an event differently each time we retell it, or remember an event that didn't really happen as if it did? These are some of the questions that cognitive psychologists have addressed in research and theories concerning the retrieval of information from long-term memory.

Mental Associations and Memory Retrieval Cues

The World Wide Web contains hundreds of millions of pages of textual information, sitting in hundreds of thousands of servers throughout the world. All that information would be useless to you and me were it not for search engines, such as Google, that can rapidly scan the pages and find what we are looking for. The long-term-memory store of your or my brain likewise contains vast quantities of information, which is useful only to the degree that we can call forth, at any given moment, the information that we need at that moment.

Retrieval of specific items from any information-storage system is dependent on how the stored information is organized. Books in a library are organized by topic, making it easier for us to find all there is about, say, the Civil War or flower gardening. Words in a dictionary are organized alphabetically, so we can find them on the basis of their spelling. Web pages on the Internet contain links to other pages that deal with related issues, and efficient search engines take advantage of those links. In the human mind, long-term memories are stored not in isolation, but in networks in which each item is linked to many others through connections referred to as **associations**. When any one memory is activated by an appropriate stimulus or thought, other memories associated with it become temporarily activated, or *primed*, to become more easily retrievable. A stimulus or thought that primes a particular memory is referred to as a **retrieval cue** for that memory.

The evidence for these ideas about memory organization and retrieval comes not from knowledge of how memories are stored physically in the brain, but from behavioral studies.

Mental Associations as Foundations for Retrieval

Speculation about mental associations goes back at least to the time of Aristotle. Aristotle considered two concepts to be associated if the thought of one tends to evoke (call forth from long-term memory) the thought of the other, and he proposed several principles of association, the most central of which are contiguity and similarity.

According to Aristotle's principle of **association by contiguity**, some concepts are associated because they have occurred together (contiguously) in the person's previous experience. Thus, *napkin* and *plate* might be associated in your mind because you have frequently seen napkins and plates together. When you see the face of someone you know, his or her name leaps to your mind (or at least often does), because you have often experienced that face and that name together in the past. The contiguity principle also accounts for our ability to bring quickly to mind the various properties of an object when we hear its name. If I hear *apple*, I can immediately think *red*, *round*, *sweet*, *tart*, *grows on trees*, *good in pies* because I have experienced all those properties of apples contiguously with apples themselves and with the word *apple*.

According to the principle of **association by similarity**, items that share one or more properties in common are linked in memory whether or not they were ever experienced together. Your thought *apple* might evoke the thought *rose*, because both are red, even if you have never seen an apple and a rose together.

In his great introductory psychology textbook, written more than a century ago, William James (1890/1950) pointed out that association by similarity can be understood as a derivative of the more primitive and fundamental principle of association by contiguity. Contiguity allows us to think of the properties of any given object and then allows us to think of other objects that have those same properties, leading to associations by similarity. Thus, my thought *apple* leads to my thought *red* (because of contiguity), and my thought *red* leads to my thought *rose* (again because of contiguity), with the result that my thought *apple* leads to *rose* (similarity). James suggested that the ability to separate mentally the various properties of objects and events from their concrete referents, and to use those properties to link objects and events that were never experienced contiguously, represents a basic

>>> 27
What do the principles of association by contiguity and association by similarity say about retrieval from long-term memory? According to James, how does the second principle depend on the first?

difference between the human mind and that of other animals. “Thoughts [in dogs] will not be found to call up their similars, but only their habitual successors,” he wrote. “Sunsets will not suggest heroes’ deaths, but supper-time. This is why man is the only metaphysical animal.”

Network Models of Memory Organization

28 What sorts of experimental results was Collins and Loftus’s spreading-activation model designed to explain? How does the model expand on the idea that mental associations provide a basis for memory and thought?

<<< Continuing the tradition begun by Aristotle, many cognitive psychologists today depict the mind’s storehouse of knowledge as a vast network of mental concepts linked by associations. Figure 9.12 illustrates such a model. Allan Collins and Elizabeth Loftus (1975) developed this specific diagram to explain the results of experiments concerned with people’s ability to recognize or recall specific words very quickly after exposure to other words. For example, a person can recognize the word *apple* more quickly if the previous word was *pear* or *red* than if it was *bus*. Collins and Loftus assumed that the degree to which one word speeds up the ability to recognize or recall another reflects the strength of the mental association between the two words or concepts.

In Figure 9.12, the strength of association between any two concepts is represented by the length of the path between them; the shorter the path, the stronger the association. Notice how the diagram incorporates the idea that common properties of objects often provide a basis for their link in memory. *Roses, cherries, apples, sunsets, and fire engines* are all linked to the concept *red*, and through that common tie they are all linked to one another. The model is called a *spreading-activation model* because it proposes that the activation of any one concept initiates a spread of activity to nearby concepts in the network, which primes those concepts, so they become temporarily more retrievable than they were before. The spreading activity declines with distance, so concepts that are closely linked with the active concept receive more priming than those that are more distantly linked. Such models are convenient ways to depict the results of many memory experiments, and they help us to visualize the patterns of associations that make up the mind.

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428. Copyright © 1975 by the American Psychological Association. Reprinted by permission.

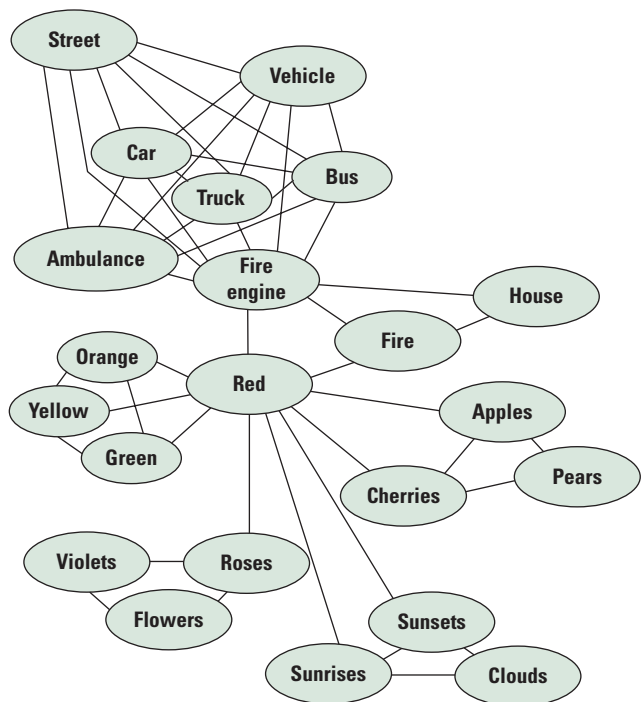


FIGURE 9.12 A network model of memory organization This diagram depicts schematically some of the links, or associations, among a tiny fraction of the thousands of different concepts that are stored in a typical person’s long-term memory. (From Collins & Loftus, 1975.)

29 How might elaborative encoding facilitate retrieval? How is this idea supported by an experiment in which subjects were tested for their memory of a long list of nouns they had been shown?

<<< Suppose you are learning for the first time that the capital of Vermont is Montpelier. You might notice, as you encode this new fact, that the syllable *mont* appears in both the capital name and the state name, and you might then think about the fact that *mont* is French for *mountain*, that Vermont is known as the Green Mountain state, and that many of the early settlers were French. More imaginatively, you might notice that *Montpelier* sounds like *mount peeler* and you might think of a strong wind peeling the snow off of one of Vermont’s mountains. Through such observations and thoughts you are setting up many possible retrieval cues for remembering, later, that the capital of Vermont is Montpelier. The last syllable of *Vermont*, or the thought of Vermont’s mountains, or of French settlers, or of wind, or of snow, or of anything that can be peeled, could prime the name *Montpelier* in your memory store.

In an experiment that demonstrated the value of retrieval cues generated by elaborative encoding, Timo Mäntylä (1986) presented 500 nouns one by one, in a

single very long session. He did not ask the subjects to memorize the nouns but asked them to write down either one or three words that they regarded as properties of the object named by each noun. For example, for the word *barn* a subject might write *large, wooden, red*. He then surprised the subjects with a test of their ability to recall all 500 nouns. As cues, he gave them either their own self-generated properties or those generated by a different subject in the same experiment. Subjects who received three self-generated properties for each word were able to recall correctly more than 90 percent of the 500 nouns. When only one property was available or when the properties had been generated by someone else, recall was much poorer (see Figure 9.13).

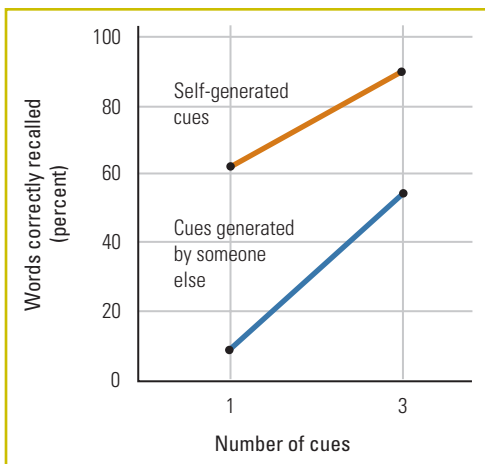


FIGURE 9.13 | Value of self-generated retrieval cues Subjects generated either one or three single-word properties related to each of 500 nouns. Later they were tested for their ability to recall each noun, using either their own self-generated properties or another subject's self-generated properties as retrieval cues. (Data from Mäntylä, 1986.)

Contextual Stimuli as Retrieval Cues

The environmental context that we find ourselves in at any give moment provides retrieval cues that prime our memories for our past experiences in that context. Such priming is generally adaptive, because our past experiences in a given context are often relevant to our future experiences in that context. Sitting behind the steering wheel of an automobile primes my memories relevant to driving, the sight and smell of a gymnasium prime my memories of basketball rules and strategies, and standing at a podium primes my memories of lecture techniques that were helpful in the past.

Many experiments have shown that people who are given facts or word lists to memorize in a particular context do better at recalling that information if tested in that same context than if tested in a different context (Smith & Vela, 2001). The context might be a particular room, or it might be a single very noticeable stimulus that seems to be incidental to the learning task. In one series of experiments, for example, people who studied a list of words in a room that smelled of chocolate or cinnamon/apple or mothballs performed better on a recall test if that same smell was present than they did if a different smell, or no smell, was present (Schab, 1990). Other experiments have shown that even instructions to imagine the context that was present during learning can facilitate recall (Smith & Vela, 2001).

>>> 30
How is the effect of context on memory adaptive? What is some evidence that retrieval is best when the retrieval context is similar to the encoding context?

Memory Construction as a Source of Distortion

Remembering is not just a process of retrieving traces that were laid down during the original encoding. The brain is not a tape recorder, video camera, or CD burner that records information at high fidelity; and remembering is not a matter of finding the right cassette or disk and playing it back. Instead, remembering is an active, inferential process guided by a person's general knowledge and intuitions about the world and by cues in the present environment. When you hear a story or experience an event, your mind encodes into long-term memory only some parts of the available information. Later, when you try to recount the story or event, you retrieve the encoded fragments and fill in the gaps through your logic and knowledge, which tell you what must have happened even if you can't quite remember it. With repeated retelling, it becomes harder to distinguish what was present in the original encoding from what was added later. Thus, memory of the story or experience is not a simple readout of the original information but a *construction* built and rebuilt from various sources. Our ability to construct the past is adaptive because it allows us to make logical and useful sense of our incompletely encoded experiences. But the process can also lead to distortions.

>>> 31
What does it mean to say that memories are constructed?

Effects of Preexisting Beliefs: Fitting Memories to Schemas and Scripts

One of the first psychologists to call attention to the role of people's general knowledge or beliefs in their more specific memories was the British psychologist Frederick Bartlett. Bartlett (1932) used the term ***schema*** to refer to one's generalized mental representation, or concept, of any given class of objects, scenes, or

events. He used the term especially in relation to concepts that may vary from culture to culture and that involve spatial or temporal relationships among the individual units of the object, scene, or event. For example, in our culture today people might share a relatively common schema of a living room, perhaps including a couch, an easy chair, and a rocking chair, all oriented around a television set, with a coffee table in front of the couch. When we enter a new living room, we recognize it as a living room and assess its unique features by comparing it with our already-existing schema. Schemas that involve the organization of events in time, rather than of objects in space, are commonly called ***scripts*** by today's cognitive psychologists (Schank & Abelson, 1977). The typical children's birthday party is a good example: There are games, followed by the presentation of the cake, the singing of "Happy Birthday," the blowing out of the candles, the eating of the cake, the opening of the presents, and then more games.



Esbin-Anderson / The Image Works

Following the script Years from now, this boy may construct a memory of his seventh-birthday party largely from his knowledge of the typical birthday-party script.

According to Bartlett and the results of many studies since his time, schemas do not just help us recognize and label the objects, scenes, and events that we encounter in daily life; they also affect the way we remember them later. We tend to remember any particular living room or birthday party as being more like the standard living room or birthday party than it really was. That is because we fill gaps in our memory of the particular with information from our schemas and scripts for the general.

32 How did Bartlett demonstrate that culture-specific schemas affect the way that people remember a story?

<<< In a classic demonstration of the effect of general knowledge on memory for the specific, Bartlett (1932) asked British university students to listen to a Native American story entitled "The War of the Ghosts" and later asked them to retell the story from memory. He found that the story often changed in the retelling, and he found certain consistencies in those changes. Details not essential to the plot tended to drop out, and those that were essential were often exaggerated. Also, points in the story that were consistent with Native American beliefs but not with the students' own beliefs were often changed to be more consistent with the latter. For example, the protagonist's obligation to certain spirits—a key component of the original story—tended to be transposed into an obligation to his parents. The changes were not deliberate; the students were trying to retell the story accurately, but they inevitably used their own ways of understanding things—their own schemas—to fill the gaps in their memory.

False Eyewitness Memories: Effects of Suggestion

Memory construction is affected not just by preexisting schemas but also by events that occur after the event being remembered was encoded. Such constructions assume more than academic importance in courtrooms or psychotherapists' offices, where they can have serious consequences.

33 What is some evidence that eyewitnesses' memories, even when very confidently expressed, are not always reliable? What is some evidence that suggestions made after the event can influence eyewitnesses' memories?

<<< Perhaps the most egregious of false memories are those that send innocent persons to prison or, sometimes, to death. In the 1990s, when DNA testing began to identify dozens of people who had been convicted of serious crimes they did not commit, the U.S. Attorney General's office ordered a study to discover the causes of those injustices. The study concluded that the great majority of those convictions came about because of highly confident eyewitness identifications (Wells &

others, 2002). In many cases the eyewitness was not confident early in the investigative process, but became so over time. An example is the case of Larry Mayes, who spent 21 years in prison for a rape he did not commit (Loftus, 2004). The victim failed to pick Mayes in two initial police lineups, but then became confident that he was the rapist after police helped her “recover” her memory of him through hypnosis. By the time of trial, the face of Larry Mayes was clearly embedded in her mind and she was certain he was the rapist.

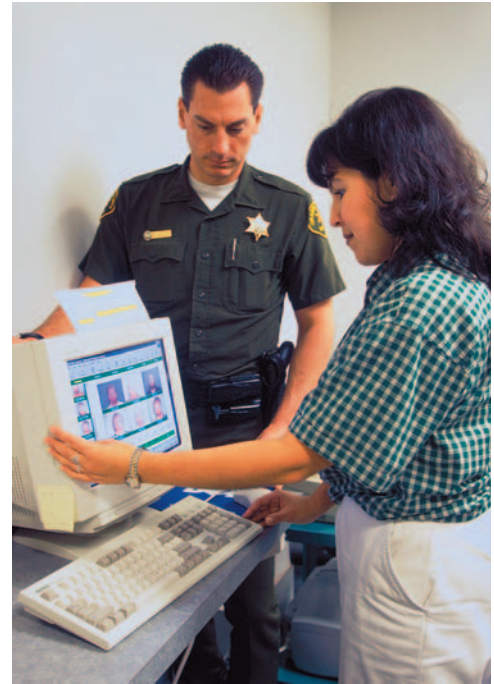
Hypnosis is a state of high suggestibility; psychologists have shown repeatedly that false memories can be created relatively easily through suggestions or encouragement made in that state (Lynn & others, 1997; Steblay & Bothwell, 1994). People who hold steadfastly to truly bizarre memories—such as having been abducted by aliens from another planet—often first recalled or became confident about those memories when questioned under hypnosis by investigators who believe in such phenomena (Newman & Baumeister, 1996). Many experiments—some involving simulated eyewitness situations—have shown that memories can also be altered or created through suggestions and encouragement without hypnosis. In one study, simply saying, “Good job, you are a good witness,” dramatically increased witnesses’ confidence in their memories of who committed the simulated crime, whether or not the memories were true (Wells & Bradfield, 1999). Other studies have shown that leading questions can alter people’s memories not just of who was involved in an incident, but also of what happened.

In a classic experiment, Elizabeth Loftus and J. C. Palmer (1974) had adults view a film depicting a traffic accident. Later, the researchers asked some of the subjects how fast the cars were going when they *hit* each other, and they asked others how fast the cars were going when they *smashed into* each other. The question with the word *smashed* elicited estimates of faster speed than did the question with the word *hit*. Moreover, when the subjects returned a week later and were asked to remember the film and say whether there was any broken glass in the accident, those who had heard the word *smashed* were more likely to say they saw broken glass (though actually there was none in the film) than were those who had heard the word *hit*. In subsequent studies, Loftus and others showed that such memory distortion is especially likely if the misinformation is introduced in just the manner that a clever but biased detective or cross-examiner is likely to introduce it—repeatedly but subtly, in the context of recounting aspects of the event that really did occur (Loftus, 1992; Zaragoza & Mitchell, 1996).

False Memories of Childhood Experiences: Effects of Suggestion and Imagination

If an adult in psychotherapy begins to recall instances of having been abused in childhood, does the memory necessarily reflect the truth, or might it have been constructed from ideas implanted by a well-intentioned but misguided therapist? In the 1980s, a number of high-profile court cases centered on this question. Some people were suing their parents for past, horrible abuses discovered in therapy, and some parents were suing therapists for implanting false memories in their offspring. The cases led to a spate of psychological research showing that techniques of suggestion, encouragement, and imagination indeed can create false memories of childhood experiences.

In one such study, Elizabeth Loftus and Jacqueline Pickrell (1995) led adults to believe that at age 5 they had been lost in a certain shopping mall and had been helped and comforted there by an elderly woman—an experience that in fact had never happened, according to the subjects’ parents and other close relatives who served as informants. Yet 25 percent of the subjects maintained—both at initial questioning and in follow-up interviews 1 and 2 weeks later—that they could remember the event. Some even elaborated upon it with details beyond those supplied by the researchers.



Spencer Grant / Photo Edit

Identifying the villain, with help from the police This woman’s judgment as to whether one or another of these men committed the crime that she witnessed may be affected by facial expressions or words of encouragement from the officer standing before her.

>>> **34** How have false memories for childhood experiences been implanted in experiments, and what evidence indicates that imagination can facilitate false-memory construction?

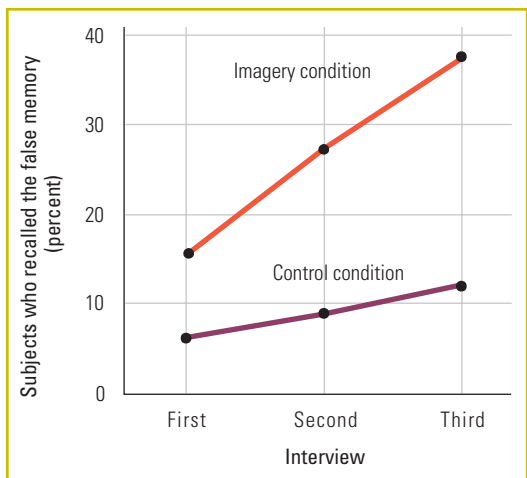


FIGURE 9.14 | Effect of imagery on development of a false memory
Adult subjects were told (falsely) about an incident that happened in their childhood. They were asked to try to remember it either by thinking about it (control condition) or by vividly imagining what might have happened (imagery condition). This procedure was repeated the next day, and the subjects' memories were again assessed in a third session, 2 days after the first. The graph shows the percentage of subjects in each condition, at each interview, who claimed to remember the incident. (Based on data in Hyman & Pentland, 1996.)

Subsequent studies showed that false-memory construction can be abetted by imagination. In one experiment, researchers told each subject (falsely) that according to the subject's parents, a particular embarrassing incident had occurred in his or her childhood (Hyman & Pentland, 1996). The made-up incident was one in which the subject, at age 5, had been running around at a wedding reception and had knocked over the punch bowl, spilling punch on the bride's parents. Subjects in the *imagery condition* were asked, in two successive sessions, to form vivid mental images of this event to help them remember it, and others, in the *control condition*, were asked just to think about the event as a way of remembering it. The result was that mental imagery sharply increased reported memory (see Figure 9.14). In an interview conducted after the second imagery session, 38 percent of the subjects in the imagery condition, compared to only 12 percent in the control condition, claimed that they could remember the punch-spilling incident.

Other experiments have shown that imagery alone, even without misleading suggestions from the researcher, can create false memories. In one, adults were simply asked to imagine a certain painful medical procedure—a procedure that in fact is never performed—and then, later, were asked to try to remember whether or not that procedure had ever been done to them in their childhood. The result was that over 20 percent of those in the imagination condition, compared to about 5 percent in the control condition, said that they could remember enduring that procedure (Mazoni & Memon, 2003).

Such research should not lead us to conclude that all our childhood memories are false constructions. The research does, however, suggest strongly that childhood memories are even more subject to distortion by suggestion and imagination than are memories acquired later in life.

Source Confusion and Social Pressure as Causes of False-Memory Construction

35

How might source confusion and social pressure contribute to false-memory construction?

<<< According to some theorists, a basic cause of most if not all cases of false-memory construction is *source confusion* (Mitchell & Johnson, 2000). We acquire information from various sources—including firsthand experiences, stories we have heard, and scenes we have imagined—and our minds reorganize the information in ways that may be meaningful but that obscure the ties between each item and its original source. Events that are conceptually linked but came from different sources may become confounded in memory. The memory of the actual traffic accident witnessed becomes confounded with the memory of what the cross-examiner implied about the accident. The memory of an actual event in childhood becomes confounded with a story heard from others or with an imagined scene constructed by free association in a psychotherapy session.

Social pressure also, no doubt, figures into many cases of false memory (Loftus, 1997). In most of the experiments just described, subjects were in one way or another led to believe that they should remember the suggested events. Similar pressure occurs in crime investigations, where the investigator acts as if any normal person should remember the incident's details. And it sometimes occurs in psychotherapists' offices, where a misguided therapist conveys the message that certain kinds of incidents, such as childhood abuse, must have happened and that the patient's task is to remember those incidents (McNally, 2003; Pendergrast, 1995). A person who feels pressured, even gently so, to come up with a memory is more likely than an unpressured person to identify a vague, possible memory as an actual memory; and the more often the memory is repeated, and the more praise the person receives for recalling it, the more confident the person becomes that the memory is true. The opposite is probably true also; a person who feels pressured by family members or an abuser to believe that certain memories are figments if his or her imagination may come to doubt memories that are accurate.

SECTION REVIEW

To use information held in long-term memory, we must be able to retrieve it.

Mental Associations and Retrieval

- Mental associations are links among items of information in long-term memory, which underlie our capacity for retrieval.
- The principles of contiguity and similarity underlie the formation of mental associations. The second of these two principles is a derivative of the first.
- Network models describe long-term memory as a vast web of associations that vary in strength. Activation can spread along associative pathways in a way that enables retrieval and thought.
- Elaborative rehearsal affects the number and meaningfulness of associations to new information and thereby its retrievability.
- The environmental context of learning can provide useful retrieval cues resulting from associations created at the time of encoding.

Memory Construction and Distortion

- Memories are not passive, complete records like photographs or sound recordings; instead, they are constructed and reconstructed.
- Schemas and scripts, which represent our general knowledge and beliefs, can affect memory construction at both encoding and retrieval, assisting memory but also sometimes distorting it.
- Leading questions, suggestions (especially those made under hypnosis), imagination, source confusion, and social pressures can distort memories or create false ones by influencing the constructive process of memory.

Multiple Memory Systems: Beyond the Modal Model

Most of the research and ideas about memory described thus far have fit at least moderately well into the framework of the modal model. The model has helped differentiate active (working) memory from passive (long-term) memory and has provided a basis for defining the processes of attention, encoding, and retrieval. But the model does not account for the entire range of phenomena that can reasonably be classified as memory.

Look again at the diagram of the modal model in Figure 9.1 (page 000). Its central feature is the box labeled “Working memory,” which is equated with conscious perception and thought. According to the model, new information can be encoded into long-term memory only if it is first perceived consciously in working memory, and information in long-term memory can influence behavior only if it is called back into working memory. Thus the model does not account for memories that can affect behavior without becoming conscious.

>>> 36
What types of memories are not accounted for by the modal model?

Distinctions Among Explicit- and Implicit-Memory Systems

Explicit memory is the type of memory that can be brought into a person's consciousness. It provides the content of conscious thought, and it is highly flexible. Explicit memories can be called to mind even in settings quite different from those in which they were acquired, and they can be combined with other explicit memories for purposes of reflection, problem solving, and planning. Such memory is called *explicit* because it is assessed through explicit tests—tests in which the person is asked to report directly (explicitly) what he or she remembers about a particular entity or event. It is also called *declarative memory* because the remembered information can be declared (stated in words).

Implicit memory, in contrast, is the type of memory that does not enter into the contents of consciousness. It consists of all the unconscious means through which previous experiences can influence a person's actions and thoughts. Such memory is called *implicit* because it is assessed through implicit tests—tests in which the memory is not reported directly but is inferred from behavioral responses. I would

>>> 37
What are the differences between explicit and implicit memory, and how is each memory type assessed? In what sense are implicit memories more context-dependent than explicit memories?



Thompson McClellan / The Picture Cube

Procedural memory The learned skill of balancing on a bicycle is retained as implicit procedural memory. You can't say just how you do it, but you don't forget it.

38 How do the two subclasses of explicit memory differ from one another?

test your memory for balancing on a bicycle not by asking you *how* to do it but by asking you *to* do it. Your good performance on the bicycle would *imply* to me that you know how to balance on it. Because people do not report in words the relevant information, implicit memory is also called *nondeclarative memory*. Implicit memories are much more closely tied to the contexts in which they were acquired than are explicit memories. Whereas explicit memories can be called forth voluntarily outside of their original context, implicit memories exert their effects automatically in the context of the specific stimuli, tasks, or problems to which they pertain.

Over time, with routine tasks, explicit memories are in some cases replaced by implicit ones. When I began commuting between my current home and job, I developed a set of explicit memories of street names and landmarks to guide me along the way. I could call up those memories, at any time, to think about the route or to tell another person how to make that trip. Now that I have made the trip hundreds of times, however, my commute is guided by implicit memories, which lead me automatically to turn my bicycle at the correct places, and I have forgotten many of the explicit memories. The description of my route that I could give you today is much less clear than the description I could have given you on my second day of commuting. (If you want a clear explanation of how to do something, ask a novice; if you want to *see* how to do it, watch an expert!)

In addition to distinguishing between explicit and implicit memories, cognitive psychologists also distinguish among subclasses of each type, along the lines depicted in Figure 9.15 (Squire & others, 1993; Tulving, 1985).

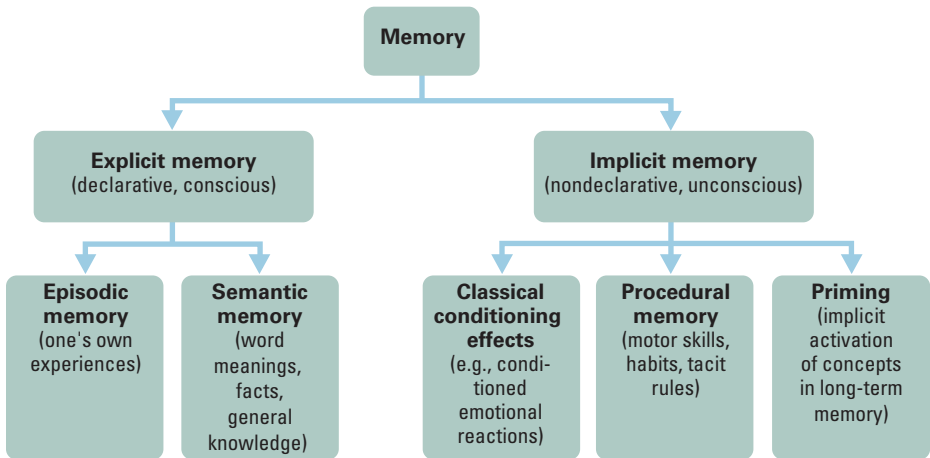
Varieties of Explicit Memory: Episodic and Semantic

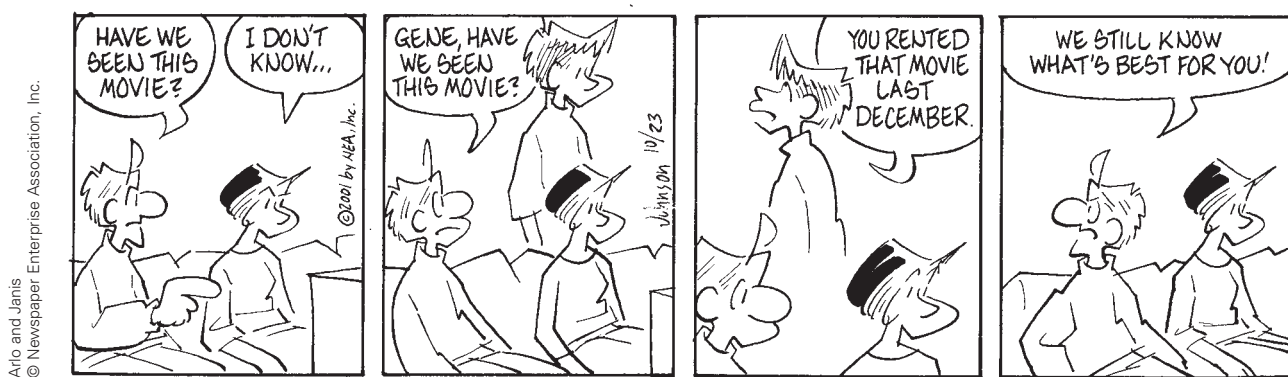
As shown in Figure 9.15, explicit memory is divisible into two subclasses. **Episodic memory** is explicit memory of one's own past experiences. Your memory of what you did and how you felt on your sixteenth birthday or of what you ate for dinner last night or of any other specific episode in your life is episodic memory. Episodic memories always have a personal quality. An integral part of your episodic memory of an event is your memory of yourself experiencing that event—as participant, witness, or learner. The experiments on constructed memories in the previous section all had to do with episodic memories.

Semantic memory, by contrast, is explicit memory that is not tied mentally to a particular past experience. It includes knowledge of word meanings (which is one definition of *semantics*) plus the myriad facts, ideas, and schemas that constitute one's general understanding of the world. It also includes knowledge of oneself that is not mentally tied to the re-experiencing of a particular episode in one's life. Your memories that apples are red, that penguins are birds, that you were born on such-and-such a date, and that psychology is the most fascinating of all academic subjects are examples of semantic memory. Of course, all such information had to have been acquired through past experiences in your life, but your memory of the information does not depend on remembering those experiences.

To remember your birth date or that penguins are birds you do not have to remember anything about the circumstances in which you learned that fact. Recalling an item from semantic memory is a bit like looking for information in an encyclopedia. To find an item that doesn't come easily to mind, you probe your mind with terms or concepts that have meaningful associations to that item. The network model of knowledge organization that you saw in Figure 9.12 has to do with the organization of semantic memory.

FIGURE 9.15 Types of long-term memory Explicit- and implicit-memory systems follow different rules and involve different neural systems in the brain. The explicit and implicit categories can be further subdivided into narrower memory categories that may also involve different neural systems. (Based on Tulving, 1985, and Squire & others, 1993.)





Here is a test question: What is classical conditioning? After answering the question, think about *how* you answered it. Did you think back to your experience of reading about classical conditioning earlier in this book, or to your experience of hearing your professor define classical conditioning, and try to reconstruct the definition from your memory of that experience? Or did you just *know* the answer, without having to reflect on any particular past experience? If the first is the case, then your memory of the definition of classical conditioning is an episodic memory (or at least partly so); if the second is the case, it is a semantic memory. In general, episodic memories are more fleeting, less stable, than semantic memories. Over time, we forget most of our memories of our specific past experiences, but the general knowledge that we extract from those experiences often stays with us.

Varieties of Implicit Memory

Implicit memory can also be divided into subclasses, and one way of doing so is depicted in the right-hand portion of Figure 9.15. One subclass consists of the memories resulting from classical conditioning—the internal changes that lead a person or animal to respond to conditioned stimuli (discussed in Chapter 4). A second subclass is a broad one referred to as **procedural memory**, which includes motor skills, habits, and unconsciously learned (tacit) rules. With practice you improve at a skill such as riding a bicycle, hammering nails, or weaving a rug. The improvement is retained (remembered) from one practice session to the next, even though you are unaware of the changes in muscle movements that make the difference. (An operant-conditioning account of such improvement is discussed in Chapter 4.) You can even learn to make decisions based on complex rules without ever becoming aware of the rules (Greenwald, 1992), and that phenomenon, too, exemplifies procedural memory.

Some experiments demonstrating rule-based procedural memories use what are called *artificial grammars* (Frensch & R nger, 2003; Reber, 1989). The grammars consist of sets of rules specifying which letters may or may not follow certain other letters in strings that are several letters long. For example, one rule might be that an *X* at the beginning of a string must be followed either by another *X* or a *V*, and another rule might be that a *J* anywhere in the middle of a string must be followed by a *B*, *K*, or *T*. Subjects are not told the rules. Instead, they are shown examples of grammatical and nongrammatical strings, labeled as such, and then are asked to categorize new examples as grammatical or not, on the basis of their “gut feelings.” The subjects typically do not learn any of the rules explicitly—they cannot state the rules—yet they learn to make correct categorizations at a rate significantly better than chance. The memories that guide their correct choices are implicit.

A third variety of implicit memory is priming (Tulving, 2000). *Priming* was defined earlier in this chapter as the activation, by sensory input, of information that is already stored in long-term memory. This activation is not experienced consciously, yet it influences subsequent conscious perception and thought and thus provides a link between implicit and explicit memory. Priming helps keep our

>>> 39
What are some examples of procedural memory, and why are such memories classed as implicit?

>>> 40
Why is priming considered to be implicit memory? What function does priming play in a person's everyday thought?

stream of thought running along consistent, logical lines. When we see or think about an object, event, or idea, those elements of our semantic memory that are relevant to that perception or thought become activated (primed) for a period of time, so they are more easily retrievable into conscious, working memory. Priming is classed as implicit memory because it occurs independently of the person's conscious memory for the priming stimulus. As noted in the discussion of attention, priming can even occur when the priming stimulus is presented in such a way that it is never consciously perceived.

Neuropsychological Evidence for Separate Memory Systems

Further evidence for multiple, distinct memory systems comes from studies of people who have impaired memory resulting from brain damage.

Implicit Memory Remains Intact in Temporal-Lobe Amnesia

41

What is the evidence that the hippocampus and nearby temporal-lobe structures are not essential for forming or using implicit memories?

<<< Earlier in the chapter you read of the severe memory deficits of H. M. and other patients with temporal-lobe amnesia. Those deficits had to do entirely with explicit memory. Such patients behave normally on all sorts of implicit-memory tests. If classically conditioned to blink their eyes in response to a conditioned stimulus, they show the conditioned response as strongly in subsequent tests as do non-amnesic subjects (Daum & others, 1989). If given practice with a new motor skill, such as tracing a pattern that can be seen only in its mirror image, they show normal improvement from session to session and retain the effects of previous learning even if months elapse between one session and the next (see Figure 9.16; Gabrieli & others, 1993; Milner, 1965). Similarly, they can learn and retain artificial grammars and tacit rules for grouping objects into categories (Knowlton & Squire, 1993; Knowlton & others, 1992), and they show as much activation of long-term semantic memories in response to priming stimuli as do normal subjects (Gabrieli, 1998; Levy & others, 2004).

In all these examples the implicit memory is manifested even when the amnesic subjects cannot consciously remember anything at all about the learning experience. In one experiment, a severely amnesic patient learned to program a computer over a series of sessions. At each session his programming ability was better than that in the previous session, even though he had no explicit memory of ever having programmed a computer before (Glisky & others, 1986).

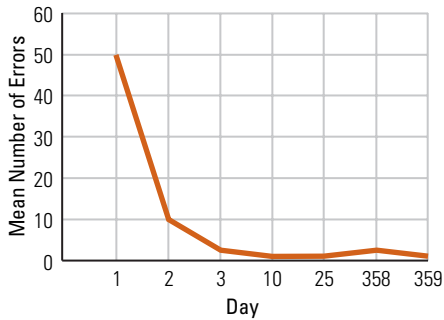
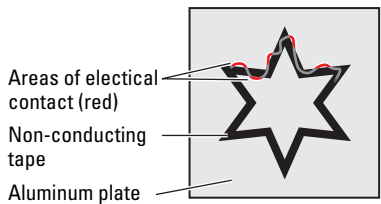


FIGURE 9.16 Implicit memory without explicit memory As shown in the graph, the temporal-lobe-amnesic patient H. M. improved from session to session in a mirror-tracing task, even though at each session he could not remember having performed the task before. The task was to trace a star under conditions in which the star and hand could be seen only in a mirror, so that movements had to be made oppositely from the way in which they appeared. An error was counted whenever the stylus moved off the star's outline. The data points on the graph represent the average number of errors per trial for the seven trials that occurred in each session. Sessions occurred on three successive days and then after delays of 1 week, 15 days, and nearly a year. (Graph adapted from Gabrieli & others, 1993.)

(Corkin, 2002). He cannot name new world leaders or entertainers, and if asked to draw a car or a radio from memory, he persists in drawing a 1940s or early 1950s version (Milner, 1984). But patients with less severe forms of amnesia typically manifest a greater deficit in episodic memory than in semantic memory (Tulving, 2002).

The most extreme differentiation between episodic and semantic memory is found in patients who suffer from a rare disorder called *developmental amnesia*. These people have bilateral damage to the hippocampus, but not to structures surrounding it, caused by temporary loss of blood flow to the brain at the time of birth or in early childhood. The hippocampus is more susceptible to permanent damage caused by lack of oxygen than is the rest of the brain. Faraneh Vargha-Khadem and her colleagues (1997, 2002) have identified and studied several young people who suffer from this disorder. All these individuals have severe deficits of episodic memory. If asked what happened a few hours ago or yesterday or at their last birthday party, they can recount little or nothing. Yet, despite this, they developed speech, reading, vocabulary, and other verbal capacities within the normal range. They all attended mainstream schools and learned and remembered facts well enough to perform passably on school tests. When they were presented with new factual information in controlled studies, they later remember the information but did not remember where or when they learned it. Their abilities are consistent with other evidence that the hippocampus is essential for episodic-memory encoding but not for semantic-memory encoding (Eichenbaum, 2003). Amnesiacs such as H. M., who lose semantic as well as episodic memory, have damage not just to the hippocampus but also to other portions of the temporal lobes.

Other Evidence of Semantic Memory Without Episodic Memory

It may at first seem surprising that people can remember new information without remembering the experience of learning that information. Yet, with a little poking around in your own semantic store, you will find many facts that you yourself know but can't relate to any episodes in your life (though it probably took you longer to forget the episodes than amnesic patients would take). I know that kumquats are a kind of fruit, but I can't recall any instance in my life of ever seeing, reading about, or hearing of kumquats. Older people are especially familiar with the phenomenon of knowing without knowing how they know. In old age, the capacity to form new episodic memories generally declines more rapidly than does the capacity to form new semantic memories (Johnson & others, 1993). Young children also show excellent semantic memory and poor episodic memory. During their first 4 years of life, children acquire an enormous amount of semantic information—including word meanings and facts about their world—that will stay with them throughout their lives. But children under 4 are relatively poor at recalling specific episodes in their lives, and none of us in adulthood can recall much about our own childhood prior to about age 4 (Eacott, 1999). Apparently the human ability for episodic-memory encoding develops more slowly and unravels more quickly than that for semantic-memory encoding.

The relatively poor episodic memory at both ends of the life span may be related to prefrontal cortical functioning (Li & others, 2005; Wheeler & others, 1997). The prefrontal cortex develops more slowly in childhood and tends to suffer more damage in old age than does the rest of the brain. People with prefrontal cortical damage typically experience a much greater loss in episodic-memory encoding than in semantic-memory encoding (Wheeler, 2000). This brain area, which is much larger in humans than in other species and is crucial for planning and complex thought, may be essential for our sense of ourselves, including our sense of our own past experiences. We are not only a conscious species but also a *self-conscious* species. We—unlike any other animal, or at least much more so than any other animal—reminisce about our past, think about our position in the present, and project ourselves into the future as we make plans and contemplate their consequences. Such abilities are intimately tied to our capacity to form episodic memories. This evolutionarily recent

>>> **42**
What is some evidence that semantic memory can occur in the absence of episodic memory and that the hippocampus is more crucial for the latter than the former?

>>> **43**
How might the relative lack of episodic memory in early childhood and old age be explained? How does episodic memory seem to distinguish us from other species?

addition to the mammalian cognitive machinery is apparently more fragile—more destructible by aging and injuries—than is the more ancient semantic-memory system or the still more ancient implicit-memory system (Tulving, 2002; Wheeler & others, 1997).

SECTION REVIEW

We must look beyond the modal model to account for some important types of

Explicit and Implicit Memory

- Information in explicit (declarative) memory can be brought into consciousness, while information in implicit (nondeclarative) memory cannot, though it can influence thought and behavior.
- Knowledge that is initially explicit can sometimes become implicit, as in the case of frequently repeated routine tasks.
- There are two subclasses of explicit memory: episodic memory, which is memory of particular past experiences, and semantic memory, which contains knowledge and beliefs at a level of generality beyond particular past experiences.
- Subclasses of implicit memory include the learning that arises from classical conditioning, procedural memories (e.g., how to play a guitar chord), and priming.

Neuropsychological Evidence

- People with temporal-lobe amnesia exhibit normal capacities to form and use all sorts of implicit memories.
- People with temporal-lobe amnesia typically show greater loss of episodic memory than of semantic memory, probably because of episodic memory's greater reliance on the hippocampus.
- Children under four and the aging are generally less able to encode episodic memories than semantic memories, which may be related to immaturity of or damage in the prefrontal cortex.

Concluding Thoughts

Memory is the central topic of cognitive psychology. It is relevant to all aspects of both conscious and unconscious mental activity. To help yourself organize, elaborate upon, and thereby encode into long-term memory the ideas in this chapter, you may find useful the following general thoughts.

- 1. The modal model as a functional representation of the mind** Throughout this chapter, the modal model served as the organizing structure for thinking about memory and the mind. You read of three memory stores, of control processes related to the stores, and of research aimed at characterizing the stores and processes. Your review and thoughts about all this will be most effective, I think, if you adopt a functionalist perspective. From that perspective, each store and process represents not a different part (or structure) of the mind but a different job that the mind performs in its overall task of acquiring and using information. As you review each mental component and process, think first of its main function—how it contributes to normal, everyday thought and behavior—and then think about how its special characteristics help it serve that function. You might apply such elaborative reasoning to each of the following:
- (a) sensory memory and differences between iconic and echoic memory;
 - (b) the process of attention, including roles played by the unconscious, automatic processing of unattended stimuli;
 - (c) working memory and differences between the phonological loop and the visuospatial sketchpad;
 - (d) the process of encoding information into long-term memory, including elaboration, organization, and visualization as encoding strategies;
 - (e) the consolidation of some but not all long-term memories into a more stable form;

- (f) the value of retrieval cues in retrieving information from long-term memory;
- (g) the roles of general knowledge and inferences in constructing memories of past experiences;
- (h) distinctions between semantic and episodic forms of long-term explicit memories; and, going beyond the modal model,
- (i) the roles of implicit-memory systems, particularly procedural memories and priming.

2. Unconscious supports for conscious thought and behavior Long ago, Sigmund Freud (1933/1964) drew an analogy between the human mind and an iceberg. Consciousness, he suggested, is the small visible tip of the mind, which is supported by massive unconscious portions of the mind that are invisible, submerged under the ocean's surface. Although Freud's concept of the functions of the unconscious mind (discussed in Chapter 15) was different from that presented here, the analogy remains apt. All we are conscious of are the perceptions and thoughts that course through our limited-capacity working memory. We are unconscious of all the preattentive analysis of information and of the top-down control of selective attention that help determine which stimuli make it into working memory. We are also unconscious at any given time of the vast store of information we have in long-term memory and of the priming processes that determine which portions of that store will, at that moment, be most available for retrieval into consciousness. And we are unconscious of the vast set of procedural memories and effects of conditioning that allow us to carry out routine tasks and respond adaptively to stimuli without conscious attention. As you review the chapter, think about all the ways in which unconscious information and processes support that small part of our mental activity that enters our consciousness.

3. The mind as a product of the brain In cognitive psychology the term *mind* refers to the entire set of processes—unconscious as well as conscious—by which information is acquired and used within a person to organize and direct the person's behavior. The mind is entirely a product of the brain. In recent times, cognitive psychology has merged increasingly with neuropsychology into what is now often called *cognitive neuropsychology*. The development of neuroimaging methods has allowed psychologists to identify which parts of the human brain become most active as people engage in specific mental tasks, and to relate those findings to the results of more traditional neuropsychological studies of deficits in people who have suffered damage to specific portions of the brain. Of course, there is a big difference between knowing *where* in the brain a particular task is accomplished and knowing *how* it is accomplished. At this point we are far from knowing how neural activity in the brain provides the basis for memories, thoughts, and decisions (but see Chapter 5 for work on how new and strengthened synapses may provide a foundation for implicit memories). The brain may be a computer, but it is vastly more complex than any non-biological computer that has yet been built. At this point the mapping of mental tasks upon brain areas is useful primarily as an adjunct to behavioral evidence in helping us to categorize

mental tasks. The idea that two mental tasks are fundamentally similar to or different from each other can be supported by evidence that they do or do not involve the same areas of the brain.

As you review the results of neuroimaging and brain-damage studies presented in this chapter, think about how each helps to validate the distinctions among the different memory systems and processes described in the chapter. Which findings regarding brain locations support each of the following ideas?

- (a) Attention involves top-down processes that magnify the neural analysis of attended stimuli relative to unattended stimuli.
- (b) Verbal working memory and visuospatial working memory are distinct from each other.
- (c) Verbal working memory is like talking and listening.
- (d) Using remembered visual images to answer questions about what and where is like looking at an actual object or scene to answer those questions.
- (e) Long-term memory is distinct from working memory.
- (f) Long-term memory exists in two forms, one more stable than the other.
- (g) Explicit memory is distinct from implicit memory.
- (h) Episodic memory is at least partly distinct from semantic memory.

Further Reading

DANIEL L. SCHACTER (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton-Mifflin.

According to Schacter, the “sins” (weaknesses) of memory are (1) transience (tendency to forget); (2) absent-mindedness (failure to shift attention); (3) blocking (thwarting of retrieval attempts); (4) misattribution (source confusion); (5) suggestibility (influence of others' statements on memory construction); (6) bias (influence of one's own beliefs and desires on memory construction); and (7) persistence (inability to forget disturbing memories). The author, who is a leading memory researcher, devotes a chapter to each “sin,” and then, in the final chapter, explains how each results from memory characteristics that are more often virtues than sins. The book is well written and easily readable by non-specialists.

PHILIP J. HILTS (1995). *Memory's ghost: The strange tale of Mr. M. and the nature of memory*. New York: Simon & Schuster. This is a literary essay on the role of memory in human existence, and it is

also the biography of a man who has no autobiography of adult life. It is the story of H. M., who at age 27 (42 years before Hilts's book) lost his ability to form new explicit, long-term memories. We learn here something of what it is like to live in a disconnected series of eyeblinks, where the past is but a few seconds long.

CHARLES P. THOMPSON, THADDEUS M. COWAN, & JEROME FRIEMAN (1993). *Memory search by a memorist*. Hillsdale, NJ: Erlbaum.

Many stories have been told about people with amazing memories. Do such people have brains that are fundamentally different from yours and mine? Is it all a matter of practice? This slim book is by three psychological researchers who performed a series of experiments on the memory abilities of Rajan Srinivasan Mahavedan, who was once listed in the *Guinness Book of World Records* for reciting 31,811 decimal digits of pi without error. (His record has since been topped by Hideaki Tomoyori, who recited 40,000 digits of pi.) The experiments described here help demystify Mahavedan's memory. The book also contains a brief dis-

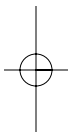
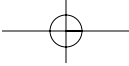
cussion of the memory abilities of 13 other memorists whose abilities were studied by psychologists.

RICHARD J. McNALLY (2003). *Remembering trauma*. Cambridge, MA: Harvard University Press.

Ever since Freud, some psychiatrists and psychologists have contended that people often repress traumatic memories, suffer from the unconscious effects of the memories, and recover in part by discovering the repressed memories in psychotherapy. In this book, McNally, a clinical psychologist and memory researcher, examines the clinical and laboratory evidence pertaining to this view and finds little support for it. The book reviews evidence concerning the poor formation of episodic memories in early childhood, the power of suggestion to distort and create memories, and the inability of real trauma survivors to forget their experiences.

FPO





Reasoning and Intelligence

HOW PEOPLE REASON I: ANALOGIES AND INDUCTION

Analogies as Foundations for Reasoning
Inductive Reasoning and Some Biases in It

HOW PEOPLE REASON II: DEDUCTION AND INSIGHT

The Concrete Nature of Deductive Reasoning
Elements of Insight: How People Solve Problems Creatively

EFFECTS OF CULTURE AND LANGUAGE ON THOUGHT

Some Cross-Cultural Differences in Perception and Reasoning
The Words of One's Language Can Affect One's Thinking

THE PRACTICE AND THEORY OF INTELLIGENCE TESTING

A Brief History of Intelligence Testing
The Validity of Intelligence Tests as Predictors of Achievement
The Concept of General Intelligence and Attempts to Explain It

GENETIC AND ENVIRONMENTAL CONTRIBUTIONS TO INTELLIGENCE

Contributions to IQ Differences Within a Cultural Group
Origins of IQ Differences Between Cultural Groups

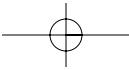
Compared with other species, we are not the most graceful, nor the strongest, nor the swiftest, nor the fiercest, nor the gentlest, nor the most long-lived, nor the most resistant to the poisons accumulating in our atmosphere. We do, however, fancy ourselves to be the most intelligent of animals; and, at least by our own definitions of intelligence, our fancy is apparently correct. We are the animals that know and reason; that classify and name the other animals; that try to understand all things, including ourselves. We are also the animals that tell one another what we know, with the effect that each generation starts off with more knowledge, if not more wisdom, than the previous one.

In the last chapter I equated the mind with memory, defining memory broadly as all the information we store, whether for long periods or only fleetingly, and all the mechanisms we have for manipulating that information. But what is the purpose of memory, thus defined? From an evolutionary perspective, there is no value in reminiscence for its own sake. What's past is past; we can't do anything about it. We can, however, influence our future. The evolutionary functions of memory—indeed, the functions of the mind—are to understand our present situation, recognize and solve problems posed by that situation, anticipate the future, and make plans that will help us prepare for and in some ways alter that future for our own well-being. Our memory of the past is useful to the degree that it helps us understand and deal adaptively with the present and the future. The processes by which we use our memories in these adaptive ways are referred to as *reasoning*, and our general capacity to reason is referred to as *intelligence*. Reasoning and intelligence are the topics of this chapter.

How People Reason I: Analogies and Induction

To a very large extent, we reason by using our memories of previous experiences to make sense of present experiences or to plan for the future (Gentner, 2003; Holyoak & others, 2001; James, 1890). To do so, we must perceive the similarities among various events that we experience. Even our most basic ability to categorize experiences and form mental concepts depends on our ability to perceive similarities (Hofstadter, 2001).

Consider, for example, our understanding of snowflakes. No two snowflakes are the same, but they are similar enough that we can have a meaningful *snowflakes* concept. From experience with past snowflakes, I can predict what will happen with future snowflakes. They will melt if I hold them in my bare hand;



they will pile up and form a white, slippery covering over the ground if they keep falling and the air temperature is low enough; masses of them can be packed to form snowballs and snow forts; and so on. My *snowflakes* concept, formed from my previous experiences with snowflakes, allows me to know how to deal with snowflakes in my present environment and allows me to make plans for future snowfalls.

The ability to see similarities among different instances of snowflakes, or snowfalls, may not seem like reasoning, because it is so easy and natural for us. In other cases, however, useful similarities are not so easily seen. Often, it is the ability to see similarities where others don't notice them that distinguishes excellent reasoners from the rest of us. The kinds of reasoning that most clearly depend on perception of similarities are analogical reasoning and, closely related to it, inductive reasoning. These are the topics of this section.

Analogies as Foundations for Reasoning

1

What is some evidence concerning the usefulness of analogies in scientific reasoning?

<<< In the most general sense of the term, an *analogy* is any perceived similarity between otherwise different objects, actions, events, or situations (Hofstadter, 2001). Psychologists, however, generally use the term somewhat more narrowly to refer to certain types of similarities and not to others. In this more restricted sense, an *analogy* refers to a similarity in behavior, function, or relationship, but not to a similarity in the identity or superficial appearance of the entities compared. Here are three examples, which may help clarify the concept.

- *A baseball glove is analogous to a butterfly net.* Here the analogy lies in the fact that both are used for capturing some category of objects (baseballs or butterflies) and both have a somewhat funnel-like shape that is useful for carrying out their function.
- *Marriage between humans is analogous to pair-bonding between geese.* Here the analogy lies in the type of relationship that is expressed by the concepts of marriage and pair-bonding.
- *Purple finches look like house finches.* This would generally not be considered an analogy, because here the similarity referred to lies in the appearance of the two species of birds rather than in some aspect of their behavior.

Use of Analogies in Scientific Reasoning

Scientists often attempt to understand and explain natural phenomena by drawing explicit analogies to other phenomena that are better understood. As I pointed out in Chapter 3, Charles Darwin came up with the concept of natural selection as the mechanism of evolution partly because he saw the analogy between selective breeding of plants and animals by humans and the selective breeding that occurs in nature. Since the former type of selective breeding could modify plants and animals over generations, it made sense to Darwin that the latter type could too. Johannes Kepler developed his theory of the role of gravity in planetary motion by drawing analogies between gravity and light, both of which can act over long distances but have decreasing effects as distance becomes greater (Gentner & Markman, 1997). Neuroscientists have made progress in understanding some aspects of the brain through analogies to such human-made machines as digital computers.

In an analysis of discussions held at weekly laboratory meetings in many different biology labs, Kevin Dunbar (1999, 2001) discovered that biologists use analogies regularly to help them make sense of new findings and to generate new hypotheses. In a typical one-hour meeting, scientists generated anywhere from 2 to 14 different analogies as they discussed their work. Most of the analogies were to other biological findings, but some were to phenomena completely outside the realm of biology.

To see if nonscientists would generate and use analogies to reason about physical systems, Dedre Gentner and Donald Gentner (1983) tested high school and col-

lege students who had little training in physical sciences on their understanding of electrical circuits. As they tried to answer questions having to do with the movement of electricity through wires, some students thought of analogies to water moving through pipes and others thought of analogies to crowds of people or cars moving through tunnels or along roads. Those who used plumbing analogies were best at answering questions about the effects of adding extra batteries to the circuit. They thought of batteries as being like pumps or like raised reservoirs, which add to the water pressure, and this helped them answer the battery questions correctly. Those thinking about traffic or crowd movement usually did not think of a good analogy to batteries, but did often think of an excellent analogy to resistors. They thought of resistors as being like gates or turnstiles, which restrict the number of cars or people that can pass through at any given time, and this analogy helped them to answer correctly questions about the effects of adding extra resistors to the circuit. Thus, the plumbing analogy was best for thinking about batteries and voltage (a measure of electrical force), but the traffic or crowd analogy was best for thinking about resistors and amperage (a measure of the amount of electricity moving through the circuit).



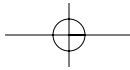
Traffic resistor The George Washington Bridge, which connects New Jersey to New York City, restricts the flow of traffic into the city in a manner analogous to the way a resistor in an electrical circuit restricts the flow of electrons. In a research study, subjects who thought of such analogies performed better in answering questions about electrical circuits than did those who failed to think of them.

Use of Analogies in Judicial and Political Reasoning and Persuasion

Lawyers, politicians, and ordinary people frequently use analogies to convince others of some claim or course of action they support. The following example is taken from a novel (Bugliosi, 1978, cited by Halpern, 1996), but it certainly could occur in real life. At the end of a trial involving much circumstantial evidence, the defense attorney, in his summation to the jury, said that evidence is like a chain; it is only as strong as its weakest link. If one piece of evidence is weak, the chain breaks and the jurors should not convict the accused. The prosecutor then stood up and told the jurors that evidence is not like a chain, but is like a rope made of many separate strands twisted together; its strength is the sum of that of the individual strands. Even if some of the weaker strands break, the strong ones still hold the rope together. The prosecutor had the more convincing analogy and won the case.

Researchers have found that university students are good at generating analogies to defend political viewpoints. In one study, in Canada, some students were asked to defend the position that the Canadian government must eliminate deficit spending even if that requires a sharp reduction in such social programs as health care and support for the needy, and other students were asked to defend the opposite position (Blanchette & Dunbar, 2000). Students arguing on either side developed many potentially convincing analogies. Not surprisingly, many of the analogies were from the closely related realm of personal finances—comparing the national debt to one's personal debt, or comparing a reduction in social programs to a failure to invest one's personal money for future gain. But many other analogies were taken from more distant realms. For instance, failure to eliminate the debt was compared to failure to treat a cancer, which grows exponentially over time and becomes uncontrollable if not treated early; and failure to provide care

2 How are analogies useful in judicial and political reasoning? What distinguishes a useful analogy from a misleading one?



for the needy was compared to a farmer's failure to care for crops that have been planted, which ruins the harvest.

The next time you become involved in, or listen to, a political discussion, tune your ears to the analogies presented and, for each, ask yourself whether it helps to clarify the issue or misleads. You may be surprised at how regularly and easily analogies slip into conversations; they are a fundamental component of human thought and persuasion. We reason about new or complicated issues largely by comparing them to more familiar or less complicated issues, where the answer seems clearer. Such reasoning is useful to the degree that the structural relationships in the analogy hold true; it is misleading to the degree that those relationships don't hold true. Good reasoners are those who are good at seeing the structural relationships between one kind of event and another.

Analogy Problems Are Used in Tests of Reasoning Ability

3

How do the Miller Analogy Test and Raven's Progressive Matrices Test assess a person's ability to perceive analogies?

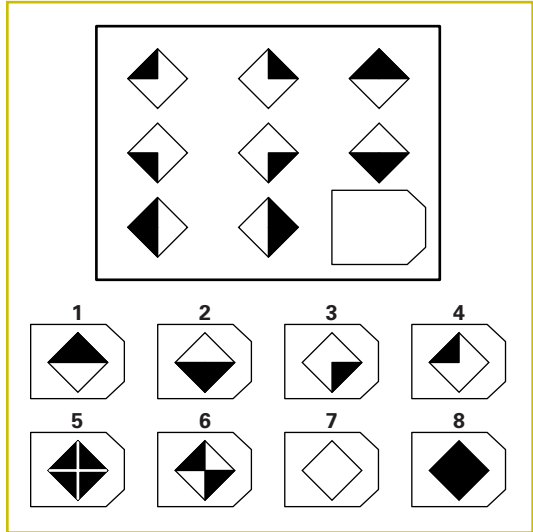
<<< Tests of reasoning ability commonly use analogy problems. The Miller Analogy Test, for example, is made up entirely of analogy problems. This test is commonly required of people applying for graduate study and for certain kinds of jobs. Many correlational studies show that a person's score on this test is a reasonably good predictor of how well he or she will perform in graduate study or in a job that requires one to take new information into account and solve complex problems (Kuncel & others, 2004). Two examples of the types of problems in the Miller Analogy Test are the following:

1. PLANE is to AIR as BOAT is to (a) submarine, (b) water, (c) oxygen, (d) pilot.
2. SOON is to NEVER as NEAR is to (a) close, (b) far away, (c) nowhere, (d) seldom.

To answer such questions correctly, you must see a relationship between the first two concepts and then apply it to form a second pair of concepts that are related to each other in the same way as the first pair. The relationship between PLANE and AIR is that the first *moves through* the second, so the correct pairing with BOAT is WATER. The second problem is a little more difficult. Someone might mistakenly think of SOON and NEVER as opposites and might therefore pair NEAR with its opposite, FAR AWAY. But *never* is not the opposite of *soon*; it is, instead, the negation of the entire dimension that *soon* lies on (the dimension of time extending from now into the future). Therefore, the correct answer is NOWHERE, which is the negation of the dimension that *close* lies on (the dimension of space extending outward from where you are now). If you answered the second problem correctly, you might not have consciously thought it through in terms like those I just presented; you may have just, intuitively, seen the correct answer. But your intuition must have been based, unconsciously if not consciously, on your deep knowledge of the concepts referred to in the problem and your understanding of the relationships among those concepts.

Another mental test that makes exclusive use of analogy problems is Raven's Progressive Matrices Test, which is often used by psychologists as a test of fluid intelligence, a concept discussed later. In this test, the items are visual patterns rather than words, so knowledge of word meanings is not essential. Figure 10.1 illustrates a typical Raven's problem. The task is to examine the three patterns in each of the top two rows to figure out the rule that relates the first two patterns in each row to the third pattern. The rows are analogous to one another in that the same rule applies to each, even though the substance of the patterns is different from row to row. Once the rule is figured out, the problem is solved by applying that rule to the bottom row. In the example shown in Figure 10.1, the rule for each row is that the first pattern is superimposed onto the second pattern to produce the third pattern. Applying that rule to the third row shows that the correct solution to this problem, chosen from the eight pattern choices at the bottom, is number 8.

FIGURE 10.1 | Sample Raven's problem This problem (from Carpenter & others, 1990, p 409) is similar to the problems in Raven's Progressive Matrices Test. The task is to infer the rule describing how the pattern changes within each of the first two rows and then to apply that rule to the third row to arrive at the correct solution. The solution is one of the eight choices at the bottom.



Inductive Reasoning and Some Biases in It

Inductive reasoning, or *induction*, is the attempt to infer some new principle or proposition from observations or facts that serve as clues. Induction is also called *hypothesis construction* because the inferred proposition is at best an educated guess, not a logical necessity. Scientists engage in inductive reasoning all the time as they try to infer rules of nature from their observations of specific events in the world. Psychologists reason inductively when they make guesses about the workings of the human mind on the basis of observations of many instances of human behavior under varied conditions. Detectives reason inductively when they piece together bits of evidence to make inferences as to who might have committed a crime. In everyday life we all use inductive reasoning regularly to make sense of our experiences or predict new ones. When I look outside in the morning, see that the ground is wet, and say, “It probably rained last night,” I am basing that statement on inductive reasoning. My past observations of relationships between rain and wet ground have led me to induce the general rule that wet ground usually implies rain.

All the examples of reasoning through the perception of analogies, discussed earlier, are also examples of inductive reasoning. In fact, in general, inductive reasoning is reasoning that is founded on perceived analogies or other similarities. The evidence from which one induces a conclusion is, ultimately, a set of past experiences that are in some way similar to one another or to the experience one is trying to explain or predict.

In general, we are very good at inductive reasoning, certainly far better at it than is any other species of animal (Gentner, 2003). However, most psychologists who study inductive reasoning have focused not on our successes but on our mistakes, and such research has led to the identification of several systematic biases in our reasoning. Knowledge of such biases is useful to psychologists for understanding the cognitive processes that are involved in reasoning and to all of us who would like to reason more effectively.

>>> 4

What is inductive reasoning, and why is it also called hypothesis construction? Why is reasoning by analogy inductive?

SPEED BUMP Dave Coverly



The Availability Bias

The availability bias is perhaps the most obvious and least surprising. When we reason, we tend to rely too strongly on information that is readily *available* to us rather than on information that is less available. When people were asked whether the letter *d* is more likely to occur in the first position or the third position of a

>>> 5

What kinds of false inferences are likely to result from the availability bias?



Contribution to an availability bias
The heavy publicity of children’s murders may lead many people to overestimate their frequency. Such misjudgments could lead parents to pay too much attention to protecting their children from murder and not enough attention to protecting them from much more common causes of death, such as traffic accidents.

word, most people said the first (Tversky & Kahneman, 1973). In reality, *d* is more likely to be in the third position; but, of course, people find it much harder to think of words with *d* in the third position than to think of words that begin with *d*. That is an example of the availability bias. As another example, when asked to estimate the percentage of people who die from various causes, most people overestimate causes that have recently been emphasized in the media, such as terrorism, murders, and airplane accidents, and underestimate less-publicized but much more frequent causes, such as heart disease and traffic accidents (Brase, 2003). The heavily publicized causes are more available to them than are the less-publicized causes.

The Confirmation Bias

Textbooks on scientific method (and this book, in Chapter 2) explain that scientists should design studies aimed at *disconfirming* their currently held hypotheses. In principle, one can never prove absolutely that a hypothesis is correct, but one can prove absolutely that it is incorrect. The most creditable hypotheses are those that survive the strongest attempts to disprove them. Nevertheless, research indicates that people's natural tendency is to try to confirm rather than disconfirm their current hypotheses (Lewicka, 1998).

6

What are two different ways by which researchers have demonstrated the confirmation bias?

<<< In an early demonstration of this *confirmation bias*, Peter Wason (1960) engaged subjects in a game in which the aim was to discover the experimenter's rule for sequencing numbers. On the first trial the experimenter presented a sequence of three numbers, such as 6 8 10, and asked the subject to guess the rule. Then, on each subsequent trial, the subject's task was to test the rule by proposing a new sequence of three numbers to which the experimenter would respond yes or no, depending on whether or not the sequence fit the rule. Wason found that subjects overwhelmingly chose to generate sequences consistent with, rather than inconsistent with, their current hypotheses and quickly became confident that their hypotheses were correct, even when they were not. For example, after hypothesizing that the rule was *even numbers increasing by twos*, a person would, on several trials, propose sequences consistent with that rule—such as 2 4 6 or 14 16 18—and, after getting a yes on each trial, announce confidently that his or her initial hypothesis was correct. Such persons never discovered that the experimenter's actual rule was *any increasing sequence of numbers*.

In contrast, the few people who discovered the experimenter's rule proposed, on at least some of their trials, sequences that contradicted their current hypothesis. A successful subject who initially guessed that the rule was *even numbers increasing by twos* might, for example, offer the counterexample 5 7 9. The experimenter's yes to that would prove the initial hypothesis wrong. Then the subject might hypothesize that the rule was *any sequence of numbers increasing by twos* and test that with a counterexample, such as 4 7 32. Eventually, the subject might hypothesize that the rule was *any increasing sequence of numbers* and, after testing that with counterexamples, such as 5 6 4, and consistently eliciting no as the response, announce confidence in that hypothesis.

In other experiments demonstrating a confirmation bias, subjects were asked to interview another person to discover something about that individual's personality (Skov & Sherman, 1986; Snyder, 1981). In a typical experiment, some subjects were asked to assess the hypothesis that the person is an *extravert* (socially outgoing), and others were asked to assess the hypothesis that the person is an *introvert* (socially withdrawn). The main finding was that subjects usually asked questions for which a yes answer would be consistent with the hypothesis they were testing. Given the extravert hypothesis, they tended to ask such questions as "Do you like to meet new people?" And given the introvert hypothesis, they tended to ask such questions as "Are you shy about meeting new people?" This bias, coupled with the natural tendency of interviewees to respond to all such questions in the affirmative, gave most subjects confidence in the initial hypothesis, regardless of which hypothesis that was or whom they had interviewed.

Why do people show the confirmation bias? One possibility, which pertains to the examples just noted, is that *yes* is more pleasant to hear than *no* in social exchanges, and so we may have learned to bias our statements in ways designed to elicit agreement. A deeper possibility, which would apply to a larger set of examples, is that in everyday life the goal of gathering information for the long term often conflicts with the goal of being rewarded, or avoiding disaster, at any given moment (Lewicka, 1998). Through evolution or learning or both, we might have developed an adaptive tendency to stick with hypotheses that seem to be working rather than to test their limits by behaving in ways that violate them. I don't honestly know that wearing my orange vest while bicycling helps keep motorists from hitting me; but I wear it each time I ride, and so far I haven't been hit. If my primary goal were knowledge rather than survival, I might perform an experiment: Wear the vest some days and not others and tally the number of times I get hit in either condition.

>>> 7
How might the confirmation bias be adaptive (useful) in everyday life?

The Predictable-World Bias

We are so strongly predisposed to find order in our world that we are inclined to “see” or anticipate order even where it doesn't exist. Superstitions often arise because people fail to realize that coincidences are often just coincidences. Some great event befalls a man when he is wearing his green shirt and suddenly that becomes his “lucky” shirt.

The predictable-world bias is most obvious in games of pure chance. Gamblers throwing dice or betting at roulette wheels often begin to think that they see reliable patterns in the results. This happens even to people who consciously “know” that the results are purely random. Part of the seductive force of gambling is the feeling that you can guess correctly at a better-than-chance level or that you can control the odds. In some games such a belief can lead people to perform *worse* than they should by chance. Here's an example:

Imagine playing a game with a 6-sided die that has 4 red sides and 2 green sides. You are asked to guess, on each trial, whether the die will come up red or green. You will get a dollar every time you are correct. By chance, over the long run, the die is going to come up red on four-sixths, or two-thirds, of the trials and green on the remaining one-third of trials. Each throw of the die is independent of every other throw. No matter what occurred on the previous throw, or the previous ten throws, the probability that the next throw will come up red is two chances in three. That is the nature of chance. Thus, the best strategy is to guess red on every trial. By taking that strategy you will, over the long run, win on about two-thirds of the trials. That strategy is called *maximizing*.

>>> 8
How do die-tossing games demonstrate the predictable-world bias?

But most people who are asked to play such a game do not maximize (Stanovich & West, 2003). Instead, they play according to a strategy referred to as *matching*; they vary their guesses over trials in a way that matches the probability that red and green will show. Thus, in the game just described, they guess red on roughly two-thirds of the trials and green on the other one-third. They know that, over the long run, about two-thirds of the throws will be red and one-third green, and they behave as if they can predict *which* ones will be red and which ones will be green. But in reality they can't predict that, so the result is that they win less money over the long run than they would if they had simply guessed red on every trial. Depending on how you look at it, you could say that the typical player in such a game is either too smart or not smart enough for his own good. Rats, in analogous situations—where they are rewarded more often for pressing a red bar than for pressing a green one—quickly learn to press the red one all the time when given a choice. Presumably, their maximizing derives from the fact that they are not smart enough to try to figure out a pattern; they just behave in accordance with what has worked best for them. Very bright people, those with the highest IQ scores, also typically maximize in these situations, but for a different reason (Stanovich & West, 2003). They do so because they understand that there is no pattern, so the best bet always is going to be red.

The predictable-world bias, really, is a tendency to engage in inductive reasoning even in situations where such reasoning is pointless because the relationship

in question is completely random. Outside of gambling situations, a tendency to err on the side of believing in order rather than randomness may well be adaptive. It may prompt us to seek order and make successful predictions where order exists, and that advantage may outweigh the corresponding disadvantage of developing some superstitions.

SECTION REVIEW

We reason largely by perceiving similarities between new events and familiar ones.	
Analogies as a Basis for Reasoning • Analogies are similarities in behavior, functions, or relationships in otherwise different entities or situations. • Scientists and even nonscientists often use analogies to make sense of observations and generate new hypotheses. • Analogies are commonly used in legal and political persuasion. • Some tests measure an individual's ability to see and apply analogies, which is a good predictor of success in graduate school or jobs requiring complex problem-solving.	Inductive Reasoning • In inductive reasoning, or hypothesis-testing, a new principle or proposition is inferred based on specific observations or facts. We are quite good at inductive reasoning, but are susceptible to certain biases. • The availability bias is our tendency to give too much weight to information that comes more easily to mind than does other relevant information. • The confirmation bias leads us to try to confirm rather than disconfirm our current hypothesis. Logically, a hypothesis cannot be proven, only disproven. • The predictable-world bias leads us to arrive at predictions through induction even when events are actually random.

9

How does deductive reasoning differ from inductive reasoning? How is it illustrated by series problems and syllogisms?

How People Reason II: Deduction and Insight

<<< **Deductive reasoning**, or *deduction*, is the attempt to derive logically the consequences that *must* be true if certain premises are accepted as true. Whereas inductive reasoning is reasoned guesswork, deductive reasoning (when done correctly) is logical proof, assuming that the premises really are true. In everyday life we reason deductively to derive the logical implications of statements that we hear. If you tell me that everyone in your family is over six feet tall, then, assuming I believe you, I know, by deduction, that everyone in your family is also over five feet tall. If you studied plane geometry, then you engaged in deductive reasoning that is more complex than most everyday examples as you tried to prove or disprove various correlates based on axioms that were given to you. In fact, all of mathematics is deduction. One starts, in mathematics, with certain givens and deduces the consequences.

Table 10.1 presents deductive problems that are typical of the types of problems used by psychologists who study such reasoning. Before looking at the answers, try solving each problem. The first is a *series problem*, which requires you to organize items into a series on the basis of a set of comparison statements and then arrive at a conclusion that was not contained in any single statement. The second is a *syllogism*, which presents a major premise, or proposition, and a minor premise that you must combine mentally to see if a particular conclusion is true, false, or indeterminate (cannot be determined from the premises). Did you get the correct answer to each problem? If you did, you deduced correctly.

The Concrete Nature of Deductive Reasoning

There was a time when many psychologists believed that deductive reasoning is, at root, a logical process best understood in mathematical terms. For example, the highly influential Swiss developmental psychologist, Jean Piaget (whose work is discussed in Chapter 11), believed that deduction by people who are roughly 13 years of age or older is best characterized by what he referred to as “formal opera-

tions,” by which he meant the understanding and application of logical principles (Inhelder & Piaget, 1958). According to that view, problems are solved by a sort of mental algebra, in which the specific contents of problems are mapped onto the Xs and Ys of logical equations, the equations are solved, and the answers are delivered.

Today relatively few psychologists accept the “abstract logic” or “formal operational” view of deductive reasoning (Evans, 2002; Shafir & LeBoeuf, 2002). One reason for rejecting that view is that study after study, with people all over the world, including highly educated people in universities, has revealed that we are much better at solving problems put to us in concrete terms than problems put to us in terms of Xs and Ys or other abstract symbols. Research has repeatedly shown that we usually solve deductive problems by reflecting on our real-world knowledge, not by thinking about laws of logic. With training, people can learn laws of logic, but even those who learn them well rarely apply them to the problems of daily life.

Bias to Attend to the Content of Deductive Problems Rather Than to the Logic

If people used formal logic to solve syllogisms, then it should not matter whether the statements in the problem are consistent with everyday experience, violate everyday experience, or are nonsensical. All that should matter is the formal structure of the problem. But numerous experiments show that the content does matter. Consider, for example, the following two syllogisms. In each, you are instructed to use logic alone, not your knowledge of the real world. In each, you are to assume that the first two propositions are valid and you are to judge whether or not the conclusion follows logically from those propositions.

- 1. All living things need water. Roses need water.
Therefore, roses are living things.
- 2. All insects need oxygen.
Mice need oxygen.
Therefore, mice are insects.

Structurally, these two problems are identical. In each case, the conclusion is not valid. It does not necessarily follow, from the first two premises of the first problem, that roses are living things, or, from the first two premises of the second problem, that mice are insects. The correct conclusion for both of these syllogisms is “indeterminate.” According to the premises, roses may be living things but they don’t have to be, and mice may be insects but they don’t have to be. When university students were given these problems, only about 30 percent got the first problem correct, but nearly all of them got the second problem correct (Stanovich, 2003). Even though the students understood that these are logic problems, not questions about real-world facts, they apparently could not resist being influenced by their knowledge of the real world. Their knowledge that roses are living things led them to believe that the conclusion to the first problem is logically valid, and their knowledge that mice are not insects led them to believe that the conclusion to the second problem is logically not valid. When the same type of problem is put to students using nonsense terms—such as “All schniezels need quistics”—an intermediate number, typically around 70 percent, get it correct (Stanovich, 2003). In that case, the content neither helps nor hinders.

The bias to use knowledge rather than logic in answering deductive reasoning questions can be construed as a bias to think inductively rather than deductively. Our natural tendency is to reason by comparing the current information with our previous experience, and, outside the mathematics classroom or psychology experiment on logic, that tendency generally serves us well.

Use of Diagrams and Mental Models to Solve Deductive Problems

A major difficulty in solving any complex deductive reasoning problem is that of representing all the problem information in a way that allows you to see all the relationships implied by that information. When you solved the series problem in

>>> **10**

How has it been shown that our strong tendency to rely on real-world knowledge can overwhelm our deductive-reasoning ability?



Sidney Harris at ScienceCartoonsPlus.com

>>> **11**

How are diagrams and mental models used to solve deductive problems? Specifically, how are Euler circles used to solve syllogisms?

TABLE 10.1
Deductive-reasoning problems

Series problem	Syllogism	
John is taller than Henry.	All chefs are violinists (major premise).	<i>Answers:</i>
John is shorter than Mary.	Mary is a chef (minor premise). Is Mary a violinist?	(indeterminate, no.)
Mary is shorter than Billy.	<i>Alternative forms, based on different minor premises:</i>	(Alternative forms: yes.)
Is Billy shorter than Henry?	Mary is a violinist. Is she a chef?	<i>Syllogism: yes.</i>
	Mary is not a chef. Is she a violinist?	<i>Series problem: no.</i>
	Mary is not a violinist. Is she a chef?	

Table 10.1, you may—either on paper or in your head—have made little stick figures to represent the relative heights of each of the four individuals—a longer stick for John than for Henry, a longer one yet for Mary, and a still longer one for Billy. Then all you had to do was examine the drawing or mental model and see that the stick for Billy is longer than that for Henry.

People who are well trained in deductive reasoning make much use of diagrams. In one study, Ph.D. mathematicians were asked to reason aloud as they solved complex mathematical problems (Stylianou, 2002). All the mathematicians drew diagrams as part of the reasoning process. Typically, they would draw an initial diagram to represent the information given in the problem, perform some calculations, draw another diagram to represent the result of those calculations, perform some more calculations, draw another diagram, and so on, until they reached a solution. A diagram allows the reasoner to visualize the problem information in a way that helps make the solution, or the steps one must take toward solution, obvious.

Most people who can solve syllogisms easily, regardless of content, have mastered a little trick that allows them to turn any syllogistic premise into a picture. One version of this trick involves *Euler circles*, named after Leonard Euler, who used them in teaching logic to a German princess (Johnson-Laird, 1983). As an illustration of this method, Figure 10.2 shows how Euler circles can be used to represent the major premise of the syllogism presented in Table 10.1, “All chefs are violinists.” The set of all chefs is represented by a circle, and the set of all violinists is represented by another circle. Since all chefs are violinists, the circle representing chefs falls entirely within the circle representing violinists. The problem information does not say whether or not all violinists are chefs, so there are two possible ways to represent the world of chefs within the world of violinists. The left-hand diagram depicts the possibility that some violinists are not chefs, and the right-hand diagram depicts the possibility that chefs and violinists are identical sets. Now, having drawn the two logically possible diagrams, it is easy to inspect them to answer all four questions posed by the syllogism in Table 10.1. For example, the question, “If Mary is a violinist, is she a chef?” must be judged as indeterminate. The left-hand diagram shows that it is possible for Mary to be in the violinist circle without being in the chef circle, but it is also possible for her to be in both circles.

Phillip Johnson-Laird (1985, 2004) contends that people who solve deductive reasoning problems without formal training and without pencil and paper typically do so by constructing mental models. To construct a good model, the reasoner must understand the premises and find a way to represent that information mentally in an easily accessible form. Once the information is well represented in a model, the reasoner can examine the model to find the answer to the problem.

Although Johnson-Laird does not claim that mental models must take the form of visual images, the easiest such models to understand are those that do. The most

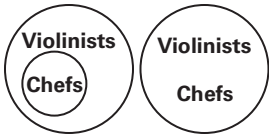


FIGURE 10.2 | Euler circles
Here the proposition “All chefs are violinists” is represented by Euler circles. Two diagrams are necessary to include both the possibility that some violinists are not chefs (left-hand diagram) and the possibility that all violinists are chefs (right-hand diagram).

efficient mental models might be mental versions of diagrams comparable to Euler circles. According to Johnson-Laird, however, the models most people construct are more closely linked to the actual content of the problem than are Euler circles. For instance, a person reasoning about the chefs and violinists syllogism might imagine a group of people in a room, some of whom are wearing chefs' hats and holding violins, some of whom are holding violins but aren't wearing chefs' hats, and some of whom have neither violins nor chefs' hats. If the image includes all the possibilities allowed by the first premise and no other possibilities, then the reasoner can inspect that image to answer all the syllogism's questions.

Consistent with Johnson-Laird's theory, the difficulty that people have with deductive problems correlates positively with the number and complexity of the models that are required to represent all the information in the premises (Johnson-Laird & others, 1994, 2000). Consistent with the idea that the models often take the form of visual images, researchers have found that performance on syllogisms correlates more strongly with visuospatial ability than with verbal ability, as measured by standard intelligence tests (Frandsen & Holder, 1969; Guyote & Sternberg, 1981). Also, fMRI studies reveal that the right hemisphere of the brain, which is involved in visuospatial reasoning, usually shows a greater increase in activity during deductive reasoning than does the left hemisphere, which is more involved in verbal reasoning (Johnson-Laird, 2004).

>>> **12**
What is some evidence that deductive reasoning often involves visual imagery?

Elements of Insight: How People Solve Problems Creatively

Sometimes a real-life problem will stymie us for hours, or days, and then suddenly, "Aha!"—we see the solution. What causes such flashes of insight? What do we do to bring them about? What might we do to achieve such insights more quickly and regularly? To address these questions, psychologists have studied people's performance on **insight problems**, problems that are specially designed to be difficult to solve until one looks at them in a way that is different from the usual way. Insight problems often entail a mix of inductive and deductive reasoning. Sometimes the crucial insight involves perceiving some similarity or analogy that one didn't perceive before, and sometimes it involves a new understanding of the problem's propositions or of the steps that could or could not lead to a solution.

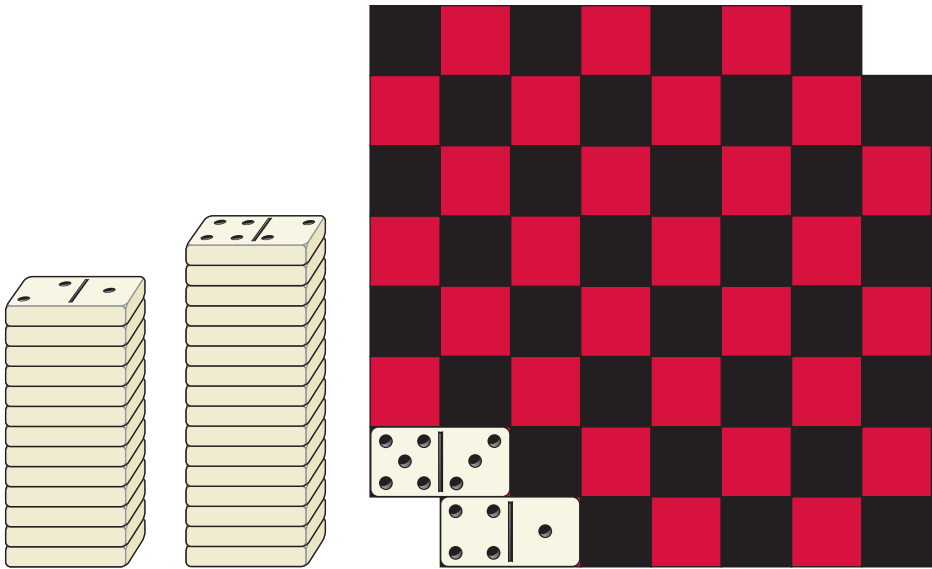
Two Examples of Insight Problems

One problem that psychologists have used in insight experiments is the *mutilated-checkerboard problem* (Robertson, 2001). Subjects are presented with a standard checkerboard that has had two of its squares, at opposite corners of the board, removed or blocked off, as shown in Figure 10.3. They are also given 31 dominos, each of which is just the right size to fit over two adjacent squares of the checkerboard. Their task is to answer correctly the following question and to explain their answer: Is it possible or impossible to set the 31 dominos on the board in such a way that they fully cover all 62 of the remaining squares of the checkerboard?

Another classic insight problem is the *candle problem*. Subjects are given a candle, a book of matches, and a box of tacks and are asked to attach the candle to a bulletin board in such a way that the candle can be lit and will burn properly (see Figure 10.4). They are allowed to use no objects other than those they have been given.

Before reading the next section, spend a few minutes trying to solve each of these problems. As you work on each, pay attention to your own thought processes. If you solved either problem, how did you do it? What thoughts led you to insight? If you did not solve either problem, join the crowd. Relatively few people solve them without hints or clues of some kind. That is especially true of the mutilated-checkerboard problem. We'll come to the solutions shortly.

FIGURE 10.3 The mutilated-checkerboard problem Is it possible to fully cover all 62 squares of this incomplete checkerboard with 31 dominos?



Breaking Out of a Mental Set: Broadening One's Scope of Perception and Thought

13 How is the concept of a mental set illustrated by the mutilated-checkerboard problem and the candle problem?

<<< Insight problems are difficult, in general, because their solution depends on abandoning a well-established habit of perception or thought, referred to as a **mental set**, and then viewing the problem in a different way.

Almost everyone presented with the mutilated-checkerboard problem sees it, at first, as a problem that can be solved by trial and error (Robertson, 2001; Kaplan & Simon, 1990). By trying all possible ways of arranging the 31 dominos over the checkerboard, with each domino covering two squares, it should be possible to find out if at least one arrangement will cover all the squares. Indeed, that is a reasonable route to solving the problem—for a computer, but not for a person. Craig Kaplan and Herbert Simon (1990) wrote a computer program to solve the problem by the trial-and-error method. The program solved the problem correctly, but had to make a total of 758,148 domino placements to do so. By eliminating, through mathematical means, those that are redundant with one another, it is possible to reduce the number of required domino placements to several thousand, but that is still far more than most people have time or patience for. (Some people, however, are more patient than others. Kaplan and Simon reported that a chemical engineering graduate student spent 18 hours trying to solve the problem through a combination of trial and error and mathematical formulas before giving up.)



So far, every person who has solved the mutilated-checkerboard problem and offered a convincing explanation for the solution, without writing a computer program, has done so by abandoning the trial-and-error approach. The problem is difficult partly because people fail to see that trial-and-error won't work, so they take a long time to give up that mental set. It is not obvious to most people that to solve it that way, they would have to run through thousands of different arrangements of the dominos and keep track of all of those arrangements so as not to repeat themselves.

The candle problem is difficult because people see the objects that they are given in too narrow a way. They see the tacks and matches as potentially useful, and they try to find ways of tacking the candle to the board, or using a match to melt wax to stick the candle to the board in such a way that it tilts out from the board so it can be lit. None of those methods work. Most subjects fail to see that the box that the tacks are in could help them solve the problem. Their mental set is to see that as just a container for the tacks, not as something that can be used to help solve the problem. Karl Duncker (1945), who invented the candle problem and performed the first experiments with it, referred to this type of mental set as *functional fixedness*, the failure to see an object as having a function other than its usual one.

Research in which subjects describe their thoughts aloud as they try to solve such problems shows that the first step to insight, generally, is to realize that the method they are trying is not likely to work and that a new method must be found. This realization leads people to reexamine the objects of the problem more broadly, to notice aspects of those objects that they didn't notice before, to think more broadly about possible ways of solving the problem, and in some cases to say, "Aha!" At this point, before reading the next section, you might try again to solve the mutilated-checkerboard problem and the candle problem if you haven't already solved them.

Discovering a Solution

Deliberate attention to aspects of the problem and materials that were not noticed before can lead to the sudden perception of a solution. Transcripts from the think-aloud study with the mutilated-checkerboard task revealed thought sequences something like the following for those who solved the problem (modified, for clarity, from Kaplan & Simon, 1990):

Hmmm. . . . Okay, each domino can fit over two squares. And no matter which way the domino is placed it always fits over one black square and one red one. Maybe that's important. Same-color squares are always on a diagonal, and the domino won't cover two squares fully on a diagonal. It never fits over two squares of the same color. I need to think about color. . . . Oh! The two squares that are removed from the board are both red. That means there are two more black squares to be covered than red ones. . . . That's it! There is no way to cover all the squares with the dominos because there are two more black squares than red ones, and the dominos can never cover more squares of one color than the other. No matter how you place them there will always be two black squares that are left uncovered!

Kaplan and Simon (1990) concluded the report of their study of the mutilated-checkerboard problem by saying, "[Good problem solvers] are good noticers." Those few who solved the problem quickly, with no or few hints, were those who noticed, early on, things about the problem that others paid no attention to. In an experiment, Kaplan and Simon (1990) showed that they could make the problem easier by making the relevant clues more salient. Instead of using a standard checkerboard, they used, for one condition, a board in which the squares were labeled "bread" and "butter" in the alternating pattern shown in Figure 10.5. Because

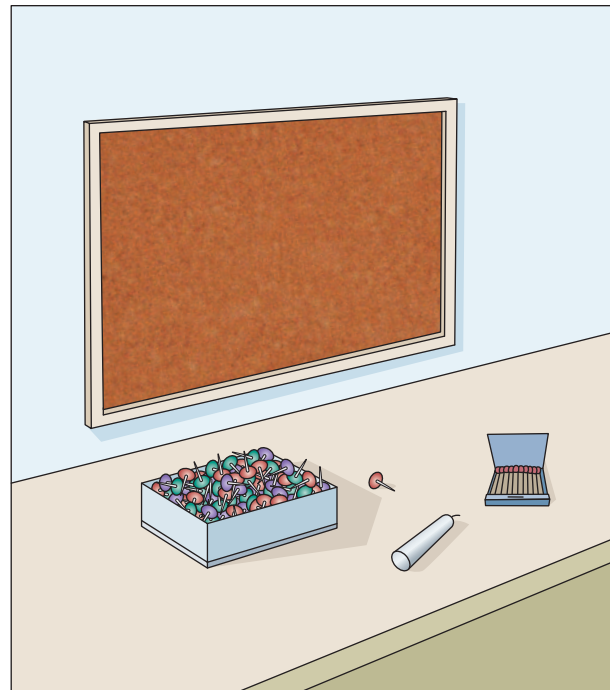


FIGURE 10.4 | The candle problem Using only the objects shown here, attach the candle to the bulletin board in such a way that the candle can be lit and will burn properly.

14 How can deliberate attempts to look at the problem materials differently lead to solution of the mutilated-checkerboard and candle problems?

bread	butter	bread	butter	bread	butter	bread	
butter	bread	butter	bread	butter	bread	butter	bread
bread	butter	bread	butter	bread	butter	bread	butter
butter	bread	butter	bread	butter	bread	butter	bread
bread	butter	bread	butter	bread	butter	bread	butter
butter	bread	butter	bread	butter	bread	butter	bread
bread	butter	bread	butter	bread	butter	bread	butter
	bread	butter	bread	butter	bread	butter	bread

FIGURE 10.5 | Variation of mutilated checkerboard The mutilated-checkerboard problem becomes easier if this bread-and-butter board is substituted for the standard checkerboard. Why?

of the unusualness of this board, subjects paid more attention to the bread and butter pattern than to the black and red pattern on a standard board, and this led them to solve the problem more quickly. Each domino covers one bread and one butter, but there are two more breads to cover than butters, so there will always be two breads that can't be covered.

With the candle problem, people might overcome functional fixedness by trying a different way of thinking about the problem. Instead of thinking, "How can I solve the problem with the materials I have?" they might step back a bit and think, "What would I need to solve this problem?" The latter question might lead them to think, "I need something like a shelf, which can be tacked to the bulletin board." This might then lead them to look again at the materials to see if there is anything that could serve as a shelf, and—Aha!—They see that the tack box will make an adequate shelf if the tacks are removed. They remove the tacks, tack the box to the bulletin board so it forms a shelf, melt some wax onto the shelf, and stick the candle to it. By thinking about the problem differently, they were able to see the analogy between the tack box and a shelf, and that solved the problem. Researchers have shown that the candle problem becomes easier if the tacks are placed next to the box rather than in it (Adamson, 1952). Seeing the tack box empty makes it more obvious

to subjects that it could be used for something other than holding tacks.

The Value of a Happy, Playful Frame of Mind

15 What evidence suggests that happiness or playfulness helps to solve insight problems? According to the "broaden-and-build" theory, how do positive emotions differ from negative emotions in their effects on perception and thought?

<<< In a series of experiments, Alice Isen and her colleagues have shown that people are better at solving insight problems if they are made to feel happy than if they are in a serious or somber mood (Isen, 2001). In one experiment, for example, college students who had just watched a comedy film were far more successful in solving the candle problem than college students who had seen either a serious film or no film (Isen & others, 1987). In another experiment, physicians were quicker to arrive at the correct diagnosis of a liver disease, based on evidence provided by the experimenter, if they had first been given a bag of candy (Estrada & others, 1997). Those who had received the candy reasoned more flexibly, took into account all the evidence more readily, and were less likely to get stuck on false leads than those who had not received candy. Other experiments have shown that a happy mood improves people's ability to see the whole picture, rather than just parts of it, in tests of visual perception, and improves performance on standard tests of creative reasoning (Fredrickson, 2003).

From the results of such research, Barbara Fredrickson (2001, 2003) has developed what she calls the *broaden-and-build theory* of positive emotions. Negative emotions, such as fear and anger, tend to narrow one's focus of perception and thought. Those emotions lead people to focus only on the specific emotion-evoking objects and to think only of routine, well-learned ways of responding. As pointed out in Chapter 6, such narrowing may be adaptive. A moment of emergency is not the right time to test creative new ideas; it is a time for action that has proven effective in the past. In contrast, according to Fredrickson, positive emotions—such as joy and interest—broaden one's scope of perception and thought and increase creativity. Positive emotions are felt when there is no imminent danger and one's immediate biological needs are relatively well satisfied. That, according to Fredrickson, is the time to think creatively and to come up with new ideas and ways of dealing with the world. From an evolutionary perspective, the building of ideas and knowledge during periods of safety and happiness is adaptive; those ideas may prove useful in satisfying future needs or preventing future emergencies.

My own interpretation (discussed more fully in Chapters 4, 11, and 12) is that it is not so much happiness per se, but the feeling of playfulness, that is conducive to

learning and creativity (Gray & Chanoff, 1984). I suspect that the comedy films and bags of candy used by Isen and her colleagues put their subjects not just in a happy frame of mind, but in a playful one. Play is a time when people regularly view objects and information in new ways. In play, a tack box can easily be something other than a box for tacks. In play, one is open to considering all the information available, not just that which would seem at first glance to be most useful. A good problem solver may be one who combines the creative spirit of play with a serious search for a solution that really works.

SECTION REVIEW

Deduction and insight contribute to problem-solving ability.	
Concrete Nature of Deductive Reasoning • Deduction is the derivation of conclusions that must be true if the premises are true. Syllogisms are classic examples of deductive-reasoning problems. • Older theories suggested that we reason by applying formal logic to the formal structure of such problems, irrespective of specific content. Newer theories recognize that we are biased toward using content knowledge even when it causes us to reason incorrectly. • Our deductive reasoning is also concrete in that we tend to construct diagrams or mental models to represent problem information and then examine them to “see” the solution.	Insight • The mutilated-checkerboard problem and the candle problem are often used to study insight, where sudden solutions come from seeing things in a new way. • Insight often derives from abandoning a mental set (a habitual way of perceiving or thinking). Often this begins with realizing that the current approach is not working. • Paying attention to aspects of the problem and materials that might otherwise be overlooked can help to lead to insight. • A happy or playful frame of mind tends to broaden the scope of perception and thought and thereby promote insight.

Effects of Culture and Language on Thought

Most studies of reasoning have been conducted in Western cultures, usually with university students as subjects. Thus, the studies may tell us more about how schooled Westerners think than about how human beings in general think. Researchers who have compared reasoning across cultures have found some interesting differences.

Some Cross-Cultural Differences in Perception and Reasoning

Responses of Unschooled Non-Westerners to Western-Style Logic Questions

Some psychologists have given standard tests of reasoning, prepared originally for Westerners, to people in non-Western cultures. A general conclusion from such research is that the way people approach such tests—their understanding of what is expected of them—is culturally dependent. Non-Westerners often find it absurd or presumptuous to respond to questions outside their realm of concrete experiences (Cole & Means, 1981; Scribner, 1977). Thus, the logic question, “If John is taller than Carl, and Carl is taller than Henry, is John taller than Henry?” is likely to elicit the polite response, “I’m sorry, but I have never met these men.” Yet the same person has no difficulty solving similar logic problems that present themselves in the course of everyday experience.

Researchers have also found that non-Westerners are more likely than Westerners to answer logic questions in practical, functional terms rather than in terms of abstract properties (Hamill, 1990). To solve classification problems, for example,

>>> 16
How do unschooled members of non-Western cultures typically perform on classification problems? Why might we conclude that differences in classification are based more on preference than on ability?

Westerners generally consider it smarter to sort things by taxonomic category than by function, but people in other cultures do not. A taxonomic category, here, is a set of things that are similar in some property or characteristic, and a functional group is a set of things that are often found together, in the real world, because of their functional relationships to one another. For instance, consider this problem: Which of the following objects does not belong with the others: ax, log, shovel, saw? The correct answer, in the eyes of Western psychologists, is *log*, because it is the only object that is not a tool. But when the Russian psychologist Alexander Luria (1971) presented this problem to unschooled Uzbekh peasants, they consistently chose *shovel* and explained their choice in functional terms: “Look, the saw and the ax, what could you do with them if you didn’t have a log? And the shovel? You just don’t need that here.” It would be hard to argue that such reasoning is any less valid than the reasoning a Westerner might give for putting all the tools together and throwing out the log.

This difference in reasoning may be one of preference more than of ability. Michael Cole and his colleagues (1971) described an attempt to test a group of Kpelle people in Nigeria for their ability to sort pictures of common objects into taxonomic groups. No matter what instructions they were given, the Kpelle persisted in sorting the pictures by function until, in frustration, the researchers asked them to sort the way stupid people do. Then they sorted by taxonomy!

An East-West Difference: Focus on Wholes Versus Parts

Richard Nisbett and his colleagues (2001) have documented a number of differences in the perception and reasoning of people in East Asian cultures, particularly in Japan and China, compared to that of people in Western cultures, particularly in North America. According to Nisbett, East Asians perceive and reason more holistically and less analytically than do Westerners. In perceptual tests, East Asians tend to focus on and remember the whole scene and the interrelationships among its objects, whereas Westerners tend to focus on and remember the more prominent individual objects of the scene as separate entities, abstracted from their background.

17 How have researchers documented a general difference between Westerners and East Asians in perception and memory? How might this difference affect reasoning?

<<< In one experiment, Japanese students at Kyoto University and American students at the University of Michigan viewed animated underwater scenes such as that depicted in Figure 10.6 (Masuda & Nisbett, 2001). Each scene included one or more large, active fish, which to the Western eye tended to dominate the scene, but also included many other objects. After each scene, the students were asked to describe fully what they had seen. The Japanese, on average, gave much more complete descriptions than did the Americans. Whereas the Americans often described just the large fish, the Japanese described also the smaller and less active creatures, the water plants, the flow of current, the bubbles rising, and other aspects of the scene. The Japanese were also much more likely than the Americans to recall the relationships among various elements of the scene. For instance, they would speak of the large fish as swimming against the water’s current, or of the frog as swimming underneath one of the water plants. Subsequently, the students were shown pictures of large fish and were asked to identify which ones they had seen in the animated scenes. Some of the fish were depicted against the same background that had existed in the original scene and some were depicted against novel backgrounds. The Japanese were better at recognizing the fish when the background was the same as the original than when it was different, whereas the Americans’ ability to recognize them was unaffected by the background. Apparently the Japanese had encoded the fish as integrated parts of the whole scene, while the Americans had encoded them as entities distinct from their background.

East Asians’ attention to background, context, and interrelationships apparently helps them to reason differently in some ways from the way Westerners do (Nisbett & others, 2001). For instance, when asked to describe why an animal or a person behaved in a certain way, East Asians more often than Americans talk about con-

textual forces that provoked or enabled the behavior. Americans, in contrast, more often talk about internal attributes of the behaving individual, such as motivation or personality. Thus, in explaining a person's success in life, East Asians might talk about the supportive family, the excellent education, the inherited wealth, or other fortunate circumstances that made such success possible, while Americans are relatively more likely to talk about the person's brilliant mind or capacity for hard work. (This East-West difference is discussed more fully in Chapter 13.)

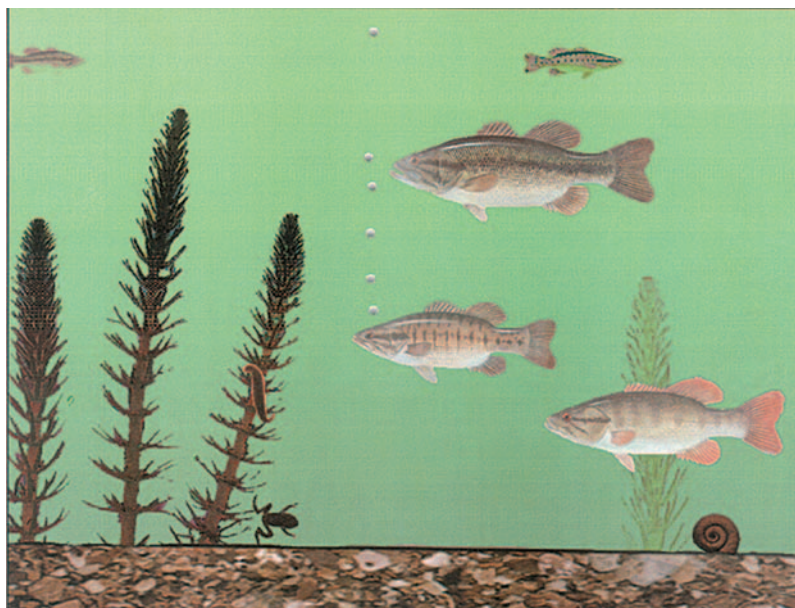
Nobody knows for sure why these differences in perception and reasoning between Westerners and East Asians came about. The difference certainly is not the result of genetic differences. The offspring of East Asians who have emigrated to North America begin to perceive and think more like other Americans than like East Asians within a generation or two (Nisbett & others, 2001). Nisbett and his colleagues believe that the roots of the difference are in ancient philosophies that underlie the two cultures. Western ways of thinking have been much influenced, historically, by ancient Greek philosophy, which emphasizes the separate, independent nature of individual entities, including individual people. East Asian ways of thinking have been much influenced by ancient Asian philosophies, such as Confucianism, which emphasize the balance, harmony, and wholeness of nature and of human society.

The Words of One's Language Can Affect One's Thinking

As you worked to solve the sample problems earlier in this chapter, you undoubtedly called forth information from long-term memory that you had originally learned through language. You may also have uttered words to yourself, subvocally, as part of your work on the problem.

We proudly conceive of ourselves as the thinking animal; but if beings from outer space were to contrast humans with other earthly creatures, they might well call us the linguistic animal. Other species can learn and use what they have learned to solve problems (discussed in Chapter 4), and they share with us a highly developed capacity for nonverbal communication (discussed in Chapter 3). But no other species has the well-developed, flexible, abstract, symbol-based mode of communication that we call *language*, which permits the conveyance of an infinite variety of facts and ideas. Language allows us to tell one another not just about the here and now but also about the past, future, far away, and hypothetical. We are effective problem solvers largely because we know so much, and we know so much because we learn not only from our own experiences but also from others' reports. As the philosopher Daniel Dennett (1994) put it, "Comparing our brains with bird brains or dolphin brains is almost beside the point, because our brains are in effect joined together into a single cognitive system that dwarfs all others. They are joined by an innovation that has invaded our brain and no others: language."

Language not only is a vehicle of communication that allows us to learn from one another but also is a vehicle of thought. To some degree, perhaps a great degree, we think with words. As the Russian psychologist Lev Vygotsky (1934/1962) pointed out, young children often speak words out loud as they think to themselves, but with time they learn to speak the words subvocally and then to use words purely mentally, in abbreviated forms that may no longer be recognized as



Taka Masuda, Ph.D.

FIGURE 10.6 | Sample scene viewed differently by Japanese and American students This is a still from one of the animated scenes used by Masuda and Nisbett (2001) to study cultural differences in perception and memory. While American students generally attended to and remembered the large "focal fish," the Japanese students generally attended to and remembered the whole scene.

words. Such thought, which uses symbols that were acquired originally in the form of words, is called *verbal thought*. Chapter 11 explains more about the nature of human language, its development in children, and its role in the development of thought. Right now our focus is on the idea that one's particular language and linguistic habits can affect the way one thinks about particular classes of problems.

18

What reasoning lies behind the idea of linguistic relativity?

<<< If language is a basis for thought, then people who speak different languages might think differently because of their different languages. This idea is most often attributed to the American linguists Edward Sapir (1941/1964) and Benjamin Whorf (1956). Sapir and Whorf both argued that language affects many of the ways that we perceive, remember, and think about the world. Whorf coined the term **linguistic relativity** to refer to all such effects. Today most psychologists dispute the strongest forms of the linguistic relativity theory, which hold that even our most basic perceptions, such as our perceptions of color, depend on language. But nobody disputes the general idea that language affects some of the ways we think (Gentner & Goldin-Meadow, 2003). Here I will describe three well-studied examples of such effects.

Effect of Spatial Terms on Spatial Reasoning

19

What is the difference between an egocentric and an absolute frame of reference? How is the use of these frames of reference dependent on language?

<<< We who speak English or other European languages regularly represent space in terms of an *egocentric frame of reference*—that is, a frame of reference that puts ourselves at the center. When looking at the objects on the table depicted in Figure 10.7, from the vantage point of the viewer shown at the left, we would say that the triangle is right of the circle, the square is left of the circle, and the oval is behind the circle. If we walked around to the other side of the table, all this would change:

Then we would say that the triangle is left of, the square is right of, and the oval is in front of the circle. Because we have the terms “right” and “left,” people growing up in our culture learn to think of locations in terms of the egocentric frame of reference.

Some languages do not have words for egocentric positions or directions. People who speak those languages regularly use an *absolute frame of reference*, even when talking about things nearby. An absolute frame is one that does not depend on point of view; it is a frame that is based on the cardinal directions, what we call north, south, east, and west. When people who speak those languages look at the table depicted in Figure 10.7, they say, in effect, that the square

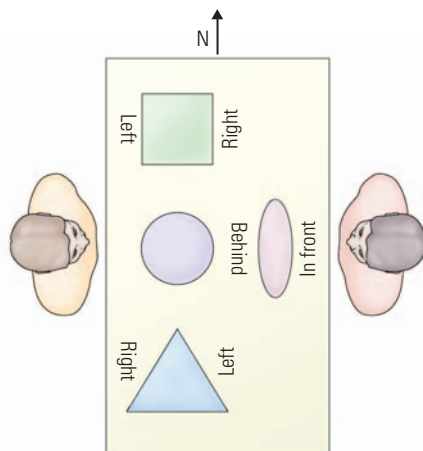


FIGURE 10.7 | Egocentric and absolute spatial relationships

When viewed from opposite sides of the table, the egocentric spatial relationships of these objects change, but the absolute spatial relationships do not.

is north of the circle, the triangle is south of the circle, and the oval is east of the circle. When they move from one side of the table to another, they do not have to change the words they use to describe the objects' positions. Because of this way of speaking, they are constantly keyed in to thinking of positions and directions in relation to the earth's coordinates, not in relation to their own momentary point of view. Reliance on the absolute frame of reference seems to have come about most often in the languages of people who make their living by hunting or fishing and must travel over broad, often unfamiliar, territories, either on land or sea, without getting lost (Levinson, 2003).

20

Why might we conclude that people who lack egocentric spatial terms develop much better absolute spatial abilities and think differently about space than do people who have such terms?

<<< Steven Levinson (1997, 2003) spent much time living with a group of Australian Aboriginal speakers of Guugu Yimithirr, which has no words for egocentric referencing. They would tell him such things as, “There is a big ant just north of your west foot” or “John is standing northeast of George and southwest of Henry.” Even when they told stories or described their dreams, they would refer to the positions of objects and directions of movements in cardinal terms. By questioning them as they went about their activities, Levinson discovered that they were always aware of the compass directions. Even when they were driven by car, along winding



Stephen Levinson of the Max-Planck Institute

A head for directions Australian speakers of Guugu Yimithirr, such as the man shown here, constantly monitor the movements and relative positions of objects in their environment in terms of an absolute frame of reference, that is, in terms of the earth's coordinates. Their language forces them to do so, as they have no words for an egocentric frame of reference.

roads, at night, they could tell him correctly, without hesitation, which way was north, any time he asked. They could also point directly to their home, from wherever they were, any time he asked. Levinson himself would have to check his compass and map to see if they were correct.

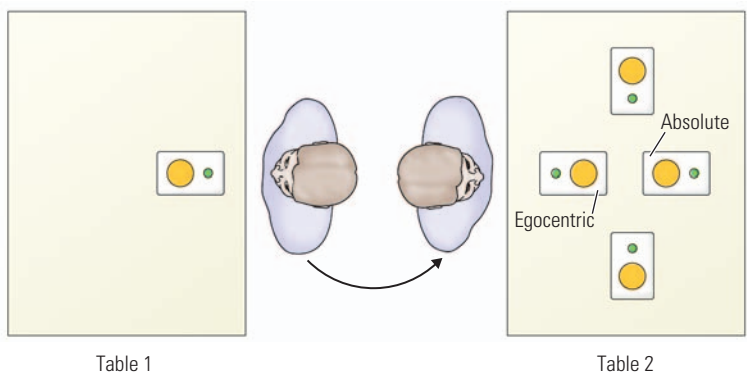
Levinson describes one case in which several Guugu Yimithirr speakers were driven 250 kilometers (155 miles) to the south from their home village to discuss land-rights issues with other Aboriginal groups. The meeting took place in a room without windows, in a building reached through a parking garage. Weeks later, without forewarning, Levinson asked each of them about the meeting room and hotel rooms they had stayed in. He found that all of them could tell him immediately the layout of the main furnishings of these rooms in terms of cardinal positions, such as, "The speaker's podium was near the middle of the southeast wall."

Levinson suggests that these people learn, beginning early in childhood, to pay attention to cues concerning cardinal direction because their language forces them to. If they did not attend to such cues, they would have no way to describe later, either to others or to themselves in verbal thought, the positions and directions of things they had experienced. Their attention to directional cues becomes so automatic that they are not very articulate in describing just which cues they use. Levinson suggests that they automatically use such cues as the positions of the sun and stars, the lichen or other plant life growing on the sides of the trunks of certain trees, the directions of the prevailing winds in particular seasons, and, when unable to see the outdoors, their own ability to keep track of the turns they have made as they move about.

People who speak languages that lack egocentric spatial terms are not only far better at knowing where they are at any given time than are those who speak European languages, but they also respond differently in certain nonverbal tests of spatial reasoning. In one experiment, for example, a group of Tzeltal speakers was compared with a group of Dutch speakers in a test using the setup depicted in Figure 10.8 (Majid & others, 2004). Tzeltal is a Mexican Mayan language that, like Guugu Yimithirr, lacks egocentric spatial terms. Each subject was first shown a sample card that contained a small green circle and a large yellow circle. Then the subject was turned around 180 degrees and looked at a table that had four test cards, all identical to the original sample but oriented in different ways. When asked to choose the card that was most similar to the one they had just seen on the other table, most of the Tzeltal speakers chose the card that was oriented identically to the original card in cardinal direction (the "absolute" choice in Figure 10.8)

FIGURE 10.8 Test of egocentric versus absolute spatial reasoning

After viewing the card on Table 1, the subject was asked to turn around and pick the card on Table 2 that is most similar to that on Table 1. Which would you have picked? Most Tzeltal speakers chose that which was identical to the original in its absolute spatial orientation, and most Dutch speakers chose that which was identical in egocentric orientation. (Adapted from Levinson, 2003, Figure 4.13, p. 160.)



and nearly all the Dutch speakers chose the card that was oriented in the same direction relative to their own bodies as the original card (the “egocentric” choice in the figure). According to Levinson and his colleagues, results such as these suggest strongly that language affects the way people think about space, not just the way they talk about it (Levinson, 2003; Majid & others, 2004). People who speak English, Dutch, or other languages that have egocentric spatial terms tend to think of space as it relates to themselves; people who speak Guugu Yimithirr, Tzeltal, or other languages that lack egocentric spatial terms tend always to think of space as it relates to the earth’s coordinates.

Effect of Number Words on Math Ability

The languages of many hunter-gatherer groups contain very few number words. For instance, the language of the Piraha tribe of Brazil has only three number words, which, translated to English, are *one*, *two*, and *more than two*. The Piraha do not count. In one experiment, tribe members were shown sets of nuts containing from 1 to 9 nuts. Then the researcher placed the nuts in a can and drew them out one by one while asking, as each nut was removed, if any nuts were left in the can. Most of the tribe members were correct for one, two, or three nuts, but were incorrect for larger sets (Gordon, 2004). If the Piraha wanted to keep exact counts of things or to think mathematically, they would have to start by inventing or learning number words.

A more subtle linguistic effect on numerical reasoning may lie in the comparison of people who speak English or certain other European languages with those who speak Asian languages such as Chinese, Japanese, and Korean. Asian children greatly outperform American and European children in mathematics at every stage of their schooling (Miller & others, 1995). Usually that effect is attributed to differences in how mathematics is taught, but some researchers have argued that the difference may stem at least partly from a difference in language.

21

How might the number words of Asian languages help children learn the base-10 number system more easily than American and European children? What evidence suggests that they do learn the base-10 system earlier?

<<< In English and other European languages the number words do not precisely mirror the base-10 number system that is used in all of arithmetic, but they do so in Chinese, Japanese, and Korean. While we count *one*, *two*, . . . , *nine*, *ten*, *eleven*, *twelve*, *thirteen*, . . . , *twenty*, *twenty-one*, . . . , the speakers of the Asian languages count (if their words were translated literally into English) *one*, *two*, . . . , *nine*, *ten*, *ten one*, *ten two*, *ten three*, . . . , *two-tens*, *two-tens one*, The words *eleven* and *twelve* give the English-speaking child no clue at all that the number system is based on groups of 10, whereas *ten one* and *ten two* make that fact abundantly clear to the Asian child. Even many English-speaking adults do not know that *teen* means “ten” (Fuson & Kwon, 1992), and children do not automatically think of *twenty* as “two tens.” Because the Asian words make the base-10 system transparent, Asian children might develop an implicit grasp of that system and thereby gain an advantage in learning arithmetic.

Consider children learning to add two-digit numbers—say, 34 plus 12. English-speaking children pronounce the problem (aloud or to themselves) as “*thirty-four*

plus *twelve*," and the words give no hint as to how to solve it. Chinese-speaking children, however, pronounce the problem (in effect) as "*three-tens four plus ten two*," and the words themselves point to the solution. In the two numbers together, there are four tens (*three tens plus one ten*) and six ones (*four plus two*), so the total is *four-tens six* (forty-six). From the perspective of Johnson-Laird's mental-model theory of reasoning, discussed earlier in the chapter, we might say that the words *three-tens four* and *ten two* lead directly to mental models that keep the tens and units separate, whereas the words *thirty-four* and *twelve* do not.

Consistent with this view, a number of experiments have shown that children who speak Asian languages have a much better implicit understanding of the base-10 system, even before they begin formal mathematics training, than do children who speak European languages. In one study, for example, 6-year-olds in the United States and in France and Sweden (where number words contain irregularities comparable to those in English) were compared with 6-year-olds in China, Japan, and Korea on a task that directly assessed their use of the base-10 system (Miura & others, 1994). All the children had recently begun first grade and had received no formal training in mathematics beyond simple counting. Each was presented with a set of white and purple blocks and was told that the white blocks represented units (ones) and the purple blocks represented tens. The experimenter explained, "Ten of these white blocks are the same as one purple block," and set out ten whites next to a purple to emphasize the equivalence. Each child was then asked to lay out sets of blocks to represent specific numbers—11, 13, 28, 30, and 42. The results were striking. The Asian children made their task easier by using the purple blocks correctly on over 80 percent of the trials, but the American and European children did so on only about 10 percent. While the typical Asian child set out four purples and two whites to represent 42, the typical American or European child laboriously counted out 42 whites. When they were subsequently asked to think of a different method to represent the numbers, most of the American and European children attempted to use the purple blocks, but made mistakes on about half the trials.

Sexism and the Generic *Man*

My final example of an effect of language on thought has to do with a revolutionary change that began to occur in English usage in the 1970s. Most of you reading this book are too young to have experienced this change firsthand, but your professor may have struggled with it at some time in his or her career. The revolution was a deliberate movement to rid English of constructions that tended to characterize males as more typically human than females. One such construction was use of the word *man* to stand for both human beings in general (the generic *man*) and human males in particular. Many reformers argued that because these two concepts had the same label, they became fused into the same schema in people's minds, leading people to think, automatically, of males as more typical examples of human beings than are females. Historically, the use of the same word to stand both for human males and for human beings in general was no accident. Historically, and still among many people today, men have been considered quite explicitly to be the primary humans and women have been considered to be secondary.

To test the degree to which *man*, used in the generic sense, elicited a mental perception of humans in general versus human males in particular, Joseph Schneider and Sally Hacker (1973) performed an experiment involving hypothetical titles for chapters of a sociology textbook. One set of titles used the generic *man* construction, which was still common in textbooks of that time: *Social Man*, *Urban Man*, *Economic Man*, and so on. Another set avoided that construction, with titles such as *Society*, *Urban Life*, and *Economic Behavior*. The researchers gave one set of titles to one group of students and the other set to a similar group and asked each student to find pictures from newspapers and magazines that would be suitable for illustrating each chapter. The result was that about 65 percent of the pictures brought

>>> 22
What logic and evidence suggest that the linguistic tradition of using the word *man* to refer to human beings in general was worth changing?

COMP: This page: first use of false Nobel-Bold ital <I> style by dummier. Please reset Usage to your custom obliqued Type 1 font.



TIME Magazine © 1997 Time Inc. Reprinted by permission.



TIME Magazine © 1999 Time Inc. Reprinted by permission.

Time catches up In recognition of the widely accepted cultural change in language, Time changed its annual designation of “man of the year” to “person of the year,” beginning in 2000.

back by students in the *man* group depicted males only (the rest contained either both genders or females only), compared with about 50 percent male-only pictures brought back by the other group. These percentages were about the same for female and male students. From this, the researchers argued that the two meanings of *man* were in fact not distinct in people’s minds. The students in this experiment knew consciously that *man* as used in those titles meant human beings in general, but their behavior suggested an unconscious mental compromise between that meaning and the male-only meaning.

Subsequent studies produced comparable results, not just for *man* but also for the use of *he* to refer to a person of unspecified gender (Harrison, 1975; Khorsroshahi, 1989; MacKay,

1980). Such research suggests that it was worth the effort for me and others of my generation to alter some of our ingrained ways of speaking and writing, which were considered proper when we were students.

SECTION REVIEW

Culture and language can affect perception, memory, and reasoning.

Cross-Cultural Differences in Perception and Reasoning

- Unschooled non-Westerners do not deal with Western-style logic problems in the same way that Westerners do; they take a more concrete, practical, and functional approach.
- Compared with analytical Westerners, people from East Asian cultures perceive more holistically—that is, in a more contextual, integrated, and relational way. This focus on wholes as opposed to parts can affect memory and reasoning.

Effects of Language

- The term *linguistic relativity* refers to differences in perceiving, remembering, and thinking that result from differences in language.
- Speakers of languages that use an absolute rather than an egocentric frame of reference think differently about spatial relations and are more able to discern and remember the absolute positions of objects.
- Compared to European languages, Asian languages use number words that more directly reflect the base-10 number system. This appears to make learning the base-10 system faster and easier for Asian children.
- The generic use of words such as *man* and *he* leads to interpretations slanted toward male examples.

The Practice and Theory of Intelligence Testing

People differ from one another in many ways. Some people are stronger, some are braver, some are more sociable, some are more dependable, some are kinder, and some are more mentally adept than others. It is that last difference that concerns us here. The variable capacity that underlies individual differences in reasoning, solving problems, and acquiring new knowledge is referred to as *intelligence*.

Psychologists have long been interested in measuring intelligence. Much of that interest derives from practical concerns. The first applied psychologists—practitioners who try to solve real-world problems using insights from psychology—were intelligence testers. School systems wanted intelligence tests to determine who could profit most from education; employers wanted them to help decide whom to hire for

jobs that require mental ability; armies wanted them to help decide how to assign recruits to ranks and tasks. Intelligence tests have long served all these functions. But psychologists have also been interested in intelligence testing for theoretical, scientific reasons. By correlating individual differences in intelligence test scores with other characteristics of people and their experiences, psychologists have aimed to understand the biological and experiential factors that contribute to intelligence.

A Brief History of Intelligence Testing

Let's begin historically. Modern ideas and controversies about intelligence and its measurement have their roots in the very different ideas of Sir Francis Galton (1822–1911), in England, and Alfred Binet (1857–1911), in France.

Francis Galton's Tests of Mental Quickness and Sensory Acuity

Galton was a wide-ranging scholar and scientist who happened to be a half-cousin of Charles Darwin. He admired Darwin's work and was interested in extending it into the realm of human mental abilities. In 1869,[OK as in list?] he published a book entitled *Hereditary Genius*, in which he argued that the capacity for high intellectual achievement is inborn and passed along biologically in family lines. In line with this idea, he defined intelligence as the biological capacity for intellectual achievement, and he initiated a research program to measure it. He hoped to show that intelligence, as he defined it, lies in certain basic characteristics of the nervous system, which manifest themselves as mental quickness and sensory acuity—the speed and accuracy with which people can detect and respond to environmental stimuli. To pursue this idea, he devised various measures of basic motor and sensory abilities, such as the ability to react quickly to a signal or to detect slight differences between two sounds, lengths, or weights.

To assess the relationship among the various measures that he devised, Galton invented a statistical method that was later refined by Karl Pearson (1920) to produce the correlational method still used today (discussed in Chapter 2). To his disappointment, using this mathematical tool, Galton found only weak correlations among his various measures of reaction time and sensory abilities (Fancher, 1985). They did not seem to be measuring a unitary underlying biological capacity.

Research by other psychologists using Galton's measures proved even more disappointing. For example, an analysis of scores collected from 300 students at Columbia University and Barnard College showed no significant correlation between either reaction time or sensory acuity and academic grades (Wissler, 1901). Because Galton's measures seemed to be only weakly related to one another and unrelated to academic achievement, most researchers in the field of intelligence lost interest in them (Fancher, 1985). As you will see later, however, in recent times some psychologists have revived Galton's approach, using more refined measures of mental quickness, and have found moderately strong correlations between such measures and other indices of intellectual ability.

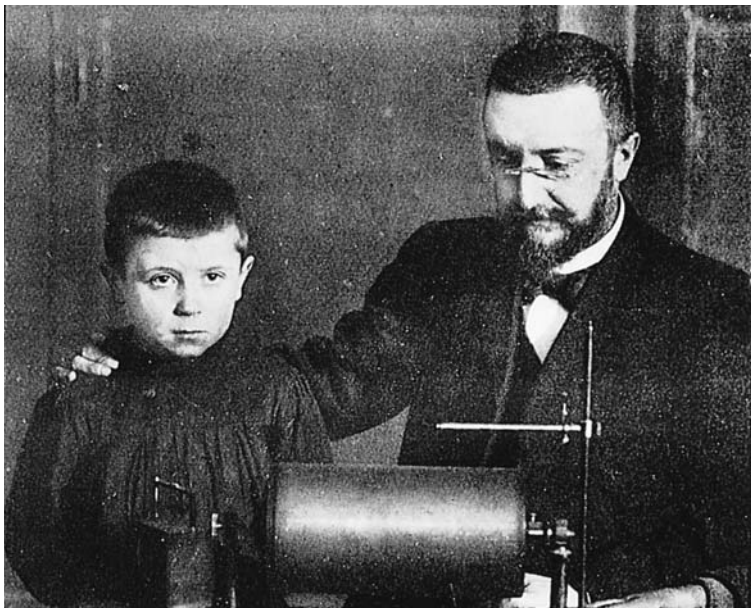
Alfred Binet's Tests of School-Related Abilities

Modern intelligence tests have their ancestry, not in Galton's measures of sensory acuity and reaction time, but in a test called the *Binet-Simon Intelligence Scale*, which was developed in France in 1905 by Alfred Binet [pronounced *Bin-ay*] (1857–1911) and his assistant Theophile Simon. Binet's view of intelligence was quite different from Galton's. He disputed Galton's claim that intelligence is closely related to sensory acuity, citing as one bit of evidence the example of Helen Keller, who had been blind and deaf since her early childhood but was universally regarded as highly intelligent (Binet & Simon, 1916/1973). Binet believed that intelligence is best understood as a collection of various higher-order mental abilities that are only loosely related to one another (Binet & Henri, 1896). He also differed from Galton on the heredity question, arguing that intelligence is nurtured through interaction with the environment and that the proper goal of schooling is

23
According to Galton, what was the biological foundation of intelligence? How did Galton attempt to measure intelligence, and how did his measures prove to be disappointing?

24
How did Binet's view of intelligence and approach to measuring it differ from Galton's? What was the purpose of Binet and Simon's intelligence test, and how did they develop the test?

Archives of the History of American Psychology, University of Akron



Alfred Binet and subject Before he was asked by the French Ministry of Education to develop a measure of intelligence, Binet experimented with various tests of motor and sensory capacities in children, using methods similar to those pioneered by Galton. The apprehensive expression on this boy's face reminds us that part of doing well on any ability test is overcoming the fear that stems from the novelty of the test situation.

to increase intelligence. In fact, the major purpose of Binet and Simon's test—developed at the request of the French Ministry of Education—was to identify children who were not profiting as much as they should from their schooling so that they might be given special attention.

Binet and Simon's test was oriented explicitly toward the skills required for schoolwork. It included questions and problems designed to test memory, vocabulary, common knowledge, use of numbers, understanding of time, and ability to combine ideas. To create the test, problems were pretested with schoolchildren of various ages, and the results were compared with teachers' ratings of each child's classroom performance (Binet & Simon, 1916/1973). Items were kept in the test only if more of the high-rated than low-rated children answered them correctly; otherwise, they were dropped. Binet was aware of the circularity of this process: His test was intended to measure intelligence better than existing measures did, but to develop it, he had to compare results with an existing measure (teachers' ratings). Yet, once developed, the test would presumably have advantages over teachers' ratings. Among other things, it would allow for comparison of children who had had different teachers or no formal schooling at all.

By 1908, the Binet-Simon test was widely used in French schools. Not long after that, English translations of the test appeared, and testing caught on in England and North America even more rapidly than it did in France.

How many objects can you name?
As part of the original Binet intelligence test, children were asked to name common household objects in this picture. Obviously, a child whose home contained such objects would do better than a child whose home did not.



From *L'Année Psychologique*, Paris, 1905. General Research Division, The New York Public Library, Astor, Lenox and Tilden Foundations

Modern Intelligence Tests Patterned After Binet's

25

What sorts of subtests make up modern IQ tests, such as the Wechsler tests, and how is IQ determined?

<<< The first intelligence test commonly used in North America was the Stanford-Binet Scale, a modification of Binet and Simon's test that was developed in 1916 at Stanford University. The Stanford-Binet Scale has been revised over the years and is still used, but the most common individually administered intelligence tests today are variations of a test that was developed by David Wechsler in the 1930s and was modeled after Binet's. The descendants of Wechsler's tests that are most widely used today are the *Wechsler Adult Intelligence Scale, Third Edition (WAIS-III)*, and the *Wechsler Intelligence Scale for Children, Third Edition (WISC-III)*.

Table 10.2 summarizes the various subtests of the WAIS-III. As indicated there, the subtests are grouped into two categories. The *verbal subtests* assess vocabulary,

TABLE 10.2 |
Subtests of the Wechsler Adult Intelligence Scale, Third Edition

Verbal subtests Figures for Chapter 10	Performance subtests
<i>Vocabulary</i> Words to be defined	<i>Block Design</i> Blocks are to be arranged to match specific designs.
<i>Similarities</i> On each trial, the person must say how two objects or concepts are alike.	<i>Picture Completion</i> The person identifies the missing elements in each picture.
<i>Information</i> Questions about generally well-known people, places, events, and objects.	<i>Picture Arrangement</i> Cartoon drawings are to be arranged in sequence to tell a logical story.
<i>Comprehension</i> Questions tapping knowledge of practical information and ability to organize that information.	<i>Object Assembly</i> Puzzle pieces are put together to form a picture.
<i>Digit Span</i> String of orally presented digits must be repeated verbatim (and, in a second phase, in reverse order).	<i>Matrix Reasoning</i> Geometric shapes that change according to some rule are presented in an incomplete grid. From a set of choices, the person selects the shape that best completes the grid.
<i>Letter-Number Sequencing</i> Letters and numbers are presented in random order and the person must repeat the numbers in ascending order followed by the letters in alphabetical order.	<i>Digit-Symbol Coding</i> The person translates a series of single-digit numbers into symbols as quickly as possible, using a code that is provided.
<i>Arithmetic</i> Arithmetic problems to be solved mentally.	<i>Symbol Search</i> On each trial a target set and search set of symbols are presented. The person says as quickly as possible whether or not a target symbol appears in the search set.

Sources: Compiled from information in A. S. Kaufman & E. O. Lichtenberger (1999), *Essentials of WAIS-III Assessment*. New York: Wiley.

ability to explain how similar concepts are similar, general knowledge and general understanding of the social and physical world, verbal short-term memory span, ability to organize verbal information quickly, and arithmetic ability. The *performance subtests* are designed to depend much less than the verbal subtests on verbal skills and already-acquired knowledge. These tests assess the abilities to match visual designs, to arrange pictures in a way that tells a story, to put puzzle pieces together to form a complete picture, to solve matrix reasoning problems (similar to the problem shown in Figure 10.1 on p. 000), to transform digits according to the rules of a code, and to spot target symbols quickly in arrays of visual symbols. The *full-scale* score is based on the sum of the scores of all the verbal and performance subtests.

The scoring system for every modern intelligence test uses results obtained from large samples of individuals who have already taken the test. These results are used as normative data to translate each individual's raw score on an intelligence test into an **IQ** (*intelligence quotient*) score. A person whose performance is exactly average is assigned an IQ score of 100. People whose performance is above or below average receive scores above or below 100, assigned in such a way that the overall distribution of scores fits the bell-shaped form known as a normal distribution, depicted in Figure 10.9. (For a more precise description of normal distributions and the method for standardizing test scores, see the Statistical Appendix at the end of the book.) For children, the comparison group for determining IQ is always a set of children who are the same age as the child being scored.



Second-grader taking the WISC IQ test In one of the subtests of the Wechsler Intelligence Scale for Children (WISC), the child must arrange blocks to match specific designs.

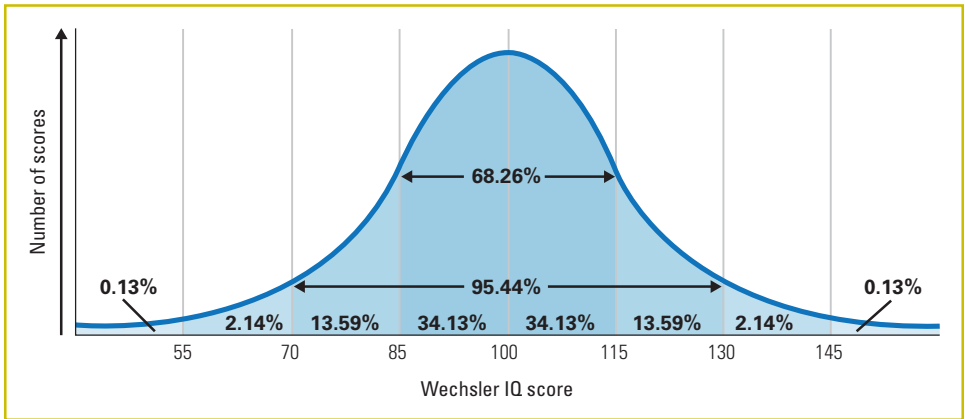
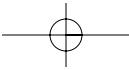


FIGURE 10.9 | Standardized scoring of Wechsler IQ tests The scoring system for Wechsler IQ tests is based on the assumption that intelligence is distributed in the population according to a normal distribution (represented by the bell-shaped curve). Thus raw scores on the test are converted into IQ scores in such a way as to produce a normal distribution. Notice that with this system about 68 percent of IQ scores are between 85 and 115 and about 95 percent are between 70 and 130.



The Validity of Intelligence Tests as Predictors of Achievement

26

How have psychologists assessed the validity of IQ tests? What are the general results of such assessments?

<<< A test is *valid* if it measures what it is intended to measure (as discussed in Chapter 2). If intelligence tests measure intellectual ability, then IQ scores should correlate with other indices of a person's intellectual ability. For the most part, researchers have considered the question of IQ validity to be directly tied to the practical question of whether or not such tests predict success in school and careers. Not surprisingly—given that most modern intelligence tests are direct or indirect descendants of Binet's and that Binet's test was explicitly designed to measure school abilities—IQ scores do correlate moderately well with grades in school; the correlation coefficients in various studies range from 0.30 to 0.70 (Brody, 1992; Jensen, 1980).

How well do IQ scores predict achievement outside of school? Many studies have shown that people with higher IQ scores are much more likely than those with lower scores to gain employment in intellectually demanding occupations such as medicine, law, science, and business management. This is true even if the comparison is just among people who were raised in the same or similar families, originating from the same socioeconomic background (Schmidt & Hunter, 2004). Still, the conclusions one can draw from such studies are limited by the fact that intellectually demanding jobs generally require high educational attainment. Thus, in theory, the relation between IQ and employment could be secondary to the fact that people with high IQs perform better in school.

A better index of the relation between IQ and career success comes from research correlating IQ with on-the-job performance as measured by supervisor's ratings, colleagues' ratings, or direct observation, for people who are already involved in a particular career. Many such studies have been conducted, and they regularly show moderate positive correlations (Schmidt & Hunter, 2004). Not surprisingly, the strength of the correlations depends on the type of job. For jobs that require relatively little judgment and reasoning—such as assembly-line work—the average correlation is about 0.2; for jobs that require a great deal of judgment and reasoning—such as scientist, accountant, and shop manager—the correlation is typically about 0.5 to 0.6 (Schmidt & Hunter, 2004). For jobs in the high mental complexity category, IQ tests are better predictors of performance than any other measures that have been developed, including measures aimed at testing specific knowledge and skills related to the job (Gottfredson, 2002; Schmidt & Hunter, 2004).

In addition to predicting school and work performance, IQ scores predict longevity. A study in Scotland revealed that people who scored high on an IQ test that was given to all 11-year-olds in the nation were significantly more likely to still be alive at age 76 than were those who scored lower on the test (Deary & Der, 2005; Gottfredson & Deary, 2004). This was true even for people who were similar in education and socioeconomic status. The numbers were such that a 15-point deficit in IQ was associated with a 21 percent decline in survival rate. The reason for longer survival, apparently, is better self-care. That study and others have shown that IQ scores correlate positively with physical fitness and healthy diets and negatively with alcoholism, smoking, obesity, and traffic accidents (Gottfredson & Deary, 2004).

The Concept of General Intelligence and Attempts to Explain It

Historically, concern with the practical uses of intelligence tests for placement in school and jobs has always been paralleled by theories and research aimed at characterizing intelligence as a basic human attribute. What is the rationale for conceiving of “intelligence” as a unitary, measurable entity? How broadly, or narrowly, should intelligence be defined? Can variations in intelligence be related to properties of the

nervous system? What is the evolutionary function of intelligence? These questions are still subject to heated debate. What I tell here, very briefly, is the most common story, the one told by those who are most centrally concerned with these questions.

General Intelligence (g)[ital] as the Common Factor Measured by Diverse Mental Tests

In the early twentieth century, the British psychologist and mathematician Charles Spearman (1927) conducted many research studies in which he gave dozens of different mental tests to people—all of whom were members of the same broad cultural group—and found that the scores always correlated positively with one another if his sample was large enough. That is, people who scored high on any one test also, on average, tended to score high on all other tests. Scores on tests of simple short-term memory span correlated positively with scores on tests of vocabulary, which correlated positively with scores on tests of visual pattern completion, which correlated positively with scores on tests of general knowledge, and so on. Most of the correlations were of moderate strength, typically in the range of 0.3 to 0.6. From this observation—coupled with a mathematical procedure that he invented, called factor analysis (discussed in Chapter 15), for analyzing patterns of correlations—Spearman concluded that some common factor is measured, more or less well, by every mental test. Spearman labeled that factor *g*, for *general intelligence*. To Spearman and to the many researchers still following in his footsteps, general intelligence is the underlying ability that contributes to a person's performance on all mental tests.

To Spearman and his followers, every mental test is partly a measure of *g* and partly a measure of some more specific ability that is unique to that test. From this point of view, the best measures of *g* are derived from averaging the scores on many diverse mental tests. That is exactly the logic that lies behind the use of many very different subtests, in standard intelligence tests such as the WISC–III, to determine the full-scale IQ score (look back at Table 10.2).

The Distinction Between Fluid Intelligence and Crystallized Intelligence

Raymond Cattell was a student and research associate of Spearman in England, until he moved to the United States in 1937. He devoted most of his subsequent long career to developing a theory of personality (described in Chapter 15), but he also developed a new theory of intelligence, which he considered to be a modification of Spearman's theory. Cattell (1943, 1971) agreed with Spearman that scores on mental tests reflect a combination of general intelligence and a specific factor that varies from test to test, but he contended that general intelligence itself is not one factor but two. More specifically, he proposed that Spearman's *g* is divisible into two separate *g*'s. He called one of them *fluid intelligence*, abbreviated *g_f*, and the other *crystallized intelligence*, abbreviated *g_c*.

Fluid intelligence, as defined by Cattell (1971), is the ability to perceive relationships among stimuli independent of previous specific practice or instruction concerning those relationships. It is best measured by tests in which people identify similarities or lawful differences between stimulus items that they have never previously experienced or between items so common that everyone in the tested population would have experienced them. An example of the first type is *Raven's Progressive Matrices Test*, discussed earlier and illustrated in Figure 10.1 (on page 000). Examples of the second type are verbal analogy problems constructed only from common words, which essentially all speakers of the language would know—such as the two problems presented back on page 000.

A verbal analogy problem that would *not* be a good measure of fluid intelligence is the following (modified from Herrnstein & Murray, 1994):

RUNNER is to MARATHON as OARSMAN is to (a) boat, (b) regatta, (c) fleet, or (d) tournament.

27
What was Spearman's concept of general intelligence, or *g*? Why did Spearman think that *g* is best measured with a battery of tests rather than with any single test?

28
What evidence led Cattell to distinguish between fluid intelligence and crystallized intelligence?

Solving this problem is limited not just by ability to perceive relationships but also by knowledge of uncommon words (*marathon* and *regatta*), which reflects crystallized intelligence.

Crystallized intelligence, according to Cattell (1971), is mental ability derived directly from previous experience. It is best assessed in tests of knowledge, such as knowledge of word meanings, of cultural practices, and of how to use particular tools. Although people may differ in the domains of their knowledge (one person may know a lot of words but little about tools, for example), Cattell considered crystallized intelligence to be a component of general intelligence. One's accumulated knowledge can be applied broadly to solve a wide variety of problems.

Like Spearman, Cattell based his theory largely on the factor analysis of scores on many different mental tests. Cattell's analysis showed him that mental tests tend to fall into two clusters: those that seem to depend mostly on raw reasoning ability and those that seem to depend mostly on previously learned information. Test scores within each cluster correlate more strongly with one another than with scores in the other cluster (as illustration, see Figure 10.10). In addition, Cattell (1971) found that measures of fluid and crystallized intelligence behave differently as a function of age. Fluid ability typically peaks at about age 20 to 25 and declines gradually after that, while crystallized ability typically continues to increase until about age 50 or even later.

FIGURE 10.10 | Hypothetical correlations among test scores, suggestive of two underlying intelligences Each coefficient in the matrix is the correlation between the two tests indicated by its row and column. Thus, 0.35 is the correlation between test 1 and test 2. All the correlations are positive. Notice, however, that the correlations among tests 1, 3, and 5 (in gold) and among tests 2, 4, and 6 (in purple) are higher than any of the other correlations. This pattern of correlations suggests that the tests measure two different but somewhat overlapping abilities. Tests 1, 3, and 5 are the best measures of one ability, and tests 2, 4, and 6 are the best measures of the other. This result could be taken as support for Cattell's theory if the items in one cluster of tests seem to measure raw reasoning (fluid intelligence) and those in the other seem to measure learned information (crystallized intelligence).

Tests	2	3	4	5	6
1	0.35	0.62	0.40	0.59	0.45
2	–	0.41	0.55	0.39	0.64
3	–	–	0.28	0.63	0.30
4	–	–	–	0.34	0.60
5	–	–	–	–	0.38

Many research studies have corroborated Cattell's conclusions about the differences between fluid and crystallized intelligence in their variation with age (McArdle & others, 2002). Figure 10.11 shows the combined results from many studies comparing people of different ages on four mental tests (Salthouse, 2004). As you can see in the figure, vocabulary (ability to identify synonyms) increases steadily until the mid-50s and then levels off or decreases slightly. In contrast, ability to solve Raven's matrix problems decreases steadily throughout adulthood. The figure also shows that short-term-memory span and mental speed decline in a manner that is essentially identical to the decline in matrix reasoning.

The evidence from the analysis of correlation patterns, and the differing effects of age, led Cattell to argue that fluid and crystallized intelligences are distinct from each other. He did not, however, think that they are entirely independent. He noted that within any given age group crystallized- and fluid-intelligence scores correlate positively. This, he suggested, is because people with higher fluid intelligence learn and remember more from their experiences than do people with lower fluid intelligence. In that sense, he claimed, crystallized intelligence depends on fluid intelligence.

Mental Speed as a Possible Basis for g

29 What findings have revived Galton's idea of mental quickness as a basis for general intelligence?

<<< As noted earlier, Galton had hoped to characterize intelligence in terms of basic properties of the nervous system, measurable in simple behavioral tests such as speed of reacting to stimuli. In recent times, researchers have revived this quest,

using more sophisticated measures and measuring devices, and have found positive correlations between a number of simple reaction-time measures and scores on conventional IQ tests (Nettelbeck, 2003; Vernon & others, 2000).

One such measure is *inspection time*—the minimal time that subjects need to look at or listen to a pair of stimuli to detect the difference between them. In one common test of inspection time, two parallel lines, one of which is 1.4 times as long as the other, are flashed on a screen and subjects must say which line is longer. The duration of the stimulus varies from trial to trial, and inspection time is the shortest duration at which a subject can respond correctly most of the time. Studies correlating fast inspection time with IQ scores have typically revealed correlation coefficients of 0.3 to 0.5 for measures of fluid intelligence and coefficients of about 0.2 for measures of crystallized intelligence (Deary, 2000; Luciano & others, 2003). The correlations are about as strong when untimed IQ tests are used as when timed tests are used (Vernon & Kantor, 1986), so they do not simply result from the fact that the subtests of some standard IQ tests require quick reaction.

How might mental speed contribute to general intelligence? One view is that it does so by contributing to the capacity of working memory.

Working-Memory Capacity as a Possible Basis for *g*

According to the standard information-processing model of the mind, described at length in Chapter 9, working memory is the center of conscious thought (look back at Figure 9.1, page 000). Information enters working memory from the environment (through the senses) and from one's long-term-memory store (through the act of retrieval). To solve a problem on an intelligence test, according to this model, you must bring the appropriate information from the senses and from long-term memory into the working-memory store and combine the information in an appropriate way to come up with a solution. But working memory can hold only a limited amount of information at any given time. While it is the center of conscious thought, it is also a bottleneck that limits the amount of information you can bring together to solve a problem. Information fades quickly from working memory when it is not being acted upon, so the faster you can process information, the more items you can bring together to make a mental calculation or arrive at a reasoned decision. This, according to some psychologists, is why mental speed correlates positively with fluid intelligence (Miller & Vernon, 1992).

Consistent with this interpretation, researchers have found moderately strong correlations between mental speed and working-memory capacity, and between working-memory capacity and fluid intelligence, as measured by such tests as Raven's matrices (Miller & Vernon, 1992; Conway & others, 2003). Working-memory capacity is measured in two quite different ways. One type of measure is referred to as *digit span* or *word span*, the number of single-digit numbers or unrelated words that a person can hold in mind and report back accurately after hearing the list just once (discussed in Chapter 9). That type of measure, as noted in Table 10.2, is one of the subtests of the Wechsler IQ tests. The other type of measure of working memory, more commonly used in recent research, is more complicated and is exemplified by a test called the *three-back test*. In this test, the person

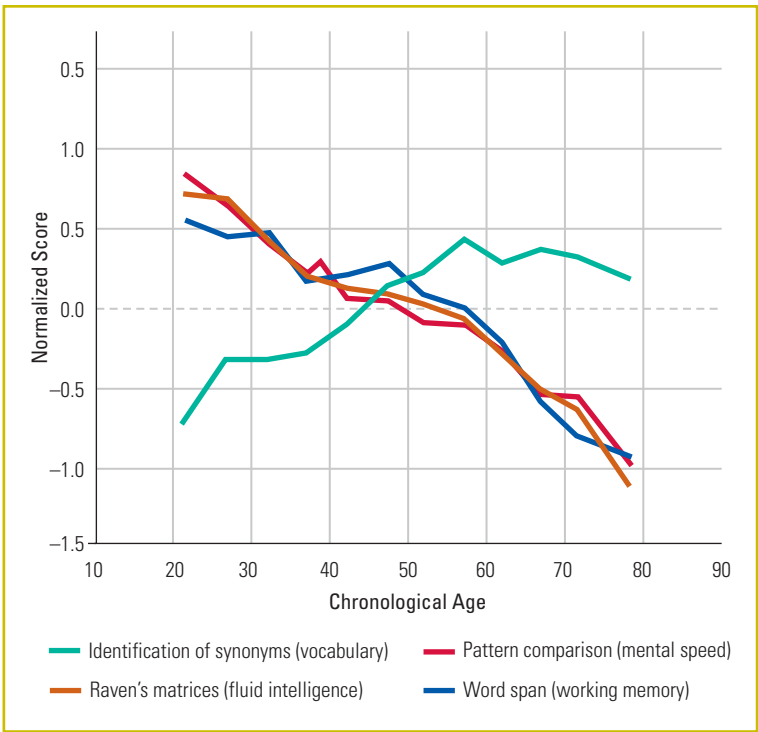


FIGURE 10.11 | Average scores on four mental tests as a function of age The scores on all of the tests are normalized such that 0 is average for all subjects combined. A score of +1 or more would place a person in the top 84 percent of the population, and a score of -1 or less would place one in the bottom 16 percent of the population. (Adapted from Salthouse, 2004, Figure 1, p. 141.)

30 How might working-memory capacity provide a basis for individual differences in intelligence, and how might mental quickness affect that capacity? What evidence supports this logical possibility?

is presented visually with words or letters in a steady stream and must indicate, on each trial, whether the current stimulus does or does not match the stimulus that had been shown three items earlier. Tests of this type require a rather high degree of mental gymnastics. The person must keep ever-changing previous items in mind while reading new ones, remembering how far back each previous item occurred, and making the appropriate item comparisons. Scores on tests of this type generally correlate more strongly with measures of fluid intelligence than do digit-span or word-span tests (Conway & others, 2003; Engle, 2002).

Mental Self-Government as a Possible Basis for *g*

31 How might the mind's central executive provide a foundation for individual differences in general intelligence?

<<< One leading intelligence researcher, Robert Sternberg (1986), has described intelligence as “mental self-government.” By this he means that people who perform well on intelligence tests are those who can control their mental resources in a way that allows for efficiency in problem solving. In a similar vein, other researchers have equated intelligence with the mind's *central executive*, conceived of as the set of mental processes that deals with overall goals and strategies and that coordinates all the rest of the mind's activities. As was noted in Chapter 9, the central executive is considered to be the supervisory component of working memory—the component that decides what information to attend to or to recall from long-term memory and that keeps the mind as a whole focused on the task at hand. The three-back test of working memory, and other tests like it, are considered to be at least partly tests of central executive function (Engle, 2002). To perform well, one must remain focused, avoid distraction, and distinguish between relevant and irrelevant information. These skills also seem to be involved in most if not all IQ tests and subtests, especially those that tap fluid intelligence more than crystallized intelligence. Some psychologists, in fact, consider fluid intelligence and central executive function to be essentially the same concept (Kane & Engle, 2002; Gray & others, 2003).

Central executive processes are involved not just in rapid, moment-to-moment judgments but also in judgments having to do with overall problem-solving strategy. Sternberg (1985) found that people who performed well on logical-reasoning problems spent relatively more time encoding the problem (thinking about the meaning of each statement in it), and relatively less time on subsequent steps, than did people who performed poorly. This is not a particularly surprising result. Understanding the problem fully is key to solving almost any problem. I have observed that students often make mistakes, on tests of all sorts, because they fail to read the questions carefully enough to know exactly what is being asked before trying to answer them. According to Sternberg, deciding how much time to devote to each step of a problem is part of mental self-government. Through practice in solving problems, people can improve their ability to make such judgments and, in doing so, can raise their intelligence.

The Role of the Prefrontal Cortex in Mental Self-Government and Fluid Intelligence

32 What is some evidence that activities of the prefrontal cortex are central to mental self-government and fluid intelligence?

<<< The mind's central executive appears to be largely a function of prefrontal portions of the cerebral cortex. As was noted in Chapter 9, the prefrontal cortex exerts top-down control over activities in other cortical areas. People who have suffered damage to the prefrontal cortex have difficulty organizing their thoughts and maintaining focus on intellectually demanding tasks, and they show a much greater decline in measures of fluid intelligence than in measures of crystallized intelligence (Kane & Engle, 2002). In neuroimaging studies, prefrontal areas show the greatest increase in activity during demanding tests of fluid intelligence and on the most demanding trials of the three-back test of working memory (Duncan & others, 2000; Gray & others, 2003). Other studies have revealed that the size of the prefrontal cortex correlates more strongly with IQ than does the size of any other cortical area, and the decline in fluid intelligence with age generally coincides with deterioration in this area of the brain (Schretlen & others, 2000). The prefrontal cortex is also the part of the brain that underwent the greatest increase in size dur-

ing human evolution. As a percentage of total cortical surface area, the human prefrontal cortex is nearly twice the size of that of our closest animal relative, the chimpanzee (Gray & others, 2003).

General Intelligence as an Evolutionary Adaptation for Novelty

Sometimes general intelligence is equated with a general ability to cope adaptively with one's environment (Snyderman & Rothman, 1987). In the opinion of most intelligence researchers, however, that is much too broad a definition. Cockroaches cope very well with their environment, but they perform poorly on IQ tests. All basic human capacities, like the basic capacities of any other species, came about through natural selection because they helped individuals adapt to the prevailing conditions of life. For example, human emotionality and sociability are valuable, evolved characteristics that help us survive in our social environment (as discussed in Chapters 6 and 12–15). People vary in measures of social and emotional competence, and these measures generally do not correlate reliably with measures of either fluid or crystallized intelligence (Kanazawa, 2004). You can be intellectually brilliant but emotionally and socially incompetent, or vice versa.

As suggested recently by Satoshi Kanazawa (2004), general intelligence seems to have evolved specifically as a means of solving problems that are evolutionarily novel. People, more than other creatures, are capable of dealing effectively with a wide range of environmental conditions, including conditions that were never regularly faced by our evolutionary ancestors. Our capacity to see analogies, to reason inductively, to deduce the logical consequences of statements, to achieve creative insights, and to predict and plan for future events all help us cope with the novelties of life and allow us to create and thrive in new environments.

>>> 33
What reasoning suggests that general intelligence is an adaptation for dealing with evolutionarily novel problems?

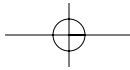


Kennan Ward / Corbis

Two intelligent species Chimpanzees and, to a much greater degree, human beings have the capacity to solve novel problems—problems that were not regularly posed by the environment in which the species evolved. Some theorists consider this capacity to be the essence of intelligence.

SECTION REVIEW

Efforts to characterize and measure intelligence have both practical and theoretical goals.		
History of Intelligence Testing - Galton proposed that intelligence reflects the biological capacity for mental quickness and sensory acuity. - Binet regarded intelligence as a loose set of higher-order mental abilities that can be increased by schooling. His tests used school-related questions and problems. - Most modern intelligence tests are rooted in Binet's approach and use a variety of verbal and nonverbal subtests.	Validity of Intelligence Testing - To gauge the validity of intelligence tests, researchers examine how well they predict school and career success. - IQ scores correlate moderately well with school grades and with job performance, especially for jobs that require considerable judgment and reasoning.	Nature of General Intelligence - Spearman proposed that general intelligence or <i>g</i> is a single factor that contributes to all types of mental performance. - Cattell believed that <i>g</i> consists of two factors rather than one—fluid and crystallized intelligence. - Modern measures of mental quickness show that it is related to general intelligence, possibly through effects on working-memory capacity. - Sternberg proposed that the efficiency of mental self-government accounts for individual differences in intelligence. - The prefrontal cortex plays a primary role in mental self-government and fluid intelligence. - General intelligence may have been selected for in nature because it helps us deal with evolutionarily novel problems.



AP / Wide World Photos

Like parents, like child Eminent people often pass on more than genes to their offspring. Anthropologist Richard Leakey, shown here discussing a fossil with his father, has benefited from the genes and the wisdom of both of his eminent parents—anthropologists Louis and Mary Leakey.

34

What is the difference between the absurd form of the nature-nurture question and the reasonable form? Why is one absurd and the other reasonable?

Genetic and Environmental Contributions to Intelligence

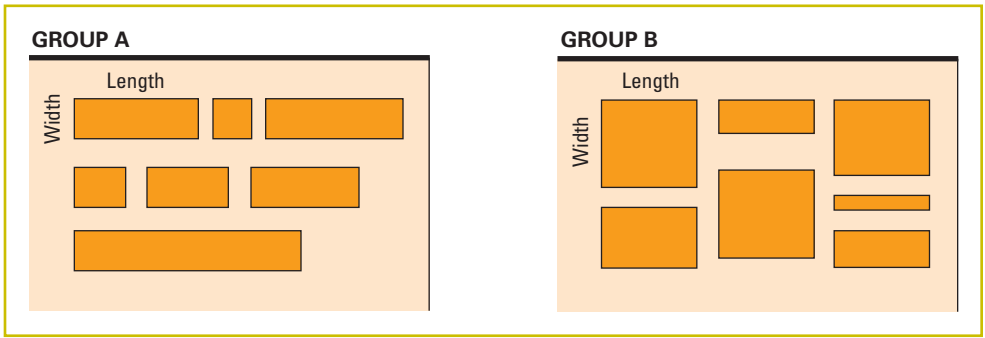
Galton's research on intelligence, over a century ago, marked the beginning of what is often called the ***nature-nurture debate***, which in various guises has continued throughout psychology's history. *Nature* here refers to a person's biological inheritance, and *nurture* refers to the entire set of environmental conditions to which the person is exposed. In essence, the question of the debate is this: *Are psychological differences among people primarily the result of differences in their genes (nature) or in their environments (nurture)?* The psychological differences that have been most often subjected to this debate are differences in personality (discussed in Chapter 15), in susceptibility to mental disorders (discussed in Chapter 16), and, especially, in intelligence as measured by IQ. As you will discover, the answer to the nature-nurture question concerning IQ is, *It depends*. It depends on just whose IQs you are comparing.

Contributions to IQ Differences Within a Cultural Group

<<< A common misunderstanding is that the nature-nurture question has to do with the degree to which a particular trait, in an individual, results from genes or environment. Some might ask, for example, "Does a person's intelligence result more from genes or from environment?" But if you think about that question, you will realize it is absurd. With no genes there would be no person and hence no intelligence, and with no environment there would also be no person and no intelligence. Given that genes and environment are both absolutely essential for any trait to develop, it would be absurd to think that one contributes more than the other to the trait. But it is not absurd to ask whether *differences* in a trait among individuals result more from *differences* in their genes or in their environments.

A useful analogy (suggested by Hebb, 1958) concerns the contribution of length and width to the areas of rectangles. It is absurd to ask whether the area of any given rectangle results more from its length or its width, because both dimensions are absolutely essential to the area. If you shrink either to zero, there is no rectangle and no area. But it is not absurd to ask whether the differences in area among a given set of rectangles result more from differences in their length or in their width. As illustrated in Figure 10.12, the answer could be length or it could be width, depending on the specific set of rectangles asked about. Similarly, when we ask whether differences in intelligence derive primarily from differences in genes or in environment, the answer might be genes for one set of people and environment for another. Logic tells us, for example, that IQ differences are likely to result more from environmental differences if the people we are studying live in very diverse environments than if they live in similar environments. If you were raised in a typical middle-class home and I were raised in a closet, the difference between us in IQ

FIGURE 10.12 A geometric analogy to heritability It is senseless to say that a given rectangle's area results more from its length than from its width, or vice versa. However, differences among rectangles in area can result primarily or entirely from differences in length (as shown in group A) or in width (as shown in group B). Similarly, one can imagine one group of people in which IQ variation results mostly from their differing genes and another group in which it results mostly from their differing environments.



would certainly be attributable mostly to differences in our environments; but if we were both raised in typical middle-class environments, then whatever difference exists between us in IQ might well result mostly from differences in our genes.

The Concepts of Heritability and Environmentality

The central concept in any modern scientific discussion of the nature-nurture question is **heritability**. Heritability is the degree to which variation in a particular trait, within a particular population of individuals, stems from genetic differences among those individuals as opposed to environmental differences. Heritability is often quantified by a statistic called the **heritability coefficient**, abbreviated h^2 . This coefficient is technically defined as the proportion of the variance of a trait, in a given population, that derives from genetic variation. As a formula, it can be written

h^2 = V_g / V_T = V_g / (V_g + V_e)

The denominator, V_T , or *total variance*, is simply a measure of the degree to which the individuals being studied differ from one another in the characteristic that was measured. It can be calculated directly from the set of scores obtained from the individuals in the group (the formula for variance can be found in the Statistical Appendix). The numerator, V_g , or *variance resulting from genetic differences*, in contrast, cannot be calculated directly but must be estimated by comparing sets of individuals who differ in known ways in their degree of genetic relationship to one another (by methods to be explained soon). The right-hand extension of the equation simply represents the fact that the total variance is equal to the variance resulting from genetic differences (V_g) plus the variance resulting from environmental differences (V_e).

Examination of the formula shows that the heritability coefficient can, in theory, vary from 0.00 to 1.00. A coefficient of 0.00 means that none of the observed variance in the characteristic results from genetic differences ($V_g = 0$); all of it results from environmental differences. A coefficient of 1.00 means that all the variance results from genetic differences; none of it results from environmental differences ($V_e = 0$). A coefficient of 0.40 means that 40 percent of the variance results from genetic differences and the remaining 60 percent results from environmental differences.

Logically, the heritability coefficient is as much a measure of the environmental contribution to variability of a trait as it is a measure of the genetic contribution. For example, a heritability coefficient of 0.40 means that 40 percent of the measured variance in the trait results from genetic variation and all the rest—60 percent—results from environmental variation. Some behavior geneticists use the term **environmentality** to refer to the proportion of the total variance of a trait that results from environmental variation. The environmentality coefficient is always 1 minus the heritability coefficient. Thus, a heritability coefficient of 0.40 corresponds to an environmentality coefficient of 0.60.

The Value of Twins in Separating Genetic and Environmental Effects

To the degree that a trait is heritable, people who are closely related to each other genetically should be more similar in the trait than are people who are less closely related genetically. A difficulty with using this logic to estimate heritability, however, is that related people also typically share similar environments. Thus, the observation that siblings are, on average, more similar to each other in IQ than are unrelated people does not, by itself, tell us the degree to which their similarity results from similarity in genes or in environment.

In an attempt to separate the contributions of genes and environment to IQ differences, behavior geneticists often compare pairs of individuals who differ from other pairs in their degree of genetic relatedness and were either raised in the same home or adopted at an early age into different homes. Twins are especially

>>> 35
How is heritability defined? Why would we expect heritability to be higher in a population that shares a similar environment than in an environmentally diverse population?

>>> 36
What is the logic of comparing identical and fraternal twins to study the heritability of traits? What difference is observed between identical and fraternal twins in IQ correlation, and how does that difference change as the pairs of twins get older?

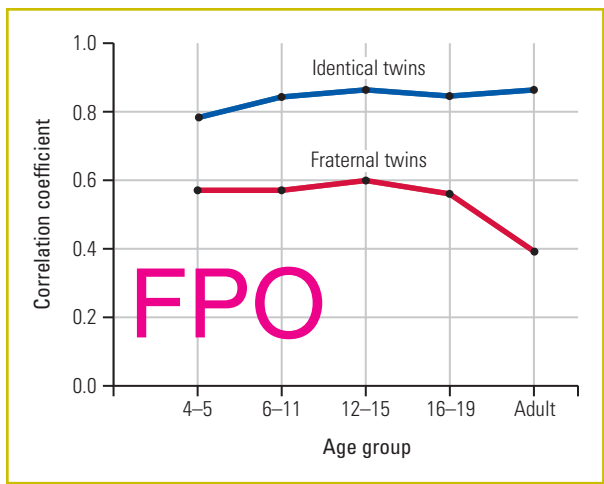


FIGURE 10.13 IQ correlations between twin pairs who were raised together The correlations in IQ between pairs of identical twins remains high throughout their lives. That between pairs fraternal twins declines as they enter adulthood and go their separate ways. The data are averages for all studies published prior to 1993. (McGue & others, 1993, p. 63.)

valuable subjects in such studies, because some pairs of twins are genetically identical and other pairs (fraternal twins) are, like non-twin siblings, only 50 percent related to each other (as explained in Chapter 3).

Galton (1876) himself pioneered the study of twins as a way to understand the contributions of heredity to psychological characteristics, and since his time many researchers have addressed the question of heritability of IQ by comparing the correlation in IQ scores of identical twins with that of fraternal twins. As explained in Chapter 2, correlation coefficients can run from 0.00 to either plus or minus 1.00. A correlation of 0.00 in this case would mean that the IQs of twins are no more alike, on average, than are those of any two people chosen at random from the population being studied; at the other extreme, a correlation of +1.00 would mean that the twins in each pair have identical IQs. In Figure 10.13, the IQ correlations between twins raised in the same home are shown, separately, for identical and fraternal twins in various age groups (McGue & others, 1993). As you can see, in each age group the correlation coefficient is considerably greater for identical twins than for fraternal twins, indicating that genes play a relatively large role. The difference between the two classes of twins is especially great for adults, because of a drop in the correlation

that occurs during adulthood for fraternal twins but not for identical twins. When fraternal twins leave home and go their separate ways, they become more different from each other in IQ than they were before, but when identical twins leave home, their IQs remain as similar as ever.

IQ Heritability Assessed from Studies of Twins

37

How can IQ heritability be estimated using the correlation coefficients for the IQs of identical and fraternal twins raised together? Using this method, what is the evidence that IQ heritability is greater for adults than for children? How can IQ heritability be estimated by studying identical twins who were adopted into different homes?

<<< Correlation coefficients are not the same as heritability coefficients, but they can be used to estimate heritability coefficients. One common method is simply to subtract the correlation for fraternal twins from that for identical twins and multiply the difference by 2 (Plomin & others, 1990). The precise logic behind this method involves more mathematical detail than most readers would desire, but you can get a sense of that logic without the mathematics. The assumption is that the environment is equally similar for the two categories of twins, so the difference between the two in IQ correlation must stem from the difference in their degree of genetic relatedness; that is, the difference must reflect heritability. Because fraternal twins are themselves 50 percent related, the difference between the two correlation coefficients is assumed to reflect only half the difference that would be observed if the comparison were between identical twins and completely unrelated individuals. Therefore, the difference is doubled to arrive at an estimate of the heritability coefficient.

If we apply this formula to the data in the figure, we arrive at a heritability coefficient of 0.46 for the 12-to-16 age group $[(0.83 - 0.60) \times 2 = 0.46]$ and 0.86 for the adults $[(0.83 - 0.40) \times 2 = 0.86]$. In other words, according to these calculations, about half of the IQ variance in the teenagers and younger children results from genetic variance, but over 80 percent of the IQ variance in the adults results from genetic variance. Apparently, the environment plays a greater role in IQ variation among younger people than it does among adults—an idea that we will return to later.

Another way to assess heritability of IQ is to locate pairs of identical twins who were adopted at an early age into separate homes and determine their IQ correlation. On the assumption that the environments of raised-apart identical twins are no more similar to each other than are the environments of any two members of the study population chosen at random, the correlation coefficient for raised-apart identical twins is itself an estimate of the heritability coefficient (Loehlin & others, 1988). The average coefficient determined from essentially all such studies published prior to 1993 is 0.73 (McGue & others, 1993). Most of the twin pairs in those

studies were adults, but a few were adolescents, and no attempt was made to determine the coefficient separately for different age groups.

Every procedure for assessing heritability involves assumptions that may not be entirely true, so any heritability coefficient should be taken as only a rough estimate. But the studies suggest, overall, that genetic differences account for about half of the IQ variance among children and adolescents and for considerably more than half of the IQ variance among adults in the populations that were studied. Contrary to what might be expected, those studies that have assessed fluid and crystallized intelligence separately indicate that the two are about equally heritable (Horn, 1985). Apparently, within the populations that have been studied, genetic differences influence the amount of factual knowledge that people have learned just as much as they influence people's ability to think quickly or to see relationships in novel stimuli. It is not hard to think of many ways by which genetic differences could have such an effect. Genes can influence curiosity, reading ability, long-term memory, or any of countless other traits that influence the acquisition and recall of facts.

The Transient Influence of the Family in Which One Is Raised

It seems reasonable to expect that the IQs of children raised in the same family would be more similar than average not just because of shared genes but also because of shared environments. After all, children in the same family are exposed to the same parents, the same neighborhoods, the same schools, and the same learning opportunities at home. But is this expectation correct?

The most direct way to assess the degree to which a shared family environment leads to similarity in IQ is to study pairs of adoptive siblings who are genetically unrelated but are raised together. Assuming that such pairs are genetically no more similar to each other than any two random people from the study population, any correlation greater than 0 in their IQs must stem from their shared environment. Several such studies have been done, and the results tell a remarkably consistent story. As long as the unrelated siblings are still children, their IQs do correlate positively with each other; but the correlation is lost completely by the time they reach adulthood. Taking all such studies together, the average IQ correlation for genetically unrelated children living in the same family is 0.25, and the average for genetically unrelated adults who had been raised in the same family is -0.01 , or essentially 0 (McGue & others, 1993). Other studies have shown that the IQ correlations for other categories of children raised in the same family also decline as the children enter adulthood, but the greater the degree of genetic relationship, the smaller the decline (Plomin & Daniels, 1987). You saw that effect in Figure 10.13, where the IQ correlation for fraternal twins declined at adulthood while that for identical twins did not. Apparently, families have a moderately strong early influence on children's IQ, but the effect fades as the children become adults.

Seeking his own path This young man's mental development may have been considerably influenced up until now by his parents and the home they provided. From now on, however, that influence will be reduced. His own dispositions will play a greater role in determining what he learns and how his mind develops. This may help explain why the heritability coefficients for IQ are greater for adults than for children.



Myrleen Ferguson Cate / PhotoEdit, Inc.

Identical twins learning together Whether they are raised together or apart, identical twins tend to exhibit similar mental abilities. Genetic dispositions to enjoy certain classes of activities may promote similarity in their development.

>>> 38 What is the evidence that the effect of a shared family environment on IQ correlations is lost in adulthood? How might this loss be explained?



David Young-Wolff / PhotoEdit, Inc.

The transient nature of the effect of the family on IQ is perhaps the most surprising result that has emerged from studies of IQ correlations. Not long ago many psychologists believed that even subtle differences in the early environments of children would give some an advantage in intellectual development that would last a lifetime. But the results of the studies just summarized indicate that the advantage or disadvantage that comes from being raised in a particular home, within the range of homes the studies sampled, disappears by early adulthood. One way to explain this finding is to assume that as children grow into adulthood, they increasingly choose their own environments, and their genetic differences influence the kinds of environments they choose (Dunn & Plomin, 1990; Scarr & McCartney, 1983). Those who are genetically similar, and therefore more similar in interests and temperament, may choose more similar environments than do those who are genetically different, and so they remain more similar in intelligence. If you think of intelligence as analogous to muscle strength, which can wax and wane depending on exercise, then you can understand why an adult's IQ may be more influenced by his or her adult environment than by his or her past, childhood environment.

39

What evidence suggests that intellectual involvement can increase a person's fluid intelligence over time?

Effects of Personality and Life Experiences on Intelligence

<<< Intelligence is maintained and strengthened through active, intellectual engagement with the world. Not surprisingly, people who score high on a personality test designed to measure openness to experience have, on average, higher IQs than do people who score lower on that personality measure (Ackerman & Heggestad, 1997; Gignac & others, 2004). Openness to experience includes the characteristics of curiosity, independence of mind, and broad interests (discussed in Chapter 15). Presumably, people who have these characteristics choose intellectually engaging styles of life, and that choice tends to raise their intelligence. Openness appears to correlate at least as strongly with measures of fluid intelligence as with measures of crystallized intelligence. Intellectual engagement apparently does not just increase one's store of knowledge, but also increases one's capacity for mental gymnastics.

More direct evidence that activities can alter intelligence is found in a long-term study, conducted by Melvin Kohn and Carmi Schooler, of the effects of men's occupations on their intellectual development. Kohn and Schooler periodically tested a large sample of men, throughout their careers, with a test of *intellectual flexibility*. The test included a number of subtests that are quite similar to those found on standard IQ tests, and it was later shown to correlate strongly with fluid intelligence as measured by standard IQ tests (Schooler, 2001). The most general finding was that the men's intellectual flexibility tended to change when their job demands changed. When they were in jobs that required them to handle a great deal of information and make complex decisions, their intellectual flexibility increased over time; when they were in routine jobs that depended more on brawn and patience than brains, their intellectual flexibility decreased over time. The researchers also observed that the job affected the ways the men spent their time at home. An intellectually challenging job led them to engage in more intellectually challenging leisure activities, and an intellectually unchallenging job had the reverse effect. Finally, the effect of an intellectually challenging job on intellectual flexibility was about twice as great for the older workers—those in the age range of 58 to 83—as it was for the younger workers (Schooler, 2001).

Here, again, the analogy between mental strength and physical strength seems to hold. Young people can maintain relatively strong muscles without much exercise, but as we get older our muscles begin to atrophy unless we increase their use. The same, apparently, is true for brain power.

Origins of IQ Differences Between Cultural Groups

The conclusions about the high heritability of IQ discussed so far in this chapter were properly qualified by the phrase “for the population that was studied.” In almost all cases that population was white, North American or European, and in the

upper two-thirds of the socioeconomic scale (Stoolmiller, 1999). Heritability coefficients are always limited to the population that was studied. The more uniform the environment of that population, the smaller is the proportion of IQ variance that stems from environmental variation and the greater is the heritability coefficient (look back at the formula on p. 000). If heritability studies included people occupying the entire range of human environments rather than just a slice of that range, the resultant heritability coefficients would be much smaller than those presented above.

Comparisons of racial or cultural groups routinely reveal average differences in IQ. The difference that has attracted the most attention, and on which we will focus here, is that between blacks and whites in the United States: Blacks, on average, score about 15 points lower than whites on standard IQ tests (Herrnstein & Murray, 1994). The question is why. Some people who have heard of the heritability studies that you have just read about assume that those studies can be applied to understand the black–white difference. They assume that if IQ is highly heritable within a group, then any IQ difference between two groups must also be largely the result of genetic differences. But that assumption is false.

The heritability of a trait within a group, in fact, tells us nothing about differences between groups. To understand why, consider the example illustrated in Figure 10.14. Imagine two wheat fields, each planted from the same package of genetically diverse wheat seeds. Imagine further that the soil fertility is relatively constant within each field but quite different between the two—one has richer topsoil than the other. Within either field, differences in the sizes of individual plants would be the result primarily of genetic differences in the seeds, yet the average difference between the two fields would almost certainly be due entirely to the difference in the environment (the richness of the soil).

To take another example, height in people is more than 90 percent heritable when measured for a given cultural group, yet group differences in height can be found that are clearly the result of the environment (Ceci, 1996). In the 1950s, for

40
Why can't heritability coefficients found within groups be used to infer the source of differences between groups?

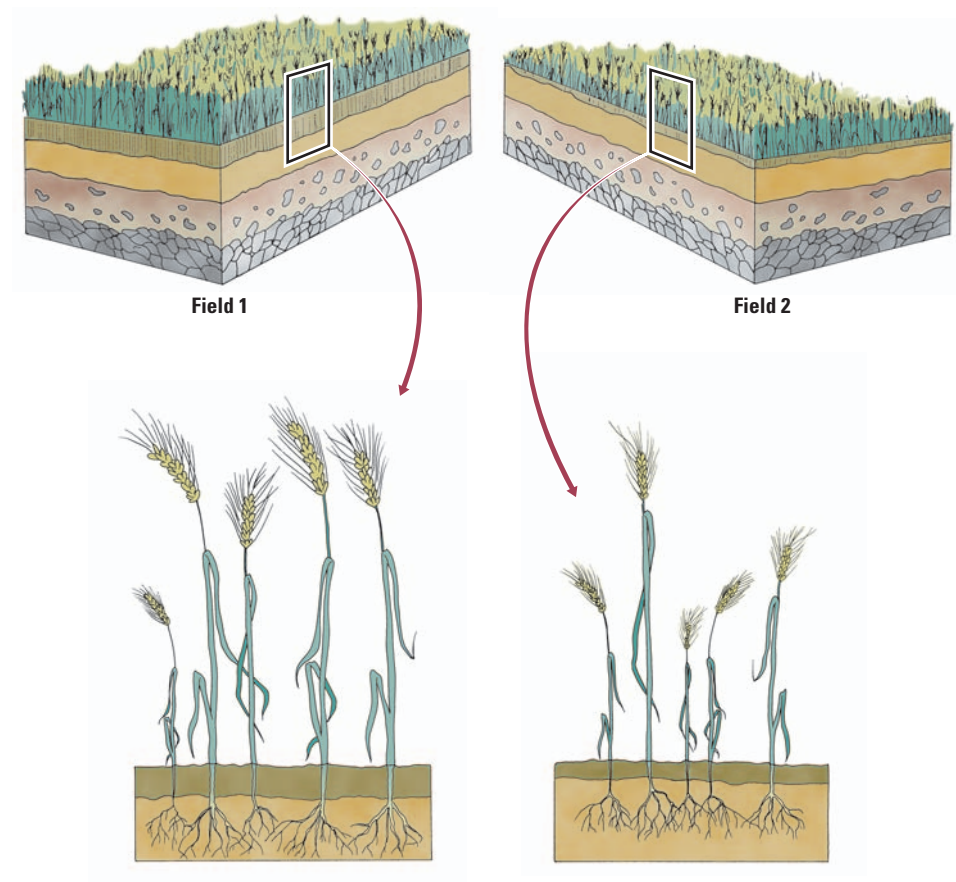


FIGURE 10.14 | Why high within-group heritability tells us nothing about group differences
The same genetically diverse mix of wheat seeds was planted in two fields. If each field is quite uniform in the environmental conditions it provides, then the differences in plant size within each must result mostly from genetic differences (high heritability). However, the difference between the two fields in average plant size in this case cannot result from genes, because genetic differences would cancel out in the averages; it must result from differences in the environment.

example, researchers found that men of full Japanese ancestry born in California were nearly 3 inches taller, on average, than Japanese men born in Japan (Greulich, 1957). That difference almost certainly resulted from differences in diet between the two groups during their childhoods.

Evidence That Black–White IQ Differences Are Cultural in Origin

In the examples just given, we could be quite certain that the group differences were environmental in origin because there was reason to believe that the members of the two groups (of wheat plants and of Japanese men) did not differ genetically, on average. In contrast, many people automatically think of differences between the so-called black and white “races” as racial differences, and they automatically assume that *racial* means “genetic.”

In the United States and many other countries, however, blacks and whites are not truly distinct races in a biological sense but, rather, are different cultural groups. We classify people as “black” who have any detectable black African ancestry, no matter how small a proportion it is. Thus, a person who is half English, one-fourth French, and one-fourth African is called black, whereas that person’s cousin, who is half English, one-fourth French, and one-fourth Polish, is called “white.” While some average genetic differences exist between the two groups—which show up in skin pigmentation, for example—the amount of genetic variation *within* each group is far greater than the average difference *between* them.

41 What evidence suggests that the average IQ difference between black and white Americans derives from the environment, not genes?

<<< Researchers who have attempted to separate the effect of black African ancestry from the effect of the social designation “black” have consistently failed to find evidence that genetic ancestry plays a role in the black–white IQ difference. In the first such study, Paul Witty and Martin Jenkins (1935) attempted to determine if high-IQ black children have more European and less African ancestry than blacks who have lower IQs. They identified a sample of black children in Chicago who had IQs in the superior range (125 or better) and then interviewed their parents to see if they had more European ancestry than the average black person. The results were negative. The proportion of European ancestry in the high-IQ black children was neither more nor less than that in the black population at large. (The highest IQ of all in that study, incidentally, was a whopping 200 scored by a girl with 100 percent black African ancestry.) More recently, other researchers have performed similar studies, using modern biochemical methods to determine the degree of black African and European ancestry (Loehlin & others, 1973; Scarr & Carter-Saltzman, 1983). Like Witty and Jenkins, they found no relationship between ancestry and IQ, and they concluded that the *social* designation of black or white, not biological ancestry, is most likely the critical variable in determining the black–white IQ difference.

Different Types of Minority Status Can Have Different Effects on IQ

42 What evidence suggests that the status of being an involuntary minority may be particularly detrimental to IQ development?

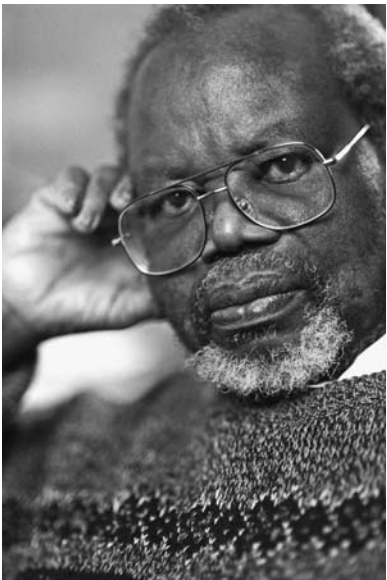
<<< How might the social designation of black or white affect IQ? Nobody knows for sure, but John Ogbu (1986; Ogbu & Stern, 2001) has suggested an interesting line of thought about it. On the basis of cross-cultural research, Ogbu distinguishes between *voluntary minorities* and *involuntary* or *castelike minorities*. Voluntary minorities are groups, such as Italian Americans and Chinese Americans, who emigrated in hopes of bettering themselves, who typically see themselves as well-off compared with those they left behind, and who see themselves as on their way up, regardless of how the dominant majority may see them. Involuntary minorities are groups, such as African Americans and Native Americans, who became minorities through being conquered, colonized, or enslaved. They are people who for long periods were, and in many ways still are, treated as if they are a separate, inferior class. According to research summarized by Ogbu, involuntary minorities everywhere perform more poorly in school, and score an average of 10 to 15 points lower on IQ tests, than the dominant majority.



Tom Wagner / SABA

A Buraku protest The Buraku people of Japan are descendants of people who worked as tanners and butchers, jobs that were traditionally believed to be unclean and worthy only of lowly people in Japan. Although they are no longer legally categorized as outcasts, they continue to be discriminated against. Like outcast groups elsewhere, they perform more poorly in school and have lower average IQ scores than other citizens of their nation. Here a group of Buraku activists protest the continuing discrimination.

Particularly informative, in Ogbu's work, is the comparison of the Buraku outcasts of Japan with blacks of the United States. The Buraku, who are a purely cultural class, not racially distinct from other Japanese, were emancipated from official outcast status by a royal edict in 1871, just 8 years after blacks in the United States were emancipated from slavery; yet both groups, to this day, often occupy menial positions and are implicitly, if not explicitly, perceived as inferior by many members of the dominant majority. The gap in school achievement and IQ between the Buraku and the majority group in Japan is about the same as that between blacks and whites in the United States; but the gap disappears when Buraku move to the United States. Most people in the United States do not know the difference between Buraku and other Japanese, and the two groups of immigrants are treated the same and perform equally well in school and on IQ tests. According to Ogbu, it is the sense that one is an outcast, and that standard routes to achievement are cut off, that oppresses caste-like minorities and depresses their scholastic achievements and IQs.



Chris Duffey

John Uzo Ogbu Ogbu was born to nonliterate farmers in a village in Nigeria, but emigrated as a young man to the United States and eventually became a renowned professor of anthropology at the University of California, Berkeley. Some of his ideas about intelligence are exemplified by his own experience. He was a voluntary minority in the United States, not an involuntary minority. Ogbu once wrote, of himself, "My intelligence changed when I moved from a village in Nigeria to a major U. S. university" (Ogbu & Stern, 2001). Ogbu died of a heart attack in 2003 at age 64.

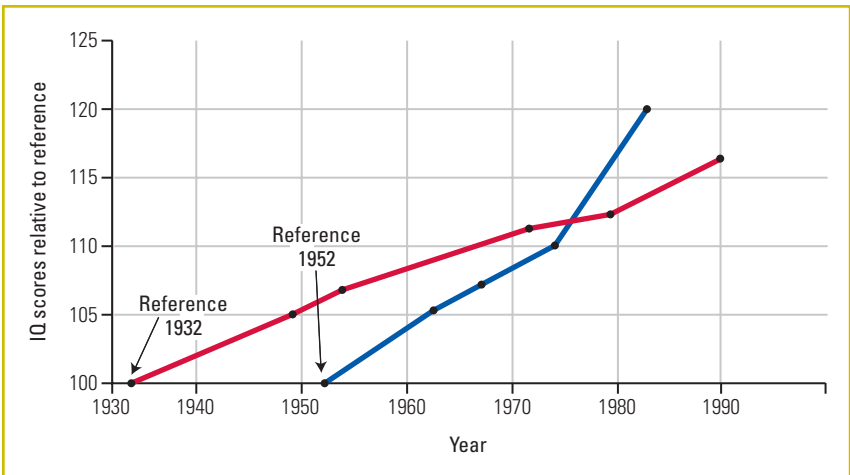
The Historical Increase in IQ

Perhaps the most dramatic evidence of cultural influence on intelligence is the improved performance on intelligence tests that has been observed worldwide, over the years since they were first invented. As you know, IQ tests are literally graded on a curve, with the average score for the population at any given time in history assigned a value of 100. But the average score keeps rising, indicating that the tests become easier for each successive generation, so researchers periodically modify the scoring system and increase the difficulty of the questions.

James Flynn (1987, 1999, 2003) has compiled data on norm adjustments for many different countries, from the dawn of IQ testing to present times, and has concluded that the increase in IQ has occurred at a rather steady rate of about 9 to 15 points every 30 years, depending on the type of test (for examples, see Figure 10.15). It has occurred for people of all races and ethnicities and has occurred in countries as varied as the United States, Belgium, Argentina, and Kenya (Daley & others, 2003[ref missing]; Flynn, 2003). The greatest increases, interestingly, are in

>>> **43** How does history provide further evidence that IQ is highly susceptible to cultural influence? On which measures has IQ increased the most?

FIGURE 10.15 Examples of the rise in intelligence-test scores The red line shows the mean full-scale IQ scores that Americans would have obtained on Stanford-Binet and Wechsler IQ tests each year if the standard that was used in 1932 had remained unchanged. The blue line shows the mean IQ that 18-year-old Dutch males would have obtained each year on Raven's Progressive Matrices—a test of fluid intelligence—if the standard that was used in 1952 had remained unchanged. (Adapted from Neisser, 1998, p. 14, Figure 2.)



the tests geared toward fluid intelligence, such as Raven's Progressive Matrices—the very tests that were originally conceived of as least affected by cultural experience and most indicative of raw reasoning ability.

What accounts for these massive gains in IQ from one generation to the next? Flynn (1999, 2003) argues against the idea that increased or improved schooling has much to do with it. He points out that the tests that are most reflective of school learning—such as the Arithmetic, Information, and Vocabulary subtests of Wechsler's IQ tests—have shown the *least* improvement. It seems likely that most of the improvement comes instead from a variety of broad cultural changes. Increased travel, increased access to information and ideas (through television and other media), and increased use of technologically complex tools (such as computers) have led people to live in increasingly rich and varied environments that require and promote abstract thinking (Greenfield, 1998). With each generation, people throughout the world are exposed to a wider variety of information, ideas, and problems, which provide more exercise for the mental capacity to deal with such stimuli. In addition, improved prenatal care and nutrition have almost certainly fostered healthier brain development (Sigman & Whaley, 1998).

SECTION REVIEW

We must be careful in how we ask and how we try to answer nature-nurture questions.

IQ Differences Within a Cultural Group

- The reasonable version of the nature-nurture question asks whether genetic or environmental variation contributes more to observed IQ differences within a population.
- Heritability is the extent to which variation in a trait (e.g., IQ), within a particular population, derives from genetic differences among the individuals.
- Twin studies have shown that, within a population, genetic variation accounts for about half of IQ variance in children and for more than that in adults.
- Effects of the shared family environment (aspects of the environment shared by children growing up in the same home) on IQ are temporary; they disappear in adulthood.
- Evidence suggests that intellectual involvement increases fluid intelligence over time, and the reverse is also true.

IQ Differences Between Cultural Groups

- Heritability coefficients for IQ within groups cannot be legitimately used to explain the source of average IQ differences between groups (e.g., racial or cultural groups).
- The average black–white IQ difference found in the United States is related to the social designation of black or white rather than to the degree of African or European ancestry.
- Involuntary minority status is particularly likely to reduce a group's IQ.
- Historical increases in IQ also suggest the strong influence of cultural factors—in this case, factors such as changes in technology and prenatal care.

Concluding Thoughts

You might find it useful to focus your review of this chapter around the following four themes:

1. The concrete nature of human reasoning Mathematics, which is formal reasoning, plays a relatively small role in the everyday reasoning of most of us. Normally we reason by comparing the current problem situation with memories that seem relevant to that situation. You might keep this idea in mind, as a unifying theme, as you review the discussions of (a) reasoning by analogy; (b) biases in inductive reasoning; (c) the content bias and use of mental models in deductive reasoning; and (d) the role of mental sets, and overcoming mental sets, in solving insight problems.

2. Cultural and other environmental influences on reasoning Consistent with the idea that people reason concretely is the idea that people whose concrete experiences are quite different from each other are likely to develop different reasoning abilities. In this chapter you read about (a) cross-cultural differences in responses to classification problems; (b) East-West differences in holistic versus analytic perception and thought; (c) evidence that scores on IQ tests vary across cultures and have been increasing over time as a result of cultural changes; and (d) evidence that a person's IQ can increase or decrease over time, depending on the work environment. You have also read of evidence that the words of one's language, which one acquires from one's culture, can influence spatial reasoning, mathematical reasoning, and one's concept of gender.

3. The intelligence concept derives from correlations Research on intelligence makes heavy use of correlational

methods. If two tests or measures correlate positively with each other—that is, if people who tend to score high, or low, on one also tend to score high, or low, on the other—then, to some degree at least, they can be considered to be measures of the same underlying characteristic. As you review the section on intelligence testing, note how patterns of correlations support (a) the concept of general intelligence; (b) the claim that intelligence tests are valid predictors of academic and employment success; (c) the distinction between fluid and crystallized intelligence; and (d) the ideas that variations in mental speed, working memory, and central executive processes may contribute to variations in intelligence.

4. Limitation of the concept of heritability Heritability is a valuable concept because it brings clarity to the long-standing nature-nurture debate. It is important, however, to understand exactly what heritability means and the limitations of the concept. Heritability does not tell us about the relative roles of genes and environment in the development of a trait within an individual, but only about their relative roles in contributing to the variability of the trait within a population. If you understand the formula for heritability, you should be able to explain (a) why heritability for a trait might be different for one population than for another; (b) why heritability decreases when the environmental diversity of the population increases; (c) why heritability of a trait can be greater for adults than for children; and (d) why high heritability within a culture at a particular time is compatible with strong environmental effects across cultures and over historical time.

Further Reading

S. IAN ROBERTSON (2001). *Problem solving*. East Sussex: Psychology Press.

This is a brief, well-written introduction to the psychology of problem solving. Topics include a survey of general problem types and problem-solving strategies, routes to insight, the problem of transfer (and lack of transfer) of learning from one problem type to another, the uses of analogy in problem solving, and development of expertise in solving particular classes of problems.

STEPHEN C. LEVINSON (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge, England: Cambridge University Press.

To read this book is to introduce yourself to cross-cultural psychology at its best. The book describes, in very readable detail, the methods and results of Levinson's research on spatial reasoning among Guugu Yimithirr speakers in

Australia and Tzeltal speakers in Mexico. The work combines the methods of direct observation and cultural immersion that characterize anthropology with controlled psychological experiments, all aimed at developing a rather full understanding of the ways that these people, who lack egocentric spatial terms, understand space and use spatial concepts in their daily lives.

IAN J. DEARY (2001). *Intelligence: A very short introduction*. Oxford: Oxford University Press.

This is a brief introduction to intelligence by a leading intelligence researcher. It is written in an entertaining, easy-to-read manner, but does not shy away from the big, somewhat difficult questions concerning intelligence research. Its 130 pages include chapters on the correlational foundation of the concept of general intelligence, effects

of age on intelligence, brain mechanisms of intelligence, heritability of IQ, validity of IQ, and historical change in IQ.

STEPHEN JAY GOULD (1997). *The mismeasure of man, 2nd ed.* Harmondsworth: Penguin.

Stephen Jay Gould was a renowned evolutionary biologist and historian of biology. In this now-classic work, originally published in 1981, Gould describes the dark side of the history of IQ testing. He documents scientific frauds and clear cases of irrational uses of IQ and other measures, by some of the scientific elite, to justify racist political theories and practices. The story told by Gould is one-sided, but it is a side that every newcomer to psychology should learn about and not forget.

FPO



GROWTH OF THE MIND AND PERSON

One way to understand any complex entity, whether it's a building under construction or a person growing, is to watch it develop over time. This is the approach taken by developmental psychologists, who study the changes through which human behavior becomes increasingly complex and sophisticated from infancy to adulthood. In this unit on developmental psychology, Chapter 11 is about the development of thought and language, and Chapter 12 is about the development of social relationships and the roles they play in promoting other aspects of development.

part 6

The Development of Thought and Language

HOW INFANTS LEARN ABOUT THE PHYSICAL WORLD

- The Infant as Explorer
- Infants' Knowledge of Core Physical Principles

TWO CLASSIC THEORIES OF COGNITIVE DEVELOPMENT: PIAGET'S AND VYGOTSKY'S

- Piaget's Theory: Role of the Child's Own Actions in Mental Growth
- Vygotsky's Theory: Role of the Sociocultural Environment in Mental Growth

DEVELOPMENT OF THE MIND'S INFORMATION-PROCESSING CAPACITIES

CHILDREN'S UNDERSTANDING OF MINDS

THE NATURE OF LANGUAGE AND CHILDREN'S EARLY LINGUISTIC ABILITIES

- Some Universal Characteristics of Human Language
- The Course of Language Development

INTERNAL AND EXTERNAL SUPPORTS FOR LANGUAGE DEVELOPMENT

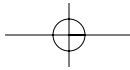
- The Idea of Special, Inborn Mechanisms for Language Learning
- The Language-Acquisition Support System
- Language Learning by Non-Human Apes

When I first saw my newborn son, my words did not match my thought. I said something like, "Oh, he's beautiful," but my thought was, "My god, will this turn into a human being?" Of course, I knew intellectually that he already was a human being, but at the moment he looked more like a cross between that and a garden slug. Over the next weeks, months, and years the little slug's mother and I watched in amazement as he grew not only to look increasingly human but also to do the things that humans everywhere do. He began to smile in response to our smiles; he began eventually to walk upright on two legs and to talk, sometimes incessantly; and, true to his species' name (*Homo sapiens*, where *sapiens* means "wise"), he manifested from early on an insatiable curiosity and soon developed a remarkable store of knowledge and theories about his world.

This chapter is about the millions of new human beings who enter the world every year. More specifically, it is the first of two chapters concerned with *developmental psychology*, the study of changes that occur in people's abilities and dispositions as they grow older. Some developmental psychologists study changes that occur in adulthood (as you will discover in Chapter 12), but most study changes that occur in infancy and childhood. They do so not only because they find infants and children fascinating and worthy of understanding for their own sake, but also because they see in infants and children the origins of adult abilities. Human thought and language in particular are extraordinarily complex abilities, and developmental psychologists have learned a great deal about them by watching them grow in infants and children.

This chapter begins by examining how infants learn about the physical world and what they know about it. Then it turns to some theories and research concerning the development of reasoning, including reasoning about people's minds as well as about physical objects. Finally, it discusses the acquisition

chapter 11



of language. A theme throughout is that the exploratory and playful activities of young people can be understood as biological adaptations serving the purpose of development. Innate tendencies and drives, which are products of human evolution, motivate the child to learn about his or her physical and social worlds, to become an increasingly sophisticated reasoner, and to learn the specifics of his or her native language. These innate tendencies and drives, operating in a responsive social environment, enable the person to develop as one who can survive and thrive in the culture into which he or she is born. This theme is also pursued in the next chapter, which focuses on the development of social relationships.

How Infants Learn About the Physical World

Infancy, roughly the first 18 to 24 months after birth, is the time of most rapid developmental change—change that lays the foundation for further development. How do babies learn about the physical world around them? What do they know, early on, about that world?

The Infant as Explorer

Babies' sensory systems all function at birth (although one sense, vision, is still quite immature). On the day of their birth, babies will turn toward sounds, turn toward anything that touches their faces, turn away from unpleasant odors, suck a nipple more readily for a sweet liquid than for a sour one, and orient their eyes toward high-contrast or moving visual stimuli (Maurer & Maurer, 1988). Within a short time after birth, babies not only respond to stimuli but do so selectively, in ways that seem well designed for learning.

Infants Look Selectively at Novel Objects

1

How do infants reveal, in their behavior, that they are actively exploring their environments with their eyes and that they remember what they have seen?



Hundreds of experiments have shown that babies gaze longer at new stimuli than at familiar ones. When shown a pattern, babies will look at it intently at first and then, over the course of minutes, look at it less and less—a phenomenon referred to as *habituation*. This decline in attention does not stem from general fatigue; if a new pattern is substituted for the old one, infants immediately increase their looking time. Similarly, if shown the new and old patterns at the same time, infants look more at the new one than the old one. This preference for novelty makes sense if we assume that infants are actively trying to learn about their world. They look at new stimuli because they have more to learn from them than from old stimuli, which they have already explored.

This preference for novelty is so reliable that developmental psychologists use it to assess infants' abilities to perceive and remember. Babies who look significantly longer at new stimuli than at ones they have already seen must perceive the difference between the new and old ones and must, at some level, remember having seen the old ones before. In one such experiment, infants as young as 1 day old perceived the difference between two checkerboards with different-sized squares, and remembered that difference over the seconds that separated one trial from the next (Friedman, 1972). This was shown by the fact that they looked longer at the checkerboard that they had not seen before when they were given a choice. Later, you will read about research in which infants' selective looking is used to assess their knowledge and expectations about the physical world.

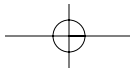
Infants Seek to Control Their Environments

2

How does infants' behavior reveal that they are motivated to control their environments and are emotionally involved in retaining control?



Within a few weeks after birth, infants begin to show a special interest in aspects of the environment that they can control. In one experiment, 2-month-olds smiled and attended much more to a mobile that moved in response to their own bodily movement than to a motor-driven mobile that they could not control (Watson,



1972). In another experiment, 4-month-olds learned quickly to make a particular movement to turn on a small array of lights but lost interest and responded only occasionally after they became good at this task. When the conditions were changed so that a different movement was needed to turn on the lights, the infants regained interest and made another burst of responses (Papousek, 1969). Their renewed interest must have been generated by the new relationship between a response and the lights, because the lights themselves were unchanged. Apparently, the babies were interested not so much in the lights per se as in their ability to control them.

In still another experiment, infants as young as 2 months old, who had learned to turn on a video and sound recording of the *Sesame Street* theme song by pulling strings tied to their wrists, showed facial expressions of anger when the device was disconnected so that they could no longer control it (Lewis & others, 1990). In subsequent, similar experiments, 4- and 5-month-old babies showed facial expressions of both anger and sadness at losing control of their ability to turn on a video and sound recording, even when the recording still came on as often as before but under the control of the experimenter rather than themselves (Sullivan & Lewis, 2003). Clearly, it was the loss of control, not the loss of opportunities to see and hear the recording, that upset them. The desire to control our environment seems to be a facet of human nature that exists in every phase of development, and its function seems obvious: We, more than any other species, survive by controlling our environment.

Infants Explore Increasingly with Hands and Eyes Together

During their first 3 or 4 months of life, babies, like puppies and other young mammals, put practically anything that they can reach into their mouths. They mouth objects in ways that seem designed to test the objects' properties. With time, however, they gradually give up their puppy ways and explore increasingly in the more uniquely human way, with hands and eyes together, rather than mouths (Rochat, 1989).

By 5 or 6 months, babies regularly manipulate and explore objects in the sophisticated manner that researchers label *examining* (Ruff, 1986). They hold an object in front of their eyes, turn it from side to side, pass it from one hand to the other, rub it, squeeze it, and in various other ways act as if they are deliberately testing its properties. Such actions decline dramatically as the infant becomes familiar with a given object but return in full force when a new object, differing in shape or texture, is substituted for the old one (Ruff, 1986, 1989). As evidence that examining involves focused mental activity, researchers have found that babies are more difficult to distract, with bright visual stimuli, when they are examining an object than at other times (Oakes & Tellinghuisen, 1994).

Infants vary their examining in ways consistent with an object's properties. They preferentially *look at* colorfully patterned objects, *feel* objects that have varied textures, and *shake* objects that make a sound when shaken (Lockman & McHale, 1989). Other experiments have shown that infants do, indeed, learn about objects' properties through such examination. In one experiment, 9-month-olds explored toys that produced interesting non-obvious effects when manipulated in certain ways, including a can that wailed when tilted and a doll that separated into two parts when pulled. The infants soon learned to produce each toy's unique effect, and when given a new toy that was similar (but not identical) to the one they had just explored, they immediately tried to produce the previously experienced effect. If the new toy differed in its basic structure from the original one, they did not try to produce the effect but explored the toy afresh, as if intent on discovering its unique properties (Baldwin & others, 1993).

Infants do not have to be taught to examine objects. They do it in every culture, whenever objects are in their reach, whether or not such behavior is encouraged. Roger Bakeman and his colleagues (1990) studied the exploratory behavior of infants among the !Kung San, a hunting-and-gathering group of people in Botswana,



Laura Dwight / Corbis

Making the cow jump over the moon Researchers have found that babies take more delight in objects they can control than in objects they can't control.

>>> 3 What is the evidence that infants' examining of objects (a) helps them learn about objects' unique properties and (b) occurs whether or not adults encourage it?



Laura Dwight / PhotoEdit

A small scientist By 5 to 6 months of age, infants learn about the properties of objects by manipulating them with their hands while watching intently to observe the effects.

Africa, who have been relatively uninfluenced by industrialized cultures. (The *!K* in *!Kung* stands for a clicklike sound that is different from the pronunciation of our *K*.) *!Kung* adults do not make toys for infants or deliberately provide them with objects for play. Yet *!Kung* babies examine objects that happen to be within their reach—such as stones, twigs, food items, and cooking utensils—in the same manner as do babies in industrialized cultures; and, as in industrialized cultures, their examining increases markedly in frequency and intensity between about 4 and 6 months of age. The adults do not encourage such behavior because they see no reason to, yet they recognize its value and do not discourage it. Their view of child development seems to be well summarized by one of their folk expressions, which can be translated roughly as “Children teach themselves” (Bakeman & others, 1990).

4

What are three ways by which infants, beginning at roughly 6 to 12 months of age, use their observations of adults’ behavior to guide their own explorations?

Although babies act upon and explore their environments independently of adult encouragement, they often use cues from adults to guide such actions. By 6 months of age, babies often mimic adults’ actions on objects. In one experiment, 6-month-olds frequently rolled a ball if they had recently seen their mother roll it, and they frequently pounded with a ball if they had seen their mother do that (Hofsten & Siddiqui, 1993).

Beginning in the latter half of their first year of life, babies regularly engage in *joint visual attention* with adults (Moore & Corkum, 1994; Woodward, 2003). A baby looks at an adult’s eyes and then directs his or her own eyes toward whatever the adult is looking at. The baby may then gaze at the object for a while or may turn back to the adult, apparently to check again the adult’s line of sight. Such behavior really does depend on attention to the eyes; if the adult’s eyes are closed or covered, the baby does not look preferentially in the direction the adult is facing



Momatiuk Eastcott / The Image Works

Inuit mother and baby In many traditional cultures, infants are carried around by their mothers throughout the day, and they learn about the world by seeing it from their mother’s vantage point.

(Brooks & Meltzoff, 2002). Such gaze following may help infants learn which objects and events are of greatest interest to their elders, which may be the most important things to attend to and learn about for survival within their culture. It may also help in language development. If the adult is naming an object, it is useful to the child to know what object the adult is naming.

By the time they can crawl or walk freely on their own (toward the end of their first year), infants engage in what is called **social referencing**—they look at their caregivers' emotional expressions for clues about the possible danger of their own actions (Walden, 1991). In an experiment with 12-month-olds, not one crawled over a slight visual cliff (an apparent 30-centimeter drop-off under a solid glass surface) if the mother showed a facial expression of fear, but most crawled over it if her expression was one of joy or interest (Sorce & others, 1985). In another experiment, 12-month-olds avoided a new toy if the mother showed a facial expression of disgust toward it, but they played readily with it otherwise (Hornik & others, 1987).

Infants' Knowledge of Core Physical Principles

You and I share certain assumptions about the nature of physical reality. We assume, for example, that objects continue to exist even when they disappear from view; that two solid objects cannot occupy the same space at the same time; and that if an object moves from one place to another, it must do so along a continuous path. Such assumptions constitute *core principles* of our knowledge of the physical world (Spelke, 2000). We expect these principles always to be true, and when they seem to be violated, we usually assume that our senses have been somehow deceived, not that the principles have really been overturned.

At what age do people begin to make these core assumptions about physical reality? This question has spurred much research in developmental psychology, partly because of its intrinsic interest and partly because of its relevance to a centuries-long philosophical debate. Empiricist philosophers, such as John Locke (1690/1975) and George Berkeley (1710/1820), argued that each person gradually acquires an understanding of core principles through use of the senses and general learning ability. In contrast, nativist philosophers, such as René Descartes (1649/1985) and Immanuel Kant (1781/1965), argued that knowledge of at least some core principles is inborn. According to the nativists, such principles are so central to human perception and thought that they must in some way be known from the beginning, in order for useful learning to occur. Without certain initial assumptions about the nature of physical reality, infants would have no way of interpreting their sensory experiences and, therefore, no way to learn from them. (This

>>> **5**
How did empiricist and nativist philosophers explain the origin of a person's knowledge of core physical principles?



Bob Daemrich / Stock, Boston

Peek-a-boo The results of selective looking experiments suggest that babies even much younger than this one know that objects continue to exist when out of view. Nevertheless, they take delight in having that understanding confirmed, especially when the object is a familiar, friendly person.

idea is discussed more fully in Chapter 1.) Psychological research has not settled the question about the origin of knowledge of core principles, but it has shown that such knowledge is manifested in the patterns of infants' exploration beginning at a very young age.

Infants Reveal Core Knowledge in Selective-Looking Experiments

Just as babies look longer at novel objects than at familiar ones, they also look longer at unexpected events than at expected ones. Researchers have capitalized on this with many experiments that have used selective looking to assess what infants expect to happen in specific conditions (Baillargeon, 2004).

6

What is the violation-of-expectancy method for studying infants' knowledge of physical principles? With this method, what have researchers discovered about the knowledge of 2- to 4-month-olds?

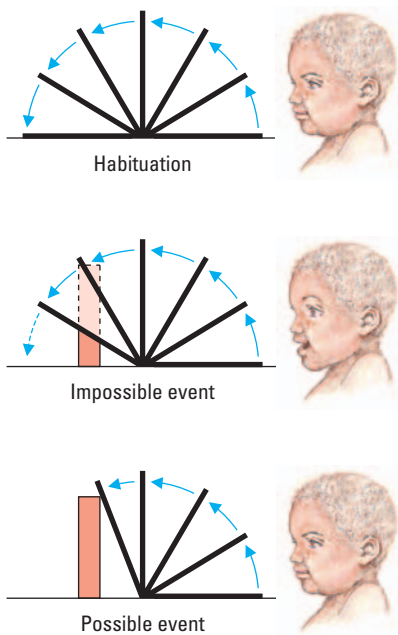


FIGURE 11.1 Evidence for object permanence in very young infants After being habituated to the back-and-forth movement of the screen, infants as young as 3½ months looked longer at the impossible event, in which the screen seemed to go through and obliterate a solid object, than they did at the possible event. (Adapted from Baillargeon, 1987.)

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology*, 23, 655–664. Copyright © 1987 by the American Psychological Association. Adapted by permission.

<<

A classic example of such a *violation-of-expectancy experiment* is illustrated in Figure 11.1. First, in the *habituation phase*, the baby is repeatedly shown a physical event until he or she is bored with it, as indexed by reduced time spent looking at it. In the example shown in the figure, that event is the back-and-forth movement of a hinged screen over a 180° arc. Then, in the test phase, the infant is shown one of two variations of the original event. One variation, the *impossible event*, is an illusion, arranged with mirrors or other trickery, that appears to violate one or more core physical principles. In the example, this event is one in which an object placed behind the rotating screen magically disappears each time the screen rotates backward, allowing the screen to go all the way back, and then reappears each time the screen rotates forward. The other event, the *possible event*, does not violate any physical principle. In the example, this event is one in which the rotating screen stops in each rotation at the place where it would bump into the object behind it.

Notice that the two test events are designed so that, on purely sensory grounds, the possible event differs more from the original habituation event than does the impossible one. Thus, if infants respond simply on the basis of sensory novelty, they should look longest at the possible event. But if they respond from knowledge of physical principles, they should look longest at the impossible event, because that differs most from what they would expect to happen. In fact, using the setup shown in Figure 11.1, Renée Baillargeon (1987) found that infants as young as 3½ months old looked longer at the impossible event than at the possible event. This result is strong evidence that even such young infants in some way understand that solid objects do not normally pass through and obliterate other solid objects.

Similar experiments have demonstrated infants' knowledge of other core principles. For instance, babies as young as 2½ to 3 months expect an object to appear behind the screen where it was originally placed, not behind a different screen (Baillargeon, 2004), expect a rolling ball to stop at a barrier rather than pass through it (Spelke & others, 1992), and expect a stationary ball to remain stationary unless pushed by another object (Spelke & others, 1994). Of course, young infants do not manifest entirely the same expectations about the physical world that adults do. Core principles may be present early on, but nuances related to them are acquired with age and experience. Baillargeon (1994, 1998) and her colleagues found that 4-month-olds expected a box to fall to the ground when it was released in midair but did not expect it to fall when it was set on the edge of a shelf with most of its weight hanging off the shelf. Only by 6 to 7 months did babies show evidence that they expected the unbalanced box to fall.

Infants Reveal Less Knowledge in Search Tasks Than in Selective-Looking Tasks

7

How did Jean Piaget test infants' understanding of object permanence? What might explain the discrepancy between Piaget's results and the results of selective-looking experiments?

<<

The findings supporting infants' early knowledge of core principles surprised many developmental psychologists. Previous research, using a different procedure, had suggested that infants under about 5 months of age lack even the most basic understanding of *object permanence*, the principle that objects continue to exist when out of view. The pioneer of that research was the famous Swiss develop-

mental psychologist Jean Piaget (1936/1963), who tested infants' understanding by having them search for hidden objects. In the simplest of his tests, the *simple hiding problem*, an attractive toy is shown to a baby and then is placed under a napkin as the baby watches. Babies younger than about 5 months typically follow the toy with their eyes as it disappears under the napkin but do not reach for it once it is there and almost immediately seem to lose interest in it. Piaget and his successors interpreted this result as evidence that babies this age completely lack the concept of object permanence.

Between about 6 and 9 months of age, most infants solve the simple hiding problem but fail the *changed-hiding-place problem*, also developed by Piaget (Wishart & Bower, 1984). In the first phase of this test, the toy is hidden under one napkin for a series of trials, and the baby retrieves it each time. Then, in the crucial phase, the toy is hidden under another napkin, right next to the first. Despite having watched the object disappear under the new napkin, the baby reaches toward the original napkin. Piaget concluded that at this age the emerging understanding of object permanence is still very fragile; when pitted against a learned motor habit (reaching toward the original hiding place), the habit wins out. Only by about 10 to 12 months do most infants solve this problem.

Why do infants younger than 5 months fail even the simplest of Piaget's tests of object permanence yet appear to understand object permanence and more in selective-looking experiments? Nobody knows for sure, but a number of researchers have suggested that the difficulty in Piaget's tests has to do with the ability to plan the correct movement to obtain the hidden object (Baillargeon, 1998; Keen, 2003). In order to retrieve a hidden object, the baby not only must know where the object is but must also be able to use that knowledge to guide his or her hand movements. Babies under 5 months of age have no difficulty reaching for objects that are in full view, but may be unable to use mental images of hidden objects to guide their reaching.

Dramatic improvement in infants' search abilities occurs shortly after they learn to crawl or in other ways move about on their own (Campos & others, 2000). In one experiment, 8-month-olds were tested in a series of search tasks, including a version of Piaget's classic changed-hiding-place problem (Kermoian & Campos, 1988). One group had learned to crawl at least 9 weeks before the tests; a second group had not learned to crawl but had at least 9 weeks of experience moving around in walkers at home; and a third group had neither learned to crawl nor been provided with walkers. Approximately 75 percent of the babies in the first two

What evidence suggests that self-produced locomotion promotes rapid development of infants' search abilities?



Myrleen Ferguson Cate / PhotoEdit

Self-produced locomotion promotes cognitive development When babies are able to move around on their own—through crawling or walking with the aid of a walker, such as that shown here—they gain new perspectives on their world and make rapid gains in their ability to solve Piagetian search problems.

groups succeeded on the changed-hiding-place problem, compared with only 13 percent of those in the third group. Self-produced locomotion requires that infants coordinate their vision with their muscular movements in new ways, to avoid bumping into objects, and affords them the opportunity to view stationary objects from varying perspectives. Such experience may well help them learn to plan all sorts of effective movements, including those involved in retrieving hidden objects.

SECTION REVIEW

Infants actively explore their physical world and know some of its core principles.

Exploring the World Around Them

- Infants prefer novel stimuli, as demonstrated by the fact that they look longer at them. This reliable tendency is used to study infant perception and memory.
- Infants exhibit a strong drive to control their environment; they become upset when control is taken away from them.
- By 5 or 6 months, independent of adult encouragement, infants examine objects with their hands and eyes, focusing on the objects' unique properties.
- From 6 to 12 months of age, infants use their observations of adults to guide their own exploration. They mimic adults' actions, look where adults are looking, and use adults' emotional expressions to identify danger or safety.

Knowledge of Core Physical Principles

- Infants' knowledge of core physical principles is revealed by the fact that they look longer at physically impossible events than at physically possible events. Such research indicates that infants as young as 2.5 to 4 months old know some core principles.
- Search tasks that involve manual reaching (such as Piaget's simple hiding problem) appear to show later development of even the rudimentary concept of object permanence, perhaps because such tasks require the infant to form a plan to obtain the hidden object.
- Experience with self-produced locomotion, either by crawling or by using a walker, promotes the ability to solve manual search problems.

Two Classic Theories of Cognitive Development: Piaget's and Vygotsky's

As children grow from infancy toward adulthood, their thinking becomes ever more logical, ever more effective in solving problems. How can these changes be characterized, and what are the processes through which they develop? Here we shall address these questions from two classic theoretical perspectives. The first is Piaget's theory, which focuses on the child's actions on the physical world as the driving force for cognitive development. The second is Lev Vygotsky's sociocultural theory, which focuses on the child's interactions with other people as the driving force.

Piaget's Theory: Role of the Child's Own Actions in Mental Growth

In his long career at the University of Geneva (from the 1920s until his death in 1980), Piaget wrote more than fifty books and hundreds of articles on children's reasoning. His goal was to understand how the adult mind, particularly its capacity for objective reasoning, develops from the child's more primitive abilities. His primary methods were those of observation, testing, and questioning. From his observations of children's conversations and play, he would extract clues about their understanding. He would also present children with specific tasks to solve and question them about their reasons for the solutions they offered. From this work, Piaget developed an elaborate, comprehensive theory of cognitive development. <<< Piaget's most fundamental idea was that mental development derives from the child's own actions on the physical environment. According to Piaget, children are constantly, in their play and exploration, striving to figure out what they can *do* with the various objects that exist in their world. By acting on objects, children de-

9 In Piaget's theory, what is a scheme, and how do schemes develop through assimilation and accommodation?

velop mental representations, called **schemes**, which are mental blueprints for actions. The earliest schemes are closely tied to specific objects and are called forth only by an object's immediate presence. Thus a young infant might have a sucking scheme most applicable to nipples, a grasping and shaking scheme most applicable to rattles, and a smiling scheme most applicable to human faces. As children grow older, they develop new, more sophisticated, more abstract schemes that are less closely tied to the immediate environment or to actual actions.

Schemes Develop Through Assimilation and Accommodation

Piaget conceived of the growth of schemes as involving two complementary processes: assimilation and accommodation.

Assimilation is the process by which new experiences are incorporated into existing schemes. Piaget was a biologist by training, and he considered the assimilation of experiences to be analogous to the assimilation of food. Two people may eat the same type of food, but the food will be assimilated into the tissues differently, depending on the inner structures involved in digestion and building the body. Moreover, just as nondigestible foods will not result in body growth, new experiences that are too different from existing schemes to be mentally digested will not result in mental growth. A calculator given to an infant will not contribute to the child's arithmetic skills, because the infant has no calculating scheme into which to assimilate the calculator's functions. Instead, the infant will probably assimilate the calculator into his or her already well-developed sucking scheme or banging scheme.

Few new stimuli perfectly fit an existing scheme. Assimilation usually requires that existing schemes expand or change somewhat to accommodate the new object or event. Appropriately, Piaget referred to this process as **accommodation**. In Piaget's theory, the mind and its schemes are not like a brick wall, which just grows bigger as each new brick (unit of knowledge) is added; they are more like a spider's web, which changes its entire shape somewhat as each new thread is added. The web accommodates to the thread while the thread is assimilated into the web. The addition of new information to the mind changes somewhat the structure of schemes that are already present.

Consider an infant who, in play, discovers that an object placed above an open box will fall into the box. Perhaps the child already had a scheme for stacking objects, which included the notion that one object placed on top of another will remain on top. But now that scheme must be modified to accommodate this new experience, to include the notion that if one object is hollow and open-topped, another object placed on it will fall inside. At the same time, other schemes that include the notion that two objects cannot occupy the same place at the same time may also undergo accommodation. In Piaget's view, children are most fascinated by experiences that can be assimilated into existing schemes but not too easily, so that accommodation is required. This natural tendency leads children to direct their activities in ways that maximize their mental growth.



Anderson / Monkneyer Press

Jean Piaget Because of his interest in the influence of the environment on children's cognitive development, Piaget preferred to observe children in natural settings. Here he is shown during a visit to a nursery school.



Bruce Plotkin / The Image Works

Accommodation This 11-month-old may be accommodating her "stacking scheme" to assimilate the experience of one block fitting inside another. From such experiences, a new "fits inside" scheme may develop.

10 In Piaget's theory, what is the special value of operations?

Reversible Actions (Operations) Promote Development

As children grow beyond infancy, according to Piaget, the types of actions most conducive to their mental development are those called **operations**, defined as *reversible actions*—actions whose effects can be undone by other actions. Rolling a ball of clay into a sausage shape is an operation because it can be reversed by rolling the clay back into a ball. Turning a light on by pushing a switch up is an operation because it can be reversed by pushing the switch back down. Young children perform countless operations as they explore their environments, and in doing so, they gradually develop *operational schemes*—mental blueprints that allow them to think about the reversibility of their actions.

Understanding the reversibility of actions provides a foundation for understanding basic physical principles. The child who knows that a clay ball can be rolled into a sausage shape and then back into a ball of the same size as it was before has the basis for knowing that the amount of clay must remain the same as the clay changes shape—the principle of *conservation of substance*. The child who can imagine that pushing a light switch back down will restore the whole physical setup to its previous state has the basis for understanding the principle of cause and effect, at least as applied to the switch and the light.

Four Types of Schemes; Four Stages of Development

Piaget conceived of four types of schemes, which represent increasingly sophisticated ways of understanding the physical environment. His research convinced him that the four types develop successively, in stages roughly correlated with the child's age.

11 In Piaget's theory, what are the four stages and the ages roughly associated with each? How are the child's capacities and limitations at each stage related to the kind of scheme that is most prominent? How does the child's behavior at each stage promote advancement to the next stage?

The Sensorimotor Stage The most primitive schemes in Piaget's theory are **sensorimotor schemes**, which provide a foundation for acting on objects that are present but not for thinking about objects that are absent. During the *sensorimotor stage* (from birth to roughly 2 years of age), thought and overt physical action are one and the same. The major task in this stage is to develop classes of schemes specific for different categories of objects. Objects the child explores become assimilated into schemes for sucking, shaking, banging, squeezing, twisting, dropping, and so on, depending on the objects' properties. Eventually the schemes develop in such a way that the child can use them as mental symbols to represent particular objects and classes of objects in their absence, and then they are no longer sensorimotor schemes.

The Preoperational Stage *Preoperational schemes* emerge from sensorimotor schemes and enable the child to think beyond the here and now. Children in the *preoperational stage* (roughly from age 2 to 7) have a well-developed ability to symbolize objects and events that are absent, and in their play they delight in exercising that ability (Piaget, 1962). Put a saucepan into the hands of a preschooler and it is magically transformed into a ray gun or a guitar—the saucepan becomes a sym-

The preoperational stage Sarah, here age 5½, does not pass Piaget's test of conservation. Even though she knows that the short glasses contain the same amount of milk and carefully watches the milk being poured into the tall glass, she still points to the tall glass when asked, "Which has more?"



Hazel Hankin

bol in the child's play. The schemes at this stage are called preoperational because, although they can represent absent objects, they do not permit the child to think about the reversible consequences of actions. According to Piaget, understanding at this stage is based on appearances rather than principles. If you roll a ball of clay into a sausage shape and ask the child if the shape now contains more than, less than, or the same amount of clay as before, the child will respond in accordance with how the clay looks. Noting that the sausage is longer than the ball was, one preoperational child might say that the sausage has more clay than the ball had. Another child, noting that the sausage is thinner than the ball, might say that the sausage has less clay.

The Concrete-Operational Stage Although (or perhaps because) preoperational children have not yet internalized an understanding of operations, they continually produce operations as they explore their environment. As they push, pull, squeeze, mix, and so on, they gradually develop **concrete-operational schemes** and eventually enter the *concrete-operational stage* (roughly from age 7 to 12). These schemes permit a child to think about the reversible consequences of actions and thereby provide the basis for understanding physical principles such as conservation of substance and cause and effect (Piaget, 1927). A concrete-operational child who has had experience with clay will correctly state that the sausage has the same amount of clay as the ball from which it was rolled because it can be rolled back into that ball (see Figure 11.2). A concrete-operational child who has had experience with bicycles will correctly say that the chain is crucial to the bicycle's movement but the fender is not, because the child can picture the reversible consequences of removing each and knows that the pedals can move the wheels only if there is a physical connection between the two.

The earliest schemes for operations are referred to as *concrete* because they are still tied closely to the child's actual experiences in the world. The child might have schemes, for example, for the conservation of clay rolled into various shapes and of fluids poured from one container to another but still lack an understanding of conservation of substance as a general principle that applies regardless of the type of substance.



credit =

The concrete-operational stage
According to Piaget, thought becomes increasingly logical during middle childhood, but it still depends on the presence of real objects. Activities such as that shown here offer opportunities to learn about cause-effect relationships.



In conservation-of-number tests, two equivalent rows of coins are placed side by side and the child says that there is the same number in each row. Then one row is spread apart, and the child is again asked if the two rows have the same number.



In conservation-of-length tests, two same-length sticks are placed side by side and the child says that they are the same length. Then one stick is moved, and the child is again asked if they are the same length.



In conservation-of-substance tests, two identical amounts of clay are rolled into similar-appearing balls and the child says that they have the same amount of clay. Then one ball is rolled out, and the child is again asked if they have the same amount.

FIGURE 11.2 Tests of children's understanding of conservation The transition from the preoperational to the concrete-operational stage, in Piaget's theory, is marked by success in tests such as these. Children usually succeed in conservation-of-number and conservation-of-length tests before they succeed in conservation-of-substance tests.

The Formal-Operational Stage During the concrete-operational stage the child begins to notice that similar operations can be performed on different entities. Thought about these similarities leads eventually to the development of **formal-operational schemes**, which represent principles that apply to a wide variety of situations. For instance, the general concept of conservation of substance—that no operation changes the total amount of a substance—is an example of a formal-operational scheme. When such schemes characterize a significant share of a person's thinking, the person is said to be in the *formal-operational stage* (which begins roughly at the onset of adolescence and continues throughout adulthood). Formal-operational schemes permit a person to think theoretically and apply principles even to actions that cannot actually be performed (Inhelder & Piaget, 1958).

You cannot really unbeat an egg, but with an appropriate formal-operational scheme you can understand the theoretical principle that an egg can be unbeaten and restored to its original form. Thus, you can answer correctly a conservation question about a beaten egg as well as about a rolled-out piece of clay: Is the amount of egg more, less, or the same after it is beaten compared with before? Formal-operational thinkers can extend principles into hypothetical realms that neither they nor anyone else has actually experienced. While the concrete-operational reasoner is limited to empirical (fact-based) science and arithmetic, the formal-operational reasoner is capable of theoretical (principle-based) science and formal mathematics.

Piaget believed that the transitions from stage to stage are gradual; each new type of scheme develops slowly, and a given child may use the more advanced type of scheme for one class of problems while still using the more primitive type for other problems. Piaget also believed that children vary in the age at which a type of scheme begins to predominate. The ages given above are only rough averages. The one constant in Piaget's stage theory is that each child, for any given category of problem, must progress through the stages sequentially: Children will not jump from the first stage to the third without going through the second, nor (barring brain damage) will they regress from one stage to an earlier one.

Limitations of Piaget's Theory

Historically, Piaget's theory was extremely important in motivating developmental psychologists to focus more closely than they had before on children's actions and on the ways that those actions promote mental growth. The concepts of assimilation and accommodation and the idea that operations contribute significantly to cognitive development are still valued by many developmental psychologists. Research and thought since Piaget's time, however, have revealed some serious limitations of Piaget's theory, including the following:

12

What are the major modern criticisms of Piaget's theory?

- It overestimates age differences in ways of thinking. Much research suggests that Piaget underestimated the abilities of young children and overestimated those of adolescents and adults. Earlier in this chapter you read about selective-looking experiments that showed that infants as young as 3 months expect objects to continue to exist when out of view and to obey certain universal physical principles. Children as young as 4 or 5 years regularly pass at least some tests indicative of operational reasoning (such as the number-conservation test illustrated in Figure 11.2) if the problems are presented clearly, without distracting information (McGarrigle & Donaldson, 1975). At the other end of the developmental scale, research with adults (discussed in Chapter 10) challenges the distinction between concrete and formal thought. Such research suggests that people who are good at solving abstract problems usually do so by thinking of analogies to more familiar problems or by turning the problem statement into a familiar visual image, thereby converting a formal task into a concrete one. In light of such studies, it seems reasonable to suggest that all of us, throughout the great bulk of our childhood and throughout adulthood, are what Piaget would call concrete-operational thinkers. Our thought advances more because

of our gradually increasing stores of knowledge and ability to manipulate that knowledge efficiently than because of fundamental revolutions in our ways of thinking.

- *It is vague about the process of change.* Some critics contend that Piaget's theory of the process by which the mind develops is too vague, too distanced from the behavioral data. What precisely is a scheme? What information must a scheme have in order, say, to permit a child to succeed in the conservation-of-clay problem? What exactly changes when assimilation and accommodation occur? Theorists in the information-processing tradition (to be described later) believe that theories of mental development should be more explicit than Piaget's in describing the links between the observed changes in behavior and the posited changes occurring in the mind.
- *It underestimates the role of the social environment.* In focusing on children's own activities and discoveries, Piaget may have paid too little attention to the role of children's social interactions, with peers and with adults, in their cognitive growth. The psychologist who most clearly developed this criticism was Lev Vygotsky, to whose theory we now turn.



Lawrence Migdale / Stock, Boston

Helping Grandma Children learn even about physical concepts, such as quantity and mass, in interaction with other people. Piaget has been criticized by contemporary developmental psychologists for having underemphasized the role of social learning.

Vygotsky's Theory: Role of the Sociocultural Environment in Mental Growth

Children do not develop in a social vacuum. They develop in a sociocultural milieu in which they interact constantly with other people and with products of their cultural history. The person most often credited with originating the sociocultural perspective on cognitive development is Lev Vygotsky, a Russian scholar who died in 1934 at age 38, after devoting just 10 years to formal research and writing in psychology. Vygotsky (1934/1962) agreed with Piaget that the main force for development is the child's active interaction with the environment, but he disagreed with Piaget's conception of the relevant environment. Whereas Piaget emphasized the child's interaction with the physical environment, Vygotsky emphasized the child's interaction with the social environment. In Vygotsky's view, cognitive development is largely a matter of internalizing the symbols, ideas, and modes of reasoning that have evolved over the course of history and constitute the culture into which the child is born.

The distinction between Vygotsky's and Piaget's perspectives can be illustrated by applying those perspectives to a story, related by Piaget (1970), about how a mathematician friend of his had, as a child, become fascinated by mathematics:

When he was a small child, he was counting pebbles one day; he lined them up in a row, counted them from left to right, and got to ten. Then, just for fun, he counted them from right to left to see what number he would get, and was astonished that he got ten again. He put the pebbles in a circle and counted them, and once again there were ten. And no matter how he put the pebbles down, when he counted them, the number came to ten. He discovered there what is known in mathematics as commutativity, that is, the sum is independent of the order.

The story is prototypically Piagetian. The child, through acting on physical objects (pebbles), discovers and is exhilarated by a core principle of mathematics (commutativity). As Piaget goes on to explain, "The knowledge that this future mathematician discovered that day was drawn, then, not from the physical properties of the pebbles, but from the actions that he carried out on the pebbles."

How might Vygotsky have reacted to this story? It was told long after Vygotsky had died, but I imagine him saying: "Where did that young boy learn to count in

>>> **13**
How does Vygotsky's perspective on cognitive development differ from Piaget's?



Ria-Novosti / Sovfoto

Lev Vygotsky A leader of Russian psychology during the early days of the Soviet Union, Vygotsky developed a theory of human development that emphasized the roles of culture and social interaction.

the first place? Of all the things he might do with pebbles, why did he decide that counting them was worthwhile? The answer lies in the boy's social environment. He was growing up in a culture where number words are in the air and people value counting. He may have discovered with pebbles that day the principle of commutativity, but his social environment had prepared him to make that discovery."

The Role of Language in Mental Development

14

In Vygotsky's view, how does the acquisition of language lead to a higher form of thought?

<<< While Piaget (1970) considered language to be more or less a side effect of the child's development of thought and not essential to it, Vygotsky (1934/1962) argued that language is the foundation for the development of higher human thought. According to Vygotsky, children first learn words as means of communicating with others but then begin to use them also as symbols for thinking. This leads to *verbal thought*, which, according to Vygotsky, is more powerful than a child's earlier, non-verbal forms of thought. Words are symbols developed over countless generations of human language use. Through internalizing language and using it for thought, children internalize a far richer set of symbols than they could possibly invent on their own.

The words of a language not only provide the building blocks for verbal thought but also direct it in ways that reflect the culture's activities and values. For example, cultures in which counting is important have developed an efficient set of number words. Children growing up in those cultures begin to recite number words sequentially even before they understand their meaning; and through that ritual act they eventually learn the meaning.

As another example, consider the word *because*. Both Vygotsky (1934/1962) and Piaget (1927) noted that children begin to use this word (or rather its Russian or French equivalent) before they fully understand its meaning. Children younger than 6 or 7 years commonly reverse cause and effect, saying such things as "Billy fell off the porch because he broke his arm." To Piaget, the word is incidental to cognitive development; its correct or incorrect usage only reflects the child's level of understanding. But to Vygotsky, the word is critical. Using the word in a communicative social context induces the child to think about its meaning, which leads the child toward a more advanced understanding of cause-effect relationships. *Because* becomes not just a tool for communication but also a tool for private thought.

Children's Private Speech: A Transition Toward Verbal Thought

15

According to Vygotsky, what is the function of children's noncommunicative, private speech? What evidence supports Vygotsky's view?

<<< In support of his view that language is critical to thought, Vygotsky (1934/1962) noted that children about 4 to 6 years old often talk aloud in a noncommunicative manner. A 4-year-old drawing a picture of a horse might say, for example, "Horses have long necks. I'll make it long." Vygotsky (1934/1962) interpreted such speech as a transition stage in the development of verbal thought. At around age 4 children discover that they can use words not just to communicate to others but also as part of planning, to direct their own behavior. By talking to themselves as they work on a task, they can maintain their focus on the task and solve problems related to it. Out-loud private speech declines at around age 7 and is replaced increasingly by *inner speech* (Vygotsky's term), or *verbal thought*. With experience, inner speech becomes increasingly abbreviated, and in older children and adults the words are usually not pronounced even to the self but are purely mental symbols.

Consistent with Vygotsky's view, researchers have since found that young children manifest more out-loud private speech when working on difficult problems than when working on easy ones (Berk, 1994; Winsler & others, 2003) and solve problems better when asked to think aloud than when asked to work silently (Kendler, 1972; Kohlberg & others, 1968). In another study, first- and second-graders who manifested the most muttering and lip movement while solving arithmetic problems showed the greatest improvement in their arithmetic ability over the course of a year (Bivens & Berk, 1990).



Elizabeth Crews

Thinking with lips and fingers
According to Vygotsky, young children talk to themselves as a means of solving problems. With time and experience, such talk becomes internalized in the form of verbal thought.

The Role of Collaboration and Dialogue in Mental Development

The most fundamental Vygotskian idea is that development occurs first at the social level and then at the individual level. People learn to converse with words (a social activity) before they learn to think with words (a private activity). People also learn how to solve problems in collaboration with more competent others before they can solve the same kinds of problems alone. Vygotsky (1935/1978) coined the term *zone of proximal development* to refer to the difference between what a child can do alone and what the same child can do in collaboration with a more competent other.

In research at an alternative, age-mixed school, I found many examples of collaboration between adolescents and younger children that nicely illustrate Vygotsky's concept of a zone of proximal development (Gray & Feldman, 2004). In one case, for example, a teenage boy helped a 5-year-old girl find her lost shoes by asking her to think of all the places she had been that day and all the places where she had already looked. Through such suggestions, the teenager added structure to the little girl's thinking, which allowed her to think and search more systematically than she would have been able to do on her own. Such collaboration not only enabled the child to find her shoes but also probably promoted her mental development by suggesting questions that she might ask herself to guide future searches for missing objects.

From a Vygotskian perspective, critical thinking—in adults as well as children—derives largely from the social, collaborative activity of dialogue. In actual dialogue, one person states an idea and another responds with a question or comment that challenges or extends the idea. In the back-and-forth exchange, the original statement is clarified, revised, used as the foundation for building a larger argument, or rejected as absurd. From many such experiences we internalize such dialogues, so that we (or what seem like voices within us that represent our friends and critics) question and extend our private thoughts and ideas and in that way improve them or throw them out. Consistent with Vygotsky's view, researchers have found that students who engage in covert dialogues with authors as they read, or who are asked to explain ideas they are studying to a real or imagined other person, acquire a more complete understanding of what they have read or studied than do other students (Chi & others, 1989, 1994; Heiman, 1987).

The Child as Apprentice

While Piaget's child can be characterized as a little scientist performing experiments on the world and discovering its nature, Vygotsky's child can be characterized as an apprentice (Rogoff, 1990). Children are born into a social world in which people routinely engage in activities that are important to the culture. Children are attracted to those activities and seek to participate. At first their roles are small, but they grow as the children gain skill and understanding. From this view, cognitive

>>> 16

According to Vygotsky, how are a child's abilities stretched and improved through collaboration with other people?

>>> 17

How does Vygotsky's "apprentice" view of the child contrast with Piaget's "scientist" view?

Apprenticeship Vygotsky pointed out that children are attracted to activities that involve them in family and community life. This Peruvian girl knitting alongside her mother is preparing for full adult participation in her culture.



development is a progression not so much from simple tasks to more complex ones as from small roles to larger roles in the activities of the social world. Barbara Rogoff (1990, 1993) has documented many ways by which children in various cultures involve themselves in family and community activities. One prediction of the apprenticeship analogy is that people who grow up in different cultures will acquire different cognitive abilities. A child surrounded by people who drive cars, use computers, and read books will not learn the same mental skills as a child surrounded by people who hunt game, weave blankets, and tell stories far into the night. The apprenticeship analogy also reminds us that logic itself is not the goal of mental development. The goal is to function effectively as an adult in one's society. To achieve that goal, children must learn to get along with other people and to perform economically valuable tasks. In our society, such tasks may involve for some people the kind of mathematical and scientific reasoning that Piaget labeled formal operational; but in another society, they may not.

SECTION REVIEW

Two classic theories focus on interaction with the environment—but different aspects of it.

Piaget: The Child as Little Scientist

- Piaget believed that cognitive development occurs through the child's actions on the physical environment, which promote the development of schemes—mental blueprints for actions.
- Piaget held that mental growth involves assimilation (fitting new experiences into existing schemes) and accommodation (modifying those schemes to fit with new experiences).
- Operational schemes—schemes for reversible actions—are particularly important to cognitive development, according to Piaget.
- Piaget described four successive stages of cognitive development (sensorimotor, preoperational, concrete-operational, and formal-operational stages), each employing an increasingly sophisticated type of scheme.

Vygotsky: The Child as Apprentice

- Vygotsky considered the child's interaction with the social and cultural environment to be the key to cognitive development, leading to internalization of symbols, ideas, and ways of thinking.
- Vygotsky, unlike Piaget, saw language as crucial to mental development, with words serving not only as a means of communication but later as the building blocks of verbal thought. Private, noncommunicative speech provides evidence of this.
- Through dialogue and collaboration with more competent others, the child learns to do something socially before being able to do it individually. This social learning takes place within the child's zone of proximal development.

Development of the Mind's Information-Processing Capacities

18

What is the information-processing perspective on cognitive development, and how does it differ from Piaget's perspective?

<<< Piaget tended to view the mind as a single entity that grows as a whole. In contrast, developmentalists who adopt the *information-processing perspective* treat the mind as a set of interacting components, any one of which might change as the child grows older. They attempt to explain specific changes in children's mental abilities in terms of specific changes in one or another of the mind's components or in terms of specific knowledge the child has acquired.

The information-processing approach to cognition, as described in Chapter 9, begins with the assumption that the mind is a system, analogous to a computer, for analyzing information from the environment. According to the standard information-processing model (see Figure 9.1, page 000), the mind's basic machinery includes attention mechanisms for bringing information in, working memory for actively manipulating (or thinking about) information, and long-term memory for passively holding information so that it can be used in the future. Together, these components are analogous to the all-purpose operating system that is built into a computer. In addition, the mind contains specific strategies and rules for analyzing particular types of information or solving particular types of problems. These strategies and rules are analogous to the specific programs, such as a word-processing program, that are loaded into a computer. From the information-processing perspective, cognitive development can stem from changes in the mind's general operating system, in its specific programs, or in both.

Development of Long-Term Memory Systems: Episodic Memory Comes Last

In Chapter 9 I described the distinction between explicit and implicit long-term memory systems. *Explicit memories* are those that people can consciously think about and report to others. When we ask people to say what they remember of an event, we are testing their explicit memory. We have no certain way to assess babies' explicit memories until they have learned to talk. *Implicit memories*, in contrast, are memories that affect behavior even though the person is unable to report them. They include procedural memories, such as how to pound with a hammer or ride a bicycle, and effects of classical conditioning, which are demonstrated in a person's nonverbal behavior. Even very young infants show, in their behavior, that they can form implicit memories. A newborn who looks at a new stimulus more than at an old one demonstrates implicit memory for the old stimulus. A 2-month-old who learns to kick her feet to move a mobile and who then, a day later, kicks her feet again as soon as the mobile appears, demonstrates implicit memory.

I also described in Chapter 9 the distinction between two categories of explicit memories—semantic and episodic. *Semantic memories* are explicit memories for facts, beliefs, and word meanings. The first clear evidence of semantic memory occurs when babies, at the age of about 10 or 12 months, begin to name objects in their environments correctly: *Mama*, *blankie*, *ball*. *Episodic memories* are explicit memories for events in one's life. To demonstrate episodic memory, one must be able to recall and describe previous experiences, and this ability develops relatively slowly. Typically by about 20 to 24 months of age, children begin to make comments about things that happened to them earlier in the day or on a previous day, but such comments are infrequent and fragmentary (Nelson & Fivush, 2004). Only by about 3 or 4 years of age do children begin to answer reliably at least some questions about their past experiences.

Research suggests that young children must develop the capacity to encode their experiences into words before they can form episodic memories of those experiences (Nelson & Fivush, 2004). At about age 3 children begin, with some reliability, to talk about their experiences as they experience them. Such talk seems to help them make sense of what they are doing, as I noted earlier, and it may also be essential to the formation of episodic memories. At first such talk depends on the existence of an older conversation partner who can help the child organize the experience in a coherent way and find the appropriate words for it. In one study, researchers recorded the conversations of mothers and their 3-year-old children at visits to a natural history museum and then, a week later, asked the children to recall what they had seen at the museum (Tessler & Nelson, 1994). The result was that the children correctly recalled only those items that had been commented on jointly by *both* the mother and child in conversation. Items that had been commented on just by the mother or just by the child were not recalled.

>>> **19**
At what ages, and in what contexts, do researchers first see evidence of (a) implicit memory, (b) explicit semantic memory, and (c) explicit episodic memory in young children?

>>> **19**
Through what developmental steps do young children develop the capacity to form episodic memories? Why do adults generally have poor memories for events that occurred early in childhood?

Forming episodic memories To form long-term episodic memories, young children must encode their experiences verbally. Such encoding is facilitated by adults who share the experience and, through conversation, help the child to find words for what he or she sees.



called and verbalized at a later date (Simcock & Hayne, 2002).

Development of Working Memory and Faster Speed of Processing

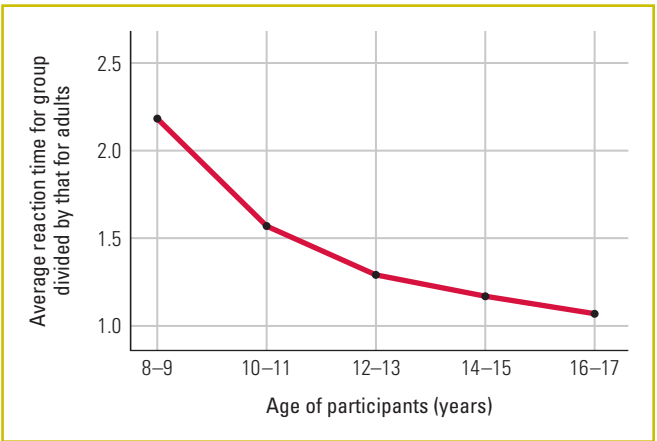
21

How do working-memory capacity and speed of processing change with age during childhood and early adolescence? How might working-memory capacity depend on speed of processing?

<<< In the standard model of the mind, working (or short-term) memory is the center of conscious thought and the place where information—from the environment and from explicit long-term memory—is combined and manipulated to solve problems. As discussed in Chapters 9 and 10, the working-memory system is limited in the amount of information it can hold and use at any given time, and this limit constrains a person's problem-solving ability. Many experiments, using many different sorts of measures, have shown that the amount of either verbal or visual information that a person can hold in working memory at any given time increases steadily throughout childhood and reaches adult levels at about age 15 (Gathercole & others, 2004). For instance, the number of digits or random single-syllable words that a person can hold in mind and repeat, after hearing them just once, increases from about three at age 4 to about seven at age 15.

Closely correlated with increased working-memory capacity, and possibly a cause of it, is increased *speed of processing*—the speed at which elementary information-processing tasks can be carried out. Speed of processing is usually assessed with reaction-time tests that require a very simple judgment, such as whether two letters or shapes flashed on a screen are the same or different or whether an arrowhead is pointing right or left. Such tests consistently reveal age-related improvement up to the age of about 15 years (Kail, 1993; see Figure 11.3).

FIGURE 11.3 Reaction time for simple tasks decreases with age Children and adolescents were tested for their speed on six different tests, including a test of elementary reaction time (releasing a button in response to a signal) and a test of picture matching (judging whether two pictures are identical or not). Each person's average time for the six tests was converted by dividing it by the average time achieved by young adults, and the results were then averaged for each age group. Note that a decline in reaction time implies an increase in speed. (Data from Kail, 1993.)



As was discussed in Chapter 10, faster processing speed permits faster mental movement from one item of information to another, which improves one's ability to keep track of (and thereby hold) a number of different items in working memory at once. Faster processing speed may result at least partly from the physical maturation of the brain that occurs throughout childhood, independent of specific experiences. Consistent with that view, 9- and 10-year-old boys who were judged as physically mature for their age—on the basis of their height as a percentage of their predicted adult height—exhibited significantly faster reaction times than did boys of the same age who were judged as physically less mature (Eaton & Ritchot, 1995).

Acquisition of Specific Rules for Specific Types of Problems

Mental development occurs not just through improvement in the all-purpose mental machinery, such as that for episodic long-term memory and working memory, but also through the acquisition of particular rules and strategies for solving particular categories of problems. Some developmental psychologists, guided by the information-processing perspective, have focused on the question of how children acquire specific rules and strategies (Siegler, 2000; Siegler & Alibali, 2005). A classic example of such research is a series of experiments conducted by Robert Siegler (1983) on the ability to solve balance-beam problems such as those depicted in Figure 11.4.

Siegler found that even people who perform poorly on such problems usually do not just guess randomly. Rather, their responses—over a large set of such problems—can be understood in terms of some rule. More specifically, Siegler identified four rules for solving balance-beam problems, which could be ranked from the simplest and least effective to the most advanced and effective as follows:

- Rule 1: Weight alone is considered; the side with more weight is predicted to go down.
- Rule 2: Weight alone is considered unless the weights on the two sides of the fulcrum are the same (as in problem A of Figure 11.4), in which case the side on which the weights are farthest from the fulcrum is predicted to go down.
- Rule 3: Both weight and distance are considered, but when weight is greater on one side and distance is greater on the other (as in problem B of Figure 11.4), the person just guesses.
- Rule 4: For each side, each weight is multiplied by its distance from the fulcrum and then the sum of these products is calculated; the side for which that sum is higher is predicted to go down.

Most 5-year-olds behaved according to rule 1, most 9-year-olds behaved according to either rule 2 or 3, and most adolescents and adults behaved according to rule 3. Only a few adolescents and adults behaved according to rule 4. In some of his experiments, Siegler allowed children to gain feedback on each trial, after they had made their prediction, by releasing the balance so that they could see which end would go down. He found that the effectiveness of such feedback depended on the specific problem for which it was given. Children who behaved according to rule 1 profited from feedback on problems that disconfirmed rule 1 (the simplest rule) and confirmed rule 2 (such as problem A in Figure 11.4). These children did not, however, profit from feedback on problems that disconfirmed both rules 1 and 2 (such as problem B in the figure). From such data, Siegler suggested that children

>>> 22
What did Siegler discover in his balance-beam experiments? What sort of feedback helped children become better at solving such problems?

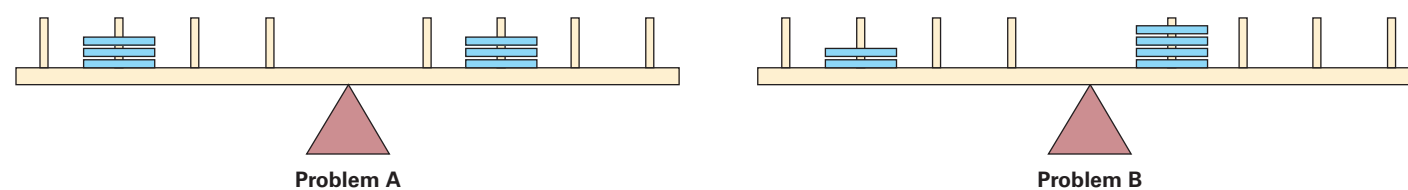


FIGURE 11.4 | Which side will go down? Siegler tested people of varying ages on problems such as these, and found that they behaved as if they were following one of four rules (described in the text).

23 How does Siegler's explanation of improvement in balance-beam problems differ from the kind of explanation that Piaget would offer?

at one level of rule application are prepared to use information needed to acquire the next rule in the sequence but cannot fathom information that would require them to skip the next rule and go directly to a more advanced one.

<<< Siegler's balance-beam research nicely illustrates distinctions between the information-processing approach and the Piagetian approach to mental development. Piaget commonly used balance-beam problems to test children's capacities for formal-operational reasoning, but he did not attempt to identify precisely, as Siegler did, the rules that people follow in attempting to solve such problems. In Piaget's account, the mind develops as a whole; a person who has reached the formal-operational stage can figure out how to solve balance-beam problems as well as a whole set of other formal problems, and a person who has not reached that stage cannot. In contrast, information-processing theorists often study *domain-specific* development—that is, development that is limited to just one class of problem. A domain might be as narrow as balance-beam problems or as broad as all of physics or mathematics. The principles that underlie and unite any given domain may include some that are innate to the human mind and others that must be acquired through experience. Developmental psychologists such as Siegler are interested in identifying the kinds of experiences that best help people acquire and understand the relevant principles.

SECTION REVIEW

The information-processing approach considers the development of separate mental components and processes.		
Long-Term Memory - Children show a capacity for implicit memory in early infancy; but we cannot assess explicit memory capacity until the child develops sufficient language skills. - Semantic memory is in evidence as soon as a child begins to use words, at about 10 to 12 months. - Episodic memory apparently requires that the child encode personal experiences verbally, which begins to happen with some regularity at about age 3.	Working Memory - The capacity of working memory increases as the child grows older, reaching adult levels at about age 15. - There is a parallel increase in processing speed over the same period, which may actually be the basis for the increase in capacity. - Faster processing may result in part from brain maturation.	Acquiring Specific Rules - Cognitive development also involves the acquisition of rules and strategies for solving particular types of problems. - Siegler found that children's solutions to balance-beam problems were based on specific rules, which could be ranked by complexity and effectiveness. - Siegler also found that feedback on the problems could help children advance only to the next level of rule.

Children's Understanding of Minds

To develop as fully functional humans, we must learn not just about the physical world around us but also about the social world around us. Most of us—adults and children alike—spend more time trying to understand other people than trying to understand inanimate objects, and we apply entirely different explanatory concepts to the two endeavors. In our explanatory frameworks, billiard balls move because they are hit by other balls or cue sticks, but people move because they want to get somewhere. We are all psychologists in our everyday lives, continually trying to account for people's behavior in terms of their minds. We attribute emotions, motives, feelings, desires, goals, perceptions, and beliefs to people, and we use those attributes to explain their actions.

David Premack (1990) has suggested that beginning at a very early age, humans automatically divide the world into two classes of entities—those that move on their own and those that don't—and attribute psychological properties to the for-

mer but not the latter. When 3- to 5-year-olds saw videos of balls moving like billiard balls, only in response to physical impacts, they described the movements in purely physical terms; but when they saw videos of balls moving and changing direction on their own, they immediately regarded the balls as representing people or animals and described the movements in mental terms (Premack, 1990). A child described one sequence of movements as one ball trying to help another ball get out of a hole.

Even Very Young Children Explain Behavior in Mental Terms

By the time children have learned language sufficiently to offer verbal explanations—that is, by about 2 to 3 years of age—they already explain people’s behavior in terms of mental constructs, especially in terms of *perceptions*, *emotions*, and *desires* (Hickling & Wellman, 2001; Lillard & Flavell, 1990). They expect others to respond to objects that they (the others) can see but not to objects that they cannot see. They describe a crying person as sad. They say that a person filling a glass with water is thirsty and wants a drink. In one experiment, 2-year-olds demonstrated an understanding that another person’s desires could be different from their own. Having learned that a particular adult preferred broccoli to crackers as a snack, they gave that adult broccoli, even though their own preference was for crackers (Repacholi & Gopnik, 1997).

In another experiment, researchers showed that even 12-month-olds can display a remarkable understanding of what is in another person’s mind (Tomasello & Haberl, 2003). In that experiment, each infant played with two adults and three new toys. One of the two adults left the room while one of the three toys was being played with (either the first toy or the third) and therefore did not see or play with that toy. Then, at the end of the play session, all three toys were brought into the room on a tray and the adult who had missed playing with one of them looked in the direction of the three toys and said, “Wow! Cool! Can you give it to me?” In response, the majority of infants gave the adult the toy that that adult had not played with before, not one of the two toys with which the adult was already familiar. To perform in this way, the infants must have known which toy was new to the adult, even though it wasn’t new to them, and must also have known that people are more excited by new things than by familiar ones.

Delay in Understanding False Beliefs

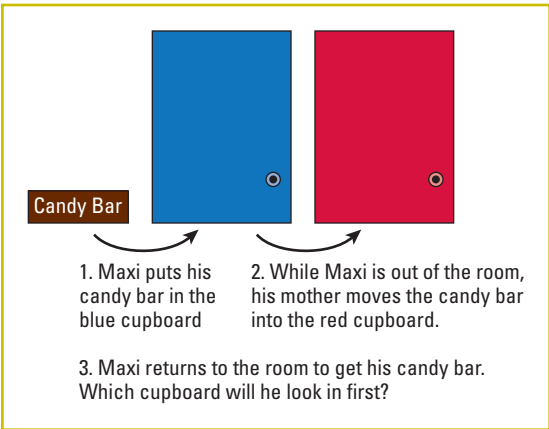
You and I explain people’s behavior not just in terms of their perceptions, emotions, and desires but also in terms of their *beliefs*, and we know that beliefs can be mistaken. For example, if we see a man carrying an umbrella on a sunny day, we might explain that he must have believed it was going to rain that day. Three-year-olds, however, rarely offer explanations in terms of beliefs (Saxe & others, 2004), and tests indicate that they do not clearly understand that beliefs can differ from reality (Wellman & others, 2001).

A typical test of false-belief understanding is the following. The child is told a story, which is also acted out with puppets for clarity, in which Maxi puts his candy bar in a blue cupboard. Then Maxi leaves the room and his mother comes in, finds the candy bar, and moves it into the red cupboard. Then Maxi re-enters the room to get his candy bar and the child is asked: “In which cupboard will Maxi look first?” Most 4-year-olds answer, just as you or I would, “In the blue cupboard,” but most 3-year-olds insist that he’ll look in the red cupboard (Wellman & others, 2001). The problem isn’t poor memory, because the 3-year-

A test of ability to understand false belief In this test, which is usually presented with the help of puppets that act out the sequence, most children under age 4 say that Maxi will look in the red cupboard.

>>> 24
What do children under 3 years old understand about other people’s minds?

>>> 25
What evidence suggests that an understanding that people can hold false beliefs usually does not develop prior to age 4? Why might false beliefs be particularly difficult for young children to understand?



olds in these experiments have no difficulty reporting accurately on all the factual details of the story when questioned. Rather, they seem not to understand that someone can believe something that isn't true.

Three-year-olds' denials of false belief apply even to their own false beliefs. In one experiment, 3-year-olds were shown a crayon box and asked to say what they believed was inside (Atance & O'Neill, 2004). They all said "crayons." Then the children (who were tested individually) were told that they could get some paper to draw on with the crayons if they wanted to. When the children returned with drawing paper, the box was opened and it proved to have candles inside rather than crayons. When asked what they thought the box had in it when they first saw it, most said "candles." When asked why they had gotten the paper if they thought the box contained candles, most had no cogent explanation.

Perhaps the concept of false belief is particularly difficult to grasp because of its inherent contradiction. False beliefs are both false and true at the same time. They are false in reality but true in the minds of the believers. In this way they differ from the products of make-believe.

Make-Believe as a Precursor to the Belief-Reality Distinction

26 What logic and evidence suggest that engagement in pretend play, especially in role play with other children, may help children acquire an understanding of false beliefs?

<<< Three-year-olds may have difficulty understanding false beliefs, but they have no difficulty understanding *pretense*. Toddlers who are 2 and 3 years old, as well as older children, engage in an enormous amount of pretend play, and researchers have found that even 2-year-olds clearly differentiate between make-believe and reality. A 2-year-old who turns a cup filled with imaginary water over a doll and says, "Oh oh, dolly all wet," knows that the doll isn't really wet (Leslie, 1994). Alan Leslie (1987, 1994) has suggested that children's understanding of false beliefs emerges from their earlier understanding of pretense. Pretense is very similar to false belief. Both, by definition, are mental conceptions that depart from reality. The only difference between the two is that pretenders know that their conception doesn't match reality, whereas believers think that theirs does. Three-year-olds, who fail false-belief tests, such as the test in which a crayon box actually holds candles, do not fail analogous tests in which they are asked to report what either they or another person had *imagined* or *pretended* was in the box before it was opened (Lillard & Flavell, 1992; Woolley, 1995).

Children everywhere engage in pretend play, whether or not they are encouraged to do so (Carlson & others, 1998). Piaget (1962) regarded such play as an expression and exercise of the child's ability to symbolize objects in their absence, but many developmental psychologists today ascribe even further significance to it. Leslie (1991) suggests that the brain mechanisms that enable and motivate pretend play came about in evolution because such play provides a foundation for understanding

nonliteral mental states, including false beliefs. A child who understands that pretense differs from reality has the foundation for understanding that people's beliefs (including the child's own beliefs) can differ from reality and that people can fool others by manipulating their beliefs.

I'm flying (but not really) A child of 2 or 3 years who pretends that a box with its flaps out is an airplane knows that it is not really an airplane. The understanding that pretense and reality are different comes well before the understanding that people can hold sincere beliefs that are false.



Scott Barrows / International Stock

Evidence for the view that pretend play promotes false-belief understanding comes from research showing strong correlations between the two. Children who have engaged in lots of pretend role-play with other children pass false-belief tests at a higher rate than do children who have engaged in less (Jenkins & Astington, 1996). Other research has shown that children who have older siblings pass false-belief tests at a much higher rate than do same-age children who lack older siblings (Ruffman & others, 1998; see Figure 11.5). Children with older siblings also engage in much more role play than do those without older siblings, because the siblings draw them into the play and teach them how to do it (Younglade & Dunn, 1995). Role play with other children may be especially valuable for development of false-belief understanding, because in such play children often say things that they know are not true in reality and must respond as if what the other person said is true: *I'm your mommy and you must obey me*, or *Bang, you're dead*.

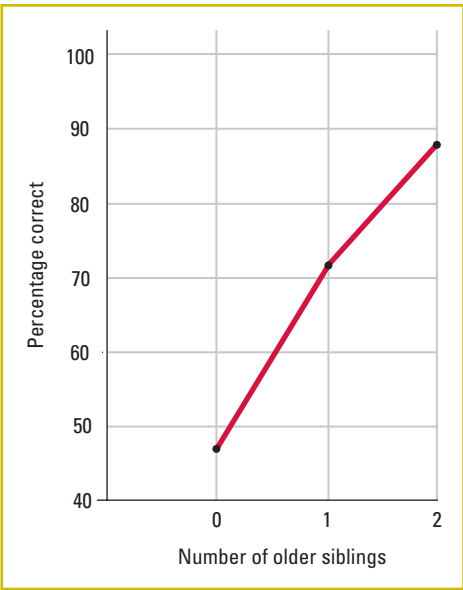


FIGURE 11.5 Older siblings promote false-belief understanding Three- and 4-year-old children with older siblings succeeded on a standard test of understanding of false belief at a much higher rate than did those who had no older siblings. (Adapted from Ruffman & colleagues, 1998, p. 164.)

Make-Believe May Also Promote Hypothetical Reasoning

Pretend play may provide a foundation not just for false-belief understanding but also for all sorts of logical reasoning that involves premises not anchored in physical reality. Consider this syllogism: *All cats bark. Muffins is a cat. Does Muffins bark?* Children under about 10 years old usually answer “no,” and their explanations show that they either can’t or won’t accept the initial premise. Yet in a series of experiments, 4-year-olds, and to a lesser extent even 2-year-olds, succeeded on syllogisms like this when they were presented in a playful voice and in words that made it clear that the world being described is a *pretend* world different from the world in which we live (Dias & Harris, 1988; Richards & Sanderson, 1999).

Children apparently have, early in their development, two separate modes of thinking: a fictional mode and a reality mode. Young children can apparently think quite well in either mode, but they do not at first combine the two. Advancement in thinking may entail, in part, the bringing together of these two modes for specific types of problems. Understanding false beliefs may be an early and quite universal result of such conjoining, and deductive reasoning from hypothetical premises may be a later, more culture-dependent result.

>>> 27
How have researchers shown that pretend play can help very young children solve syllogisms that are based on counterfactual premises?

SECTION REVIEW

Children begin quite early to understand not only physical reality but also the mind.

Using Mental Constructs

- Young children seem to automatically divide entities into two classes—those that move on their own and those that do not—and ascribe psychological characteristics to the former.
- Well before the age of 3, children use such mental constructs as perception, emotion, and desire to explain people’s behavior.
- The understanding that beliefs can be false—that is, not congruent with reality—takes longer to develop, appearing at about age 4.

Make-Believe

- Children everywhere engage in make-believe play; even toddlers can distinguish between reality and pretense.
- Pretend play, especially role-playing, and the knowledge that pretense is not reality may provide a foundation for the later understanding of false beliefs.
- Experience with make-believe may also help children to develop the ability to reason on the basis of hypothetical or counterfactual premises.

The Nature of Language and Children's Early Linguistic Abilities

Of all the things that people can do, none seems more complex than understanding and speaking a language. Thousands of words and countless subtle grammatical rules for modifying and combining words must be learned. Yet nearly all people master their native language during childhood; in fact, most are quite competent with language by the time they are 3 or 4 years old. How can children too young to tie their shoes or to understand that 2 plus 2 equals 4 succeed at such a complex task? Most developmentalists agree that language learning requires innate mechanisms that predispose children to it, coupled with an environment that provides adequate models and opportunity to practice. In this section, we will first consider briefly the question of what language is and then chart the normal course of language development. Then, in the next section, we will explore the innate and environmental requirements for language development.

Some Universal Characteristics of Human Language

Just what is it that children learn when they learn a language? Linguists estimate that at least 3,000 separate languages exist in the world today, all distinct enough that the speakers of one cannot understand those of another (Grimes, 2000). Yet these languages are all so fundamentally similar to one another that we can speak of *human language* in the singular (Pinker & Bloom, 1992).

All Languages Use Symbols (Morphemes) That Are Arbitrary and Discrete

28

What are the universal characteristics of morphemes? How do morphemes differ from nonverbal signals?

<<< Every language has a vocabulary consisting of a set of *symbols*, entities that represent other entities. The symbols in a language are called **morphemes**, defined as the smallest meaningful units of a language—that is, the smallest units that stand for objects, events, ideas, characteristics, or relationships. In all languages except the sign languages used by the deaf, morphemes take the form of pronounceable sounds. Most morphemes are words, but others are prefixes or suffixes used in consistent ways to modify words. Thus, in English, *dog* is both a word and a morpheme, *-s* is a morpheme but not a word, and *dogs* is a word consisting of two morphemes (*dog* and *-s*). The word *antidisestablishmentarianism* contains six morphemes (*anti-dis-establish-ment-arian-ism*), each of which has a separate entry in an English dictionary.

In any language, we can distinguish two general classes of morphemes. One class, **content morphemes**, includes nouns, verbs, adjectives, and adverbs—the morphemes that carry the main meaning of a sentence. The other class, **grammatical morphemes**, includes (in English) articles (*a*, *an*, *the*), conjunctions (such as *and*, *but*), prepositions (such as *in*, *of*), and some prefixes and suffixes (such as *re-*, *-ed*). These serve primarily to fill out the grammatical structure of the sentence, although they also contribute to meaning.

Morphemes in any language are both arbitrary and discrete. A morpheme is *arbitrary* in that no similarity need exist between its physical structure and that of the object or concept for which it stands. Nothing



AP Photo / Ng Han Guan

The flexibility of language Any human language—unlike nonverbal forms of communication—can be used to express an infinite number of different ideas and observations. Moreover, any idea or observation that can be expressed in one language can be translated into any other language, although in some cases new words may have to be invented and defined to facilitate the translation.

about the English morpheme *dog*, or the French morpheme *chien*, naturally links it to the four-legged, barking creature it represents. Because morphemes are arbitrary, new ones can be invented whenever needed to stand for newly discovered objects or ideas or to express newly important shades of meaning. This characteristic gives language great flexibility. A morpheme is *discrete* in that it cannot be changed in a graded way to express gradations in meaning. For example, you cannot say that one thing is bigger than another by changing the morpheme *big*. Rather, you must add a new morpheme to it (such as *-er*) or replace it with a different morpheme (such as *huge*).

In contrast, nonverbal signals, used by other animals as well as humans (discussed in Chapter 3), are neither arbitrary nor discrete. Nonverbal signals typically develop from and bear physical resemblance to such actions as fighting or fleeing, and the signals communicate intentions to engage in the actions that they resemble. Moreover, nonverbal signals can be presented in a graded manner. One expresses *more* surprise, anger, or whatever nonverbally by presenting the signal more vigorously or with greater amplitude. In everyday speech, you might communicate that one thing is bigger than another by saying, “This one is big but that one is *big*,” but the vocal emphasis placed on the second *big* is a nonverbal addition. In speech, we commonly mix nonverbal with verbal communication to get a point across (which is one reason why speaking is easier than writing).

All Languages Are Hierarchically Structured in a Similar Way

In addition to commonalities in their symbol systems, all languages share a particular hierarchical structure of units (see Figure 11.6). The top level (largest unit) is the sentence, which can be broken down into phrases, which can be broken down into words or morphemes, which can be broken down into elementary vowel and consonant sounds called **phonemes**. The power of this four-level organization is that the relatively few phonemes (anywhere from 15 to 80 occur in any given language) can be arranged in different ways to produce an enormous number of possible words, which themselves can be arranged in different ways to produce a limitless number of possible phrases and sentences.

Every language has rules—collectively referred to as the **grammar** of the language—that specify permissible ways to arrange units at one level to produce the next higher level in the hierarchy. Grammar includes rules of *phonology*, which specify how phonemes can be arranged to produce morphemes; rules of *morphology*, which specify how morphemes can be combined to form words; and rules of **syntax**, which specify how words can be arranged to produce phrases and sentences. These rules differ from language to language, but every language has them, and similarities exist across all languages in the fundamental nature of the rules (Pinker & Bloom, 1992).

>>> 29
How can any sentence, in any language, be described as a four-level hierarchy? How can rules of grammar be described in relation to that hierarchy?

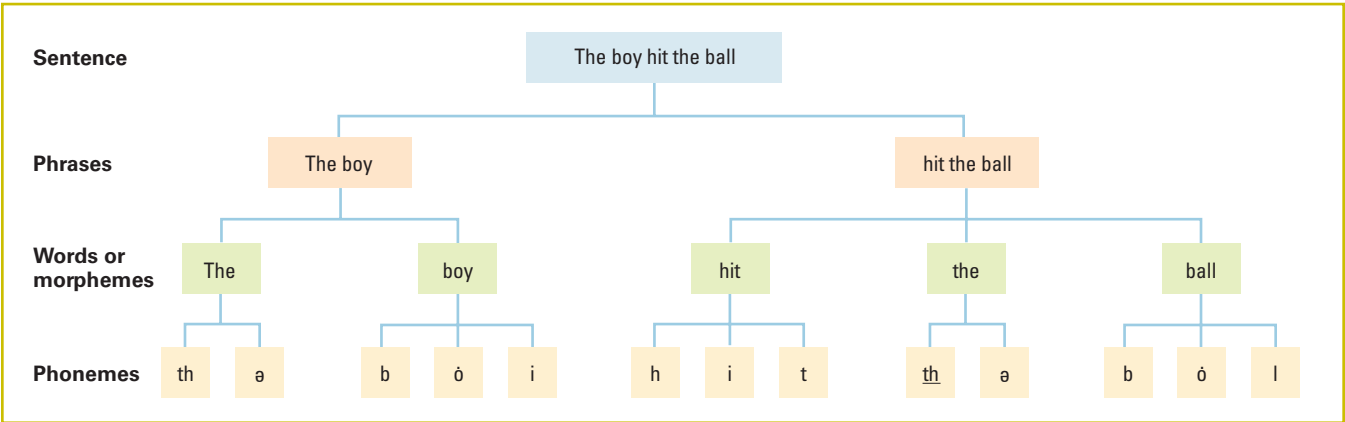
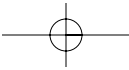


FIGURE 11.6 The hierarchical structure of language These four levels of organization characterize every spoken language.



Grammatical Rules Are Usually Learned Implicitly, Not Explicitly

30

What does it mean to say that knowledge of grammar is usually implicit rather than explicit?

<<< People often think of grammar as something they learned (or tried to learn) in elementary school (which used to be called *grammar* school, perhaps for that very reason). But grammar is learned implicitly, without conscious effort, long before formal schooling. The fact that 4-year-olds can carry on meaningful conversations with adults, producing and understanding new and unique sentences, indicates that by age 4 children have already acquired much of the essential grammar of their native language. Four-year-olds can't name or describe the rules of grammar (nor can most adults), yet they use them every day. Indeed, even professional linguists have yet to describe the full grammar of any of the world's languages (Jackendoff, 2003). Grammatical rules in this sense are like the rules that underlie the sequence and timing of specific muscle movements in walking or running; both sets of rules are generally encoded in *implicit* rather than explicit memory. We generally can't state them, but we use them when we walk, run, or carry on a conversation.

People's implicit knowledge of grammar is demonstrated in their ability to distinguish acceptable from unacceptable sentences. Nearly every English speaker can identify *The mouse crawled under the barn* as a grammatical sentence and *The crawled barn mouse the under* as nongrammatical, although few can explain exactly why. The ability to distinguish grammatical from nongrammatical sentences is not based simply on meaning. As the linguist Noam Chomsky (1957) pointed out, English speakers recognize *Colorless green ideas sleep furiously* as grammatically correct but absurd.

The Course of Language Development

Early Perception of Speech Sounds

31

What is the evidence that humans pay special attention to speech from birth on, perhaps even from before birth?

<<< Infants seem to treat speech as something special as soon as they are born and maybe even before. When allowed to choose between tape-recorded sounds by sucking in different ways on a nipple, 3-day-olds chose to listen to human speech rather than to other sounds such as instrumental music (Butterfield & Siperstein, 1974). In other experiments, babies as young as 2 hours old, who had not heard their mother speak since their birth, chose their own mother's voice over that of another woman (DeCasper & Fifer, 1980; Querleu & others, 1984). The mother's voice is audible in the womb, and during the last few weeks before birth her voice has a quieting effect on the fetus, as indexed by reduced muscle movements (Hepper & others, 1993).

32

How have researchers shown that very young infants can distinguish between subtly different speech sounds? What abilities to distinguish between sounds do infants lose over time, and what is the value of those losses?

<<< The ability of very young infants to hear the differences among speech phonemes has been demonstrated in many experiments. One technique is to allow an infant to suck on a pacifier that is wired to trigger the playing of a particular sound each time a sucking response occurs. When the baby becomes bored with a sound, as indicated by a reduced rate of sucking, the sound is changed (maybe from *pa* to *ba*). Typically, the rate of sucking increases immediately thereafter, which indicates that the infant hears the new sound as different from the previous one. Another method, which can be used with infants 5 months old and older, involves rewarding the baby with an interesting sight for turning his or her head when the sound changes. This method is illustrated and described further in Figure 11.7.

The results of such experiments suggest that babies younger than 6 months old hear the difference between any two sounds that are classed as different phonemes in any of the world's languages. After 6 months, however, they begin to lose the ability to distinguish between sounds that are classed as the same phoneme in their native language but as distinct phonemes in another language (Kuhl, 1987; Werker & Tees, 1999). For example, regardless of what culture they are growing up in, babies under 6 months old can distinguish among the subtly differ-

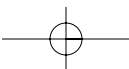




FIGURE 11.7 | Test of an infant's ability to detect changes in a speech sound The baby attends to a toy held by the researcher while a speech sound is played repetitively through a loudspeaker. Occasionally the speech sound changes (perhaps from *pa* to *ba*), and shortly after each such change an attractive toy to the baby's right is lit up and begins to move. The baby soon learns to look to the right each time the sound changes, and this response serves as an index that the baby distinguishes the new sound from the old one. The mother and researcher wear headphones that prevent them from hearing the critical sound and inadvertently cuing the infant to turn. (From Werker & Tees, 1992.)

ent /t/ sounds that constitute different phonemes in Hindi but not in English; but older children and adults in English-speaking cultures do not hear those differences (Werker & Tees, 1999; Werker & others, 1981). Similarly, children growing up in a Japanese culture lose the distinction between the English /l/ and /r/, which belong to the same phoneme category in Japanese (Eimas, 1975). The sharpest loss occurs between 6 and 12 months of age, which is when babies are beginning to recognize and occasionally mimic words in their native language. The loss is best understood as an adaptive change in how the sounds are automatically categorized and heard (Werker & Desjardins, 1995). Presumably, it is useful to hear different phonemes within one's native language as different, so we retain the ability to hear those differences; but it is confusing to hear different versions of the same phoneme as different, so we learn not to hear those differences.

Cooing and Babbling

Beginning at birth, infants can cry and produce various other vocal signs of distress, but at about 2 months they begin to produce a new, more speechlike category of sounds called *cooing*, which consists of repeated drawn-out vowels (*oooh-oooh*, *eeeh-eeeh*). At about 6 months, cooing changes gradually to *babbling*, which consists of repeated consonant-and-vowel sounds such as *paa-paa-paa* or *tooda-tooda* (Masataka, 2003). Cooing and babbling occur most often when the infant is happy. They seem to be forms of vocal play that have evolved to help the infant exercise and refine the complex muscle movements needed to produce coherent speech.

Coos and the earliest babbles do not depend on the infant's hearing spoken sounds. Deaf infants coo and begin to babble at about the same age and in the same manner as hearing infants (Lenneberg, 1969), and early babbles are as likely to contain foreign-language sounds as native-language sounds (Locke, 1983). By 8 months of age, however, hearing infants begin to babble in ways that mimic the rhythm and pitch patterns of the language they hear around them; the babbling of a French baby becomes recognizably French, and that of a British baby becomes recognizably British (de Boysson-Bardies, 1999). Beginning at about 10 months of age, hearing infants produce babbled sounds that increasingly resemble syllables and words of their native language (de Boysson-Bardies, 1999; Locke, 1983). Also, at this age, deaf infants who are exposed to a sign language begin to babble with their hands—repeating over and over hand movements that are similar in form

>>> 33 What is the distinction between cooing and babbling? What are the reasons for thinking that these vocalizations are precursors to language production?

Jeffrey DeBelle / © Dr. Laura Ann Petitto



Manual babbling Deaf or hearing infants whose deaf parents communicate by sign language go through a stage of babbling with their hands. Their babbling gestures resemble the signs of the language but do not yet express meaning.

and rhythm to those of the language they see around them (Petitto & others, 2001). Eventually, recognizable words appear in the hearing infant's vocal babbling and the deaf infant's manual babbling.

Word Comprehension Precedes Word Production

34

What is the evidence that babies understand at least some words before they begin to speak?

<<< During the babbling phase of life, before the first production of recognizable words, infants begin to show evidence that they understand some words and phrases that they hear regularly. In one experiment, 6-month-olds were shown side-by-side videos of their parents. When they heard the word “Mommy” or “Daddy,” they looked reliably more at the video of the named parent than at the unnamed parent (Tincoff & Jusczyk, 1999). Other experiments have revealed that 9-month-olds can respond to a number of common words by looking at the appropriate object when it is named (Balaban & Waxman, 1997) and can follow simple verbal commands, such as “Get the ball” (Benedict, 1979).

Naming and Rapid Vocabulary Development

Typically at about 10 to 12 months, babies produce their first recognizable words (Ganger & Brent, 2004). Such words are most often produced in a playful spirit. The child at first uses words to point things out, or simply to name them for fun, not generally to ask for them (Bloom & Lahey, 1978). For instance, my son used the word *ba* joyfully to name his bottle, after he received it, weeks before he began using it to ask for his bottle when he wanted it.

New words come slowly at first, but then, typically at about 15 to 20 months of age, the rate begins to accelerate. Between the ages of 2 and 17 years, the typical person learns about 60,000 words, an average of about 11 new words per day (Bloom, 2001). Relatively few of these are explicitly taught; most often, the child must infer the meaning of a new word from the context in which others use it. How do young children, early in the language-learning process, draw these inferences?

35

How do young children make the link between new words that they hear and appropriate referents in their environments?

<<< Most of the earliest words learned, in any language culture, are nouns that refer to categories of objects in the child's environment (Golinkoff & others, 1994). Young children's tendency to look at whatever an older person is looking at (discussed earlier) no doubt helps them identify objects that the older person is referring to when speaking. Experiments have shown that infants are especially likely to follow an adult's gaze when the adult is labeling an object in the environment (Baldwin, 2000). In addition, young children seem to have a number of cognitive biases, or built-in assumptions, that help them narrow down the likely referent to a new word

they hear (Golinkoff & others, 1994; Regier, 2003). One of these is a strong tendency to link new words with objects for which they do not already know a name.

Other things being equal, young children assume that a new word is not a synonym for a word they already know but a label for something whose name they don't know. In one experiment, 3- and 4-year-olds were presented with toy animals that they could name (a pig, a sheep, and a cow) plus one that they could not name (a tapir). When they heard the novel word *gombe* in the presence of these objects, all the children applied it to the novel animal (Clark, 1987). Other research indicates that toddlers begin to manifest this bias at about the same time—in their second year of life—at which their rate of vocabulary learning begins to increase rapidly (Mervis & Bertrand, 1994). Although the bias leads to some mistakes, it is apparently much more helpful than harmful to their acquisition of words.

By the time they can understand multiword sentences, young children are able to use their tacit knowledge of grammar to help them infer the meaning of new words, including verbs and other parts of speech as well as nouns (Fisher, 2000). Thus if they are shown a videotaped scene and told, “The duck and bunny are biffing,” 2-year-olds—who understand *duck* and *bunny* and know implicitly that words ending in *-ing* refer to actions—infer that *biffing* means whatever the duck and bunny are both doing (Naigles, 1990). If told, instead, “The duck is biffing the bunny,” they infer that *biffing* means whatever the duck is doing to the bunny. As another example, 2-year-olds who heard “Mommy feeds the ferret” inferred that a ferret is an animal, not an inanimate object (Goodman & others, 1998).



Laura Dwight / Peter Arnold

First words Naming objects is an important step in the mastery of language. It is also a source of great delight to infants and their caregivers. Here 1-year-old Genevieve uses her favorite word, *light*.

Extending Words to Fit Appropriate Categories

In addition to linking a new word to its immediate referent, children must learn how to *extend* it to new referents. Common nouns such as *ball* refer to categories of objects, and a full understanding is demonstrated when a child applies the word to all members of the category and not to nonmembers. Researchers have found that young children, including even infants as young as 12 months, behave as though they assume that a newly heard label applies not just to the specific object that has been labeled but also to other objects that are perceptually like the original one (Golinkoff & others, 1995; Waxman & Markow, 1995). In other words, infants are biased toward assuming that labels are common nouns, not proper nouns. The bias leads to some mistakes, as when a child refers to all men as *Daddy*, but is useful overall because the vast majority of nouns to be learned are common nouns.

By the time they are 2 years old, children can use the grammatical context of a sentence to discern whether a name for an object is a proper name or a category name. In one experiment, 24-month-old children were shown a stuffed animal and were told either “This is *a* ZAV” or “This is ZAV.” Those in the first group subsequently applied the label to other stuffed animals that looked like the original one, and those in the second group did not (Hall & others, 2001).

Children sometimes *overextend* common nouns, using them more broadly than adult usage would allow. On the basis of an analysis of a large collection of examples, Eve Clark (1973) proposed that overextension results when a child implicitly defines a new word in terms of just one or a few of the prominent features of the original referent object. Thus a child who hears *ball* in association with a specific ball might take the object's most prominent feature—its roundness—as the defining characteristic and subsequently refer to all round objects, including oranges and full moons, as balls. In other cases, overextensions may not represent errors at all but simply derive from children's attempts to communicate about objects that they have not yet learned to name (Clark, 1995). A toddler who says, “Look doggie,”

>>> **36**
How do infants “decide” whether a new name that they hear is a proper noun or a common noun?

>>> **37**
What are two reasons why children might overextend common nouns that they have learned?

Ohh, Kermit. When young children overextend nouns and proper nouns to apply to objects beyond the appropriate category they may not always be making mistakes. A child who sees an alpine marmot for the first time and says, “Ohh, Kermit,” may really be saying, “Ohh, there is an animal that looks like Kermit the Frog from Sesame Street.” The drive to speak and communicate is so strong that, lacking the appropriate word, children will come up with the closest approximation they know.



blickwinkel / Alamy

while pointing to a cat, may in essence be saying, “Look at that thing that is something like a doggie.” The same child, when asked which animals in a set of pictures are dogs, may pick the actual dogs and not the cats (Thomson & Chapman, 1977).

Children also sometimes *underextend* words, applying them to more narrowly defined categories than do adults. In fact, some research suggests that underextension is more common than overextension (MacWhinney, 1998). But underextensions do not result in errors in children’s spontaneous speech, so they are difficult to detect and have been little studied.

The remarkable fact, however, is not that children make errors in their extensions but that they usually extend new words appropriately (Bloom, 2001). Somehow they figure out quickly the categories that are referred to by the words they hear adults and older children using.

Using Grammatical Rules

38 How do children demonstrate knowledge of grammatical rules in their early speech? How do some of their “mistakes” in grammar confirm that they know the rule and are not just mimicking?

<<< All children go through a relatively prolonged period during which each of their utterances is only one word long. When they do begin to put words together, typically at about 18 to 24 months of age, they at first use content words almost exclusively, especially nouns and verbs, and usually arrange them in the grammatically correct sequence for simple, active sentences (Brown, 1973). For an English-speaking child, this means that subjects are placed before verbs, and verbs before objects. A child at the two-word stage will say “Billy kick” to mean that Billy is kicking something, and “Kick Billy” to mean that someone is kicking Billy.

When children acquire a new grammatical rule, such as adding *-ed* to the end of a verb to create the past tense, they almost invariably *overgeneralize* it at first (Kuczaj, 1977; Marcus & others, 1992). The 3-year-old who says “kicked,” “played,” and “laughed” also says “goed,” “thinked,” and “swimmed.” Similarly, children who have just learned to add *-s* to pluralize nouns will talk about many *mouses*, *sheeps*, and *childs*. This overgeneralization confirms that children really know the rule. If they followed the rule only when adults did, their usage might be attributed to simple imitation. As further evidence suggesting that their grammar is based on rules, young children have been shown to use the rules with made-up words that they had never heard before, as illustrated in Figure 11.8.

Children are not taught the rules of grammar explicitly; nobody sits a 2-year-old down and tries to explain how to create infinitives, possessives, or past-tense

FIGURE 11.8 One wug and two _____. With this test, Jean Berko found that children who had just begun to use the rule of forming plurals by adding *-s* would use the rule correctly even for words they had never heard before. (From Berko, 1958.)



This is a wug.



Now there is another one.
There are two of them.
There are two ____.

verbs. Some parents correct their children's grammar, but even this is rare (Brown & Hanlon, 1970), and long-term experiments in preschools have shown that deliberate programs of correcting grammar have little effect on rule acquisition (de Villiers & de Villiers, 1979). Through their own devices, children actively (and mostly unconsciously) infer grammatical rules from examples of rule-based language spoken around them and to them.

SECTION REVIEW

Children rapidly acquire the complex tools of language beginning in infancy.

Linguistic Universals • All languages include a set of symbols called morphemes, which are a language's smallest meaningful units; morphemes are arbitrary and discrete. • All languages are hierarchically structured, with sentences at the top of the hierarchy and phonemes (elementary vowel and consonant sounds) at the bottom. • Every language has a grammar—a set of rules that specify the permissible ways to combine units at one level of the hierarchy to create a unit at the next higher level. • A person's knowledge of the grammar of his or her language is generally more implicit than explicit.	The Course of Language Development • Babies under 6 months can distinguish any two phonemes of any language, but after 6 months they retain only the ability to hear distinctions that represent different phonemes in their native language. • Infants coo and later babble as a form of vocal play that helps to prepare the vocal musculature for speech. By 8 months, their babbling starts to mimic their native language. • The first recognizable words appear at about 10 to 12 months; vocabulary growth accelerates soon after and continues for years, aided by innate biases and knowledge of grammar, though children sometimes over- or underextend words. • Children first combine words at about 18 to 24 months, demonstrating knowledge of word-order rules. Knowledge of other grammatical rules is demonstrated in overgeneralizations of them (such as saying <i>deers</i> or <i>goed</i>).
---	--

Internal and External Supports for Language Development

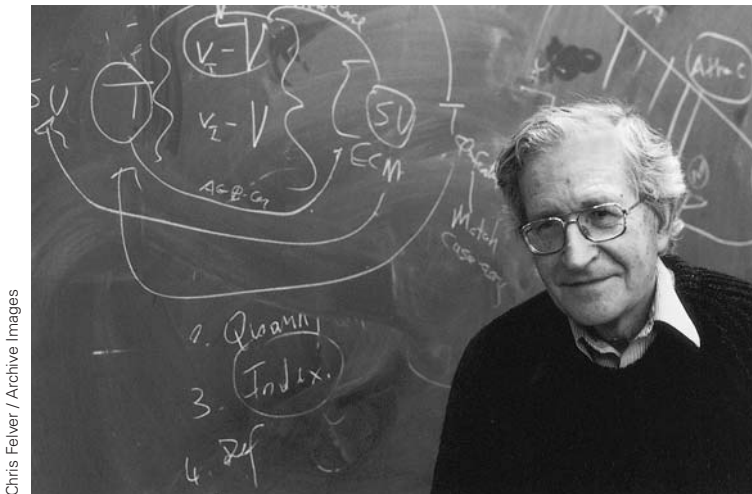
How is it that children can learn language so early, with so little apparent conscious effort and without deliberate training? There is no doubt that humans enter the world equipped in many ways for language. We are born with (a) anatomical structures in the throat (the larynx and pharynx) that enable us to produce a broader range of sounds than any other mammal can produce; (b) brain areas specialized for language (including Broca's and Wernicke's areas, discussed in Chapter 5); (c) a preference for listening to speech and an ability to distinguish among the basic speech sounds of any language; and (d) mechanisms that cause us to exercise our vocal capacities through a period of cooing and babbling regardless of experience. There is also no doubt that most of us are born into a social world that provides rich opportunities for learning language. We are surrounded by language from birth on, and when we begin to use it, we achieve many rewards through this extraordinarily effective form of communication.

The Idea of Special Inborn Mechanisms for Learning Language

Chomsky's Concept of an Innate Language-Learning Device

The linguist Noam Chomsky, more than anyone else, was responsible for drawing psychologists' attention to the topic of language. In his highly influential book *Syntactic Structures* (1957), Chomsky characterized grammatical rules as fundamental

39
How did Noam Chomsky link the study of grammar to psychology? What did he mean by a language-acquisition device?



Chris Felver / Archive Images

Noam Chomsky In his book *Syntactic Structures*, Chomsky called attention to the idea that producing a verbal statement is a creative act guided by implicit mental rules. This work and the revolution in psycholinguistics that it inspired have enriched research and theory in cognitive psychology.

vary from one language to another, they are all, according to Chomsky, based on certain fundamental principles, referred to as *universal grammar*, that are innate properties of the human mind. These properties account for the universal characteristics of language (discussed earlier) and for other, more subtle language universals (Pinker, 1994). To refer to the entire set of innate mental mechanisms that enable a child to acquire language quickly and efficiently, Chomsky coined the term **language-acquisition device**, or **LAD**. The LAD includes the inborn foundations for universal grammar plus the entire set of inborn mechanisms that guide children's learning of the unique rules of their culture's language. Support for the concept of an innate LAD comes partly from observations of language-learning deficits in people who have suffered damage to particular brain areas (discussed in Chapter 5) or who have a particular genetic disorder (specific language impairment, discussed in Chapter 3).

Children's Invention of Grammar

Further support for the LAD concept comes from evidence that young children invent grammar when it is lacking in the speech around them. Here are two examples of such evidence.

40 How have studies of creole languages and studies of deaf children in Nicaragua supported the idea that children invent grammar in the absence of a preexisting grammatical language?

Children's Role in Development of Creole Languages New languages occasionally arise when people from many different language cultures simultaneously colonize an area and begin to communicate with one another. These first-generation colonists communicate through a primitive, grammarless collection of words taken from their various native languages—a communication system referred to as a **pidgin language**. Subsequently, the pidgin develops into a true language, with a full range of grammatical rules, at which point it is called a **creole language**. Derek Bickerton (1984) studied creole languages from around the world and found that at least some of them are developed into full languages within one generation by the children of the original colonists. Apparently, the children imposed grammatical rules on the pidgin they heard and used those rules consistently in their own speech—powerful evidence, in Bickerton's view, that children's minds are innately predisposed to grammar.

Deaf Children's Invention of a Grammatical Sign Language Bickerton's evidence of children's role in imposing grammar on new languages is necessarily indirect, as he was studying languages that emerged many years ago. More recently, Ann Senghas and her colleagues have documented the emergence of a new sign language among deaf children in Nicaragua (Senghas & Coppola, 2001; Senghas & others, 2004). Prior to 1977, deaf Nicaraguans had little opportunity to meet other deaf people. There was no deaf community or common sign language, and the deaf were typically treated as though they were retarded. In 1977 the first Nicaraguan school for the deaf was founded, so deaf children for the first time came into ex-



Ann Senghas

Creating a language When deaf children in Nicaragua were brought together for the first time, in a school community, they gradually created a new sign language. The youngest children contributed most to the grammatical structure of the language.

tended contact with one another. On the basis of outmoded ideas, the schools did not at first teach any sign language but, instead, attempted to teach the deaf to speak and lip-read the nation's vocal language (Spanish), an approach that rarely succeeds. Despite the official policy, the students began to communicate with one another using hand signs.

At first their signing system was a manual pidgin, an unstructured and variable amalgam of the signs and gestures that the individuals had been using at home. But over the course of a few years the signs became increasingly regularized and efficient, and a system of grammar emerged. All this occurred naturally, with no formal teaching, simply through the students' desire to communicate with one another. Of most significance for our discussion, the new grammar was produced not by the oldest, wisest members of the community but by the youngest. In fact, those who were more than about 10 years old when the deaf community was formed not only failed to contribute to the development of a grammar but learned little of the one that eventually did develop. The sign language invented by the children has since become the official sign language of Nicaragua. It is a true language, comparable to American Sign Language, in which the morphemes are elementary hand movements and grammatical rules stipulate how the morphemes can be combined and sequenced into larger units.

Children's improvement of grammar may be viewed by us as a creative act, but it is almost certainly not experienced as that by the children themselves. Children tacitly assume that language has grammar, so they unconsciously read grammar into language even where it doesn't exist. Just as children learning English overgeneralize grammatical rules when they say things like *goed* and *mouses*, and thereby temporarily make English grammar more consistent and elegant than it really is, children exposed to a pidgin language or to an early version of the Nicaraguan sign language may accept the slightest random hint of a grammatical rule as a true rule and begin to use it as such. In the case of a newly emerging language—unlike the case of a long-established language such as English—the children's rules are incorporated into the language and are learned by the next generation.

Critical Period for Learning of First Language

The LAD apparently functions much more effectively during the first 10 years of childhood than later in life. Children who are deprived of the opportunity to hear and interact with a language during their first 10 years have great difficulty learning language later on, and never master the grammar of the language they learn.

>>> 41
What evidence supports the view that grammar is learned more readily in early childhood than later in life?

That is one of the lessons learned from deaf children who were not exposed to a true language until adolescence or late in childhood (Mayberry & others, 2002; Senghas & others, 2004). Much more rarely, a hearing person is discovered who was deprived of language throughout childhood, and in every such case that has been documented, the person was subsequently unable to learn language fully (Curtiss, 1989).

The most thoroughly studied language-deprived hearing person is a woman known as Genie. She was rescued in 1970, at age 13, from the inhuman conditions in which her deranged, domineering father and partly blind mother had raised her (Curtiss, 1977; Rymer, 1993). From shortly after birth until her rescue, Genie had been locked in a tiny bedroom and exposed to almost no speech. At the time of her rescue she understood a few words but could not string words together and had learned no grammar. She was then placed in a foster home where she was exposed to English much as infants normally are and received tutorial help. In this environment, she eventually acquired a large vocabulary and learned to produce meaningful, intelligent statements; but even after 7 years of language practice, at age 20, her grammar lagged far behind other indices of her intelligence (Curtiss, 1977). A typical sentence she produced was “I hear music ice cream truck,” and she often misunderstood sentences whose meaning depended on grammar. Over subsequent years, her grammar showed little improvement (Rymer, 1993).

Learning within the critical period is much less crucial for second-language learning than for first-language learning. People who learned their first language within the critical period can learn a second language reasonably well at any time in life, although not quite as well as they could have if they had learned it earlier. Those who learn a second language after the age of about 10 or 11 almost always speak with an accent and generally do not acquire the grammar of the language as fully or easily as do those who learn the language at a younger age (Johnson & Newport, 1989; Au & others, 2002).

The Language-Acquisition Support System

42

How do parents in our culture modify their speech to infants?

<<< Children come into the world predisposed to acquire language, but they do not acquire it in a social vacuum. Neither Genie nor any other child has ever invented language alone. Normal language development requires not just the LAD but also the **LASS**—the **language-acquisition support system**—provided by the social world into which the baby is born (Bruner, 1983).

In our culture and many others, adults regularly simplify their speech to infants and young children in ways that might help the children learn words and some aspects of grammar. They enunciate more clearly than when speaking to adults

Taking account of the listener

Most people automatically simplify their speech when talking to infants and young children. Even young children addressing younger children deliberately slow their rate of speech, choose simple words and grammatical structures, and gesture broadly.



Lew Merrin / Monkneyer

(Kuhl & others, 1997), use short sentences that focus on the here and now, repeat salient words, and use gestures to help convey meaning (Snow, 1984). A 6-month-old playing with a ball might be told, “Oh, you have a *ball*. A nice *ball*. What a pretty *ball*.” Such speech is often referred to as *motherese* (although it is not just mothers who speak this way to babies). Adults also frequently treat infants’ early vocalizations *as if* they were verbal statements. For instance, a mother or father might respond to the baby’s *ba-ba-ga-goo-goo* with, “Oh, yes, tell me more!” Such responsiveness can lead to back-and-forth, conversation-like exchanges between infant and adult.

Mothers’ Speech to Infants and Rates of Language Acquisition

Many studies have revealed correlations between mothers’ styles of communicating with their infants and the rates at which the infants acquire language. An infant’s rate of language acquisition has been found to correlate positively with (a) the extent of the mother’s verbal responsiveness to the infant’s early vocalizations (Tamis-LeMonda & others, 2001); (b) the degree to which the mother uses simplified motherese during early stages of language learning (Furrow & others, 1979); and (c) the total amount of time that the mother speaks to her child and the number of different words she uses in later stages of language learning, at around age 2 and later (Hoff & Naigles, 2002). A rarely acknowledged problem with such studies, however, is that the observed correlations might derive more from genetic similarities between mothers and their children than from differences in the language environments that the different mothers provide. Genetically verbal mothers may produce genetically verbal children.

I have been able to find only one study that separated environmental from genetic factors in mothers’ influences on children’s language, a study involving infants who were adopted at birth (Hardy-Brown & others, 1981). In that study, adopted infants’ rates of language development correlated more strongly with their biological mothers’ verbal abilities than with their adoptive mothers’ verbal abilities, but the linguistic environments provided by the adoptive mothers also played a significant role. In particular, infants whose adoptive mothers often responded verbally to their early vocalizations developed language more rapidly than did those whose adoptive mothers were less responsive.

>>> 43
What evidence suggests that differences in the language environments provided by mothers may affect the rates at which infants acquire language?

Cross-Cultural Differences in the LASS

Cross-cultural research shows that children all over the world acquire language at roughly the same rate, despite wide variations in the degree and manner of adults’ verbal interactions with infants (Ochs & Schieffelin, 1995). Bambi Schieffelin and Elinor Ochs (1983) found, for example, that the Kalikuli people of the New Guinea rain forest believe that there is no reason to speak to babies who cannot yet speak themselves. These babies hear no motherese, and little speech of any kind is directed toward them. However, they go everywhere with their mothers and constantly overhear the speech of adults and children around them, and they begin to speak at about the same age that children in our culture do. Apparently, large variations can occur in the LASS without impairing infants’ abilities to learn language.

>>> 44
What light has been shed on the LASS by cross-cultural research?

Language Learning by Non-Human Apes

To what degree does language depend on special, inborn, language-learning mechanisms that are unique to humans, and to what degree does it depend on more general learning abilities that we share with other animals? That question has motivated much research on the abilities of non-human animals to learn human-like languages.

Our closest relatives, chimpanzees and bonobos (discussed in Chapter 3), have brains that are structurally quite similar to ours. Although neither species communicates in its natural environment with anything like the symbol-based, grammar-

>>> 45
What has motivated researchers to study the ability of non-human animals to learn human languages? Why did the Gardners try to teach their chimp, Washoe, a sign language rather than spoken English?



Manuela Hartling / Reuters / Corbis

Fetch the cuddly This border collie, Rico, has been trained to fetch at least 200 different objects by name. When presented with a new name, which he hasn't heard before, he will pick the one new object, not an object that he has already been taught to fetch (Kaminski & others, 2004). Here he fetches a pokemon cuddly.

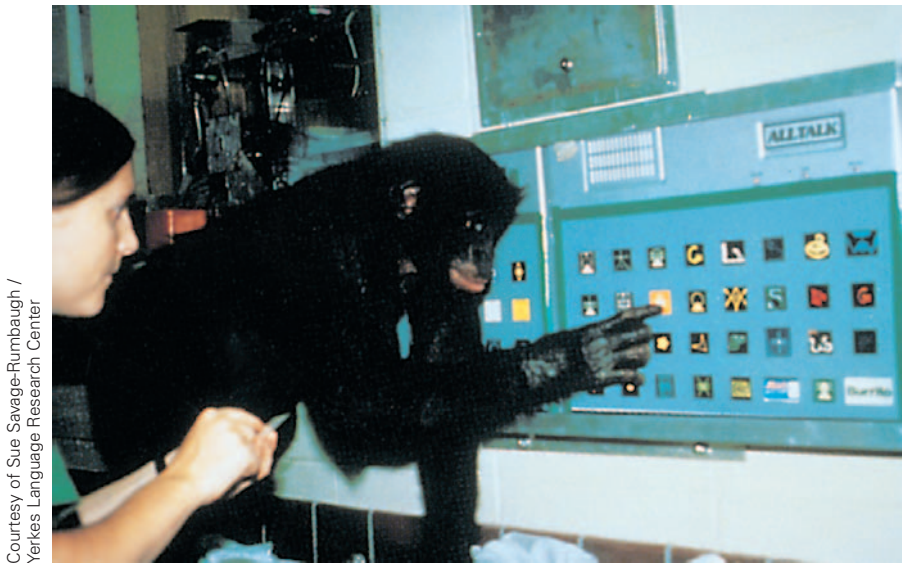
chimp named Washoe in the constant presence of people who directed language toward her as well as toward one another. Because the Gardners hoped that Washoe would learn to produce as well as understand language, they and Washoe's other caretakers used a modified version of American Sign Language rather than a vocal language. Chimpanzees lack the vocal apparatus needed to produce the sounds of human speech, but they have flexible fingers capable of producing manual signs. Washoe's success in acquiring some of the elements of language inspired many subsequent ape-language projects. Researchers have since studied language learning in a number of other chimpanzees, several bonobos, a gorilla, and an orangutan (Miles, 1983; Patterson & Linden, 1981; Savage-Rumbaugh & Fields, 2000).

The Accomplishments of Kanzi

46 What LASS was supplied for the bonobo Kanzi? What are Kanzi's linguistic accomplishments?

<<< The most linguistically accomplished ape to date is a bonobo named Kanzi, whose learning has been nurtured and documented by Sue Savage-Rumbaugh and her associates. Before working with Kanzi, Savage-Rumbaugh had been using operant conditioning to teach chimpanzees to communicate through an invented language in which the words are geometric figures, called *lexigrams*, that are arranged on a keyboard (see Figure 11.9). In 1981, Savage-Rumbaugh was, with little success, training Kanzi's mother in the lexigram system when she noticed that young Kanzi, who had been allowed to run free in the laboratory, was learning to use some of the lexigrams correctly just by watching, even though he wasn't being rewarded. This inspired Savage-Rumbaugh to try a different method with Kanzi: immersion in a language culture rather than systematic training. Kanzi was continuously free, but not forced, to communicate, using a stationary keyboard in the lab and a portable one outdoors. Kanzi's caregivers used both spoken English and lexigrams with and around him, and they responded to Kanzi's lexigrams and gestures as parents might to the communicative attempts of their children. Their task was not to teach Kanzi language, but to communicate with him as best they could as they and he went about their daily activities. My inference from the reports and films on Kanzi is that he received even more linguistic attention and heard even more motherese than does a typical human child.

Kanzi soon learned to use all the lexigrams appropriately, and his behavior indicates that he does not simply use them to get rewards (Savage-Rumbaugh & others, 1986, 1993). For example, he uses the lexigram for *apple* after he already has an apple. He also often announces his intentions before acting, even though he doesn't need permission to act. Before taking a trip to his tree house, he presses the lexigram for *tree house*. In addition, he regularly combines lexigrams with nonverbal gestures in order to get across more complex ideas than would be possible with the limited set of lexigrams on his board. For example, to get caretaker A to chase caretaker B, he presses the lexigram for *chase* and then pushes A's hand in the direction of B.



Courtesy of Sue Savage-Rumbaugh / Yerkes Language Research Center

FIGURE 11.9 | Kanzi using the lexigram keyboard Each symbol on the keyboard is a word. When Kanzi, the bonobo, depresses a key, the symbol is illuminated both on the keyboard and on a larger screen that his caretakers can see. Kanzi also has a portable keyboard that he carries with him as he roams a forested area with his caretakers.

Even more impressive than Kanzi's use of lexigrams is his understanding of spoken English. His behavior suggests that he knows the meanings of at least 500 English words (Savage-Rumbaugh & Fields, 2000). When Kanzi was 8 years old, he was tested for his ability to carry out spoken requests that he had never heard before, such as "Put the egg in the noodles" and "Hammer the ball." The requests were made by a caretaker who stayed out of view so that nonverbal cues could not guide Kanzi's behavior. Kanzi carried out most of the requests successfully, performing at a rate comparable to that of a 2½-year-old human girl who was given the same requests (Savage-Rumbaugh & others, 1993). Kanzi also responded appropriately to requests whose meaning depended on word order, such as "Make the doggie bite the snake" and "Make the snake bite the doggie" (using a toy dog and toy snake as props). More recently, two other bonobos and a chimpanzee in Savage-Rumbaugh's care have been exposed to language in the same manner that Kanzi was, and their abilities are reported to approach that of Kanzi (Savage-Rumbaugh & Fields, 2000).

Tentative Conclusions from the Ape-Language Studies

Has Kanzi or any other ape learned human language? Some scholars argue about that as if it were a meaningful question, but it isn't. The answer depends entirely on how one chooses to define *human language*. A more sensible question is this: What aspects of human language have non-human apes been able to acquire, and what aspects, at least so far, seem beyond them? Taking all the studies together, apes appear to be relatively adept at acquiring a vocabulary, and systematic tests suggest that they can use signs and lexigrams as true symbols, standing for referents, and not merely as operant responses to obtain rewards (Sevcik & Savage-Rumbaugh, 1994). Moreover, Kanzi can apparently use word order to decipher the meanings of some simple multiword sentences. However, on the basis of the published research, there is little or no evidence that any ape to date has acquired or invented a rule for distinguishing plural from singular nouns, for marking the tense of verbs, or for marking any words by grammatical class. Apparently, the brain mechanism that makes grammar so easy and natural for human children came about in our evolution sometime after we split off from the line leading to chimpanzees and bonobos.

To me, the most interesting general conclusion from the ape-language research has to do with the conditions that optimize learning. Attempts at deliberate, focused training, involving rewards for correct language usage, have generally failed (for example, Terrace, 1985). In those cases the animals learned to produce word-

>>> **47**
What conclusions may be tentatively drawn from the ape-language studies?

like gestures as ways to get rewards rather than truly as symbols to label their experiences. Success has occurred when the focus was on communication rather than training. It has occurred when the researchers entered the ape's world for long periods and brought the ape into their world, using words and sentences along the way (Shanker & King, 2002). This is similar to the conditions in which most human children learn language. The function of language is communication, and a major prerequisite for acquiring language is a desire to communicate.

SECTION REVIEW

Inborn mechanisms and the social context jointly support children’s acquisition of language.

Innate Mechanisms for Language Acquisition

- Chomsky hypothesized the existence of an innate language-acquisition device (LAD), consisting of a universal grammar and mechanisms that guide native-language learning.
- The reality of the LAD is supported by children’s imposition of grammatical rules to create creole languages and, in Nicaragua, a sign language.
- The Lad appears to function most effectively in the first 10 years of life. Children deprived of language during that period do not fully learn language later.

External Support for Language Acquisition

- The social context provides children with a language-acquisition support system (LASS).
- Caregivers often assist language acquisition by speaking motherese and by being responsive to early linguistic efforts.
- Children acquire language at roughly the same rate everywhere, despite wide cross-cultural variation in the LASS.

Language Learning by Apes

- Washoe, a chimpanzee, successfully learned some American Sign Language, inspiring other ape-language studies.
- Kanzi, a bonobo that has learned to use lexigrams and gestures to express meaning, understands at least 500 English words and uses word order to interpret meaning.
- Successful efforts like those involving Washoe and Kanzi have not used systematic training but rather immersion in an environment rich with linguistic communication, like those that children experience.
- Apes are much better at acquiring vocabulary than grammar.

Concluding Thoughts

Thinking and talking are—to use a term introduced in Chapter 3—*species-typical* activities of humans. We are biologically predisposed for thought and language, just as we are predisposed to smile and walk on two legs. The human brain has been specialized for these activities in the course of evolution; people everywhere, in every culture, think and talk in ways that are recognizably human and distinguishable from the activities of any other species. Yet, despite their universality, these abilities must develop anew in each human being.

1. Why thought and language must develop Why aren't these species-typical abilities fully present at birth? An immediate answer is that birth occurs in humans at a relatively early stage in the body's growth. The brain, including the pathways required for thought and language, continues to mature after birth. A more profound answer, however, lies in the ultimate functions of thought and language. Thought serves to make sense of the environment so that we can navigate safely through it and use parts of it to promote our survival and reproduction. We can survive in a wide range of conditions because we can think of ways to modify them to suit our needs. To do so, we must consider not just the universals of every environment, such as the fact that unsupported objects fall to the earth, but also the par-

ticulars of our specific environment, such as the foods that are available and the tools that are needed. Similarly, to communicate effectively in our particular cultural group, we must acquire the specific linguistic elements (the words and grammar) that have emerged in our group's history and represent concepts that are crucial to survival in that group. In other words, to be effective, thought and language must be fine-tuned to the unique physical and social environment in which each individual must survive, and such adjustment can be accomplished only when thought and language develop in that environment.

2. Innate foundations for development Evolution could not endow human beings with the specific knowledge and skills needed to survive in every human environment, but it could and did endow the species with a solid foundation for acquiring them. This theme, of innate foundations for development, can help you tie together many of the ideas and research findings that you have read about in this chapter. As you review, consider the following questions: What inborn assumptions may help each new person make sense of his or her physical, social, and linguistic worlds? What universal aspects of babies' and children's explorations and play may help them learn about the particulars of their native environments and languages?

The findings and ideas to which you can relate these questions include (a) infants' preferences for novelty, their mode of examining objects, their drive to control their environment, and their knowledge of core physical principles; (b) Piaget's ideas about the roles of action, assimilation, and accommodation in development, and other researchers' theories about domain-specific development; (c) young children's understanding of other people's minds, their delayed understanding of false beliefs, and the possible role of make-believe in acquiring that understanding; (d) the roles of early attention to language, of vocal play, and of inborn biases and drives affecting vocabulary and grammar acquisition; and (e) Chomsky's concept of a language-acquisition device, and other researchers' studies of children who invented language.

3. Social supports for development The innate tendencies and drives that promote development are useless without a re-

sponsive environment. Babies need solid surfaces against which to exercise their muscles in order to learn to walk, objects to explore in order to learn about the physical world, and people to listen to and with whom to exercise their linguistic play in order to learn to talk. Young children may, as Piaget contended, choose what to assimilate from the smorgasbord of information around them, but adults help provide the smorgasbord. As you review the chapter, consider the specific ways in which the child uses and is influenced by the social environment. In particular, consider (a) infants' employment of mimicry, joint visual attention, and social referencing to guide their explorations; (b) Vygotsky's theory of the roles of language, dialogue, and cooperative action in the development of thought; and (c) the role of the language-acquisition support system in language development. In the next chapter, on the development of social relationships, you will read much more about ways in which variation in the social environment can influence development.

Further Reading

ROBERT S. SIEGLER & MARTHA W. ALIBALI (2005). *Children's thinking* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

This is a well-written midlevel paperback textbook on cognitive development primarily from an information-processing perspective. Three early chapters are devoted to three theories of development—Piaget's theory, information-processing theory, and sociocultural theory. These are followed by chapters on the development of perception, language, memory, concepts, social cognition, problem solving, and academic skill.

KANG LEE (Ed.) (2000). *Childhood cognitive development: The essential readings*. Malden, MA: Blackwell.

This is a collection of 19 articles published by contemporary developmental psychologists. They were selected because they are representative of current theory and research in the field of cognitive development and are relatively easy for nonspecialists to read. Included

are articles on children's development of perception, motor control, attention, language, understanding of minds, mathematics, scientific reasoning, physical knowledge, and moral understanding.

JEAN PIAGET (1929). *The child's conception of the world*. London: Routledge & Kegan Paul.

Many of Piaget's books are difficult to read, but this one, written early in his career, is an exception. In the introduction, he spells out his method of learning about children's understanding through interviewing them. The book is about children's thoughts on such issues as where the names of things come from, where rain comes from, and what it means to think. The book provides a historical foundation not just for Piaget's subsequent work but for the whole subfield of psychology concerned with cognitive development.

LEV S. VYGOTSKY (1934; reprinted 1962). *Thought and language* (E. Haufmann &

G. Vaker, Eds. and Trans.). Cambridge, MA: MIT Press.

This book was first published in the Soviet Union in 1934, shortly after Vygotsky's death from tuberculosis at age 38, but was not translated into English until 1962. It is a fascinating collection of papers in which Vygotsky sets forth his theory that the internalization of language is critical to the development of human thought.

STEVEN PINKER (1994). *The language instinct*. New York: Morrow.

Pinker is not only a leading expert on language but a gifted user of it. In delightfully clear and humorous prose, he elaborates on essentially all the ideas about language and its development that are touched on in this chapter, plus much more. Have you ever wondered why baseball announcers say the batter "fled out" the last time up, rather than "flew out"? Pinker answers this and dozens of other questions about language.

FPO



Social Development

INFANCY: USING CAREGIVERS AS A BASE FOR GROWTH

- Attachment to Caregivers
- Cross-Cultural Differences in Infant Care

CHILDHOOD I: CONTINUING INTERACTIONS WITH CAREGIVERS

- Behavioral and Emotional Foundations for Morality
- Parents' Styles of Discipline

CHILDHOOD II: ROLES OF PLAY AND GENDER IN DEVELOPMENT

- Developmental Functions of Play
- Gender Differences in Social Development

ADOLESCENCE: BREAKING OUT OF THE COCOON

- Shifting from Parents to Peers for Intimacy and Guidance
- Increased Rates of Recklessness and Delinquency
- An Expanded Moral Vision and Moral Sense of Self
- Sexual Explorations

ADULTHOOD: FINDING SATISFACTION IN LOVE AND WORK

- Love
- Employment
- Growing Old

CONCLUDING THOUGHTS

The natural human environment is a social environment. We are adapted to survive with the help of others. Evolution has endowed us with brain mechanisms that help us make the kinds of connections with other people that are essential to our survival and reproduction.

Over the span of our lives we are involved continuously in interpersonal relationships that sustain, enhance, and give meaning to our existence. As *infants* we depend physically and emotionally on adult caregivers. As *children* we learn to get along with others and to abide by the rules and norms of society. As *adolescents* we begin to explore romantic relationships and, in other ways, seek a niche in the adult world. As *adults* we assume responsibility for the care and support of others and contribute, through work, to the broader society.

Social development refers to the changing nature of our relationships with others over the course of life. What characterizes our ties to other people at each phase of life? How do those relationships promote our survival and influence our subsequent development? How variable is social development from culture to culture and between males and females? These are some of the principal questions of this chapter, which begins with infancy and proceeds, section by section, through childhood, adolescence, and adulthood.

Infancy: Using Caregivers as a Base for Growth

In the mid-twentieth century, the psychoanalytically oriented child psychologist Erik Erikson (1963) developed an elaborate theory of social development. The theory posited that each stage of life is associated with a particular problem or set of problems to be resolved through interactions with other people. It also posited that the way the person resolves each problem influences the way he or she approaches subsequent life stages. In infancy, according to Erikson, the primary problem is that of developing a sense of *trust*, that is, a secure sense that other people, or certain other people, can be relied upon for care and help.

Other theorists, too, have focused on the infant's need for care and on the psychological consequences of the manner in which care is provided. One such theorist was John Bowlby, a British child psychologist who, in the mid-twentieth century, brought an evolutionary perspective to bear on the issue of early child development. Bowlby (1958, 1982) contended that

the emotional bond between human infant and adult caregiver—especially the mother—is promoted by a set of instinctive tendencies in both partners. These include the infant's crying to signal discomfort, the adult's distress and urge to help on hearing the crying, the infant's smiling and cooing when comforted, and the adult's pleasure at receiving those signals.

Human infants are completely dependent on caregivers for survival, but, consistent with Bowlby's theory, they are not passively dependent. They enter the world biologically prepared to learn who their caregivers are and to elicit from them the help they need. By the time they are born, babies already prefer the voices of their own mothers over other voices (discussed in Chapter 11), and shortly thereafter they also prefer the smell and sight of their own mothers (Bushnell & others, 1989; Macfarlane, 1975). Newborns signal distress through fussing or crying. By the time they are 3 months old, they express clearly and effectively their interest, joy, sadness, and anger through their facial expressions, and they respond differentially to such expressions in others (Izard & others, 1995; Lavelli & Fogel, 2005). Through such actions, infants help build an emotional bond between themselves and those on whom they most directly depend, and then they use those caregivers as a base from which to explore the world. In the 1950s, Bowlby began to use the term *attachment* to refer to such emotional bonds, and since then the study of infants' attachment to caregivers has become a major branch of developmental psychology.

Attachment to Caregivers

At about the same time that Bowlby was beginning to write about attachment, Harry Harlow initiated a systematic program of research on attachment with rhesus monkeys. Harlow controlled the monkeys' living conditions experimentally to learn about the developmental consequences of raising infants in isolation from their mothers and other monkeys. Most relevant to our present theme are experiments in which Harlow raised infant monkeys with inanimate surrogate (substitute) mothers.

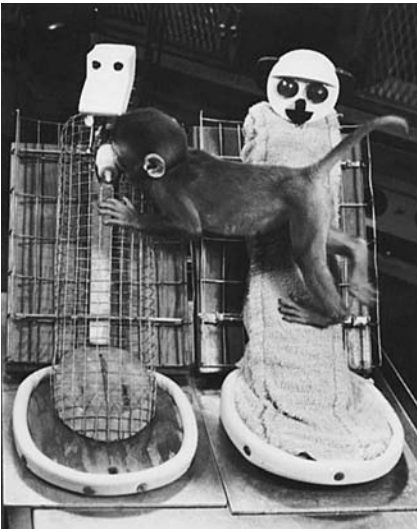
Harlow's Monkeys Raised with Surrogate Mothers

1

What was Harlow's procedure for studying attachment in infant monkeys, and what did he find?

<<< In one experiment, Harlow (1959) raised infant monkeys individually in isolated cages, each containing two surrogate mothers—one made of bare wire and the other covered with soft terry cloth (see Figure 12.1). The infants could feed themselves by sucking milk from a nipple that for half of them was attached to the wire surrogate and for the other half was attached to the cloth surrogate. Harlow's purpose was to determine if the infants would become attached to either of these surrogate mothers as they would to a real mother; he also wanted to know which

FIGURE 12.1 Harlow's motherless monkeys Harlow raised infant monkeys with two surrogate (substitute) mothers, one wire and one cloth. The monkey shown here received its nourishment from a bottle attached to the wire surrogate but went to the cloth surrogate for comfort, affection, and reassurance.



Courtesy of University of Wisconsin Primate Lab, Madison, WI



Monkmeyer Press

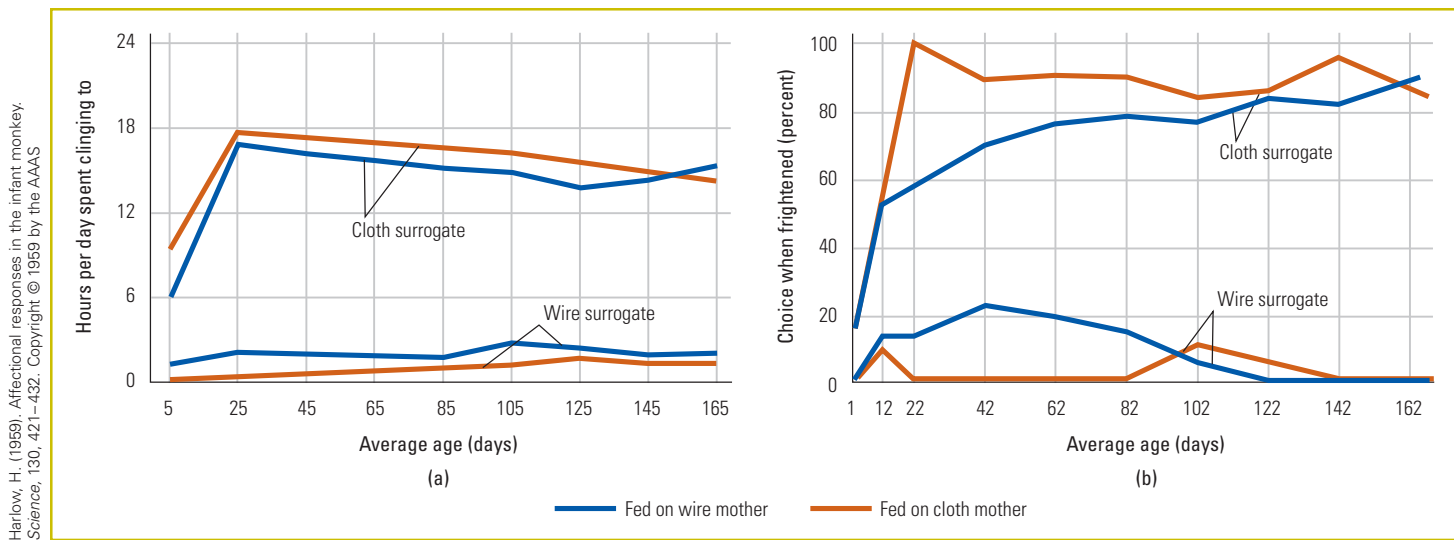


FIGURE 12.2 Evidence of infant monkeys' preference for the cloth surrogate These graphs show (a) the average number of hours that Harlow's monkeys spent on each of the surrogate mothers and (b) the percentage of time that they ran to each when they were frightened by a strange object. Notice that the preference for the cloth surrogate was as strong for the monkeys that were fed on the wire surrogate as for those fed on the cloth surrogate. (Adapted from Harlow, 1959.)

characteristic—the milk-providing nipple or the soft cloth exterior—was most effective in inducing attachment.

Harlow's main finding was that regardless of which surrogate contained the nutritive nipple, all the infant monkeys treated the cloth-covered surrogate, not the wire one, as a mother. They clung to it for much of the day and ran to it when threatened by a strange object (see Figure 12.2). They also were braver in exploring an unfamiliar room when the cloth surrogate was present than when it was absent, and they pressed a lever repeatedly to look at it through a window in preference to other objects. This work demonstrated the role of contact comfort in the development of attachment bonds, and it helped to revolutionize psychologists' thinking about infants' needs. Provision of adequate nutrition and other physical necessities is not enough; infants also need close contact with comforting caregivers.

The Form and Functions of Human Infants' Attachment

Bowlby (1958, 1982) observed attachment behaviors in young humans, from 8 months to 3 years of age, that were similar to those that Harlow observed with monkeys. He found that children exhibited distress when their mothers (the objects of their attachment) left them, especially in an unfamiliar environment; exhibited pleasure when reunited with their mothers; exhibited distress when approached by a stranger unless reassured or comforted by their mothers; and were more likely to explore an unfamiliar environment in the presence of their mothers than alone. Many research studies have since verified these general conclusions (Crittenden & Claussen, 2003).

Bowlby contended that attachment is a universal human phenomenon with a biological foundation that derives from natural selection. Infants are potentially in danger when out of sight of caregivers, especially in a novel environment. During our evolutionary history, infants who scrambled after their mothers and successfully protested their mothers' departure, and who avoided unfamiliar objects when their mothers were absent, were more likely to survive to adulthood and pass on their genes than were those who were indifferent to their mothers' presence or absence. Evidence that similar behaviors occur in all human cultures (Kagan, 1976; Konner, 2002) and in other species of mammals (Kraemer, 1992) supports Bowlby's evolutionary interpretation.

Also consistent with the evolutionary explanation is the observation that attachment begins to strengthen at about the same age (6 to 8 months) that infants begin to move around on their own. A crawling or walking infant, who, for good evolutionary reasons, is intent on exploring the environment, can get into more danger than an immobile one. For survival, exploring and learning must be balanced by

2 According to Bowlby, what infant behaviors indicate strong attachment, and why would they have come about in natural selection?

3 From an evolutionary perspective, why does attachment strengthen at about 6 to 8 months of age?

appropriate fears, and exploration is safer if the caregiver is present than if not. As was noted in Chapter 11, infants who can crawl or walk exhibit much *social referencing*; that is, they look to their caregivers for cues about danger or safety as they explore. To feel most secure in a novel situation, infants require not just the presence of their attachment object but also that person's emotional availability and expressions of reassurance. Infants who cannot see their mother's face typically move around her until they can (Carr & others, 1975), and infants who are approached by a stranger relax more if their mother smiles cheerfully at the stranger than if she doesn't (Broccia & Campos, 1989).

The Strange-Situation Measure of Attachment Quality

4

How does the strange-situation test assess the security of attachment? What are some limitations of this measure?

<<< In order to assess attachment systematically, Mary Ainsworth—who originally worked with Bowlby—developed the **strange-situation test**. In this test the infant and mother (or another caregiver to whom the infant is attached) are brought into an unfamiliar room with toys. The infant remains in the room while the mother and an unfamiliar adult move out of and into it according to a prescribed sequence. The infant is sometimes with just the mother, sometimes with just the stranger, and sometimes alone. Infants are classed as *securely attached* if they explore the room and toys confidently when their mother is present, become upset and explore less when their mother is absent (whether or not the stranger is present), and show pleasure when the mother returns (see Figure 12.3). Other response patterns are taken to indicate one or another form of insecure attachment. An infant who avoids the mother and seems to act coldly toward her is said to have an *avoidant attachment*, and an infant who does not avoid the mother but continues to cry and fret despite her attempts to comfort is said to have an *anxious attachment*. By these criteria, Ainsworth and other researchers have found that about 70 percent of middle-class North American infants in their second year of life are securely attached to their mothers, about 20 percent are avoidant, and about 10 percent are anxious (Ainsworth & others, 1978; Waters & others, 2000).

Like most measures of psychological attributes, the strange-situation test has limitations. The test assesses fear-induced aspects of attachment but fails to capture the harmonious caregiver-child interactions that typically occur in less stressful situations (Field, 1996). Moreover, the test is designed to assess the infant's reactions to the caregiver in a *mildly* fearful situation, but for some infants, because of innate temperament (Kagan & others, 1992) or past experience, the situation may be either too stressful or insufficiently stressful to induce the appropriate responses. Some researchers in Japan have concluded that the strange situation is not a valid measure of attachment in their culture (Nakagawa & others, 1992; Rothbaum & others, 2000). Japanese infants are rarely separated from their mothers, even briefly, and their abandonment in the strange situation leads many of them to become inconsolable by the time the mother returns—and to present the superficial appearance of a high rate of anxious attachment.

FIGURE 12.3 | The strange-situation test In this test of attachment, the mother (or other person to whom the infant is attached) moves into and out of an unfamiliar room, leaving the infant either with a stranger or alone. Here we see one infant, Brian, at different stages in the test: In (a) he plays confidently when his mother is present, in (b) he cries when she leaves, and in (c) he is comforted when she returns. Brian's behavior indicates that he is securely attached to his mother.



(a)



(b)



(c)

Courtesy of Mary Ainsworth

Quality of Care Affects Quality of Attachment

Ainsworth hypothesized that an infant will become securely attached to a mother who provides regular contact comfort, responds promptly and sensitively to the infant's signals of distress, and interacts with the infant in an emotionally synchronous manner (see Figure 12.4). Consistent with that hypothesis, Ainsworth (1979) found, in home observations, that mothers of infants who showed secure attachment in the strange-situation test were on average more attentive and comforting to their infants than were mothers of infants who showed insecure attachment. Many subsequent studies have revealed similar correlations (Posada & others, 2004; Seifer & others, 1996).

As you by now well know, however, we must be cautious about inferring causal relationships from correlations. The assumption that the mothers' behavior causes the infant's attachment status may underestimate the power of infants to influence their mother's behavior. Maybe some infants are, for genetic reasons, easier to care for than others, and maybe that temperament leads them (a) to show the "secure" pattern in the strange situation and (b) to succeed in eliciting sensitive attention from their mothers at home. To test most directly Ainsworth and Bowlby's causal hypothesis, one would have to conduct an experiment: get some mothers to behave in a more sensitive manner toward their infants than other mothers do, and then assess their infants for attachment security.

In the Netherlands, Dymphna van den Boom (1994) performed just such an experiment. She focused on mothers with temperamentally irritable babies—babies who by disposition are unusually fussy, easily angered, and difficult to comfort. Van den Boom (1991) had previously observed that mothers of such babies tended to withdraw emotionally from them, which seemed to set off a spiral of decline in the mother–infant relationship. For her experiment, she recruited 100 mother–infant pairs, in each of which the infant had been judged by a standard test of temperament to be highly irritable. Half the mothers, randomly chosen, participated in a 3-month training program, beginning when their infants were 6 months old, designed to help and encourage the mothers to perceive and respond appropriately to their babies' signals, especially signals of distress. When the infants were 12 months old, all of them were tested with their mothers in the strange situation. The result was that 62 percent of the infants with trained mothers showed secure attachment and only 22 percent of the other infants did. More recently, a number of other maternal-training experiments have shown similar, though less dramatic, results (Bakermans-Kranenburg & others, 2003).

Secure Attachment Correlates with Positive Later Adjustment

Many psychologists have theorized that the quality of an infant's early attachment to the mother or another caregiver strongly influences the child's later development. For instance, Erikson (1963) proposed that secure attachment in infancy results in a general sense of trust of other people and oneself, which permits the infant to enter subsequent life stages in a confident, growth-promoting manner. Bowlby (1973) and Ainsworth (1989) postulated that infants develop an internal "working model," or cognitive representation, of their first attachment relationship and that this model affects their subsequent relationships throughout life. Many research findings are consistent with such theories. Infants judged to be securely attached through the strange-situation test have been found to be on average more confident, better problem solvers, emotionally healthier, and more sociable in later

5
What early caregiver behaviors are predictive of secure attachment? How did Bowlby and Ainsworth interpret the correlations, and how else might they be interpreted?



Julie O'Neill / The Picture Cube

FIGURE 12.4 | Interactional synchrony Interactions such as this during the first 3 months of life correlate with secure attachment measured several months later.

6
What experimental evidence supports Ainsworth and Bowlby's theory that caregivers' responsiveness and warmth promote secure attachment?

7
What evidence supports the theory that infant attachment affects subsequent emotional and social development, and what are the limitations of that evidence?



By permission of Dave Coverly and Creators Syndicate, Inc.

childhood than those judged to be insecurely attached (Ainsworth, 1989; McElwain & Volling, 2004; Schneider & others, 2001).

Secure attachment in infancy clearly helps predict positive development later in life. Again, however, correlation does not tell us about the direction of causality. While it seems logical that secure attachment would help produce the positive later effects, other explanations of the correlation are possible. Perhaps the correlation derives from the child's innate temperament. The same temperament that predisposes secure attachment in childhood might also predispose successful social interactions later in life. Or perhaps it derives from the continuity of parental behavior. Parents who are affectionate toward their children typically remain so throughout their children's development (Levitt, 1991), and the continuing emotional support may be more influential than the quality of infant attachment in promoting positive development. Consistent with this view are many observations showing that children who were emotionally deprived during infancy and adopted later into affectionate homes can develop warm relationships with their adoptive parents and adjust positively to life's subsequent challenges (Chisholm, 2003; Tizard & Hodges, 1978).

High-Quality Day Care Also Correlates with Positive Adjustment

The earliest studies of attachment focused exclusively on infants' relationships with mothers, but subsequent studies showed that infants can also develop secure attachments with fathers, grandparents, older siblings, and day-care providers (Cox & others, 1992; Goossens & van Ijzendoorn, 1990). During the decades after Bowlby developed his theory, the use of day-care providers increased dramatically. In the United States today, roughly 80 percent of preschool children are in some form of day care, usually with caregivers who are not relatives, for an average of 40 hours per week (Marshall, 2004). Concern about the developmental effects of such care has led to many research studies aimed at correlating variations of day care with children's subsequent behavioral development.

8

What have researchers found about possible developmental effects of day care?

<<< The results, taken as a whole, suggest that high-quality day care correlates with positive development just as high-quality home care does (Marshall, 2004). High-quality day care is generally defined as that in which children are with the same familiar caregivers each day; the caregivers are warm, sensitive, responsive, and happy; and developmentally appropriate, interesting activities are available to the children. The higher the quality of day care during infancy and toddlerhood, by these measures, the better is the social and academic adjustment of children in their early school years (Hrdy, 2005; Marshall, 2004). Researchers have also found that day care generally does not interfere with the abilities of infants and toddlers to maintain secure attachments with their mothers and fathers, and that parental attachments are better predictors of positive subsequent adjustment than are attachments with day-care providers, regardless of the amount of time spent in day care (Marshall, 2004).

Cross-Cultural Differences in Infant Care

Beliefs and practices regarding infant care vary considerably from culture to culture. Our Western, Euro-American culture is in many ways less indulgent of infants' desires than are some other cultures. Are such differences in care associated with any differences in infant social behavior and subsequent development?

Sleeping Arrangements

Our culture is unusual in the expectation that infants and young children should sleep alone. A survey of 90 other cultures revealed that, in every one, infants and



Steven Rubin / The Image Works

Having a good day. Researchers may argue about the long-term effects of day care on children's social and emotional development, but high-quality settings like the one shown here seem to offer an excellent environment for development. Such centers provide toddlers a safe setting, with opportunities for exploration, and responsive, consistent, nurturing care.

toddlers slept in the same room as the mother or another closely related adult, usually in the same bed (Barry & Paxson, 1971). Young chimpanzees, bonobos, and gorillas—our closest nonhuman relatives—sleep in direct contact with their mothers during their first few years of life (Mosko & others, 1993; de Waal, 1997).

Interviews with Mayan mothers in rural Guatemala and middle-class mothers in the United States revealed that differing beliefs and values support the differing sleeping practices (Morelli & others, 1992). The Mayan mothers, who all slept with their infants and toddlers, emphasized the comforting value of physical closeness. When told that young children in the United States often sleep alone, in their own rooms, they expressed dismay and pity at what seemed to them a heartless practice. In contrast, the U.S. mothers, none of whom slept regularly with their youngsters, expressed concern that to do so would establish a hard-to-break habit, would foster dependence, and might even be physically dangerous because the infant could be smothered. [Concerning this last point, there is, indeed, evidence that sleeping with infants can in rare cases cause accidental smothering (Scheers & others, 2003). Parents who wish to consider this practice are advised to review that evidence and take precautions against smothering.] Moreover, the U.S. mothers saw bedtime as a time for intimate relations with their husbands, whereas the Mayan mothers saw it as a time for family togetherness. Not surprisingly, the task of putting young children to bed was much easier for the Mayan mothers than for the U.S. mothers, who commonly engaged in elaborate bedtime rituals.

Some researchers who began with the view that sleeping with parents is likely to be developmentally harmful—a view held by most pediatricians in the United States (Lozoff & others, 1984)—have concluded just the opposite. In one study, conducted in Spain, adult women who had slept with their parents up to the age of 4 or 5 years showed somewhat healthier, more mature personalities (higher “ego strength”) on a standard paper-and-pencil test of personality than did otherwise-comparable women who had slept alone during infancy and early childhood (Crawford, 1994). In another study, conducted with U.S. military families in which the father was often away, children who had been diagnosed with behavioral or psychological problems were significantly *less* likely than other children to have slept regularly with their mothers (Forbes & Weiss, 1992). In still another study, conducted in the United States, children who had slept regularly with their parents during their first few years were reported to be more self-reliant and to have better relationships with playmates than was the case for children who had usually slept alone (Keller & Goldberg, 2004). Again, of course, these are correlational studies, so we don't know for sure what is cause and what is effect.

9 How do adults in different cultures explain their practice of sleeping or not sleeping with their young children? What evidence suggests that co-sleeping may be beneficial?

Security In Western cultures, babies and young children are commonly required to sleep alone. As a consequence, they often develop a strong attachment to a blanket, teddy bear, or other soft inanimate object that they sleep with.



M. Edrington / The Image Works

10

How might our culture's typical sleeping arrangements explain infants' attachments to "security blankets"?

<<< One intriguing consequence of requiring infants and toddlers to sleep alone, apparently, is a high rate of attachment to inanimate objects—the "security blanket" or the special doll or teddy bear. Several studies, some across cultures and some comparing different families within the same culture, have shown that infants who sleep alone are much more likely to develop such attachments than are those who sleep with or in the presence of an adult (Ahluvalia & Schaefer, 1994; Honig & Townes, 1976; Wolf & Lozoff, 1989). Perhaps, to relieve their fear at night, infants who sleep alone—much like Harlow's monkeys—attach themselves to the softest, most comforting mother substitute available. Unlike Harlow's monkeys, of course, the human infants are in contact with their mothers and other caregivers during the day, so the attachment to the inanimate object is in addition to other attachments.

Caregiving in Hunter-Gatherer Societies

During the great bulk of our species' history, our ancestors lived in small groups, made up mostly of close relatives, who survived by gathering and hunting food cooperatively in the forest or savanna. The biological underpinnings of infants' and caregivers' behaviors evolved in the context of that way of life. Partly for that reason, an examination of infant care in the few hunter-gatherer cultures that remain is of special interest.

11

What observations suggest that hunter-gatherers are highly indulgent toward infants? What parenting styles distinguish the !Kung, Efe, and Aka?

<<< Melvin Konner (1976, 2002) studied the !Kung San people, who live in Africa's Kalahari Desert. He observed that !Kung infants spend most of their time during their first year in direct contact with their mothers' bodies. At night the mother sleeps with the infant, and during the day she carries the infant in a sling at her side. The sling is arranged in such a way that the infant has constant access to the mother's breast and can nurse at will—which, according to Konner, occurs on average every 15 minutes. The infant's position also enables the infant to see what the mother sees and in that way to be involved in her activities. When not being held by the mother, the infant is passed around among others, who cuddle, fondle, kiss, and enjoy the baby. According to Konner, the !Kung never leave an infant to cry alone, and usually they detect the distress and begin to comfort the infant before crying even begins.

Studies of other hunter-gatherer cultures have shown a similarly high degree of indulgence toward infants (Hewlett & others, 1998), but the people who provide the care can vary. Among the Efe [*Eff ay*]*—*who gather and hunt in the Ituri forest of central Africa—infants are in physical contact with their mothers for only about half the day (Morelli & Tronick, 1991; Tronick & others, 1992). During the rest of the day they are in direct contact with other caregivers, including siblings, aunts, and unrelated women. Efe infants nurse at will, not just from their mothers but also from other lactating women in the group. However, at about 8 to 12 months of age—the age at which research in many cultures has shown that attachment strengthens—Efe infants begin to show increased preference for their own mothers. They are less readily comforted by other people and will often reject the breast of another woman and seek the mother's.

In no culture yet studied is the average father nearly as involved as the average mother in direct care of infants and young children; but, in general, paternal involvement appears to be greater in hunter-gatherer cultures than in agricultural or industrial cultures (Hewlett, 1988; Konner, 1981). The record on this score seems to be held by the Aka of central Africa, a hunter-gatherer group closely related to the Efe (Hewlett, 1988). Aka fathers have been observed to hold their infants an average of 20 percent of the time during daylight hours and to get up with them frequently at night. Among the Aka, the whole family—mother, father, infant, and other young children—sleep together in the same bed.

12

What evidence suggests that indulgence of infants' desires does not spoil them?

<<< Books on infant care in our culture (for example, Spock & Rothenberg, 1985) commonly advise that indulgent parenting—including sleeping with infants and offering immediate comfort whenever they cry—can lead infants to become ever



Courtesy of Gilda Morelli

An Efe girl and her infant brother In hunter-gatherer cultures, infants are in direct physical contact with a caregiver almost constantly. Efe infants spend about half of each day with their mothers and the rest of the day with other members of the group.

more demanding and prevent them from learning to cope with life's frustrations. Konner (1976, 2002), however, contends that such indulgence among the !Kung does not produce overly demanding and dependent children. In a cross-cultural test, !Kung children older than 4 years explored more and sought their mothers less in a novel environment than did their British counterparts (Blurton-Jones & Konner, 1973). By the time they are teenagers, !Kung children must do their share of providing food and other necessities in an environment that is full of dangers (such as wild animals) and hardships (such as prolonged droughts). If their indulgent child-rearing practices had produced uncooperative or incompetent adults, the !Kung would not have survived in their harsh environment. Konner's view is also supported by the study, mentioned earlier, in which U.S. children who had slept regularly with their parents were by some measures more autonomous—that is, less dependent on adults to help them solve problems—than were otherwise comparable children who had been required to sleep alone (Keller & Goldberg, 2004).

Some researchers speculate that high indulgence of infants' desires, coupled with a complete integration of infants and children into the community's social life, in non-Western societies may foster long-lasting emotional bonds that are stronger than those developed by our Western practices (Rogoff, 2003; Tronick & others, 1992). The result is apparently not dependence, in the sense of fear when alone or inability to function autonomously, but *interdependence*, in the sense of strong loyalties or feelings of obligation to a particular set of people. In cultures such as the !Kung San and the Efe, where people depend economically throughout their lives on the cooperative efforts of their relatives and other members of their tribe, such bonds may be more crucial to each person's well-being than they are in our mobile society, where the people with whom we interact change as we go from home to school to work or move from one part of the country to another.

In our culture, where nuclear families live in relative isolation from one another, most parents cannot indulge their infants to the degree seen in many other cultures. !Kung mothers are constantly in contact with their young infants, but they are also constantly surrounded by other adults who provide social stimulation and emotional and material support (Konner, 1976). A survey of infant care in 55 cultures in various parts of the world revealed a direct correlation between the degree of indulgence and the number of adults who live communally with the infant (Whiting, 1971). Indulgence is greatest for infants who live in large extended families or close-knit tribal groups, and least for those who live just with one or both parents.

>>> **13**
How might indulgence of infants be related to interdependence and to living in extended families?

SECTION REVIEW

Infants develop emotional attachment bonds with the caregivers on whom they depend.

Attachment to Caregivers ● Harlow found that infant monkeys became attached to a cloth surrogate mother but not to a wire one, even if the latter provided milk. They turned to the cloth “mother” for contact comfort and explored the environment more fully in its presence. ● Bowlby found that human infants also exhibit attachment behaviors, such as expressing distress when their mothers leave them. He considered attachment to be a universal, evolution-based human phenomenon. ● The strange-situation test is commonly used to determine whether an infant’s attachment to a caregiver is secure or insecure (e.g., avoidant). ● A caregiver’s attentive, responsive, emotionally sensitive behavior toward the infant promotes secure attachment. ● Secure attachment to caregivers in infancy and high-quality day care are both correlated with positive behaviors later in life (e.g., greater sociability).	Infant Care in Different Cultures ● Co-sleeping with infants and young children, which is unusual in North America, is the norm in non-Western cultures and is associated with positive social and emotional development. ● Hunter-gatherer societies such as the !Kung treat infants with extraordinary indulgence, keeping them in nearly constant physical contact, permitting nursing at will, and responding quickly to signs of distress. ● The indulgent approach taken by hunter-gatherers does not appear to produce demanding or overly dependent individuals but rather strong interdependence and group loyalty.
--	--

Childhood I: Continuing Interactions with Caregivers

As children grow from infancy into toddlerhood and beyond, they become increasingly mobile and able to initiate and carry out a wide variety of actions in their physical and social worlds. In his life-span theory of social development, Erikson (1963) divided the years from age 1 to 12 into three successive stages, concerned, respectively, with the development of *autonomy* (self-control), *initiative* (willingness to initiate actions), and *industry* (competence in completing tasks). These characteristics are all closely related to one another; they have to do with the child's ability to control his or her own actions.

As Erikson pointed out, however, children's actions frequently bring them into conflict with caregivers and others around them. According to Erikson, caregivers' responses to children's actions, and the ways by which caregivers and children resolve their conflicts, influence children's social development. On the positive side, children may develop the ability to behave appropriately, with confidence, in ways that are satisfying both to themselves and to others. On the negative side, children may develop feelings of shame, doubt, and inferiority that interfere with autonomy, initiative, and industry. The psychologically healthy, moral person, in Erikson's theory, is one who responds appropriately to others' needs without sacrificing his or her own sense of self-control.

In the years since Erikson developed his theory, much research has been conducted on child-caregiver interactions and the possible effects these have on children's development. Much of this research has to do, broadly, with the development of morality. This section, then, is about ways by which children develop morally through their interactions with caregivers.

Behavioral and Emotional Foundations for Morality

Long ago, philosophers who thought about children's social development tended to view young children as akin to either wild animals that must be tamed or lumps of clay that must be shaped into socially acceptable humans. Most present-day developmental psychologists, in contrast, consider children to be the primary conductors of their own development, with caregivers and others in their environments serving as the substrate with which they work. Of course, that substrate is important. Without feedback from others or good models of acceptable social behavior, children would be hard put to create acceptable social selves. Here we shall examine some instinctive tendencies of children that seem to have been designed, by natural selection, to promote their development as social, moral individuals.

The Young Child's Natural Tendency to Give and Help

14

What evidence suggests that young children naturally enjoy giving? How does the !Kung culture use that enjoyment for moral training?

<<< Giving appears to be a human instinct, much like smiling and babbling. Near the end of their first year of life, infants routinely, without any special encouragement, begin to give objects to their caregivers and to delight in games of give-and-take, in which the child and caregiver pass an object back and forth to each other. In a series of experiments conducted in the United States, nearly every one of more than 100 infants, aged 12 to 18 months, spontaneously gave toys to an adult during brief sessions in a laboratory room (Hay & Murray, 1982; Rheingold & others, 1976). They gave not just to their mothers or fathers but also to unfamiliar researchers, and they gave new toys as frequently as familiar ones. They gave when an adult requested a toy by holding a hand out with palm up (the universal begging posture), and they gave when no requests were made. Infants in a !Kung hunter-gatherer community were likewise observed to give objects regularly, beginning near the end of their first year of life (Bakeman & others, 1990).

In our culture, young children's enjoyment of giving is known to most parents but is not much commented on. Books on child care, for example, rarely mention

it, and very few psychologists have studied it. In contrast, the !Kung treat a child's early acts of giving as a crucial developmental milestone, much as we treat a child's first words. The sharing of food and supplies often marks the difference between life and death for these hunting-and-gathering people, and the culture as a whole ascribes great value to its system of giving and receiving, referred to as *hxoro*. Grandmothers are especially charged with the task of initiating infants into *hxoro* by encouraging their giving and by guiding the infants' hands in the giving of beads to relatives (Bakeman & others, 1990; Wiessner, 1982). Whether or not such training increases the !Kung's tendency to give is unknown, but it surely must help alert young !Kung to the special significance of such behavior in their culture.

In addition to liking to give, young children enjoy helping with adult tasks. In one study, children between 18 and 30 months old were frequently observed joining their mothers, without being asked, in such household tasks as making the bed, setting the table, and folding laundry (Rheingold, 1982). Not surprisingly, the older children helped more often, and more effectively, than the younger ones.

One might argue that young children's early giving and helping are largely self-centered rather than other-centered, motivated more by the child's own needs and wishes than by the child's perception of others' needs and wishes. But the very fact that such actions seem to stem from the child's own wishes is evidence that our species has evolved prosocial drives, which motivate us—with no feeling of sacrifice—to involve ourselves in positive ways with other people. Giving and helping become increasingly other-focused as they become linked to another apparently innate capacity in the child, that of *empathy*, the capacity to perceive and feel the emotions that another person is feeling.

The Early Emergence of Empathy and Empathic Comforting

Martin Hoffman (2000) has compiled much evidence concerning the development of empathy. Newborn babies, as young as 2 or 3 days old, reflexively cry and show other signs of distress in response to another baby's crying. Hoffman suggests that this instinctive tendency to feel discomfort in response to another's expressed discomfort is a foundation for the development of empathy. Over time, the response gradually becomes less reflexive and more accompanied by thought. By about 6 months of age, babies may no longer cry immediately in response to another's crying, but rather turn toward the distressed individual, look sad, and whimper.

Until about 15 months of age, the child's distress when others are distressed is best referred to as *egocentric* empathy (Hoffman, 2000). The distressed child seeks comfort for himself or herself rather than for the other distressed person. At about 15 months, however, children begin to respond to another's discomfort by attempting to comfort that person, and by 2 years of age they begin to succeed at such comforting. For example, Hoffman (2000) described a case in which 2-year-old David first attempted to comfort his crying friend by giving him his (David's) own teddy bear. When that didn't work, David ran to the next room and returned with his friend's teddy bear and gave it to him. The friend hugged the bear and immediately



Peter Southwick / Stock, Boston

Giving is natural Infants and young children all over the world delight in games of give and take with adults or older children. Sharing appears to be an aspect of human nature that begins to be exercised as soon as the child's capacity for motor coordination makes it possible.

15 According to Hoffman, how does empathy develop during infancy and early toddlerhood?



© E. Ann Kaplan

Martin Hoffman and Grandchild Hoffman has documented that even infants respond empathically to others' emotional feelings. That capacity, he argues, lies at the root of moral development. As children become linguistically competent, parents can help them develop morally by drawing attention to the effects that their actions have on others' feelings.

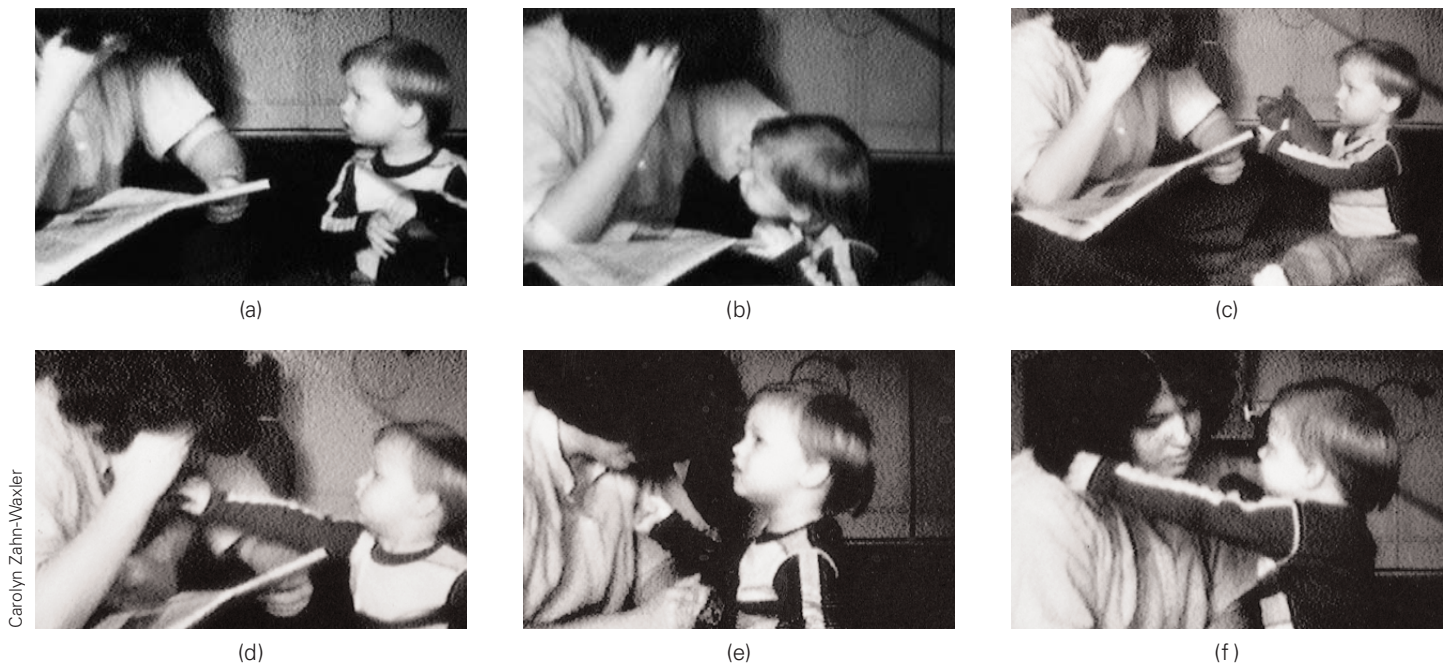


FIGURE 12.5 | A toddler's empathic concern In this sequence, a 21-month-old (a) notices and (b) examines his mother's simulated expression of sadness; (c) tries to cheer her with a puppet; (d) expresses concern; (e) asks the researcher to come to help; and (f) gives his mother a hug while uttering consoling sounds and statements. (From Zahn-Waxler & Radke-Yarrow, 1990, pp. 116–117.)

stopped crying. To behave in such an effective manner, the child must not only feel bad about another's discomfort, but must also understand enough about the other person's mind to know what will provide comfort. Another example of empathic comforting is illustrated in Figure 12.5.

In their relationships with caregivers, young children are most often on the receiving end of acts of giving, helping, and comforting. They therefore have ample opportunity not only to witness such behaviors but also to feel their pleasurable and comforting consequences. Correlational studies indicate, not surprisingly, that children who have received the most sensitive care, and who are most securely attached to their caregivers, also demonstrate the most giving and comforting to others (Bretherton & others, 1997; Kiang & others, 2004).

Guilt as a Force for Moral Development

Of course, young children are not always giving, helpful, and caring, no matter how sensitively they are treated. Very often they are disruptive, selfish, and even hurtful to others around them. Toddlers under the age of about 2 years often look to caregivers for guidance about what is appropriate, but between their second and third birthdays—the period known to many parents as the “terrible twos”—they become increasingly willful and ready to challenge authority. At this stage a challenge of good parenting is to correct the child's disruptive or hurtful behaviors without damaging the child's positive sense of autonomy and initiative. A developmental event that tempers the 2-year-old's behavior and can serve as a powerful force for moral development is the gradual emergence of the capacity for *guilt*.

16

How might the capacity for empathy-based guilt serve as a force for moral development in young children?

<<< Until relatively recently, most developmental psychologists tended to think of guilt primarily in negative terms. For instance, Erikson (1963) considered guilt to be a harmful outcome of interactions between toddlers and guilt-inducing caregivers. Some forms of guilt certainly are harmful, as when children feel guilty about events that they cannot control (Donenberg, 1998) or about every attempt to exert their will. But guilt about one's own hurtful actions, which one can control,

may be crucial to healthy social development. Hoffman (1982, 1998) distinguishes between *empathy-based guilt*, which is constructive, and *anxiety-based guilt*, which can be harmful. He suggests that empathy-based guilt emerges from the capacity for empathy coupled with the child's growing understanding of the relationship between his or her own actions and others' feelings. A child who can sense another's hurt, and who knows that the hurt came from his or her actions, can feel discomforting guilt. Guilt helps curb aggression and selfishness and promotes acts of kindness and helping.

Consistent with Hoffman's theory, a number of studies have shown that expressions of guilt by young children, when they break something or hurt someone, are *positively* correlated with sensitive parenting and secure attachment to parents (Kochanska & others, 2002). According to Hoffman (2000), sensitive parents can draw on their child's capacity for guilt to help the child become a moral person.

Parents' Styles of Discipline

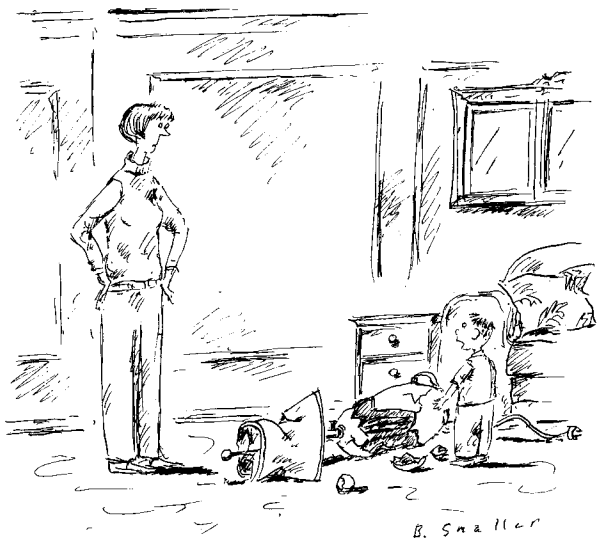
Discipline refers to the methods by which parents and other caregivers attempt to stop or correct their children's misbehavior. Of course, parents do not *just* discipline their children; they also play with them, converse with them, interpret the world to them, and serve as models through their own behavior. Nevertheless, a high percentage of the interactions between parents and young children are disciplinary in nature, and most of the controversy about what constitutes good or bad parenting has to do with styles of discipline.

Hoffman's Theory About the Role of Discipline in Moral Development

Hoffman (1983, 2000) has categorized disciplinary techniques into three classes. The first—which he favors—is *induction*, a form of verbal reasoning in which the parent induces the child to think about the harmful consequences of the child's actions, from the point of view of the person who is hurt. Induction draws on the child's capacity for empathy and empathy-based guilt. For example, a parent may say to a young child, "When you hit Billy, it hurts him and makes him cry," or to a somewhat older child, "When you tease Susan and call her names, it makes her feel that nobody likes her." Such statements demonstrate respect for the child's desire to act responsibly. The child who then stops hitting or teasing can attribute the change in behavior to his or her own moral standards: "I don't want to hurt anyone." According to Hoffman, induction helps the child develop a set of personal moral standards that can promote good behavior even when no authority figure is watching.

The second category of disciplinary style in Hoffman's taxonomy is *power assertion*—the use of physical force, punishment, or (less often) rewards or bribes to control a child's behavior. The third is *love withdrawal*, which occurs when parents, often unwittingly, express disapproval of the child, rather than just of the child's specific action. Love withdrawal may take the form of verbal statements, such as "You are worthless," or nonverbal actions, such as coldly ignoring the child.

Induction In the discipline technique of induction, the parent explains to the child how her actions have hurt another person. The technique draws on the child's capacity to empathize with the hurt person and to feel empathy-based guilt.



"What do I think is an appropriate punishment? I think an appropriate punishment would be to make me live with my guilt."

The New Yorker Collection © 2000 Barbara Smaller from cartoonbank.com. All rights reserved.

>>> 17 What are Hoffman's three categories of discipline, and why does he favor the category called *induction*?



Daraina / Corbis

Like many other psychologists, Hoffman points to the harmful effects of power assertion and love withdrawal. He argues that power assertion focuses attention on the punishments and rewards rather than on reasons why an action is wrong or right. A child so disciplined may continue the undesirable behavior when there is little chance of being caught and may behave well only when someone is sure to notice and provide a reward. Moreover, both punishment and love withdrawal elicit negative emotions in the child (anger in the case of punishment, anxiety in the case of love withdrawal), which interfere with the child's ability to experience empathy for the person who was hurt and to grow morally from that experience. Anger and anxiety can also weaken the parent-child relationship and promote further misbehavior.

Hoffman acknowledges that sometimes a degree of power assertion may be necessary to stop a child from engaging in a seriously wrong or dangerous action; in that case, according to Hoffman, the additional use of induction can blunt the potentially harmful effects of the power assertion. The child who understands that he has been forcibly removed from the playground because he has been hurting another child, not because his parent hates him, has grounds for thinking morally.

Correlations Between Disciplinary Styles and Children's Behavior

Many research studies have shown correlations between parenting styles and children's behavior that are consistent with Hoffman's theory (Eisenberg, 2000). The best-known, pioneering study of this sort was conducted many years ago by Diana Baumrind (1967, 1971), before Hoffman had developed his theory.

18

What three parental discipline styles did Baumrind identify? How do her findings support Hoffman's theory?

<<< Baumrind assessed the behavior of young children by observing them at nursery schools and in their homes, and she assessed parents' behaviors toward their children through interviews and home observations. On the basis of the latter assessments, she classed parents into three groups:

- *Authoritarian* parents strongly value obedience for its own sake and use a high degree of power assertion to control their children.
- *Authoritative* parents are less concerned with obedience for its own sake and more concerned that their children learn and abide by basic principles of right and wrong. In Hoffman's terms, they prefer inductive discipline but couple it with power assertion when they feel it is needed.
- *Permissive* parents are most tolerant of their children's disruptive actions and least likely to discipline them at all. The responses they do show to their children's misbehavior seem to be manifestations of their own frustration more than reasoned attempts at correction.

Consistent with what one would expect from Hoffman's theory, Baumrind found that children of authoritative parents exhibited the most positive qualities. They were friendlier, happier, more cooperative, and less likely to disrupt others' activities than were children of either authoritarian or permissive parents. In a follow-up study of the same children, the advantages for those with authoritative parents were still present at age 9 (Baumrind, 1986).

Other researchers have also found correlations consistent with Hoffman's theory. In one study, 4-year-olds whose mothers used much power assertion were less likely to comply with rules when they thought they could not be seen, and exhibited less guilt when they broke a toy lent to them by the experimenter, compared to 4-year-olds whose mothers used less power assertion and more induction (Kochanska & others, 2002).

The Cause-Effect Problem in Relating Discipline to Children's Behavior

As always, we must be cautious in drawing causal inferences from correlational research. It is tempting to conclude from studies such as Baumrind's that the positive

parenting style caused the good behavior of the offspring, but the opposite causal relationship may be just as plausible. Some children are temperamentally, for genetic reasons, less cooperative and more disruptive than others, and that behavior may elicit harsh, power-assertive discipline and reduced warmth from parents. Several studies have shown that children with different innate temperaments do, indeed, elicit different disciplinary styles from their parents (Jaffee & others, 2004; O'Connor & others, 1998).

The best evidence that disciplinary styles influence children's development comes from experiments that modify, through training, the styles of one group of parents and then compare their offspring to those of otherwise similar parents who did not receive such training (Collins & others, 2000). In one experiment, divorced mothers of 6- to 8-year-old sons were assigned either to a training condition, in which they were taught how to use firm but kind methods of discipline, or to a comparison condition, in which no such training was given. Assessments a year later showed that the sons whose mothers had undergone training had better relationships with their mothers, rated themselves as happier, and were rated by their teachers as friendlier and more cooperative than was the case for the sons of the comparison mothers (Forgatch & DeGarmo, 1999). Apparently, good parenting does make a difference in children's behavior.

19
How have experiments supported the idea that parents' disciplinary styles influence children's development?

SECTION REVIEW

The child's inborn drives and emotions, and interactions with caregivers, promote social development.

The Beginnings of Morality

- Young children have an inborn predisposition to give; they give objects spontaneously to others beginning near the end of their first year.
- The development of empathy during the second year of life causes the child, increasingly, to base actions of giving, helping, and comforting on an understanding of and concern for others' needs and feelings.
- Empathy-based guilt emerges as children connect their own actions with others' pain or sorrow. Such guilt provides a foundation for moral development.

Styles of Discipline

- Hoffman contends that the style of parental discipline referred to as induction is most conducive to the child's moral development.
- Baumrind found that children of parents with an authoritative disciplinary style were happier, friendlier and more cooperative than children of parents with either authoritarian or permissive styles.
- Though Baumrind's study was correlational, experimental research also supports Baumrind's and Hoffman's ideas about effective disciplinary styles.

Childhood II:
Roles of Play and Gender in Development

Parents play important roles in children's social development, but so do peers. Indeed, if developmental psychology were an endeavor pursued by children rather than by adults, I suspect that research would focus more than it currently does on children's relationships with one another and less on their relationships with adults. Parents and other caregivers provide a base from which children grow, but peers are the targets toward which children are oriented and about which they often have the most conscious concerns. A mother and father want their daughter to wear a certain pair of shoes, but the neighborhood children think the shoes are "geeky" and a different style is "cool." Which shoes will the girl want to wear? From an evolutionary perspective, children's strong orientation toward peers—that is, toward members of their own generation—makes sense. After all, it is the peer group, not the parental group, that will provide the child's most direct future collaborators in life-sustaining work and reproduction.

Across cultures and over the span of history, a child's social world is and has been largely a world of other children (Harris, 1995). In most cultures for which data are available, children beyond the age of 4 or 5 years spend more of their day-time hours with other children than with adults (Konner, 1975; Whiting & Edwards, 1988). What are they doing together? Mostly they are playing. In every culture that has been studied, children play when they have the opportunity, and their play takes certain universal forms (Pellegrini & Bjorklund, 2004; Power, 2000; Schwartzman, 1978). It is also true that in every culture that has been studied, children tend to segregate themselves by sex when they play: Boys play mostly with boys, and girls play mostly with girls (Maccoby, 1998, 2002; Whiting & Edwards, 1988). Through playing with others of their own sex, children develop the gender-specific skills and attitudes of their culture.

Developmental Functions of Play

20

What are some universal forms of human play, and what developmental functions do they seem to serve?

<<< Play Is a Vehicle for Acquiring Skills

Chapter 4 presents evidence that play evolved in mammals as a means to ensure that the young of the species will practice and become expert at skills that are necessary for their long-term survival and reproduction. Young mammals play at fleeing, chasing, fighting, stalking, and nurturing. The play of young humans is often like that of other young mammals and appears to serve many of the same functions. Children all over the world play chase games, which promote physical stamina, agility, and the development of strategies to avoid getting caught. Play nurturing, often with dolls or other infant substitutes rather than real infants, and play fighting are also universal, and everywhere the former is more prevalent among girls and the latter is more prevalent among boys (Eibl-Eibesfeldt, 1989).

Other universal forms of human play are specific to our species and help children develop human-specific skills (Power, 2000). Children everywhere become good at making things with their hands through constructive play (play at making things), become skilled with language through word play, and exercise their imaginations and planning abilities

through social fantasy play. In cultures where children can observe directly the sustenance activities of adults, children focus much of their play on those activities. For instance, young boys in hunter-gatherer cultures spend enormous amounts of time at play hunting, such as shooting at butterflies with bows and arrows, and develop great skill in the process (Konner, 1975).

<<< An example of how children's play reflects and perhaps helps transmit a culture's values and skills is found in a study by Douglas Fry (1992) of two Mexican villages. The two villages were observed to be alike in many ways, and the similarity was reflected in aspects of the children's play. In both communities boys made toy plows with sticks and used them to furrow the earth as their fathers worked at real plowing in the fields, and girls made pretend tortillas. In one respect, however, the two communities differed markedly. For generations in La Paz the people had prided themselves on their peacefulness and nonviolence, but the same was not true in San Andrés (the villages' names are pseudonyms). Often in San Andrés, but rarely in La Paz, children saw their parents fight physically, heard of fights or even murders among men stemming from sexual jealousy, and themselves were victims of beatings administered by their parents. In his systematic study of the everyday activities of 3- to 8-year-olds, Fry observed that the San Andrés children engaged in about twice as much serious fighting, and about three times as much play fighting, as the La Paz children. When fighting is common in a culture, fighting will apparently be understood intuitively by children as a skill to be practiced not just in anger but also in play, for fun.

21

How do observations of two Mexican villages illustrate the role of play in the transmission of cultural skills and values from one generation to the next? How might play promote cultural advancement?



Tag, you're it Children, like all young mammals, enjoy chase games. Such play promotes physical development and skill in escaping from predators and enemies.

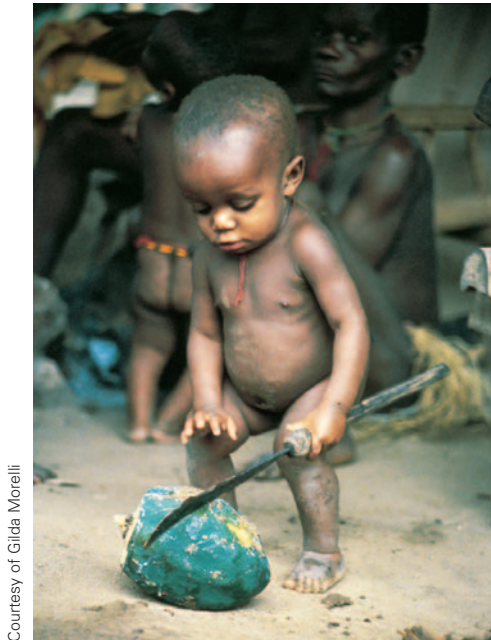
Children's play can help create and advance culture as well as reflect it. When the computer revolution began in North America, children, who usually had no use for computers other than play, were in many families the first to become adept with the new technology. They taught their parents how to use computers. Some of the same children—as young adults, but still in the spirit of play—invented new and better computers and computer programs. Many years ago the Dutch cultural historian Johan Huizinga (1944/1970) wrote a book contending that much, if not most, of what we call “high culture”—including art, literature, philosophy, and legal systems—arose originally in the spirit of play where play was extended from childhood into adulthood.

Play as a Vehicle for Learning About Rules and Acquiring Self-Control

In addition to practicing species-typical and culturally valued skills, children may, through play, acquire more advanced understandings of rules and social roles and greater self-control. Both Jean Piaget and Lev Vygotsky—the famous developmental psychologists introduced in Chapter 10—developed theories along these lines.

In his book *The Moral Judgment of the Child*, Piaget (1932/1965) argued that unsupervised play with peers is crucial to moral development. He observed that adults use their superior power to settle children's disputes, but when adults are not present, children argue out their disagreements and acquire a new understanding of rules based on reason rather than authority. They learn, for example, that rules of games such as marbles are not immutable but are human contrivances designed to make the game more interesting and fair and can be changed if everyone agrees. By extension, this helps them understand that the same is true of the social conventions and laws that govern life in democratic societies. Consistent with Piaget's theory, Ann Kruger (1992) found that children showed greater advances in moral reasoning when they discussed social dilemmas with their peers than when they discussed the same dilemmas with their parents. With peers, children engaged actively and thoughtfully in the discussions, which led to a higher level of moral reasoning; with parents they were far more passive and less thoughtful.

>>> 22 What ideas did Piaget and Vygotsky present concerning the value of play in social development? What evidence supports their ideas?



Courtesy of Gilda Morelli

Play with a purpose In their self-initiated play, children in all cultures exhibit their fascination with the tools and actions they see their elders use. This 11-month old Efe boy is trying to cut a papaya with a huge, sharp knife.



Laura Dwight

Lets pretend In sociodramatic play, children plan out situations and abide by the rules and roles they invent.



© Time Hall / Getty Images

A helping hand In age-mixed play, older children enable younger children to do things that they would not be able to do alone. As a result, younger children acquire new physical and intellectual skills and older children acquire nurturing skills. Age-mixed play is also more frequently gender-mixed than is same-age play.

means of practicing self-discipline of the sort that is needed to follow social conventions and rules.

Consistent with Vygotsky's view, researchers have found that young children put great effort into planning and enforcing rules in their social fantasy play (Furth, 1996; Garvey, 1990). Children who break the rules—who act like themselves rather than the roles they have agreed to play—are sharply reminded by the others of what they are supposed to do: “Dogs don't sit at the table; you have to get under the table.” Also in line with Vygotsky's view, researchers have found positive correlations between the amount of social fantasy play that children engage in and subsequent ratings of their social competence and self-control (Connolly & Doyle, 1984; Elias & Berk, 2002).

The Special Value of Age-Mixed Play

In age-graded school settings, such as recess, children play almost entirely with others who are about the same age as themselves. But in neighborhood settings in our culture (Ellis & others, 1981), and even more so in cultures that don't have age-graded schools (Whiting & Edwards, 1988), children often play in groups with age spans of several years. Indeed, as Konner (1975) has pointed out, the biological underpinnings of human play must have evolved under conditions in which age-mixed play predominated. Hunter-gatherer communities are small and births within them are widely spaced, so a given child rarely has more than one or two potential playmates who are within a year of his or her own age.

23

What features of age-mixed play may make it particularly valuable to children's development?

<<< Psychologists have paid relatively little attention to age-mixed play, but the work that has been done suggests that such play is often qualitatively different from play among age mates. One difference is that it is less competitive; there is less concern about who is best (Feldman, 1997; Feldman & Gray, 1999). An 11-year-old has nothing to gain by proving himself or herself stronger, smarter, or more skilled than a 7-year-old, and the 7-year-old has no chance of proving the reverse. Konner (1972) noted that age-mixed rough-and-tumble play among the !Kung children whom he observed might better be called “gentle-and-tumble,” because an implicit rule is that participants must control their movements so that they do not hurt a younger child.

I have studied age-mixed play at an alternative school in the United States, where children from age 4 to 18 intermingle freely. At this school, age-mixing appears to be a primary vehicle of education (Gray & Feldman, 2004; D. Greenberg, 1992). Young children acquire more advanced interests and skills by observing older children, and older children develop skills at nurturing and consolidate some of their own knowledge by helping younger children. We observed this happening in a wide variety of forms of play, including rough-and-tumble, sports of various types, board games, computer play, constructive play (such as play with blocks or with art materials), and fantasy play. In each of these contexts, to make the game more fun, older children helped younger children understand rules and strategies. Studies of play among siblings who differ by several years in age have yielded similar findings (Brody, 2004). Unfortunately, our age-graded system of schooling, coupled with the decline in neighborhood play that has occurred in recent times, deprives many children in our culture of a powerful natural source of education—the opportunity to interact closely, over prolonged periods, with other children who are substantially older or younger than themselves.

Gender Differences in Social Development

Life is not the same for girls and boys. That is true not just in our culture but in every culture that has been studied (Maccoby, 1998; Whiting & Edwards, 1988). To some degree, the differences are biological in origin, mediated by hormones (discussed in Chapter 6). But, as the anthropologist Margaret Mead (1935) noted long ago, the differences also vary from culture to culture and over time within a given culture in ways that cannot be explained by biology alone.

The words *sex* and *gender* are in some contexts used as synonyms, but in typical psychological usage *sex* refers to the clear-cut biological basis for categorizing people as male or female, while *gender* refers to the entire set of differences attributed to males and females, which can vary across cultures (Deaux, 1985). You were born one sex or the other, but from the moment of birth—from the moment that someone announced, “It’s a girl!” or “It’s a boy!”—you also had a gender. Immediately, in the minds of all who heard the announcement, you were tied through a web of mental associations to all kinds of “girlish” or “boyish” traits, activities, and material belongings. Biological sex differences combined with cultural models and expectations produce the gender differences observed in any given culture.

Gender Differences in Interactions with Caregivers

Even in early infancy, boys and girls, on average, behave somewhat differently from each other. Newborn boys are more irritable and less responsive to caregivers’ voices and faces than are newborn girls (Hittelman & Dickes, 1979; Osofsky & O’Connell, 1977). By 6 months of age, boys squirm more and show more facial expressions of anger than do girls when confined in an infant seat, and girls show more facial expressions of interest and less fussing than do boys when interacting with their mothers (Weinberg & others, 1999). By 13 to 15 months of age, girls are more likely than are boys to comply with their mothers’ requests (Kochanska & others, 1998), and that difference persists throughout childhood and adolescence (Maccoby, 1998).

Parents and other caregivers behave differently toward girls and boys, beginning at birth. They are, on average, more gentle with girls than with boys; they are more likely to talk to girls and to jostle boys (Maccoby, 1998). Such differences in treatment may in part reflect caregivers’ sensitivity and responsiveness to actual differences in the behaviors and preferences of the infant girls and boys, but it also reflects adult expectations that are independent of the infants’ behaviors and preferences. In one study, mothers were asked to hold a 6-month-old female infant, who in some cases was dressed as a girl and introduced as Beth, and in other cases was dressed as a boy and introduced as Adam (Culp & others, 1983). The mothers



Jeffrey W. Myers / Stock, Boston

My tough little boy Researchers have found that fathers generally play more vigorously with their infant sons than with their infant daughters.

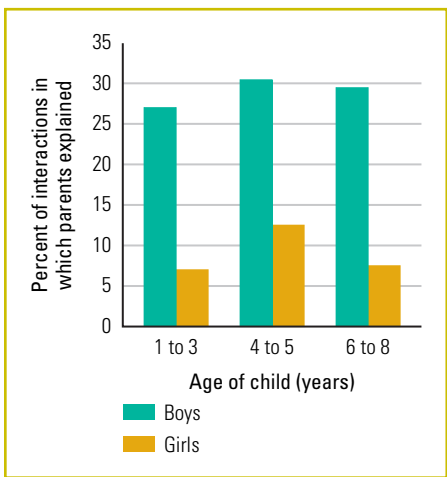
>>> **24** What are some of the ways that girls and boys are treated differently by adults in our culture, and how might such treatment promote different developmental consequences?

talked to Beth more than to Adam, and gave Adam more direct gazes unaccompanied by talk.

Other research suggests that, regardless of the child's age, adults offer help and comfort more often to girls than to boys and more often expect boys to solve problems on their own. In one experiment, college students were quicker to call for help for a crying infant if they thought it was a girl than if they thought it was a boy (Hron-Stewart, 1988). In another study, mothers of 2-year-old daughters helped their toddlers in problem-solving tasks more than did mothers of 2-year-old sons (Hron-Stewart, 1988). Theorists have speculated that the relatively warmer treatment of girls and greater expectations of self-reliance for boys may lead girls to become more affectionate and sociable and boys to become more self-reliant than they otherwise would (Dweck & others, 1978; MacDonald, 1992).

Adults' assumptions about the different interests and abilities of girls and boys may play a role in the types of careers that the two sexes eventually choose. In particular, a number of studies have shown that parents and teachers provide more encouragement and instruction in physical sciences to boys than to girls, appar-

ently because they assume that boys are more interested and capable in these areas than are girls (Tenenbaum & Leaper, 2003). In one such study, parents and their young children were observed at interactive exhibits at a science museum (Crowley & others, 2001). The main finding was that parents, both fathers and mothers, were far more likely to explain something about the workings of the exhibits, or the underlying principles being demonstrated, to their sons than to their daughters. This occurred regardless of the age of the child (see Figure 12.6) and despite the fact that the researchers could find no evidence, from



analyzing the children's behavior and questions, that the sons were more interested in the exhibits or in the principles being demonstrated than were the daughters. Perhaps such differential treatment helps explain why many more men than women choose careers in the physical sciences.

FIGURE 12.6 | More explanations to boys than to girls At a science museum, parents were much more likely to explain the mechanisms or principles of interactive exhibits to sons than to daughters, regardless of age. (Adapted from Crowley & others, 2001.)

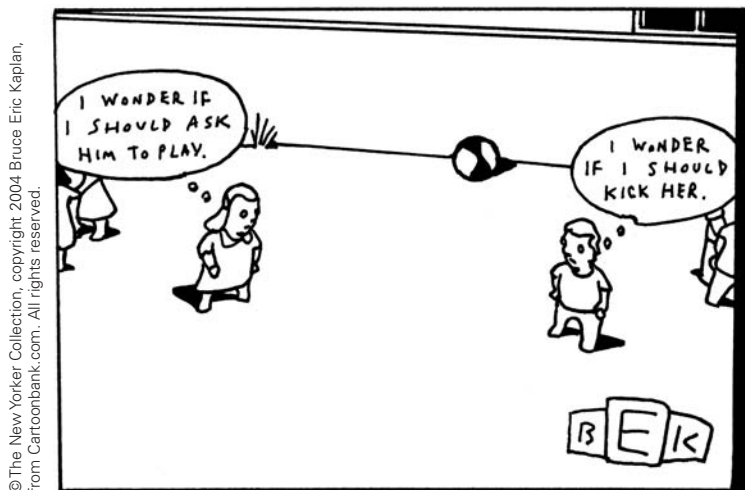
Gender Identity and Children's Self-Imposed Gender Segregation

25

How do children mold themselves according to their understanding of gender differences?

<<< Children's gender development may be influenced by adults' differential treatment, but children also actively mold themselves to behave according to their culture's gender conceptions. By the age of 4 or 5, most children have learned quite clearly their culture's stereotypes of male and female roles (Martin & Ruble, 2004; Williams & Best, 1990) and recognize that they themselves are one gender or the other and always will be, an understanding referred to as **gender identity** (Kohlberg, 1966). Once they have this understanding, children in all cultures seem to become concerned about projecting themselves as clearly male or female. They attend more closely to people of their own gender and model their behavior accordingly, often in ways that exaggerate the male-female differences they see. When required to carry out a chore that they regard as gender-inappropriate, they often do so in a style that clearly distinguishes them from the other gender. For example, in a culture where fetching water is considered women's work, young boys who are asked to fetch water carry it in a very different manner from that employed by women and girls (Whiting & Edwards, 1988).

From a biological perspective, gender is not an arbitrary concept but is linked to sex, which is linked to reproduction. A tendency toward gender identity may well have evolved as an active assertion of one's sex as well as a means of acquiring



culture-specific gender roles. By acting “girlish” and “boyish,” girls and boys clearly announce that they are on their way to becoming sexually viable women and men. Researchers have found that young children often create or overgeneralize gender differences (Martin & Ruble, 2004). For example, when preschoolers were told that a particular boy liked a certain sofa and a particular girl liked a certain table, they generalized this to assume that all boys would like the sofa and all girls would like the table (Bauer & Coyne, 1997). Children’s active imposition of gender roles upon their world has been likened to their active imposition of grammatical rules upon the language that they hear (Martin & others, 2002).

In all cultures that have been studied, boys and girls play primarily with others of their own sex (Maccoby, 1998), and, at least in our culture, such segregation is more common in activities structured by children themselves than in activities structured by adults (Berk & Lewis, 1977; Maccoby, 1990). Even 3-year-olds have been observed to prefer same-sex playmates to opposite-sex playmates (Maccoby & Jacklin, 1987), but the peak of gender segregation, observed in many different settings and cultures, occurs in the age range of 8 to 11 years (Gray & Feldman, 1997; Hartup, 1983; Whiting & Edwards, 1988). In their separate play groups, boys practice what they perceive to be the masculine activities of their culture, and girls practice what they perceive to be the feminine activities of their culture.

In some but not all settings (Gray & Feldman, 1997; Thorne, 1986, 1993), children reinforce gender segregation by ridiculing those who cross gender lines. But this ridicule is not symmetrical. Boys who play with girls are much more likely to be teased and taunted by both gender groups than are girls who play with boys (Petitpas & Champagne, 2000; Thorne, 1993). Indeed, girls who prefer to play with boys are often referred to with approval as “tomboys” and retain their popularity with both sexes. Boys who prefer to play with girls are not treated so benignly. Terms like “sissy” or “pansy” are never spoken with approval. Adults, too, express much more concern about boys who play with girls or adopt girlish traits than they do about girls who play with boys or adopt boyish traits (Martin, 1990). Perhaps the difference reflects the culture’s overall view that male roles are superior to female roles, which might also help explain why, over the past several decades, many women in our culture have moved into roles that were once regarded as exclusively masculine, while relatively fewer men have moved into roles traditionally considered feminine.

What function might children’s self-segregation by gender serve? In our culture, why might boys avoid playing with girls more than the reverse?

Sex-segregated play Even in pre-school, girls tend to play with girls and boys with boys. The little girl looking at the truck might want to join the two boys, but she will probably stay with the girls.



Gender Differences in Styles of Play

27

In what ways can girls' and boys' peer groups be thought of as separate subcultures? Why might differences in boys' and girls' play be greater in age-segregated settings than in age-mixed settings?

<<< Girls and boys tend to play differently as well as separately, and the differences are in some ways consistent from culture to culture (Geary & others, 2003; Whiting & Edwards, 1988). Some social scientists consider boys' and girls' peer groups to be so distinct that they constitute separate subcultures, each with its own values, directing its members along different developmental lines (Maccoby, 1998, 2002; Maltz & Borker, 1982). The *world of boys* has been characterized as consisting of relatively large, hierarchically organized groups in which individuals or coalitions attempt to prove their superiority through competitive games, teasing, and boasting. A prototypical boys' game is king of the hill, the goal of which is to stand on top and push everyone else down. The *world of girls* has been characterized as consisting of smaller, more intimate groups, in which cooperative forms of play predominate and competition, although not absent, is more subtle. A prototypical girls' game is jump rope, the goal of which is to keep the action going as long as possible through the coordinated activities of the rope twirlers and the jumper.

The "two-worlds" concept may exaggerate somewhat the typical differences in the social experiences of boys and girls. Most research supporting the concept has

been conducted in age-segregated settings, such as school playgrounds and summer camps. At home and in neighborhoods children often play in age-mixed groups. As noted earlier, age-mixed play generally centers less on winning and losing than does same-age play, so age mixing reduces the difference in competitiveness between boys' and girls' play. Moreover, several studies indicate that boys and girls play together more often in age-mixed groups than in age-segregated groups (Ellis & others, 1981; Gray & Feldman, 1997).

Whether or not they play together, boys and girls are certainly interested in each other. On school playgrounds, for example, the separate boys' and girls' groups interact frequently, usually in teasing ways (Thorne, 1993). That interest begins to peak, and interactions between the sexes take on a new dimension, as children enter adolescence.



A lack of symmetry We are much more accepting of the little girl who dresses and acts like a boy (left-hand photo) than we are of the little boy who dresses and acts like a girl. Why?



© Margaret Ross / Stock, Boston

SECTION REVIEW

The child's development is affected by play with other children and by gender.

Developing Through Play

- Children everywhere play in ways that promote the development of skills needed for their survival. These include culture-specific skills, acquired by observing adults, as well as universal skills.
- Piaget contended that children learn about rules and become better moral reasoners through play, and Vygotsky contended that children develop self-control through play. Contemporary research supports these ideas.
- Age-mixed play appears to be less competitive and more conducive to teaching, learning, and development of nurturing skills than is play among age-mates.

Gender and Social Development

- Adults treat girls and boys differently beginning at birth, at least partly on the basis of socially grounded beliefs about gender. This may help to create or widen some gender differences.
- Once gender identity is established, by age 4 or 5, children attend to and mimic the culturally appropriate behaviors for their gender. They also exaggerate gender stereotypes and play increasingly with same-sex peers.
- In their peer groups, boys and girls can be thought of as inhabiting different subcultures. Boys tend to play competitively, in relatively large, hierarchical groups. Girls play more cooperatively, in smaller, more intimate groups. These differences may be muted in age-mixed play.

Adolescence: Breaking Out of the Cocoon

Adolescence is the transition period from childhood to adulthood. It begins with the first signs of *puberty* (the physical changes leading to reproductive capacity), and it ends when the person is viewed by himself or herself, and by others, as a full member of the adult community. Defined this way, adolescence in our culture begins earlier and ends later than it did in times past or still does in many other cultures.

The physical changes of puberty include a growth spurt in height, as well as changes in sexual organs and body shape toward their adult forms. In present-day North America the average age at which the growth spurt begins is 10 in girls and 12 in boys (Eveleth & Tanner, 1990). In contrast, 125 years ago in North America it began an average of 4 years later in both sexes, and even today it begins about 4 years later in many pre-industrial cultures (Eveleth & Tanner, 1990). Likewise, *menarche*—a girl's first menstruation—typically occurs between ages 12 and 13 in North America today but occurred most often between ages 16 and 17 in the nineteenth century and still occurs at that age or later in hunter-gatherer groups and some other pre-industrial cultures (Worthman, 1999; Zacharias & Wurtman, 1969). The early onset of puberty in industrialized cultures is believed to stem primarily from more food intake and less disease.

Acceptance by self and others into adulthood—the end of adolescence—comes gradually in our culture and has no clear-cut markers. In traditional societies, where adult roles are clearly defined and are learned through the child's direct involvement in the adult world, the transition to adulthood may coincide with one or another of the physical changes near the end of puberty and be officially marked by rites of passage or other celebrations. But our culture lacks such rites, and our laws dole out adult privileges and responsibilities inconsistently over a wide age span. In most states of the United States, young people are first allowed to drive a car at 16, vote at 18, and purchase alcohol at 21. The legal age for marriage varies from 12 to 19, depending on sex, place of residence, and whether parental permission has been granted. More important, the age at which people actually begin careers or families, often seen as marks of entry into adulthood, varies greatly. The average age for first marriage is currently about 27 in the United States and many other industrialized countries (Dahl, 2004). If you prefer to define adolescence as the teenage years, as many people do, you can think of this section as about adolescence and youth, with *youth* defined as a variable period after age 20 that precedes one's settling into routines of career or family.

In Erikson's life-span theory, adolescence is the stage of *identity crisis*, the goal of which is to give up one's childhood identity and establish a new identity—including a sense of purpose, a career orientation, and a set of values—appropriate for entry into adulthood (Erikson, 1968). Many developmental psychologists today disagree with Erikson's specific definition of identity and with his idea that the search for identity necessarily involves a "crisis," but nearly all agree that adolescence is a period in which young people either consciously or unconsciously act in ways designed to move themselves from childhood toward adulthood. With an eye on the ways that they affect adolescents' emerging self-identities, we shall look here at adolescents' breaking away from parents and forming closer relationships with peers; their risk-taking, delinquency, bravery, and morality; and their sexual explorations.

Shifting from Parents to Peers for Intimacy and Guidance

Breaking Away from Parental Control

"I said, 'Have a nice day,' to my teenage daughter as she left the house, and she responded, 'Will you please stop telling me what to do!'" That joke, long popular among parents of adolescents, could be matched by the following story, told by an adolescent: "Yesterday, I tried to really communicate with my mother. I told her

>>> **28**
Why does adolescence last longer today in our culture than in times past or in some other cultures?

29

What is the typical nature of the so-called adolescent rebellion against parents?

how important it is that she trust me and not try to govern everything I do. She responded, ‘Oh, sweetie, I’m so glad we have the kind of relationship in which you can be honest and tell me how you really feel. Now, please, if you are going out, wear your warmest coat and be back by 10:30.’”

<<< Adolescence is often characterized as a time of rebellion against parents, but it rarely involves out-and-out rejection. Surveys taken at various times and places over the past several decades have shown that most adolescents admire their parents, accept their parents’ religious and political convictions, and claim to be more or less at peace with them (Offer & Schonert-Reichl, 1992; Steinberg, 2001). The typical rebellion, if one occurs at all, is aimed specifically at some of the immediate controls that parents hold over the child’s behavior. At the same time that adolescents are asking to be treated more like adults, parents may fear new dangers that can accompany this period of life—such as those associated with sex, alcohol, drugs, and automobiles—and try to tighten controls instead of loosening them. So adolescence is often marked by conflicts centering on parental authority.

For both sons and daughters, increased conflict with parents is linked more closely to the physical changes of puberty than to chronological age (Steinberg, 1989). If puberty comes earlier or later than is typical, so does the increase in conflict. Such conflict is usually more intense in the early teenage years than later (Steinberg, 2001); by age 16 or so, most teenagers have achieved the balance of independence and dependence they are seeking.

Establishing Closer Relationships with Peers

30

How do young people’s friendships change during their teenage years?

<<< As adolescents gain more independence from their parents, they look increasingly to their peers for emotional support. In one study, for example, fourth-graders indicated that their parents were their most frequent providers of emotional support, seventh-graders indicated that they received almost equal support from their parents and friends, and tenth-graders indicated that they received most of their emotional support from friends (Furman & Buhrmester, 1992). When asked to describe the meaning of friendships, adolescents of both genders talk about the sharing of thoughts, feelings, and secrets, whereas younger children are more likely to talk about playing together and sharing material things (Berndt, 1992; Damon & Hart, 1992).

Breaking Down the Gender Barriers

31

How do adolescents’ peer groups promote the emergence of romantic relationships?

<<< In addition to providing emotional support, adolescent peer groups help break down the gender barriers that were erected in childhood. This is one of the main conclusions of a classic study conducted by Dexter Dunphy (1963) in Sydney, Australia. By gaining the confidence of a large number of adolescents, Dunphy was able to chart the membership of many peer groups and study them by direct observation as well as by questionnaires. He found two varieties of such groups, which he labeled *cliques* and *crowds*. A clique was a close-knit group, consisting of from three to nine individuals, who saw themselves as best friends and who spent great

amounts of time together throughout the week. A crowd was a larger group consisting of several cliques who got together for parties or other planned occasions, usually on weekends.

Younger adolescents’ cliques were exclusively of one sex or the other, but their crowds usually consisted of a combination of girls’ and boys’ cliques. The clique leaders were usually the earliest to begin dating, and the crowds were often made up of cliques to which dating couples belonged. In the crowd, boys and girls could interact in a setting made more secure by the presence of their same-sex friends. In later adolescence, as more individuals began to date regularly, single-sex cliques tended to break apart and new, more loosely associated groups, consisting of two or more pairs of dating couples, would take their place.

Good friends During adolescence, friends become increasingly important as confidants and sources of emotional support.



Brand X Pictures / Alamy

Recent studies in North America suggest that processes similar to those documented by Dunphy still operate (Steinberg, 2002; Steinberg & Morris, 2001). One study, which used questionnaires to track the experiences of a large sample of high school students from grades 9 through 11, revealed that over time the number of opposite-sex peers in an adolescent's social network tended to increase, and, at any given time, that number was positively correlated with the likelihood that the adolescent would become involved in a romantic relationship by the end of that year (Connolly & others, 2000).

Conforming to Peers

As children become teenagers, they become even more concerned than they were before about looking and behaving like their peers (Steinberg, 2002; Steinberg & Silverberg, 1986). The early teenage years are, quite understandably, the years when parents worry most about possible negative effects of peer pressure. Indeed, teenagers who belong to the same friendship groups are more similar to one another with regard to such risky behaviors as smoking, drinking, drug use, and sexual promiscuity than are teenagers who belong to different groups (Steinberg, 2002). Such similarity, however, is at least partly the result of selection; people tend to choose friends who have interests and behaviors similar to their own. Still, a number of research studies have shown that, over time, friends become more similar to one another in frequency of risky or unhealthful behaviors than they were originally (Curren & others, 1997; Jaccard & others, 2005).

Western parents and researchers tend to emphasize the negative influences of peers, but adolescents themselves, when questioned, often describe positive peer pressures, such as encouragement to avoid unhealthful behaviors and engage in healthful ones (Steinberg, 2002). On the basis of extensive studies in China, Xinyin Chen and his colleagues (2003) report that peer pressure is viewed there, by parents and educators as well as by adolescents, much more as a positive force than as a negative one. In China, according to Chen, young people as well as adults place great value on academic achievement, and adolescent peer groups do homework together and encourage one another to excel in school. In the United States, in contrast, peer encouragement for academic achievement is rare (Steinberg, 1996).

Increased Rates of Recklessness and Delinquency

On a statistical basis, throughout the world, people are much more likely to engage in disruptive or dangerous actions during adolescence than at other times in life. In Western cultures, rates of theft, assault, murder, reckless driving, unprotected sex, illicit drug use, and general disturbing of the peace all peak between the ages of 15 and 25 (Archer, 2004; Arnett, 1995, 2000). The adolescent peak in recklessness and delinquency is particularly sharp in males, as exemplified in the graphs in Figure 12.7.

What causes the increased recklessness and delinquency? Some psychologists have addressed this question by attempting to discern the underlying cognitive and motivational characteristics of the adolescent's mind that differentiate it from either the child's or the adult's. Adolescents have been described as having a *myth of invulnerability*—that is, a false sense that they are

What evidence suggests that peer pressure can have negative and positive effects? What difference in attitude about peer pressure is reported to exist in China compared to the United States?

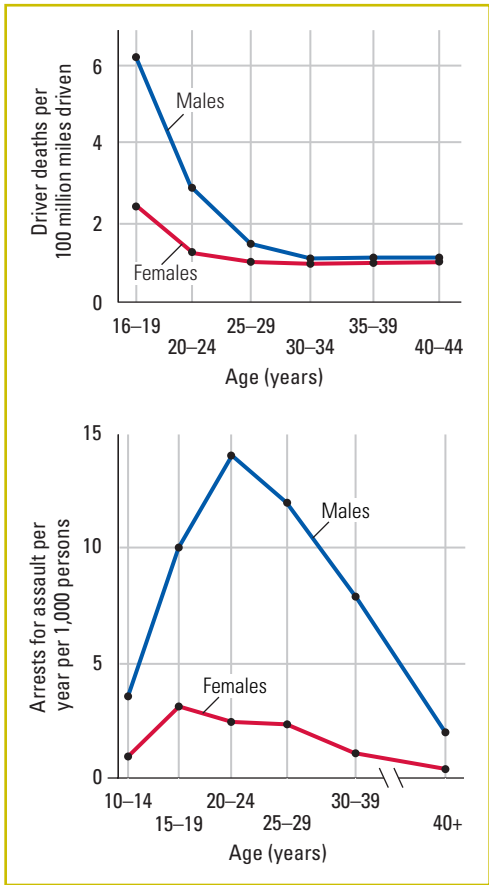


FIGURE 12.7 Evidence of heightened recklessness and aggressiveness in adolescence and youth The top graph is based on an analysis of traffic fatalities in the United States in 1970 by Wilson and Daly (1985, p. 69). The high rate of death for young drivers is probably due to inexperience as well as recklessness, but inexperience can't explain why the rate is so much higher for young men than for young women. The bottom graph, from Campbell (1995, p. 100), shows the rate of arrests for assault in 1989 in the United States, as reported by the U.S. Department of Justice.

protected from the mishaps and diseases that can happen to other people (Elkind, 1978). They have also been described as *sensation seekers*, who enjoy the adrenaline rush associated with risky behavior, and as having heightened *aggressiveness*, which leads them to be easily provoked. Reasonable evidence has been compiled for all these ideas (Arnett, 1992, 1995; Bradley & Wildman, 2002; Martin & others, 2004), but such concepts leave one wondering why adolescents have such seemingly maladaptive characteristics, which can bring grief to themselves and others.

Explanations That Focus on Adolescents' Segregation from Adults

33

What are two theories about how adolescents' segregation from adults might contribute to their recklessness and delinquency?

<<< One line of explanation focuses on aspects of adolescence that are relatively unique to modern, Western cultures. Terrie Moffitt (1993), for example, suggests that the high rate of delinquency is a pathological side effect of the early onset of puberty and delayed acceptance into legitimate adult society. She cites evidence that the adolescent peak in violence and crime is greater in modern cultures than in traditional cultures, where puberty usually comes later and young people are more fully integrated into adult activities. According to Moffitt, young people past puberty, who are biologically adults, are motivated to enter the adult world in whatever ways are available to them. Sex, alcohol, and crime are understood as adult activities. Crime, in particular, is taken seriously by adults and brings adolescents into the adult world of lawyers, courtrooms, and probation officers. Crime can also bring money and material goods that confer adultlike status.

Moffitt's theory makes considerable sense, but it does not account well for risky adolescent activities that are decidedly not adultlike. Adults do not “surf” on the tops of fast-moving trains or drive around wildly in stolen cars and deliberately crash them, as adolescents in various cities have been observed to do (Arnett, 1995).

In a theory that is in some ways the opposite of Moffitt's, Judith Harris (1995, 1998) suggests that adolescents engage in risky and delinquent activities not to join the adult world but to set themselves apart from it. Just as they dress differently from adults, they also act differently. According to Harris, their concern is not with acceptance by adults but with acceptance by their own peers—the next generation of adults. Harris agrees with Moffitt that our culture's segregation of adolescents from adult society contributes to adolescents' risky and sometimes delinquent behavior, but she disagrees about the mechanism. To Moffitt it does so by reducing the chance that adolescents can find safe, legitimate ways to behave as adults; to Harris it does so by producing adolescent subcultures whose values are relatively unaffected by those of adults. Moffitt's and Harris's theories may each contain part of the truth. Perhaps adolescents seek adultlike status while, at the same time, identifying with the behaviors and values of their adolescent subculture.

An Evolutionary Explanation of the “Young-Male Syndrome”

Neither Moffitt nor Harris addresses the question of why risky and delinquent activities are so much more readily pursued by young males than by young females or why they occur, at least to some degree, in cultures that do not segregate adolescents from adults. Even in hunter-gatherer communities, young males take risks that appear foolish to their elders and die at disproportionate rates from such mishaps as falling from trees that they have climbed too rapidly (Hewlett, 1988). To address such issues, Margo Wilson and Martin Daly (1985; Daly & Wilson, 1990) have discussed what they call “the young-male syndrome” from an evolutionary perspective, focusing on the potential value of such behavior for reproduction.

34

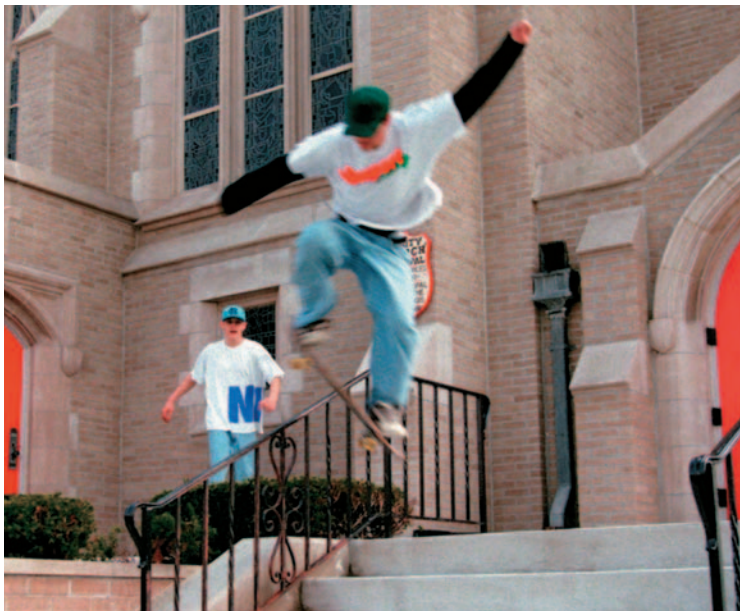
How have Wilson and Daly explained the recklessness and delinquency of adolescent males in evolutionary terms?

<<< As discussed in Chapter 3, among mammals in general, the number of potential offspring a male can produce is more variable than the number a female can produce and is more closely tied to status. In our species' history, males who took risks to achieve higher status among their peers may well have produced more offspring, on average, than those who didn't, so genes promoting that tendency may have been passed along. This view gains credibility from studies in which young women report that they indeed are sexually attracted to men who succeed in risky,

adventurous actions, even if such actions serve no social good (Kelly & Dunbar, 2001; Kruger & others, 2003). In hunter-gatherer times, willingness to take personal risks in hunting and defense of the tribe and family may well have been an especially valuable male trait. According to Wilson and Daly, train surfing, wild driving, careless tree climbing, and seemingly senseless acts of violence are best understood as ways in which young men gain status by demonstrating their fearlessness and valor.

In support of their thesis, Wilson and Daly point to evidence that a high proportion of violence among young men is triggered by signs of disrespect or challenges to status. One young man insults another, and the other responds by punching, knifing, or shooting him. Such actions are more likely to occur if other young men are present than if they aren't. No intelligent person whose real goal was murder would choose to kill in front of witnesses, but young men who commit murder commonly do so in front of witnesses. Young men also drive more recklessly when another young person is in the car than when they drive alone, while young women's driving appears to be unaffected by the presence or absence of passengers (Jackson & Gray, 1976). Of course, not all young males are reckless or violent, but that does not contradict Wilson and Daly's thesis. Those who see safer paths to high status—such as college, inherited wealth, or prestigious jobs—have less need to risk their lives for prestige and are less likely to do so.

Females also exhibit a peak in violence during adolescence and youth, although it is a much smaller peak than men's (see Figure 12.7). Anne Campbell (1995, 2002) has argued that when young women do fight physically, they, like young men, do so for reasons that can be understood from an evolutionary perspective. According to Campbell's evidence, young women fight most often in response to gossip or insults about their alleged sexual activities, which could tarnish their standing with men, and in instances when one woman appears to be trying to attract another's boyfriend.



© Jacksonville Journal Center / The Image Works

Taking a chance Risk taking reaches a peak in adolescence and young adulthood, especially among males.

An Expanded Moral Vision and Moral Sense of Self

Adolescence seems to bring out both the worst and the best in people. Adolescents can be foolhardy and violent, but they can also be heroic and work valiantly toward making the world better. Adolescence is, among other things, a period of rapid growth in the sophistication of moral reasoning and a time in which many people develop moral self-images that guide their actions.

Advancement on Kohlberg's Scale of Moral Reasoning

Over the past 40 years or so, most research on moral development has made use of a theory and methods developed originally by Lawrence Kohlberg. Kohlberg assessed moral reasoning by posing hypothetical dilemmas to people—primarily to adolescents—and asking them how they believed the protagonist should act and why. In one dilemma, for example, a man must decide whether or not to steal a certain drug under conditions in which that theft is the only way to save his wife's life. To evaluate the level of moral reasoning, Kohlberg was concerned not with whether people answered yes or no to such dilemmas but with the reasons they gave to justify their answers. Drawing partly on his research findings and partly on concepts gleaned from the writings of moral philosophers, Kohlberg (1984) proposed that moral reasoning develops through a series of stages, which are outlined in Table 12.1.

>>> 35
How did Kohlberg assess moral reasoning? How can his stages be described as the successive broadening of one's social perspective? How does research using Kohlberg's system help explain adolescent idealism?

TABLE 12.1 |
Kohlberg's stages of moral reasoning

The quotations in each stage description exemplify how a person at that stage might justify a man's decision to steal an expensive drug that is needed to save his wife's life.

Stage 1: Obedience and punishment orientation

Reasoners in this stage focus on direct consequences to themselves. An action is bad if it will result in punishment, good if it will result in reward. "If he lets his wife die, he will get in trouble."

Stage 2: Self-interested exchanges

Reasoners here understand that different people have different self-interests, which sometimes come into conflict. To get what you want, you have to make a bargain, giving up something in return. "It won't bother him much to serve a little jail term if he still has his wife when he gets out."

Stage 3: Interpersonal accord and conformity

Reasoners here try to live up to the expectations of others who are important to them. An action is good if it will improve a person's relationships with significant

others, bad if it will harm those relationships. "His family will think he's an inhuman husband if he doesn't save his wife."

Stage 4: Law-and-order morality

Reasoners here argue that to maintain social order, each person should resist personal pressures and feel duty-bound to follow the laws and conventions of the larger society. "It's a husband's duty to save his wife. When he married her he vowed to protect her."

Stage 5: Human-rights and social-welfare morality

Reasoners here balance their respect for laws with ethical principles that transcend specific laws. Laws that fail to promote the general welfare or that violate ethical principles can be changed, reinterpreted, or in some cases flouted. "The law isn't really set up for these circumstances. Saving a life is more important than following this law."

Note: The quotations are based on examples in Kohlberg (1984) and Rest (1986). A sixth stage, which emphasized universal ethical principles almost to the exclusion of other considerations, has been dropped in current versions, because of failure to find people who reason in accordance with it.

As you study the table, notice the logic underlying Kohlberg's theory of moral reasoning. Each successive stage takes into account a broader portion of the social world than does the previous one. The sequence begins with thought of oneself alone (Stage 1) and then progresses to encompass other individuals directly involved in the action (Stage 2), others who will hear about and evaluate the action (Stage 3), society at large (Stage 4), and, finally, universal principles that concern all of humankind (Stage 5).

According to Kohlberg, the stages represent a true developmental progression in the sense that, to reach any given stage, a person must first pass through the preceding ones. Thinking within one stage and discovering the limitations of that way of thinking provide the impetus for progression to the next. Kohlberg did not claim that everyone goes through the entire sequence; in fact, his research suggested that few people go beyond Stage 4 and many stop at Stage 2 or Stage 3. Nor did he link his stages to specific ages; but he did contend that adolescence and young adulthood are the times when advancement to the higher stages is most likely to occur. Figure 12.8 illustrates the results of one long-term study that supports Kohlberg's claim about the rapid advancement of moral reasoning during adolescence.

Kohlberg's theory is about moral reasoning, which is not the same thing as moral action. Kohlberg recognized that one can be a high-powered moral philosopher without being a moral person, and vice versa, yet he argued that the ability to think abstractly about moral issues does help account for the

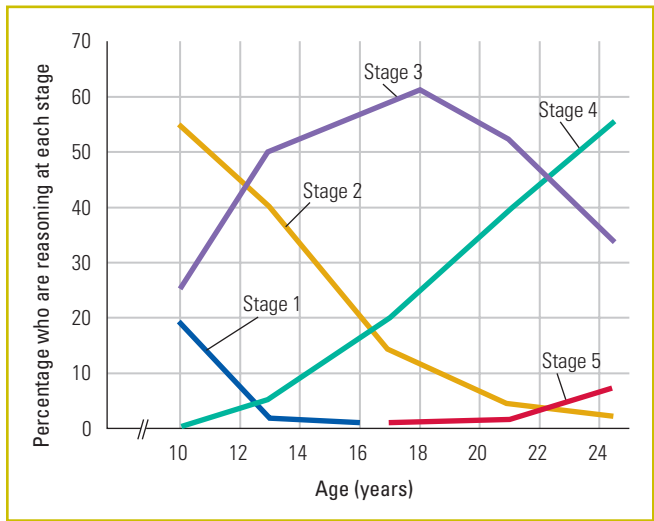


FIGURE 12.8 | Changes in moral reasoning from age 10 to 24 This graph, based on a longitudinal study of males from a relatively high socioeconomic background, shows the percentage of subjects at each age who were reasoning at each of Kohlberg's stages. Notice the sharp decline of Stages 1 and 2 reasoning during early adolescence, the rise and fall of Stage 3 reasoning, and the consistent increase of Stage 4 reasoning during adolescence and young adulthood. Little Stage 5 reasoning was found, but it, too, increased with age. (Adapted from Colby & others, 1983.)

idealism and moral commitment of youth. In line with this contention, in several research studies adolescents who exhibited the highest levels of moral reasoning were also the most likely to help others, to volunteer to work for social causes, and to refrain from taking part in actions that harm other people (Haan & others, 1968; Kuther & Higgins-D'Alessandro, 2000; Muhlberger, 2000).

Developing a Moral Image of Oneself

Another way to study moral development is to identify individuals who have distinguished themselves through moral actions and then to try to understand how their experiences and thoughts contributed to their moral commitment. Daniel Hart and his colleagues (1995) took this approach in a study of 15 exceptionally morally committed young people residing in low-income, inner-city neighborhoods in Camden, New Jersey. They had resisted the temptations of crime, drugs, and violence around them and were volunteering large amounts of time to improving their community and the lives of those within it. Extensive interviews revealed that these adolescents were motivated not by a selfless sense of duty or abstract principles of right and wrong, as Kohlberg's theory might predict, but, rather, by their intense personal investment in doing the right thing. Being moral and setting good examples for others were core aspects of their self-images, so their community services were acts of self-assertion, not of self-denial. It felt good to them to help, and it would hurt them not to.

Some of these young people developed their moral self-images in reaction to specific acts of violence they saw in their neighborhoods. For instance, one 17-year-old African American youth became committed to civic action and personal morality after his older sister was murdered. He developed the view that young people saw too many bad examples, and he became determined to be a good example himself. Perhaps the most common influence for all these people, however, was the presence of at least one adult, usually a parent, who exemplified moral virtue and had a warm relationship with the adolescent. Other studies of morally committed people, of various ages and from a variety of backgrounds, have likewise revealed the value of good parental models (Colby & Damon, 1995; Snyder & Omoto, 1992; White & Matawie, 2004). Young people may reject the specific controls that their parents attempt to exert over them, yet may guide their own actions by the values and virtues they perceive in their parents.

Sexual Explorations

Adolescence is first and foremost the time of sexual blooming. Pubertal hormones act on the body to make it reproductively functional, and on the brain to heighten greatly the level of sexual desire (discussed in Chapter 6). Girls and boys who previously watched and teased each other from the safety of their same-sex groups become motivated to move closer together, to get to know each other, to touch in ways that aren't just teasing. The new thoughts and actions associated with all these changes can bring on fear, exhilaration, dread, pride, shame, and bewilderment—sometimes all at once.

Problems Associated with Emerging Sexual Relationships

In industrialized cultures, where puberty comes early and full acceptance into adulthood comes late, most young people are sexually fully mature long before they are considered fully mature in other ways. As a culture, we glorify sex and present highly sexual images of teenagers in advertisements and movies; yet, at the same time, we typically disapprove of sex among the real teenagers of everyday life. Teenage sex is associated in the public mind with delinquency, and, indeed,

>>> 36
What ideas about moral development have been derived from studying morally committed adolescents?



Strat Douthat / AP / Wide World Photos

Remembering young heroes In the 1960s many young Americans, black and white together, risked their lives in the effort to end segregation in the South. Among those murdered in that effort were the three young men remembered here.

Gay pride For those whose sexual orientation is not the culture's ideal, the sexual awakening of adolescence can be especially difficult. Increasingly, support groups have been established by and for members of the gay and lesbian community, which offer opportunities for education, empowerment, comfort, and recreation to homosexual men and women.



© AP Photo / Michael Dwyer

the youngest teenagers to have sexual intercourse are often those most involved in delinquent or antisocial activities (Capaldi & others, 1996). Earliest intercourse usually occurs surreptitiously, often without benefit of adult advice, and often without a condom or other protection against pregnancy and sexually transmitted infections.

37

What correlations have been observed between sex education and rates of teenage pregnancy?

<<< Unprotected Sex and Teenage Pregnancy In the United States today, roughly 8 percent of teenage females aged 15–19 become pregnant in any given year (Guttmacher Institute, 2004). About half of these pregnancies are terminated by abortion, and most of the rest result in births. Most teenage pregnancies occur outside of marriage, often to mothers who are not in a position to care for their babies adequately. The good news is that the rate of teenage pregnancy has fallen in recent years, from a peak of about 12 percent reached in the late 1980s. The decline apparently stems primarily from a sharp increase in teenagers' use of condoms and other birth control methods, which itself appears to be a result of increased sex education in the schools and parents' greater willingness to discuss sex openly with their children and teenagers (Abma & others, 2004).

The United States, however, still lags behind the rest of the industrialized world in sex education and in use of birth control by teenagers, and still has the highest teenage pregnancy rate of any industrialized nation. Such countries as Germany, France, and the Netherlands have teenage pregnancy rates that are less than one-fourth of the rate in the United States (Abma & others, 2004), despite the fact that their rates of teenage sexual activity are equivalent to the rate in the United States.

38

How might the double standard regarding boys' and girls' sexuality contribute to the incidence of date rape?

<<< The Double Standard and Sexual Pressuring of Teenage Girls Another problem arises from the persistent double standard regarding sexuality for girls and boys, which probably stems from biological as well as cultural influences. Not just in our culture, but apparently everywhere, boys are more often encouraged in their sexual adventures, and more likely to feel proud of them, than are girls (Gordon & Gilgun, 1987; Michael & others, 1994). Boys, more often than girls, say that they are eager to have sex for the sheer pleasure of it, and girls, more often than boys, equate sex with love or say that they would have intercourse only with someone they would marry. One index of this difference is depicted in Figure 12.9, which shows results of a survey conducted in the United States in 2002 (Abma & others, 2004). Far more young women than young men reported that they did not want their first sexual intercourse to happen at the time that it did happen. A separate analysis showed that the younger a teenage female was when she had her first intercourse, the less likely she was to have wanted it to happen.

One consequence of the sex difference in sexual eagerness is the high frequency of date rape. In a survey taken some years ago, 6 percent of incoming college women said that they had been raped during their high school or junior high years, and an additional 33 percent said that they had experienced at least one serious form of sexual victimization short of full rape (Himelein & others, 1994). Other research suggests that rates of rape of non-college-bound young women are even higher (Rickert & others, 2004). To some degree, such events may be attributable to social pressure experienced by young men to have sexual intercourse. They may also have heard from peers, and from society at large, that a woman's resistance is a natural part of the process, or even that women secretly want to be forced to have sex—a notion belied by studies showing that all forms of sexual coercion are extremely negative, and often traumatic, experiences for the victim (Gordon & Gilgun, 1987).

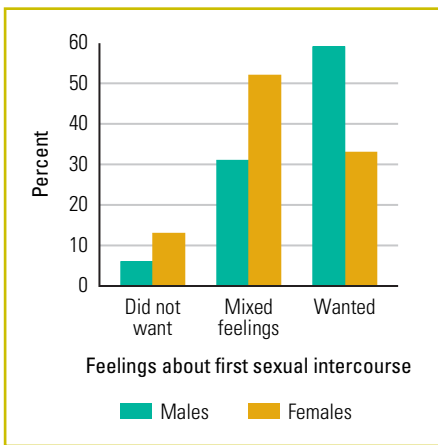


FIGURE 12.9 A difference in desire This graph shows differences in the degree to which young males and females wanted their first sexual intercourse to happen when it did happen, according to their own reports. (Adapted from Abma & others, 2004.)

Sexual Restraint and Promiscuity as Products of Evolution and Life Conditions

Evolutionary Explanation of Sex Differences in Sexual Eagerness In culture after culture, young men are more eager than young women to have sexual intercourse without a long-term commitment (Buss, 1994b, 1995; Schmitt, 2003). Why? The standard evolutionary explanation is founded on the *theory of parental investment*, developed by Robert Trivers (1972) to account for sex differences in courtship and mating in all animal species. According to the theory (discussed more fully in Chapter 3), the sex that pays the greater cost in bearing and rearing young will—in any species—be the more discriminating sex in choosing when and with whom to copulate, and the sex that pays the lesser cost will be the more aggressive in seeking copulation with multiple partners.

The theory can be applied to humans in a straightforward way. Sexual intercourse can cause pregnancy in women but not in men, so a woman's interest frequently lies in reserving intercourse until she can afford to be pregnant and has found a mate who will help her and their potential offspring over the long haul. In contrast, a man loses little and may gain much—in the economics of natural selection—through uncommitted sexual intercourse with many women. Some of those women may succeed in raising his children, sending copies of his genes into the next generation, at great cost to themselves and no cost to him. Thus natural selection may well have produced an instinctive tendency for women to be more sexually restrained than men.

How Teenage Sexuality May Depend on Conditions of Rearing A key word in the last sentence of the preceding paragraph is *tendency*. Great variation on the dimension of sexual restraint versus promiscuity exists within each sex, both across cultures and within any given culture (Belsky & others, 1991; Small, 1993). As in any game of strategy, the most effective approach that either men or women can take in courtship and sex depends very much on the strategy taken by the other sex. In communities where women successfully avoid and shun men who seek to behave promiscuously, promiscuity proves fruitless for men and the alternative strategy of fidelity works best. Conversely, in communities where men rarely stay around to help raise their offspring, a woman who waits for “Mr. Right” may wait forever. Cross-cultural studies have shown that promiscuity prevails among both men and women in cultures where men devote little care to young, and sexual restraint prevails in cultures where men devote much care (Barber, 2003; Draper & Harpending, 1988; Marlowe, 2003).

39 How can the sex difference in desire for uncommitted sex be explained in evolutionary terms?

40 How can sexual restraint and promiscuity, in both sexes, be explained as adaptations to different life conditions? What evidence suggests that the presence or absence of a father at home, during childhood, may tip the balance toward one strategy or the other?



"I'm focusing less on girls and more on accumulating the things girls will eventually want."

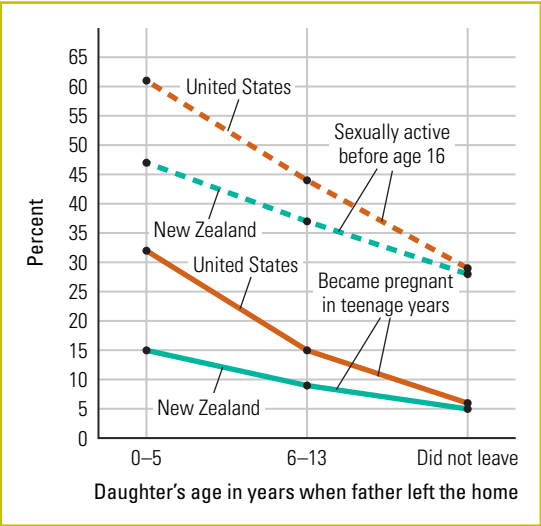
Some researchers have theorized that natural selection may have predisposed humans to be sensitive to cues in childhood that predict whether one or the other sexual strategy will be more successful. One such cue may be the presence or absence of a caring father at home. According to a theory originated by Patricia Draper and Henry Harpending (1982), the presence of a caring father leads girls to grow up assuming that men are potentially trustworthy providers and leads boys to grow up assuming that they themselves will be such providers; these beliefs promote sexual restraint and the seeking of long-term commitments in both sexes. If a caring father is not present, according to the theory, girls grow up assuming that men are untrustworthy "cads" rather than "dads," and that assumption leads them to flaunt their sexuality to extract what they can from men in short-term relationships; and boys grow up assuming that long-term commitments to mates and care of children are not their responsibilities, and that assumption leads them to go from one sexual conquest to another. These assumptions

may not be verbally expressed or even conscious, but are revealed in behavior.

In support of their theory, Draper and Harpending (1982, 1988) presented evidence that even within a given culture and social class, adolescents raised by a mother alone are generally more promiscuous than those raised by a mother and father together. In one early study, teenage girls who were members of the same community playground group and were similar to one another in socioeconomic class were observed for their degrees of flirtatiousness, both with boys on the playground and with an adult male interviewer. Girls who were raised by a mother alone—after divorce early in the girl's childhood—were, on average, much more flirtatious than girls who still had a father at home (Hetherington, 1972). Other, more recent studies have revealed that girls raised by a mother alone are much more likely to become sexually active in their early teenage years and to become pregnant as teenagers than are girls raised by a mother and father (Ellis, 2004; Ellis & others, 2003). The results of two studies, conducted in the United States and New Zealand, are depicted in Figure 12.10 (Ellis & others, 2003).

Researchers have also found that girls raised in father-absent homes tend to go through puberty earlier than do those raised in homes where a father is present

FIGURE 12.10 Effect of fathers' absence on daughters' rates of early sexual activity and teenage pregnancy These data came from two large-scale studies, one in the United States and one in New Zealand. The rates of early sexual activity and teenage pregnancy were statistically adjusted to factor out possible confounding variables, including race, socioeconomic class, the mother's age at first birth, family life stress, parental discipline style, degree of parental monitoring, and neighborhood dangers. (Adapted from Ellis & others, 2003.)





Culver Pictures



Stock, Boston

(Ellis, 2004). Early puberty may be part of the mechanism through which experiences at home affect the onset of sexual activity.

Draper and Harpending's theory is still controversial. The evidence for it comes entirely from correlational studies, which, as you know, can support but rarely prove specific cause-effect relationships. Statistical controls have revealed that the correlations between father absence and girls' sexual development probably do not stem from differences in race, socioeconomic status, or degree of parental monitoring, and there is at least some evidence that they cannot be accounted for by genetic differences between parents who divorce early and those who divorce later or do not divorce (Ellis, 2004; Ellis & others, 2003). Regardless of whether or not the theory eventually proves to be true, it nicely illustrates the attempt of many contemporary psychologists to understand the life course in terms of alternative strategies that are at least partly prepared by evolution but are brought selectively to the fore by life experiences.

A matter of time and place
Victorians were probably somewhat less concerned with propriety and contemporary Western couples are probably somewhat more ambivalent about sexual freedom than these illustrations suggest. Nevertheless, permissible sexual behavior varies greatly from culture to culture. Across cultures, sexual promiscuity correlates positively with a high divorce rate and low rate of paternal support of children.

SECTION REVIEW

Adolescence is a period of breaking away and developing an adult identity.			
Shifting from Parents to Peers - Adolescent conflict with parents generally centers on the desire for greater independence from parental control. - Increasingly, adolescents turn to one another rather than parents for emotional support. - Adolescent peer groups serve to break down the gender barriers of childhood and lead to romantic relationships. - Peer pressure can have positive as well as negative influences.	Recklessness and Delinquency - Risky and delinquent behaviors are more frequent in adolescence than in other life stages. - Segregation from adults may promote delinquency by allowing adolescents few positive adult ways to behave or by creating an adolescent subculture divorced from adult values. - Risky and delinquent behavior is especially common in young males. It may serve to enhance status, ultimately as part of competition to attract females.	The Moral Self - In Kohlberg's theory, moral reasoning develops in stages, progressing in breadth of social perspective. Rapid advancement in moral reasoning often occurs in adolescence. - A study of morally committed adolescents suggests that their moral actions were rooted in their self-images as moral persons who set good examples for others.	Sexual Explorations - Across cultures, rates of teenage pregnancy are inversely related to the availability of sex education and contraceptives. - Sex differences in eagerness to have sex can be explained in terms of parental investment. - An evolution-based theory, supported by correlational research, suggests that the presence or absence of a father at home may affect the sexual strategy—restraint or promiscuity—chosen by offspring.

Adulthood: Finding Satisfaction in Love and Work

In his life-span theory, Erikson (1963) proposed that establishing intimate, caring relationships and finding fulfillment in work are the main tasks of early and middle adulthood. In this respect he was following the lead of Sigmund Freud (1935/1960), who defined emotional maturity as the capacity to love and to work. Some psychologists believe that adult development follows a predictable sequence of crises or problems to be resolved (Erikson, 1963; Levinson, 1986), while others contend that the course of adulthood in our modern culture is extraordinarily variable and unpredictable (Neugarten, 1979, 1984). But in essentially every psychological theory of adult development, caring and working are the two threads that weave the fabric of adulthood.

Love

We are a romantic species. In every culture for which data are available, people describe themselves as falling in love (Fisher, 2004; Jankowiak & Fischer, 1991). We are also a marrying species. In every culture, adults of child-producing age enter into long-term unions sanctioned by law or social custom, in which the two members implicitly or explicitly promise to care for each other and the offspring they produce (Fisher, 1992; Rodseth & others, 1991), although cultures vary greatly in the degree to which people abide by those promises. Love and marriage do not necessarily go together, but they often do, and in most cultures their combination is considered ideal. In some cultures people fall in love and then get married; in others they get married—through an arrangement made by the couple's parents—and then, if fate works as hoped for, fall in love. Researchers have attempted to understand the underlying psychological elements of romantic love and to learn why some marriages are happy and others are not.

Romantic Love Viewed as Adult Attachment

41

How is romantic love like infant attachment? What evidence suggests continuity in attachment quality between infancy and adulthood?

<<< Romantic love is similar in form, and perhaps in underlying mechanism, to the attachment that infants develop with their parents (Hazan & Shaver, 1994; Diamond, 2004). Close physical contact, caressing, and gazing into each other's eyes are crucial to the early formation of both types of relationships, and cooing and baby talk are common. A sense of fusion with the other reigns when all is well, and a feeling of exclusivity—that the other person could not be replaced by anyone else—prevails. The partners feel most secure and confident when they are together and may show physiological evidence of distress when separated (Feeney & Kirkpatrick, 1996). Research with animals, and some preliminary research with humans, indicates that the neural and hormonal mechanisms of mating bonds are similar to those of the bonds between infants and caregivers (Diamond, 2004; Fisher, 2004).

The emotional bond is not simply a by-product of shared pleasures. In long-married couples it may exist even when the two have few interests in common. Sometimes the bond reveals its full intensity only after separation or divorce or the death of one partner. Common experiences in such cases are intense anxiety, depression, and feelings of loneliness or emptiness that are not relieved even by highly supportive friends and an active social life (Stroebe & others, 1996; Weiss, 1975).

As with infants' attachments with their caregivers, the attachments that adults form with romantic partners can be classed as *secure* (characterized by comfort), *anxious* (characterized by excessive worry about love or lack of it from the partner), or *avoidant* (characterized by little expression of intimacy or by ambivalence about commitment). Studies using questionnaires and interviews have revealed continuity between people's descriptions of their adult romantic attachments and their recollections of their childhood relationships with their parents (Fraley, 2002;



Frank Siteman / Stock, Boston



T. J. Florian / Rainbow

Cultural connections Research on successful marriage may misrepresent the lives of couples in traditional non-Western societies. The meaning of marriage, and definition of marital success, may vary from culture to culture.

Hazan & Shaver, 1987, 1994). People who recall their relationships with parents as warm and secure typically describe their romantic relationships in similar terms, and those who recall anxieties and ambiguities in their relationships with parents tend to describe analogous anxieties and ambiguities in their romances. Such continuity could stem from a number of possible causes, including a tendency for adult experiences to color one's memories of childhood. Most attachment researchers, however, interpret it as support for a theory developed by Bowlby (1982), who suggested that people form mental models of close relationships based on their early experiences with their primary caregivers and then carry those models into their adult relationships (Fraley & Brumbaugh, 2004).

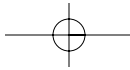
Ingredients of Marital Success

In the sanitized version of the fairy tale (but not in the Grimm brothers' original), love allows the frog prince and the child princess to transform each other into perfect human adults who marry and live happily ever after. Reality is not like that. Between half and two-thirds of new marriages in North America are predicted to end in divorce (Gottman, 1998), and even among couples who don't divorce, many are unhappy in marriage. Why do some marriages work while others fail?

In interviews and on questionnaires, happily married partners consistently say that they *like* each other; they think of themselves not just as husband and wife but also as best friends and confidants (Buehlman & others, 1992; Lauer & Lauer, 1985). They use the term *we* more than *I* as they describe their activities, and they



© Hilary B. Price. Reprinted with permission of King Features Syndicate



42

What are some characteristics of happily married couples? Why might marital happiness depend even more on the husband's capacity to adjust than on the wife's?

tend to value their interdependence more than their independence. They also talk about their individual commitment to the marriage, their willingness to go more than halfway to carry the relationship through difficult times.

<<< Happily married couples apparently argue as often as unhappily married couples do, but they argue more constructively (Gottman, 1994; Gottman & Krokoff, 1989). They genuinely listen to each other, focus on solving the problem rather than “winning” or proving the other wrong, show respect rather than contempt for each other's views, refrain from bringing up past hurts or grievances that are irrelevant to the current issue, and intersperse their arguments with positive comments and humor to reduce the tension. Disagreement or stress tends to draw them together, in an effort to resolve the problem and to comfort each other, rather than to drive them apart (Murray & others, 2003).

In happy marriages, both partners are sensitive to the unexpressed feelings and needs of the other (Gottman, 1998). In contrast, in unhappy marriages, there is often a lack of symmetry in this regard; the wife perceives and responds to the husband's unspoken needs, but the husband does not perceive and respond to hers (Gottman, 1994, 1998). Perhaps this helps explain why, in unhappy marriages, the wife typically feels more unhappy and manifests more physiological distress than does her partner (Gottman, 1994; Levinson & others, 1993). Such observations bring to mind the differing styles of interaction and communication shown by girls and boys in their separate play groups. Researchers have found that in all sorts of relationships, not just marriages, women are on average better than men at attending to and understanding others' unspoken emotions and needs (Thomas & Fletcher, 2003). Success in marriage may often depend on the husband's willingness and ability to acquire some of the intimacy skills that he practiced less, as a child, than did his wife.

Employment

Most adults in the United States, both men and women, work outside the home even if they have young children. From the perspective of the past half century of American history, this fact represents a social change. Only 19 percent of married women with children under 6 years old were employed outside the home in 1960, compared with 64 percent in 2002 (Bureau of Labor Statistics, 2004). From a broader and longer-term perspective, however, there is nothing new about women's going out to work. In our evolutionary history as hunter-gatherers, women apparently left camp daily to forage for food for their families—sometimes taking their young children with them, and sometimes leaving them in others' care at camp (Konner, 1982; Morelli & Tronick, 1991). In most subsequent cultures, women as well as men have worked at jobs other than child care and domestic chores.

Work, of course, is first and foremost a means of making a living, but it is more than that. It occupies an enormous portion of adult life. Work can be tedious and mind-numbing or exciting and mind-building. At its best, work is for adults what play is for children. It brings them into social contact with their peers outside the family, poses interesting problems to solve and challenges to meet, promotes development of physical and intellectual skills, and is fun.

In surveys of workers, people most often say they enjoy their work if it is (a) complex rather than simple, (b) varied rather than routine, and (c) not closely supervised by someone else (Galinsky & others, 1993; Kohn, 1980). Sociologist Melvin Kohn refers to this much-desired constellation of job characteristics as *occupational self-direction*. A job high in occupational self-direction is one in which the worker makes many choices and decisions throughout the workday. Research suggests that jobs of this sort, despite their high demands, are for most people *less* stressful—as measured by effects on workers' mental and physical health—than are jobs in which workers make few choices and are closely supervised (Spector, 2002). Research also suggests that such jobs promote certain positive personality changes.

Effects of Occupational Self-Direction on Workers and Their Children

In a massive, long-term study, conducted in both the United States and Poland, Kohn and his colleague Kazimierz Slomczynski found that workers who moved into jobs that were high in self-direction, from jobs that were low in that quality, changed psychologically in certain ways, relative to other workers (Kohn & Slomczynski, 1990). Over their time in the job, they became more intellectually flexible, not just at work but in their approaches to all areas of life. They began to value self-direction more than they had before, in others as well as in themselves. They became less authoritarian and more democratic in their approaches to child rearing; that is, they became less concerned with obedience for its own sake and more concerned with their children's ability to make decisions. Kohn and Slomczynski also tested the children. They found that the children of workers in jobs with high self-direction were more self-directed and less conforming than children of workers whose jobs entailed less self-direction. Thus the job apparently affected the psychology not just of the workers but also of the workers' children.

All these effects occurred regardless of the salary level or prestige level of the job. In general, blue-collar jobs were not as high in self-direction as white-collar jobs, but when they were as high, they had the same effects. Kohn and Slomczynski contend that the effects they observed on parenting may be adaptive for people whose social class determines the kinds of jobs available to them. In settings where people must make a living by obeying others, it may be sensible to raise children to obey and not question. In settings where people's living depends on their own decision making, it makes more sense to raise children to question authority and think independently.

Balancing Out-of-Home and At-Home Work

Women, more often than men, hold two jobs—one outside the home and the other inside. Although men today spend much more time at housework and child care than they did 30 years ago (Barnett & Hyde, 2001), the average man is still less involved in these tasks than the average woman. Even in families in which the husband and wife work equal hours outside the home, the wife still usually does more than half of the child rearing, grocery shopping, cooking, and housecleaning (Calasanti & Bailey, 1991; Deutsch, 2001). Career women, more than career men, feel torn between career and family. In one study, women business executives and college professors rated themselves worse as parents than their husbands did, even though the women in fact spent more time at parenting than their husbands did (Biernat & Wortman, 1991).

Despite the traditional stereotype of who “belongs” where—or maybe because of it—research suggests that wives enjoy their out-of-home work more than their at-home work, while the opposite is true for husbands. Reed Larson and his colleagues (1994) asked 55 working-class and middle-class married parents of school-age children to wear pagers as they went about their daily activities. The pagers beeped at random times during the daytime and evening hours, and at each beep the person filled out a form to describe his or her activity and emotional state just prior to the beep. Over the whole set of reports, wives and husbands did not differ in their self-rated happiness, but wives rated themselves happier at work than at home and husbands rated themselves happier at home than at work. These results held even when the specific type of activity reported at work or at home was the same for the two. When men did laundry or vacuumed at home, they said they enjoyed it; when women did the same, they more often said they were bored or angry. Why this difference?

43
What evidence suggests that the type of job one has can alter one's way of thinking and style of parenting, and can influence the development of one's children?



Donald Dietz / Stock, Boston

Occupational self-direction Self-employed people, and others who make most of their own decisions throughout a typical workday, learn to think independently and question authority, and they tend to raise their children to do the same.

44
What difference has been found between husbands' and wives' enjoyment of out-of-home and at-home work? How did the researchers explain that difference?



"Oh, no. If you're going to be primary breadwinner, then I get to be primary caregiver."

The men's greater enjoyment of housework might stem simply from the fact that they do it less than their wives, but that explanation doesn't hold for the women's greater enjoyment of out-of-home work. Even the wives who worked out of home for as many hours as their husbands, at comparable jobs, enjoyed that work more than their husbands did. Larson and his colleagues suggest that both differences derived from the men's and women's differing perceptions of their choices and obligations. Men enjoyed housework because they didn't really consider it their responsibility; they did it by choice and as a gallant means of "helping out." Women did housework because they felt that they had to; if they didn't do it, nobody else would, and visitors might assume that a dirty house was the woman's fault. Conversely, the men had a greater sense of obligation, and reduced sense of choice, concerning their out-of-home work. They felt that it was their duty to support their family in this way. Although the wives did not experience more actual on-the-job choice in Kohn's sense of occupational self-direction than their husbands did, they apparently had a stronger global belief that their out-of-home work was optional. As Larson and his colleagues point out, the results do match a certain traditional stereotype: Men "slave" at work and come home to relax, and women "slave" at home and go out to relax. What's interesting is that the very same activity can be slaving for one and relaxing for the other, depending apparently on the feeling of obligation or choice.

Growing Old

According to current projections of life expectancy, and taking into account the greater longevity of college graduates compared with the rest of the population, the majority of you who are reading this book will live past your eightieth birthday. How will you change as you grow beyond middle adulthood into old age?

Some young people fear old age. There is no denying that aging entails loss. We lose gradually our youthful looks and some of our physical strength, agility, sensory acuity, mental quickness, and memory. We lose some of our social roles (especially those related to employment) and some of our authority (as people take us less seriously). We lose loved ones who die before we do, and, of course, with the passing of each year we lose a year from our own life span.

Yet, if you ask the elderly, old age is not as bad as it seems to the young. In one study, the ratings that younger adults gave for their expected life satisfaction in late adulthood were much lower than the actual life-satisfaction scores obtained from people who had reached old age (Borges & Dutton, 1976). Several research studies, using a variety of methods, have revealed that elderly people, on average, report greater current enjoyment of life than do middle-aged people, and middle-aged people report greater enjoyment than do young adults (Mroczek, 2001; Sheldon & Kasser, 2001). This finding has been called the "paradox of aging": Objectively, life looks worse in old age—there are more pains and losses—but subjectively, it feels better. As we age, our priorities and expectations change to match realities, and along with losses there are gains. We become in some ways wiser, mellower, and more able to enjoy the present moment.

A Shift in Motivation for Activities and Relationships

In the past, much research and writing on aging centered on the extent to which people choose to disengage themselves from activities and social involvements as they grow older. Some researchers supported a *disengagement theory* of aging, arguing that



Jodi L. Buren Photo

elderly people gradually and by choice withdraw from active involvement with the world around them and attend increasingly to their subjective inner worlds of memories, thoughts, dreams, and feelings (Cumming & Henry, 1961; Jung, 1969). Others held an *activity theory* of aging, contending that elderly people prefer to remain active and pointing to evidence that those who are most involved in the external world are also, by their own reports, the happiest (Botwinick, 1984). According to that theory, when disengagement occurs, it usually comes not from choice but from constraints such as illness or forced retirement that limit opportunities for engagement.

Increased Focus on the Present and the Positive More recently, research has shifted from the question of whether or not elderly people prefer to be active to more subtle questions about the kinds of activities they choose and the motives underlying those choices. From such research, Laura Carstensen (1992; Löckenhoff & Carstensen, 2004) has developed what she calls the *socioemotional selectivity theory* of adult development and aging. According to Carstensen, as people grow older—or, more precisely, as they see that they have fewer years left—they become gradually more concerned with enjoying the present and less concerned with activities that function primarily to prepare for the future. Young people are motivated to explore new pathways and meet new people, despite the disruptions and fears associated with the unfamiliar. Such activities provide new skills, information, social contacts, and prestige that may prove useful in the future. But with fewer years left, the balance shifts. The older one is, the less sense it makes to sacrifice present comforts and pleasures for possible future gain. According to Carstensen, this idea helps us understand many of the specific changes observed in the elderly.

As people grow older, they devote less attention and energy to casual acquaintances and strangers and more to people with whom they already have close emotional ties (Löckenhoff & Carstensen, 2004; Fung & others, 1999, 2001). Long-married couples grow closer. Husbands and wives become more interested in enjoying each other and less interested in trying to improve, impress, or dominate each other, and satisfaction with marriage becomes greater (Levenson & others, 1993). Ties with children, grandchildren, and long-time friends also grow stronger and more valued with age, while broader social networks become less valued and shrink in size. Such changes have been observed not just among the elderly but also among younger people whose life expectancy is shortened by AIDS or other terminal illnesses (Carstensen & Fredrickson, 1998). People who continue working into old age typically report that they enjoy their work more than they did when they were younger (Levinson, 1978; Rybash & others, 1995). They become, on average, less concerned with the rat race of advancement and impressing others and more concerned with the day-to-day work itself and the pleasant social relationships associated with it.

Selective Attention to and Memory for the Positive In several experiments, Carstensen and her colleagues have shown that older people, unlike younger people, attend more to emotionally positive stimuli than to emotionally negative stimuli and show better memory for the former than the latter (Kennedy & others, 2004; Mather & Carstensen, 2003). In one experiment (Charles & others, 2003), for example, young adults (aged 18–29), middle-aged adults (41–53) and old adults (65–80) were shown pictures of positive scenes (such as happy children and puppies), negative scenes (such as a plane crash and garbage), and neutral scenes (such as a runner and a truck). Then they were asked to recall and briefly describe, from memory, as many of the pictures as they could. One predictable result was that older people recalled fewer of the scenes, overall, than did younger people; memory for all information declines as we age (discussed in Chapters 9 and 10). The

How does the socioemotional selectivity theory differ from the earlier disengagement and activity theories of aging? What observations about elderly people's behaviors and memories are explained by the theory?



Macduff Everton / The Image Works

Socioemotional satisfaction As people grow older, they tend to become more interested in enjoying the present and less future-oriented than they were when they were younger.

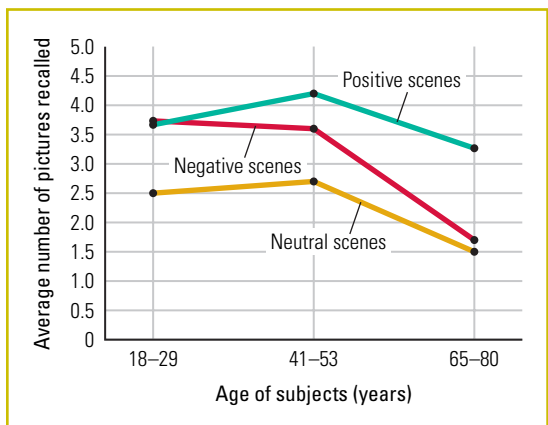


FIGURE 12.11 | Effect of age on memory for positive, negative, and neutral scenes Memory for depictions of emotionally negative scenes declined much more sharply between middle age and old age than did memory for depictions of positive or neutral scenes. (Adapted from Charles & others, 2003.)

other, more interesting finding was that the decline in memory was much sharper for the negative scenes than for the positive scenes (see Figure 12.11). Apparently, selective attention and memory is one means by which older people regulate their emotions in a positive direction.

Approaching Death

The one certainty of life is that it ends in death. Surveys have shown that fear of death typically peaks in the person's fifties, which is when people often begin to see some of their peers dying from such causes as heart attack and cancer (Karp, 1988; Riley, 1970). Older people have less fear of death (Cicirelli, 2001). They are more likely to accept it as inevitable; and death in old age, when a person has lived a full life, seems less unfair than it did earlier.

Various theories have been offered regarding the stages or mental tasks involved in preparing for death. On the basis of her experience with caring for dying patients, Elisabeth Kübler-Ross (1969) proposed that people go through five stages when they hear that they are incurably ill and will soon die: (1) *denial*—“The diagnosis can't be right,” or “I'll lick it,” (2) *anger*—“Why me?” (3) *bargaining*—“If I do such-and-such, can I live longer?” (4) *depression*—“All is lost,” and (5) *acceptance*—“I am prepared to die.” Another theory is that preparation for death consists of reviewing one's life and trying to make sense of it (Butler, 1975).

As useful as such theories may be in understanding individual cases, research shows that there is no universal approach to death. Each person faces it differently. One person may review his or her life; another may not. One person may go through one or several of Kübler-Ross's stages but not all of them, and others may go through them in a different order (Kastenbaum, 1985). The people whom I have seen die all did it in pretty much the way they did other things in life. When my mother discovered that she would soon die—at the too-young age of 62—it was important to her that her four sons be around her so that she could tell us some of the things she had learned about life that we might want to know. She spent her time talking about what, from her present perspective, seemed important and what did not. She reviewed her life, not to justify it but to tell us why some things she did worked out and others didn't. She had been a teacher all her life, and she died one.

SECTION REVIEW

Love and work are the major themes of adulthood.

Love

- Romantic love has much in common with infant attachment to caregivers. The attachment style developed in infancy—secure, anxious, or avoidant—seems to carry forward into adult attachments.
- Happy marriages are generally characterized by mutual liking and respect between partners, individual commitment to the marriage, and constructive arguing.
- Wives are generally better than husbands at perceiving and responding to their spouse's unspoken needs, so marital happiness often depends on the husband's developing those abilities.

Employment

- Jobs that permit considerable self-direction are enjoyed more than other jobs. Workers with such jobs become more self-directed in their overall approach to life and may, through their parenting, pass this trait on to their children.
- When husbands and wives both work outside the home, wives generally enjoy the out-of-home work more, and husbands enjoy the at-home work more. Perhaps the non-stereotypical task seems more a matter of choice, which promotes greater enjoyment.

Growing Old

- The elderly generally report greater life satisfaction than do middle-aged and young adults, despite the objective losses that accompany aging.
- Elderly people focus more on the present and less on the future than do younger people. They also attend to and remember emotionally positive stimuli more than negative ones.
- Though various theories make general statements about what people do as they approach death, it is really a highly individual matter.

Concluding Thoughts

This chapter has run the life course, from birth to death. As you review it, you might organize your thinking around the following two themes:

1. Development as a product of evolution The universal characteristics of human beings at each life phase are evolved adaptations that promote survival and reproduction, either immediately within that phase or through helping the individual prepare for challenges ahead. In your review, think about how this idea might apply to (a) the nature of infants' attachments to caregivers; (b) the characteristics of toddlers that influence their sharing, helping, and complying with requests; (c) the nature of children's play, including differences between girls' play and boys' play; (d) the tendency of children to attend to gender differences and to exaggerate them; (e) the heightened risk-taking and, sometimes, aggressiveness of adolescence and youth; (f) the differences between young men and young women in courtship style; (g) the attachment bonds that form between adults in love; and (h) the changes in priorities that arise in old age. As you think about each of these topics, you may well decide that the evolutionary accounts offered for some are more complete or compelling than those offered for others. You may think of alternative explanations, either founded in evolutionary theory or not.

2. Development as adaptation to the specifics of one's social environment Evolution has endowed us not with rigid instincts but with tendencies, drives, and alternative strategies that are facilitated, inhibited, selected, or redirected in accordance with our experiences in our social environments. Researchers have found many reliable correlations between aspects of a person's social environment and the person's development. How might these correlations be understood as adaptive changes that help the person to meet the specific social conditions of his or her life? Think about this question as it might apply to behavioral changes or differences associated with (a) caregiver responsiveness or unresponsiveness to infants' signs of distress; (b) cultural and individual differences in sleeping arrangements for infants and young children; (c) differing styles of parental discipline of children; (d) age-mixed compared with age-segregated play environments; (e) cultural differences in the degree to which children witness violence among adults; (f) social-class differences that alter the chance of adolescents' achieving status through non-risky means; (g) good parental models for moral development in adolescence; (h) conditions that may alter sexual strategy toward either promiscuity or restraint; and (i) occupational self-direction in adulthood. As you review, you may find that some of the explanations just called for are lacking or incomplete, or you may decide that the correlational evidence does not warrant the cause-effect inference implied. In such cases, how might future research fill the gaps and establish causal relationships?

Further Reading

WENDY CRAIG (2000). *Childhood social development: The essential readings*. Malden, MA: Blackwell.

The 13 articles in this collection are representative of current theory and research about social development and are quite readable by the non-specialist. The topics include infant-mother attachment, parenting, friendships and peer relations, gender differences in development, moral reasoning, aggression, and development of antisocial behavior.

BARBARA ROGOFF (2003). *The cultural nature of human development*. Oxford, UK: Oxford University Press.

A thesis of this book, by a leading cross-cultural researcher, is that throughout development people look to others in their culture to see how to behave. The book examines cross-cultural differences in disciplinary styles, sleeping arrangements, children's segregation from or

integration into the adult community, concepts of independence and interdependence, responsibilities at each life stage, concepts of adolescence, initiation into adulthood, and education.

JUDY DELOACHE & ALMA GOTTLIEB (2000). *A world of babies: Imagined childcare guides for seven societies*. Cambridge, UK: Cambridge University Press.

Each chapter of this book is written as an advice-to-parents manual from the point of view of a culture very different from our own. The cultures include seventeenth-century Puritan New England and various indigenous and traditional cultures scattered throughout the world. The manuals are delightful to read, emotionally moving, and thought-provoking.

ELEANOR MACCOBY (1998). *The two sexes: Growing apart, coming together*. Cambridge, MA: Harvard University Press.

This book is about the ways in which a person's gender influences his or her life from infancy to adulthood. The first seven chapters are concerned with the biological, social, and cognitive factors that lead boys and girls to play separately and differently and to develop differently. The final three chapters are about the coming together of the two genders in adolescence and adulthood.

JOHN M. GOTTMAN (1994). *What predicts divorce? The relationship between marital processes and marital outcomes*. Hillsdale, NJ: Erlbaum.

A leading researcher summarizes the results of many years' research on the psychological variables that influence, or at least correlate with, marital happiness, strife, and divorce.



THE PERSON IN A WORLD OF PEOPLE

We humans are social beings through and through. We are motivated to understand others; we are concerned about what others think of us; and our understanding of ourselves is strongly affected by our perception of what others think of us. This two-chapter unit is on social psychology, the study of human thought and behavior in relation to the social contexts in which they occur. Chapter 13 is about the mental processes involved in understanding others, ourselves, and the social world in general. Chapter 14 is about some of the ways in which other people influence our behavior.

part 7

Social Perception and Attitudes

FORMING IMPRESSIONS OF OTHER PEOPLE

- Making Attributions from Observed Behavior
- Effects of Prior Information and Physical Appearance on Person Perceptions
- Forming Impressions on the Internet

PERCEIVING AND EVALUATING THE SELF

- Seeing Ourselves Through the Eyes of Others
- Actively Constructing Our Self-Perceptions
- East-West Differences in Self-Perceptions

PERCEIVING OURSELVES AND OTHERS AS MEMBERS OF GROUPS

- Shifting Between Personal and Social Identities
- Stereotypes and Their Automatic Effects on Perception and Behavior

ATTITUDES: THEIR ORIGINS AND THEIR EFFECTS ON BEHAVIOR

- Attitudes as Products of Conditioning and Reasoning
- Attitudes as Social Norms
- Attitudes as Rationalizations to Attain Cognitive Consistency
- Attitudes as Guides to Behavior

We are intensely social animals. We are designed, by natural selection, to depend on one another for even our most basic needs. We are not good, for example, at obtaining food alone; we need the help of others and the knowledge that is shared by members of a human community. Throughout our evolutionary history, to be thrown out of the tribe was tantamount to death. We are also *thinking* social animals. Most of what we think about is other people, ourselves, our relationships with other people, and the social conventions and norms that are essential aspects of life in any human society.

This is the first of a two-chapter sequence on **social psychology**, the subfield of psychology that deals most explicitly with how people view one another and are influenced by one another. This chapter focuses on *person perception*, the processes by which people perceive and understand one another and themselves, and on *attitudes*, the evaluative beliefs that people have about their social world and the entities within it. The next chapter focuses on the effects that those perceptions and beliefs have on people's emotions and actions.

Forming Impressions of Other People

One of social psychology's pioneers, Fritz Heider (1958), pointed out long ago that human beings are natural psychologists—or *naïve psychologists*, to use his term. Heider contended that humans are naturally interested in assessing the personality characteristics and attitudes of other humans they encounter. From an evolutionary perspective, this drive to understand others has clear adaptive functions. Other individuals can help us or hurt

chapter 13

1 In what sense are people natural psychologists?

us in our life endeavors. Understanding others helps us predict their behavior and decide how to interact with them. Consistent with Heider's general view, researchers have found that people untrained in psychology are nonetheless often remarkably accurate and quick at assessing others' personalities by observing their behavior (Ambady & others, 2000; Borkenau & others, 2004).

Yet, as Heider himself pointed out, the accuracy of our judgments of others sometimes suffers from certain consistent mistakes, or *biases*. These biases occur most often when we are not using our full mental resources, or have only limited information with which to reason, or have unconscious motives for reaching particular conclusions. Such biases interest social psychologists for two reasons. First, they provide clues about the mental processes that contribute to accurate as well as inaccurate perceptions and judgments. In this regard, social psychologists' interest in biases is analogous to perceptual psychologists' interest in visual illusions, which (as discussed in Chapter 8) provide clues to understanding normal, accurate visual perception. Second, an understanding of biases can promote social justice. By helping people understand the psychological tendencies that contribute to prejudice and unfair treatment of other people, social psychological findings can help people overcome such biases.

Making Attributions from Observed Behavior

Actions are directly observable and thoughts are not. Therefore, our judgments about the personalities of people we encounter are based largely on what we observe of their actions. As “naïve psychologists” we intuitively, in our everyday experiences, form impressions of people's personalities on the basis of their actions. For example, if a new acquaintance smiles at you, you do not simply register the fact that she smiled; rather, you interpret the smile's meaning and use that interpretation to infer something about her personality. Depending on the context and any prior information you have about her, you might decide that the smile represents friendliness, or smugness, or guile. What you carry away from the encounter is not so much a memory that the person smiled as a memory that she was friendly, smug, or deceitful. That memory is added to your growing impression of her and may affect your future interactions with her.

Any such judgment about another person is, in essence, a claim about causation. It is an implicit claim that the person's behavior is caused in part by some more or less permanent characteristic of the person, such as friendliness or deceitfulness. In common English usage, any claim about causation is called an *attribution*. In the study of person perception, an **attribution** is a claim about the cause of someone's behavior. As Heider (1958) pointed out, a difficulty in judging someone's personality on the basis of his or her action lies in assessing the degree to which the action truly represents something unique and lasting about the person as opposed to being a normal human response to a particular situation or set of circumstances.

The Logic of Attributing Behavior to the Person or the Situation

If you see a man running and screaming and then see that a tiger is chasing him, you might logically attribute his fear to the situation rather than to any special aspect of his personality; almost anyone would be afraid of a loose and charging tiger. To build a useful picture of a person on the basis of his or her actions, you must decide which actions imply something unique about the person and which actions would be expected of anyone under similar circumstances. Heider noted that when behavior is clearly appropriate to

Is it the person or the situation?

When we see someone behaving in a certain way, or expressing a particular emotion, we may attribute the behavior or emotion primarily to something unique about the person or to something about the situation. Here, we would need more information before making a confident judgment.



© Brooke Fasani / Corbis

the environmental situation, people commonly attribute the behavior to the situation rather than to the behaving person's personality.

In line with Heider's general ideas about attributions, Harold Kelley (1967, 1973) developed a logical model for judging whether a particular action should be attributed to some characteristic of the acting person or to something about the immediate environment. The essence of the model is that we consider three questions in making an attribution:

1. *Does this person regularly behave this way in this situation?* If the answer is yes, we have grounds for attributing the behavior to some stable characteristic of either the person or the situation. If the answer is no, then this particular instance of the behavior may be a fluke that tells us little about either the person or the situation.
2. *Do many other people regularly behave this way in this situation?* If the answer is yes, we have grounds for attributing the behavior more to the situation than to the person. If the answer is no, then this behavior may tell us something unique about the person.
3. *Does this person behave this way in many other situations?* If the answer is yes, we have grounds for making a relatively general claim about the personality of the observed person. If the answer is no, then any personality claim we make about the person is limited to the particular situation.

As an illustration, imagine that we are caught in a traffic jam and Susan, our driver, is expressing a great deal of anger. Does her anger tell us something useful about her as a person? If we have observed that Susan regularly gets very angry in traffic jams (yes to Question 1) and that most other people don't get so angry in traffic jams (no to Question 2), then we might appropriately attribute her anger to her personality. Given these answers to Questions 1 and 2, our answer to Question 3 will allow us to assess the generality of the personality attribute we can reasonably infer. If Susan also gets very angry in many other situations (yes to Question 3), we might logically conclude that she is an easily angered person and we should be careful around her in all situations. Conversely, if Susan rarely gets angry in other situations (no to Question 3), a reasonable conclusion would be that she is not generally an angry person but just cannot tolerate traffic jams.

Notice that there is nothing surprising in this model. It is simply a statement of the logic that you or I or anyone else—with sufficient motivation and information—would use in deciding whether or not an observed bit of behavior tells us something interesting about the person. It states explicitly the logic that leads us to conclude that a man's repeated fearful reaction to puppies and housecats tells us more about the man than does his fearful reaction to a loose and raging tiger.

Not surprisingly, a number of research studies have shown that when people are asked to explain the cause of a particular behavior and are given sufficient information to answer the three questions, they usually do make attributions that accord with the model just described (McArthur, 1972). But often people lack the information, the time, or the motivation to make a logical attribution. In that case they may take shortcuts in their reasoning, which may result in certain consistent errors, or biases.

The Person Bias in Attributions

In his original writings about attribution, Heider (1958) noted that people tend to give too much weight to personality and not enough to the environmental situation when they make attributions about others' actions. In our example of Susan and the traffic jam, people tend to ignore the traffic jam as cause and to attribute Susan's anger too heavily to her personality. Many researchers have confirmed the existence of this *person bias* in attribution.

Some of the most dramatic examples of the person bias occur in situations in which a person is socially pressured or required to behave in a certain way. In one

2
According to Kelley's logical model, when should behavior be attributed to the person and when should it be attributed to the situation?

3
What evidence supports the existence of a person bias in attributions?



Matt Wendelsohn / Corbis

A victim of bias? Leonard Nimoy called his autobiography *I Am Not Spock*. He has apparently often encountered the fundamental attribution error.

4

Why is the person bias often called the “fundamental attribution error”? In what conditions does the bias most often occur?

experiment, for example, college students listened to a student who they were told was assigned to read a political statement written by someone else (Gilbert & Jones, 1986). Even when the assignment was made by the observers themselves, so they could be sure that the reader had not chosen it himself, observers tended to rate the reader as politically liberal when the statement he read was liberal and as politically conservative when the statement was conservative. Although there was no logical reason to assume that the statement had anything to do with the reader's own political beliefs, the students made that attribution.

Other research has shown that a person's social role can have undue effects on the attributions that others make about that person. When we observe a police officer, nurse, teacher, or student carrying out his or her duties, we tend—in accord with the person bias—to attribute the action to the individual's personality and to ignore the constraints that the role places on how the person can or must act. We might develop quite different impressions of the same person if we saw him or her in out-of-role situations. In one experiment demonstrating this effect of roles, Ronald Humphrey (1985) set up a simulated corporate office and randomly assigned some volunteer subjects to the role of manager and others to that of clerk. The managers were given interesting tasks and responsibilities, and the clerks were given routine, boring tasks. At the end of the study, the subjects rated various aspects of the personalities of all subjects, including themselves. Compared with those in the clerk role, those in the manager role were judged by others more positively; they were rated higher in leadership, intelligence, assertiveness, supportiveness, and likelihood of future success. In keeping with the person bias, the subjects apparently ignored the fact that the role assignment, which they knew was random, had allowed one group to manifest characteristics that the other group could not. The bias did not hold when the subjects rated themselves, but it did hold when they rated others who had been assigned to the same role as themselves.

<<< By the mid-1970s so much evidence appeared to support the person bias that Lee Ross (1977) called it the *fundamental attribution error*, a label designed to signify the pervasiveness and strength of the bias and to suggest that it underlies many other social-psychological phenomena. That label is still in use despite growing evidence that the bias may not be as fundamental as Ross and others thought. People are much more likely to make this error if their minds are occupied by other tasks or if they are tired than if they devote their full attention to the task (Gilbert, 1989). People are also more likely to make the error if they make their judgments very soon after observing the action than if a longer period of time elapses between the action and the judgment (Burger, 1991; Truchot & others, 2003). Also, in some cases, the apparent demands of the experiment may artificially produce the person bias. Research subjects who are told that their task is to judge someone's personality are much more likely to exhibit the person bias than are those who are asked to explain the observed behavior in whatever terms they wish (Malle & others, 2000).

The Actor–Observer Discrepancy in Attributions

The person bias breaks down, or is even reversed, when people make attributions about their own behavior. This difference between the kinds of attributions people make about their own and others' behavior is referred to as the **actor–observer discrepancy**. The person who performs an action (the *actor*) commonly attributes the action to the situation—“I am whistling because it is a beautiful day,” or “I read those political statements because I was asked to read them.” In contrast, another person (the *observer*) who sees the same action is likely to attribute it to the actor's internal characteristics—“She is whistling because she is a cheerful person” or “He read those statements because he is politically liberal.”

5

What are two hypotheses as to why people commonly attribute their own behavior to the situation and others' behavior to personality? What evidence supports each hypothesis?

<<< One possible explanation of the actor–observer discrepancy is that it has to do with the degree to which we know ourselves compared to others. We know that our own behavior changes a good deal from situation to situation, but we do not have as much evidence that the same is true for other people. You may assume that your psychology professor's calm demeanor in the classroom is indicative of his or



Getty Images / Brand X

The actor–observer discrepancy When we as observers see this scene (left), we are likely to focus on the woman's face, which predisposes us to explain her fear in terms of her personality. She, however, does not see her own face, but sees a drop-off (right), so she is more likely to explain her fear in terms of the situation.



Photodisc Red / Getty Images

her behavior everywhere and thus attribute it to personality, but this may be only because you haven't seen your professor at home, in traffic court, or on the softball diamond. Consistent with this *knowledge-across-situations hypothesis*, people usually judge the behavior of their close friends as more flexible—more determined by the situation and less by unvarying personality traits—than the behavior of people whom they know less well (Prentice, 1990; Sande & others, 1988).

Another hypothesis holds that the actor–observer discrepancy stems from a basic characteristic of visual perception: Our eyes point outward, away from ourselves. When we watch someone *else* perform an action, our eyes focus on the actor, not the environment to which the actor is responding, so we tend to see the act as caused by the person rather than the situation. But when we ourselves perform an action, we see the surrounding environment, not ourselves, so we tend to attribute causal properties to the situation. Consistent with this *visual-orientation hypothesis*, one researcher found that the actor–observer discrepancy was reversed when the actor and the observer watched videotaped replays of the action from reversed visual orientations (Storms, 1973). When people watched themselves on videotape, they attributed relatively more of their behavior to their own traits and less to the situation. When people watched a videotape of another person's performance from that person's perspective—so that they saw the environment as that person would see it—they attributed relatively more of the behavior to the situation and less to the person.

A Cross-Cultural Difference in Attributions

Prior to the 1980s, social-psychological studies of attributions had been conducted only in Western cultures, mostly in North America and Western Europe. This observation led to the suggestion that the person bias in attributions may be a product of a predominantly Western way of thinking. Western philosophies, religions, and political ideologies tend to emphasize the idea that people are in charge of their own destinies, so people growing up in Western cultures may learn to attribute behavior more to the person than to the situation (Jellison & Green, 1981). If so, then in Eastern cultures—such as those of India, China, and Japan, where philosophies and religions emphasize the role of fate or circumstances in controlling one's destiny—people might make relatively fewer person attributions and more situation attributions.

6 What logic and evidence suggest that the person bias may be a product of Western culture and may not exist in Eastern cultures?

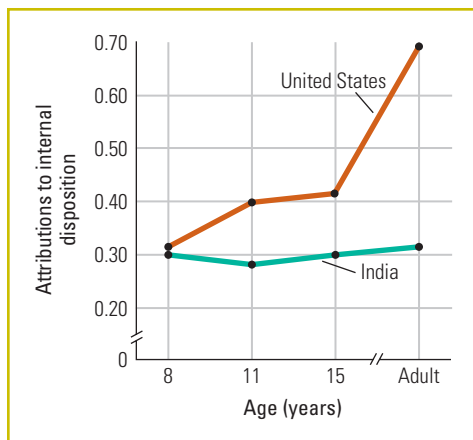


FIGURE 13.1 Cultural difference in making attributions When asked to explain another person's behavior, the proportion of attributions to internal disposition (personality or attitude) was greater among people in the United States than it was among Hindus in India, and this difference was greater for adults than for children. (The proportions were determined by dividing the number of person attributions by the total number of person plus situation attributions for each group. Data from Miller, 1984.)

To test this theory, Joan Miller (1984) asked middle-class children and adults in the United States and in a Hindu community in India to think of an action by someone they knew and then to explain why the person had acted in that way. As predicted, the Americans made more attributions to personality and fewer to the situation than did the Indians. This difference was greater for adults—who would presumably have incorporated the cultural norms more strongly—than it was for children (see Figure 13.1).

In the years since Miller's pioneering work, similar results have been found in dozens of studies comparing attributions made by people raised in North America with those raised in various Far Eastern countries, including China, Japan, and Korea (Lehman & others, 2004; Norenzayan & Nisbett, 2000). The difference has been observed not just in the attributions made by research subjects in experiments, but also in attributions found in the literature and journalism of the different cultures. In one study, for example, Michael Morris and Kaiping Peng (1994) analyzed the content of every article published in the *New York Times* and in the *World Journal*, a Chinese-language newspaper published in New York City, concerning two mass murders that took place in 1991. The researchers found that the articles in the *Times* focused most heavily on personality characteristics of the murderers—their traits, attitudes, character flaws, mental disorders, and so on. In contrast, the articles in the Chinese newspaper focused most heavily on the life situations of the murderers—their living conditions, social relationships, and frustrations that might have provoked their actions.

Effects of Prior Information and Physical Appearance on Person Perceptions

Thus far we have been considering how people judge the characteristics of another individual on the basis of their impressions of the causes of that individual's behavior. But judgments relating behavior to character go in both directions. People not only assess character from behavior; they also interpret behavior in terms of what they already know—or think they know—about the person's character. A person's physical features, or information that we have heard about a person, can influence our interpretations of the person's actions in ways that can be quite unfair.

Using a Preexisting Schema to Interpret a Person's Actions

The organized set of information or beliefs that we have about any entity or event is referred to by cognitive psychologists as a *schema* (discussed in Chapter 9). Thus, to know a person, even slightly, accurately or inaccurately, is to have a mental schema for that person. Even before you meet a person, you may already have a schema of him or her that is based on what you have heard and supplemented by your imagination. That schema can influence your interpretation of the person's behavior and can bias the attributions you make.

7 How did an experiment at MIT illustrate the biasing effect of a preexisting schema on perception of a person? Why do first impressions often resist change?

<<< A classic experiment demonstrating this biasing effect involved perception of a guest lecturer by students in a course at MIT (Kelley, 1950). Before the guest arrived, half the students received a written biographical sketch of the guest that included the statement, "People who know him consider him to be a very warm person, industrious, critical, practical, and determined." The other half received the same sketch except that the words "very warm" were replaced by "rather cold." After the guest had appeared and led a 20-minute discussion, the students were asked to fill out a form evaluating his performance. The main results were that the students who had received the sketch containing the words "very warm" took greater part in the discussion and rated the guest and his performance more positively than did the students who had received the sketch containing the words "rather cold." Thus, the two groups responded differently to the same lecturer and made different attributions about his performance depending on the initial schema set up for them.

First impressions are often hard to change, because people tend to interpret new information so as to make it consistent with their already-formed impressions (Asch, 1946; Park, 1986). In the last U.S. presidential election, your a priori beliefs about each candidate may have colored your interpretation of everything that each candidate said or did, in ways that tended to confirm your initial beliefs. The same type of effect occurs to some degree in our daily interactions with other people. Our smile looks friendly to those who think we are kind, smug to those who think we are aloof, and deceitful to those who think we are untrustworthy.

The Attractiveness Bias

Often the first impression we gain of another person derives purely from the person's physical features, and, like any other first impression, it can bias our subsequent judgments. Consistent with folktales in which the good people (the princesses and princes) are beautiful and the bad people (the witches and ogres) are ugly, experiments have shown that physically attractive people are commonly judged as more intelligent, competent, sociable, and moral than less attractive people (Dion, 2002; Langlois & others, 2000).

In one experiment, fifth-grade teachers were given report cards and photographs of children whom they did not know and were asked to rate each child's intelligence and achievement. The teachers rated physically attractive children as brighter and more successful than unattractive children with identical report cards (Clifford & Walster, 1973). In a similar experiment, adults more frequently attributed a child's misbehavior to environmental circumstances if the child was physically attractive and to the child's personality if the child was not attractive (Dion, 1972). In yet another study, which analyzed actual court cases, judges regularly gave longer prison sentences to unattractive persons than to attractive persons convicted of comparable crimes (Stewart, 1985).

Maxims such as "Don't judge a book by its cover" and "Beauty is only skin deep" have arisen for a reason. We do automatically tend to judge one another by our covers and to assume that beauty on the skin extends to the core of one's character, so we need reminders to counter such unfair tendencies. On the brighter side, there is evidence, too, that judgments of personality can affect judgments of physical appearance (Zebrowitz & Rhodes, 2002). In one experiment, an instructor who behaved in a warm and friendly way to one group of students and in a cold and aloof way to another group was judged as better looking by the first group than by the second (Nisbett & Wilson, 1977). There is also evidence that East Asians are less susceptible to the attractiveness bias than are Westerners (Dion, 2002). Just as they are less inclined than Westerners to judge a person's character from a brief glimpse of the person's behavior, they are also less inclined to judge it from the person's physical attractiveness. Apparently, the attractiveness bias is at least partly a result of the influence of Western culture.

The Baby-Face Bias

Another pervasive, although less well known, bias concerns a person's facial maturity. Some people, regardless of their age, have facial features that resemble those of a baby: a round rather than elongated head, a forehead protruding forward rather than sloping back, large eyes, and a small jawbone (see Figure 13.2). In a series of experiments conducted in both the United States and Korea, baby-faced adults were perceived as more naïve, honest, helpless, kind, and warm than mature-faced adults of the same age and sex, even though the perceivers could tell that the baby-faced persons were not really younger (McArthur & Berry, 1987; Zebrowitz & others, 1993).

>>> **8**
How have researchers documented biasing effects of (a) physical attractiveness and (b) facial maturity on perceptions of personality?

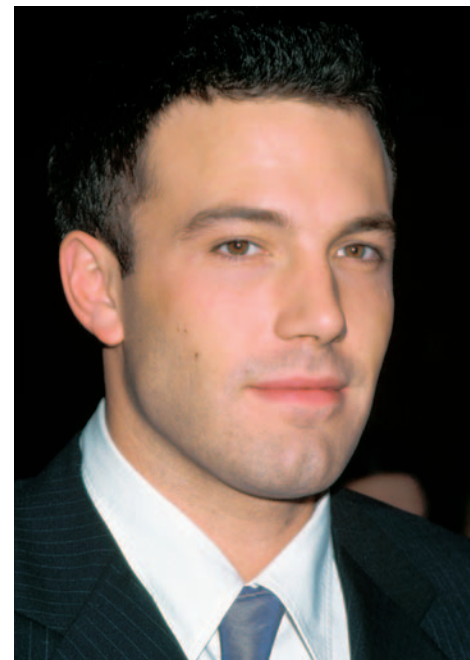


© Catherine Ledner / Getty

Luckily for him, he's cute

Researchers have found that good-looking children are less likely to be blamed for their misbehavior than are children who are not so good looking. The teacher is horrified by this boy's action, but, on reflection, she may excuse it as something caused by the situation—perhaps by the merriment of his classmates.

FIGURE 13.2 Who would deceive you? Adults whose faces are babyish (left) are commonly seen as more naïve, honest, helpless, kind, and warm than are mature-faced adults (right). The characteristics of a baby face include a round head, large forehead, large eyes, short nose, and small chin.



In one study, Leslie Zebrowitz and Susan McDonald (1991) found that the baby-face bias influenced the outcomes of actual cases in a small-claims court. Baby-faced defendants were much more frequently found innocent in cases involving intentional wrongdoing than were mature-faced defendants, but they were neither more nor less frequently found innocent in cases involving negligence (such as performing a contracted job incompetently). Apparently, judges find it hard to think of baby-faced persons as deliberately causing harm but do not find it difficult to think of them as incompetent or forgetful.

9 How might sex differences in facial features contribute to perceived psychological differences between men and women?

<<< Zebrowitz (1996) has also found evidence that differences in facial maturity between men and women may contribute to differences in how the two sexes are perceived. Women, on average, are more baby-faced than men, and women are also, on average, judged as kinder, more naïve, more emotional, and less socially dominant than men. In an experiment, Zebrowitz and her colleagues presented college students with schematic drawings of men's and women's faces in which facial maturity was varied by altering the size of the eyes and the length of the jaw (Friedman & Zebrowitz, 1992). When the typical differences between men's and women's faces were present, students judged the men as more dominant and less warm than the women. But when the faces were equivalent on the maturity dimensions, the students judged the two as equal in dominance and warmth. Zebrowitz did not suggest that facial features are the sole determinant of the different perceptions people have of women and men under more natural conditions, but she did suggest that such features may contribute to that difference.

Konrad Lorenz (1943, 1971) suggested long ago that human beings instinctively respond to infants' facial features with feelings of compassion and care, a characteristic that helps promote the survival of our offspring. He also noted that the infant-like features of some animal species (such as rabbits and pandas) lead us to perceive them as particularly cute, innocent, and needing care, regardless of the animals' actual needs or behaviors. The work of Zebrowitz and others suggests that we generalize this response not just to babies and animals but also to adult humans whose faces resemble those of babies. For another example, see Figure 13.3.

From an evolutionary perspective, it is noteworthy that human adults of both sexes are much more baby-faced than the adults of our closest primate relatives. The typical adult human face looks much more like that of an infant chimpanzee than like that of an adult chimpanzee. This difference is generally attributed to the



FIGURE 13.3 | The evolution of innocence in Mickey Mouse Mickey Mouse began life, in the 1928 cartoon *Steamboat Willie*, as a mischievous rodent who delighted in cranking a goat's tail and playing bagpipes on a pig's nipples. In response to social pressure from citizens concerned about Mickey's moral effects on children, the Walt Disney Company made him look and act increasingly innocent over the years. As part of this change, Disney artists made him look more innocent and cute by augmenting his juvenile features. Over a 50-year period Mickey's eye size increased markedly, and the apparent vault of his forehead increased as a result of the migration of his front ear toward the back of his head. (Based on S. J. Gould, 1980, pp. 96–97.)

expanded cranial cavity that came with enlargement of the brain in humans. But I wonder if that is the whole explanation. In the course of human evolution, individuals who had babyish faces may have been treated more benignly than those who had more mature faces, and perhaps this helped promote our species' evolution toward baby-facedness. This is speculation, but it is supported by evidence that baby-faced children and adolescents are less often physically abused than are those who have more mature faces (McCabe, 1984) and by evidence that baby-faced adults receive more unsolicited help from strangers than do mature-faced adults (Keating & others, 2003). As further speculation, perhaps the protective effect was of greater value for girls and women than for boys and men—for reasons having to do with the general sex differences in strength and aggressiveness—and perhaps this difference led to the sex difference in baby-facedness observed today.

Forming Impressions on the Internet

The Internet, like the telephone before it and postal services before that, has added new dimensions to human communication. More than any other communication tool, the Internet allows people to locate and “meet” other people who have similar interests—through chat rooms, news groups, special-interest e-mail lists, dating services, and the like. Surveys indicate that many close friendships and valued acquaintanceships today involve people who first met in these ways (Bargh & McKenna, 2004). Some of these relationships remain confined to cyberspace, but in other cases Internet friends eventually meet in person and the friendship extends into the tangible world.

In several experiments, randomly formed pairs of opposite-sex college students who did not know each other were assigned to participate in get-acquainted meetings, either on-line (through an Internet chat room) or in face-to-face encounters. The most striking general result is that those who had met on-line reported more liking of each other than did those who had met face-to-face (Bargh & McKenna, 2004). In one such experiment, this first meeting was followed by a second meeting, in which all pairs met face-to-face (McKenna & others, 2002). As you can see in Figure 13.4, the degree of liking between those who first met on the Internet increased even more in the face-to-face meeting, while the lesser degree of liking between those who first met face-to-face was not affected by the second meeting.

How can such results be explained? Researchers have found that get-acquainted meetings over the Internet are more intimate, more revealing of what each person considers to be his or her “true self,” than are such meetings conducted face-to-face (Bargh & McKenna, 2004; Bargh & others, 2002). Apparently, the relative anonymity of the Internet, along with the lack of visual and auditory contact, reduces social anxiety and frees people to reveal more about themselves than they would if they met face-to-face. Also, without knowledge of the physical features of the other person, the biasing effects of attractiveness, or lack thereof, are

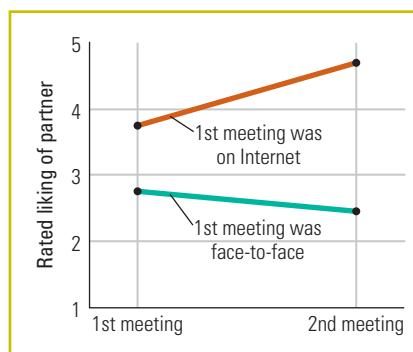


FIGURE 13.4 | Getting acquainted on the Internet Randomly selected opposite-sex pairs of students met for 20 minutes either on the Internet or face-to-face, and then all pairs met face-to-face for a second 20-minute period. After each meeting, subjects anonymously rated their liking of their partners using a scale in which 7 was maximal liking. (Data from McKenna & others, 2002, Study 3.)

10 What evidence suggests that strangers who meet on the Internet like each other more than do strangers who meet in person? How might this phenomenon be explained?

absent. Communication is not shut down by early negative judgments or anxieties based on physical features. When and if the two partners do meet, they already know a good deal about each other and may feel something of an emotional bond, which may lead them to see each other as more attractive than they would have if they were complete strangers.

Used wisely, the Internet apparently is a valuable tool for making friends. But, as is often pointed out, the Internet's seductive nature also creates potential dangers. People can easily create false impressions of themselves over the Internet, and early face-to-face meetings with Internet acquaintances should be held only in safe, public places.

SECTION REVIEW

We are constantly forming impressions of others and judging the causes of their behavior.		
Basing Attributions on Observed Behavior - Logically, we might attribute a person's behavior primarily to characteristics of the person or the situation. - The person bias is the tendency to give undue weight to personality and not enough to the situation in making attributions. - The person bias may disappear or even be reversed when we make attributions about our own behavior, a phenomenon known as the actor–observer discrepancy. - The person bias appears not to hold true for Asian cultures.	Effects of Prior Information and Physical Appearance - We use schemas (organized sets of beliefs) about people to interpret their actions, even when the information in our schema is inaccurate or irrelevant. - A common bias is that we tend to see physically attractive people as more intelligent, social, competent, and moral than less attractive people. - We also tend to see baby-faced individuals as more honest, naïve, helpless, warm, and kind than otherwise comparable people with mature faces. This may partly account for our differing perceptions of men's and women's personalities.	Forming Impressions on the Internet - In experiments, people who met initially on the Internet liked each other more than people who initially met face-to-face. - This tendency may result from people on the Internet being less anxious, more intimate, and freed from the biasing effects of physical features.

Perceiving and Evaluating the Self

11 How does the rouge test of self-recognition work? What animals besides humans have passed this test?

<<< Self-awareness is often described as one of the hallmarks of our species. At about 18 months of age, human infants stop reacting to their image in a mirror as if it were another child and begin to treat it as a reflection of themselves (Lewis & Brooks-Gunn, 1979; Lewis & Ramsay, 2004). If a researcher surreptitiously places a bright red spot of rouge on the child's nose or cheek before placing the child in front of the mirror, the 18-month-old responds by touching his or her *own* nose or cheek to feel or rub off the rouge; a younger child, by contrast, touches the mirror or tries to look behind it to find the child with the red spot. The only other animals besides ourselves who have passed the rouge test of self-recognition are the other apes—chimpanzees, bonobos, orangutans, and at least one gorilla and one gibbon (Parker & others, 1994; Ujhelyi & others, 2000). There is also some evidence, based on other tests, that dolphins may be capable of self-recognition (Sarko & others, 2002). Other animals, including all species of monkeys tested so far, continue to treat

Dr. Michael Lewis, Institute for the Study of Child Development, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, East Brunswick, NJ.



It's me By pointing to the spot of rouge on her own cheek, this girl demonstrates her understanding that the mirror image is indeed of her. This photo is from the original study of mirror self-recognition by Lewis and Brooks-Gunn (1979).

the mirror image as another animal—a creature to threaten and try to chase away—no matter what their age or how much experience they have had with mirrors.

Research with chimpanzees suggests that for them social interaction is crucial for self-recognition. Chimpanzees raised in isolation from others of their kind did not learn to make self-directed responses to their mirror images, whereas those raised with other chimps did (Gallup & others, 1971). Many psychologists and sociologists have argued that the self-concept, for humans as well as chimps, is fundamentally a *social* product. To become aware of yourself, you must first become aware of others of your species and then become aware, perhaps from the way others treat you, that you are one of them. In humans, self-awareness includes awareness not just of the physical self, reflected in mirror images, but also of one's own personality and character, reflected psychologically in the reactions of other people.

Seeing Ourselves Through the Eyes of Others

Many years ago the sociologist Charles Cooley (1902/1964) coined the term *looking-glass self* to describe what he considered to be a very large aspect of each person's self-concept. The "looking glass" to which he referred is not an actual mirror but other people who react to us. He suggested that we all naturally infer and imagine what others think of us from their reactions, and we use those inferences and images to build our own self-concepts. Cooley's basic idea has been supported by much research showing that people's opinions and attitudes about themselves are very much affected by the opinions and attitudes of others.

According to Cooley, what is the "looking glass" with which we evaluate ourselves?

Effects of Others' Appraisals on Self-Understanding and Behavior

The beliefs and expectations that others have of a person—whether they are initially true or false—can to some degree create reality by influencing that person's self-concept and behavior. Such effects are called *self-fulfilling prophecies* or *Pygmalion effects*.

What are Pygmalion effects in psychology, and how were such effects demonstrated in elementary school classrooms?

Pygmalion was the mythical Roman sculptor, in Ovid's story, who created a statue of his ideal woman and then brought her to life by praying to Venus, the goddess of love. More relevant to the point being made here, however, is George Bernard Shaw's revision of the myth, in his play *Pygmalion* (upon which the musical *My Fair Lady* was based). In the play, an impoverished Cockney flower seller, Eliza Doolittle, becomes a fine lady largely because of the expectations of others. Professor Higgins assumes that she is capable of talking and acting like a fine lady, and Colonel Pickering assumes that she is truly noble at heart. Their combined actions toward Eliza lead her to respond in ways that transform their assumptions into realities. Psychological research indicates that such effects are not confined to fiction.

Pygmalion in the Classroom

In a classic experiment, Robert Rosenthal and Lenore Jacobson (1968) led elementary school teachers to believe that a special test had revealed that certain students would show a spurt in intellectual growth during the next few months. Only the teachers were told of the supposed test results, not the students. In reality, the students labeled as "spurters" had been selected not on the basis of a test score but at random. Yet when all the students were tested 8 months later, the selected students showed significantly greater gains in IQ and academic performance than did their classmates. These were real gains, measured by objective tests, not just perceived gains. Somehow, the teachers' expectations that certain children would show more intellectual development than other children created its own reality.

Subsequent replications of this *Pygmalion in the classroom* effect provided clues concerning its mechanism. Teachers became warmer toward the selected students, gave them more time to answer difficult questions, gave them more challenging work, and noticed and reinforced more often their self-initiated



Warner Bros. / The Kobal Collection

A Pygmalion effect In the movie *My Fair Lady*, based on the play *Pygmalion*, an impoverished Cockney woman begins to behave as an upper-class "fine lady" because she is treated as one.

efforts (Cooper & Good, 1983; Rosenthal, 1994). In short, either consciously or unconsciously, they created a better learning environment for the selected students than for other students. They also, through their treatment of them, changed the selected students' self-concepts. The students began to see themselves as more capable academically than they had before, and this led them to work harder to live up to that perception (Cole, 1991; Jussim, 1991).

14

What evidence supports and delimits the idea that a person's self-concept and behavior are shaped by others' beliefs or claims about the person?



Changing Others' Behavior by Altering Their Self-Concepts In the experiments just described it is impossible to know whether the children's improved performance resulted primarily from changes in their self-concepts or from changes in the learning opportunities that the teachers provided for them. Other experiments, however, have shown that simply telling others that they are a certain kind of person can, in some cases, lead them to behave in ways that are consistent with the attribute that they are told they have.

In one such experiment, some elementary school children were told in the course of classroom activity that they *were* neat and tidy (*attribution condition*); others were told that they *should be* neat and tidy (*persuasion condition*); and still others were given no special treatment (*control condition*). The result was that those in the attribution condition became significantly neater, as measured by the absence of littering, than did those in either of the other conditions (Miller & others, 1975). Similarly, children who were told that they were good at math showed greater subsequent improvements in math scores than did those who were told that they should try to become good at math (see Figure 13.5). In these experiments the change in behavior presumably occurred because of a direct effect of the appraisals on the children's self-concepts, which they then strove to live up to.

Of course, people's self-concepts are not always as moldable as the experiments just cited might suggest. The effects are strongest with young children and with characteristics for which people do not already have firm self-beliefs. Adolescents and adults often respond to such appraisals in ways that seem designed to correct what they perceive to be another person's misperception of them—Pygmalion in reverse (Collins & Zebrowitz, 1995; Swann, 1987).

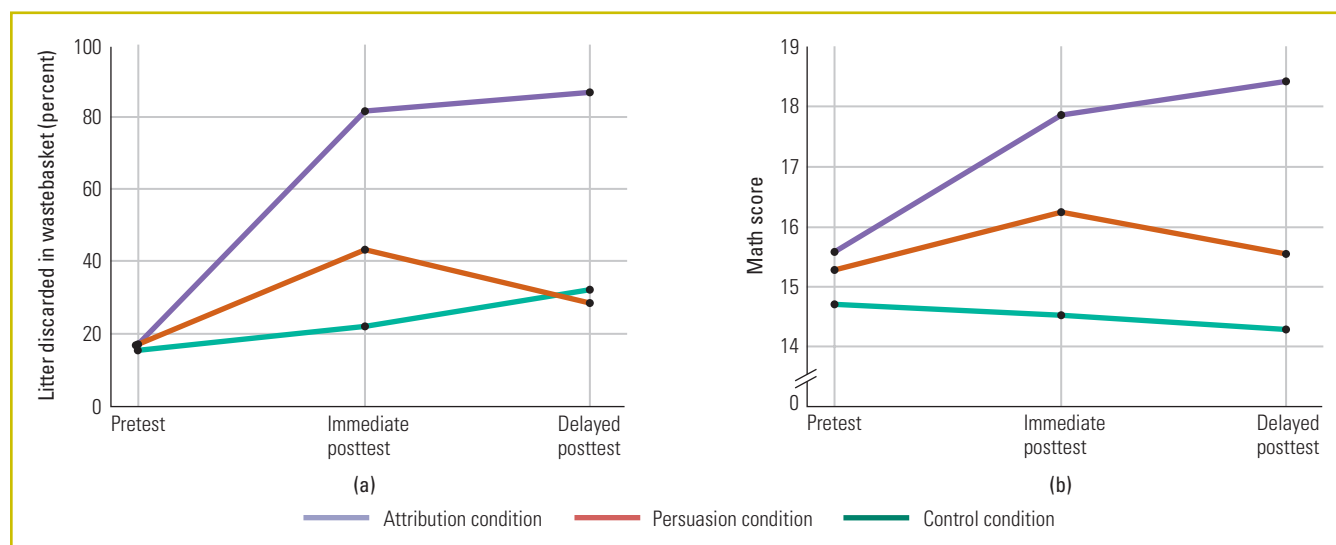
FIGURE 13.5 Effect of attribution compared with persuasion

(a) Fifth graders who were repeatedly told that they were neat and tidy (attribution condition) showed greater gain in use of the wastebasket than did those in the other conditions. (b) Second graders who were repeatedly told that they were good at math (attribution condition) showed greater improvement in math scores than did those in the other conditions. The delayed posttests were conducted several weeks after the pretests. (Adapted from Miller & others, 1975.)

Copyright © 1975 by the American Psychological Association. Adapted by permission.

Self-Esteem as an Index of Others' Approval and Acceptance

Self-esteem, by definition, is one's feeling of approval and acceptance of oneself. It is generally measured with questionnaires in which people rate the degree to which they agree or disagree with such statements as, "On the whole, I am satisfied with myself" and "I feel that I have a number of good qualities" (Rosenberg, 1965; Tafarodi & Milne, 2002).



Self-esteem is experienced as deriving from one's own self-judgment, but, according to a theory proposed by Mark Leary (1999; Leary & Baumeister, 2000), this judgment is based largely on the person's sense of others' attitudes toward him or her. The theory is referred to as the *sociometer theory*, because it proposes that self-esteem acts like a meter to inform a person, at any given time, of the degree to which he or she is likely to be accepted or rejected by others. According to the sociometer theory, what you experience as your self-esteem largely reflects your best guess about the degree to which other people, whom you care about, hold you in esteem. As partial support of the sociometer theory, Leary and his colleagues have cited the following lines of evidence:

- Individual differences in self-esteem correlate strongly with individual differences in the degree to which people believe that they are generally accepted or rejected by others (Leary & others, 2001).
- When people were asked to rate the degree to which particular real or hypothetical occurrences in their lives (such as rescuing a child, winning an award, or failing a course) would raise or lower their self-esteem, and also to rate the degree to which those same occurrences would raise or lower other people's opinions of them, the two sets of ratings were essentially identical (Leary & others, 1995).
- In experiments, and in correlational studies involving real-life experiences, people's self-esteem increased after praise or other satisfying social experiences and decreased after evidence of social rejection (Baumeister & others, 1998; Leary & others, 2001; Tafarodi & Milne, 2002).
- Feedback about success or failure on a test had greater effects on self-esteem if the person was led to believe that others would hear of this success or failure than if the person was led to believe that the feedback was private and confidential (Leary & Baumeister, 2000).

The sociometer theory is designed to offer an evolutionary explanation of the function of self-esteem. In our real-world interactions with other people, other people's views of us matter a great deal. Our survival depends on others' acceptance of us and willingness to cooperate with us. An inflated self-view that is greatly out of sync with how others see us could be harmful. If I see myself as a highly capable and trustworthy person, but nobody else sees me that way, then my own view will seem foolish to others and will not help me in my dealings with them. A major evolutionary purpose of our capacity for self-esteem, according to the sociometer theory, is to motivate us to act in ways that promote our acceptance by others. A decline in self-esteem may lead us to change our ways in order to become more socially acceptable, or it may lead us to seek a more compatible social group that approves of our ways. Conversely, an increase in self-esteem may lead us to continue on our present path, perhaps even more vigorously than before.

Actively Constructing Our Self-Perceptions

Although other people's views of us play a large role in our perceptions of ourselves, we do not just passively accept those views. We actively try to influence others' views of us, and in that way we also influence our own self-perceptions. In addition, we compare ourselves to others as a way of defining and evaluating ourselves, and we often bias those comparisons by giving more weight to some pieces of evidence than to others.

Social Comparison: Effects of the Reference Group

In perception everything is relative to some frame of reference, and in self-perception the frame of reference is other people. To see oneself as short, or conscientious, or good at math, is to see oneself as that *compared with other people*. The process of comparing ourselves with others in order to identify our unique characteristics and evaluate our abilities is called **social comparison**. A direct consequence

15 What is the sociometer theory of self-esteem, and what evidence supports it?

16

What is some evidence that people construct a self-concept by comparing themselves with a reference group? How can a change in reference group alter self-esteem?

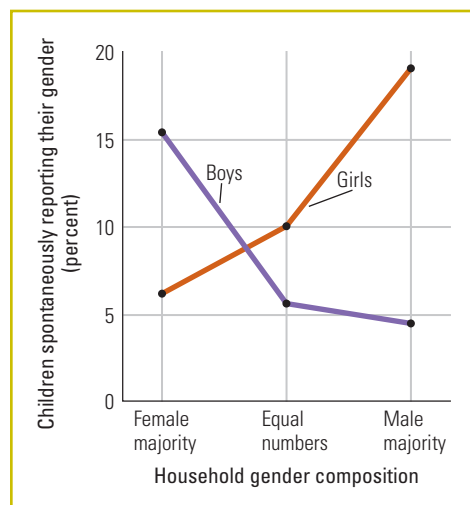


FIGURE 13.6 Evidence that children define themselves in terms of differences from their reference group As shown here, children were more likely to mention their gender when describing themselves if their gender was in the minority in their household than they were if it was in the majority. (Adapted from McGuire & McGuire, 1988.)

of social comparison is that the self-concept varies depending on the **reference group**, the group against whom the comparison is made. If the reference group against which I evaluated my height were made up of professional basketball players, I would see myself as short, but if it were made up of jockeys, I would see myself as very tall.

Effect of the Reference Group on Self-Descriptions In one series of studies that illustrates the role of the reference group, children's self-descriptions were found to focus on traits that most distinguished them from others in their group (McGuire & McGuire, 1988). Thus, children in racially homogeneous classrooms rarely mentioned their race, but those in racially mixed classrooms quite commonly did, especially if their race was in the minority. Children who were unusually tall or short compared with others in their group mentioned their height, and children with opposite-gender siblings mentioned their own gender more frequently than did other children (see Figure 13.6). Such evidence suggests that people identify themselves largely in terms of the ways in which they perceive themselves to be different from those around them.

Effect of the Reference Group on Self-Evaluations The evaluative aspect of social comparison can be charged with emotion. We are pleased with ourselves when we feel that we measure up to the reference group and distressed when we don't. A change of reference group, therefore, can dramatically affect our self-esteem. Many first-year college students who earned high grades in high school feel crushed when their marks are only average or less compared with those of their new, more selective reference group of college classmates. Research conducted in many different countries has shown that academically able students at nonselective schools typically have higher academic self-concepts than do equally able students at very selective schools (Marsh & Hau, 2003), a phenomenon aptly called the *big-fish-in-little-pond effect*. The difference reflects the difference in the students' reference groups.

William James (1890/1950), reflecting on extreme instances of selective comparison, wrote: "So we have the paradox of the man shamed to death because he is only the second pugilist [boxer] or second oarsman in the world. That he is able to beat the whole population of the globe minus one is nothing; he has 'pitted' himself to beat that one and as long as he doesn't do that nothing else counts."

In a follow-up of James's century-old idea, Victoria Medvec and her colleagues (1995) analyzed the televised broadcasts of the 1992 Summer Olympics for the amounts of joy and agony expressed by the silver and bronze medalists after each

The agony of being second best in the world Isabella Ochichi (on the left), the silver medalist in the women's 5000-meter footrace of the 2004 World Olympics, seems much less happy about her medal than does bronze medalist Tirunesh Dibaba (on the right). A possible explanation, in terms of the reference group in social comparison, is offered in the text.



© Mike Blake / Reuters / Corbis

event. The main finding was that the silver medalists (the second-place finishers) showed less joy and more agony than did the bronze medalists (the third-place finishers), whom they had defeated. This seemingly paradoxical finding makes sense if we assume that the groups were implicitly making different comparisons. The silver medalists had almost come in first, so the prominent comparison to them—after the contest if not before—was likely that of themselves to the gold medalists, and in that comparison they were losers. In contrast, the bronze medalists had barely made it into the group that received a medal at all, so the prominent comparison in their minds was likely that of themselves to the non-medalists, and in that comparison they were winners.

James (1890/1950) wrote that a person's self-esteem is equal to his achievements divided by his pretensions. By pretensions, he was referring to one's self-chosen goals and reference groups. This formula nicely explains why high achievers typically do not have markedly higher levels of self-esteem than do people who achieve much less (Crocker & Wolfe, 2001). With greater achievements come greater aspirations and ever more elite reference groups for social comparison.

Enhancing Our Views of Ourselves

The radio humorist Garrison Keillor describes his mythical town of Lake Wobegon as a place where “all the children are above average.” We smile at this statistical impossibility partly because we recognize the same bias in ourselves. Repeated surveys have found that most college students rate themselves as better students than the average college student, and in one survey 94 percent of college instructors rated themselves as better teachers than the average college instructor (Alicke & others, 1995; Cross, 1977). Indeed, at least in North America and Western Europe, people tend to rate themselves unduly high on practically every dimension that they value (Mezulis & others, 2004). It is useful to have relatively accurate views of ourselves, but it feels good to think well of ourselves, so most of us skew our self-evaluations in positive directions. We maintain our unduly high self-evaluations by treating evidence about ourselves differently from the way we treat evidence about others.

Attributing Our Successes to Ourselves, Our Failures to Something Else One way that we maintain a high view of ourselves is to systematically skew the attributions we make about our successes and failures. Earlier in this chapter you read of the actor–observer effect in attributions, the tendency for people to attribute their own actions to the situation and others' actions to inner qualities of the person. That effect is best assessed with actions that are neutral on the dimension of success or failure. When success or failure enters the picture, another bias comes into play, the **self-serving attributional bias**—the tendency of people to attribute their successes to their own inner qualities and their failures to the situation. This bias has been demonstrated in countless experiments, with a wide variety of different kinds of tasks (Mezulis & others, 2004).

In one demonstration of this bias, students who performed well on an examination attributed their high grades to their own ability and hard work, whereas those who performed poorly attributed their low grades to bad luck, the unfairness of the test, or other factors beyond their control (Bernstein & others, 1979). In another study, essentially the same result was found for college professors who were asked to explain why a paper they had submitted to a scholarly journal had been either accepted or rejected (Wiley & others, 1979). My favorite examples of the self-serving attributional bias come from people's formal reports of automobile accidents, such as the following (quoted by Greenwald, 1980): “The telephone pole was approaching. I was attempting to swerve out of its way when it struck my front end.” Clearly, a reckless telephone pole caused this accident; the skillful driver just couldn't get out of its way in time.



William Hamilton © 1991 from The New Yorker Collection. All rights reserved.

“Of course you’re going to be depressed if you keep comparing yourself with successful people.”

>>> 17

What are three means by which people build and maintain inflated views of themselves?



The self-serving attributional bias

When people make a serious mistake, they tend to blame the situation—or someone else—rather than attribute the error to their own misjudgment.

Accepting Praise at Face Value Most of us hear many more positive statements about ourselves than negative statements. Norms of politeness as well as considerations of self-interest encourage people to praise each other and inhibit even constructive criticism. “If you can’t say something nice, say nothing at all” is one of our mores. Since we build our self-concepts at least partly from others’ appraisals, we are likely to construct positively biased self-concepts to the degree that we believe the praise we hear from others. A number of experiments have shown that most people tend to discount praise directed toward others as insincere flattery or ingratiation, but accept the same type of praise directed toward themselves as honest reporting (Vonk, 2002).

Defining Our Own Unique Criteria for Success In many realms of activity the criteria for success are vague, and people may evaluate themselves by those measures that are most favorable to themselves (Suls & others, 2002). The majority of people may truly be “above average” if the criteria are allowed to vary from person to person in accordance with each person’s unique view of the task. One student considers himself an above-average scholar because he plays such a constructive role in class discussions, another because she relates what she learns in class to practical problems in her life, another because he gets high scores on tests, and yet another because she disdains “grubbing for grades” and spends time reading books that are not assigned. One instructor sees herself as better than

average because she treats her students as individuals, another because he explains the subject matter clearly, another because she has high standards and is the toughest grader in her department, and yet another because he has never failed anyone.

Consistent with this interpretation of people’s inflated self-ratings, researchers have found such inflation to be stronger for endeavors in which the criteria for success are not well defined than for those (such as running speed in the 100-meter dash) in which the criteria are more uniform and objectively measurable (Dunning & others, 1989). It is hard to know to what degree people define their criteria differently in order to view themselves in a better light, or to what extent their differing goals and criteria stem from truly differing conceptions of the task. Although psychologists have tended to focus mostly on the former possibility, it seems likely that both factors are involved.

East-West Differences in Self-Perceptions

As noted earlier in this chapter, the psychology of North Americans and Western Europeans is not necessarily the psychology of people everywhere. This appears to be particularly true in relation to people’s perceptions of themselves. Much research indicates that people in East Asian countries—including Japan, Korea, and China—conceive of the self quite differently from the way people do in the West (Lehman & others, 2004).

Greater Focus in the East on Relationships and Roles, Less on the Independent Self

18

How does Triandis characterize individualist and collectivist cultures? What differences have been found between the two in people’s self-descriptions?

<<< Harry Triandis (1995), one of the pioneers of cross-cultural research in psychology, has drawn a rough distinction between two categories of cultures: individualist and collectivist. *Individualist cultures* predominate in Western Europe, North America, and Australia, where philosophical and political traditions emphasize personal freedom, self-determination, and individual competition. *Collectivist cultures* exist in much of Africa and Latin America, but have been most fully studied in East Asia. While individualist cultures emphasize the *independence* of each person, collectivist cultures emphasize the *interdependence* of people with others in their families, communities, and social groups.



© James Marshall / Corbis

Respect your elders As part of their general emphasis on interdependence and proper social relationships, East Asian cultures encourage respectful behavior toward elders that is quite unlike that shown in the West. Here two children in Bangkok greet an adult acquaintance.

Consistent with Triandis's view are numerous studies indicating that East Asians describe themselves differently from the way North Americans do (Kanagawa & others, 2001; Markus & Kitayama, 1991, 1994). The former describe themselves more often in terms of their social groups and roles and less often in terms of consistent personality traits that cut across their groups and roles. Asked to describe themselves, they are more likely to make such statements as "I am a student at University X" or "I am the eldest daughter in my family" and less likely to make such statements as "I am easygoing" or "I am ambitious." When they do describe themselves with personality traits, they typically delimit the traits to particular social contexts—"I am easygoing with my friends," "I am ambitious at work" (Cousins, 1989). They are more likely than are Westerners to see themselves as different in different social contexts (Kanagawa & others, 2001; Tafarodi & others, 2004). They are also more likely to attribute their achievements to their families or other social groups than to themselves as individuals (Chen & others, 1998). Such differences in self-descriptions are found even in young children (Wang, 2004).

Less Self-Enhancement in the East, or Different Criteria for Self-Evaluation

Particularly in Japan, but also in other East Asian countries, the ideal person is a modest person. The ideal person is not someone who thinks highly of himself or herself, but someone who is aware of his or her deficiencies and is working hard to overcome them (Heine & others, 1999). Consistent with this observation are research findings indicating that self-enhancing biases are generally small or nonexistent in East Asia and may even be reversed in Japan (Lehman & others, 2004; Mezulis & others, 2004).

In one classic study, for example, Michael Bond and Tak-Sing Cheung (1983) asked university students in the United States, Hong Kong, and Japan to describe themselves in an open-ended way by completing 20 statements, each of which began with the words *I am*. They then analyzed the statements for evaluative content and found that the ratio of positive to negative self-statements was nearly 2 to 1 for the American students, 1 to 1 for the Hong Kong students, and 1 to 2 for the Japanese students. In other words, the Japanese students showed a *self-effacing bias* that was as strong as the Americans' self-enhancing bias.

Such findings are apparently not just the result of Easterners' attempts to *appear* modest to others. In most studies, the questionnaires are filled out anonymously in such a way that it is clear that nobody, including the researchers, can know whose

>>> 19 What evidence suggests that East Asians show less self-enhancement than Westerners do? What evidence suggests that East Asians may self-enhance on characteristics that are most valued in collectivist cultures?

questionnaire is whose. Moreover, similar results are found in tests of implicit modesty, in which subtle behavioral measures are used in such a way that the subjects do not know that modesty is being assessed. In one experiment, Japanese students revealed, through various nonverbal behavioral indices, more surprise and disbelief upon hearing that they had performed better than the average student at their university on a test of intelligence than upon hearing that they had performed below average on the test, while the opposite was true for Canadian students (Heine & others, 2000).

Several studies, however, indicate that on certain traits—traits that are particularly valued in collectivist cultures—Japanese do show self-enhancing biases. The typical Japanese person does not rate himself or herself as more intelligent or physically skilled than average, but, according to one study, does rate himself or herself above average in agreeableness, cooperativeness, respectfulness to others, patience, loyalty, and being a good listener (Sedikides & others, 2003). Such findings have led some researchers to suggest that self-enhancing biases are universal, but the traits that people most value, and on which they self-enhance, vary across cultures (Brown & Kobayashi, 2002; Sedikides & others, 2003).

SECTION REVIEW

The social world around us profoundly affects our understanding of ourselves.		
Seeing Ourselves Through Others' Eyes - Classroom experiments have demonstrated Pygmalion effects, in which adults' expectations about children's behavior created the expected behavior. Such effects occur at least partly by altering the children's self-concepts. - A variety of evidence supports the sociometer theory, which states that self-esteem reflects the level of acceptance or rejection we believe we can expect from others.	Active Construction of Self-Perceptions - We understand ourselves largely through social comparison—comparing ourselves to others. Our judgments and feelings about ourselves depend on the reference group to which we compare ourselves, as illustrated by the big-fish-in-little-pond effect. - At least in North America and Western Europe, people tend to enhance their views of themselves through such means as making self-serving attributions (attributing success to the self and failure to the situation), accepting praise at face value, and defining their own criteria for success.	East–West Differences in Self-Perception - In the collectivist cultures of East Asia, people tend to describe themselves more in terms of their social groups and roles than in terms of individual personality traits. - East Asians are less likely than Westerners to self-enhance; the Japanese may even show a self-effacing bias. However, East Asians may still self-enhance on traits that they value highly, such as modesty itself.

Perceiving Ourselves and Others as Members of Groups

In the previous section you read of evidence that the self-concept is social in that others are involved in its construction: We see ourselves reflected in others' reactions to us, and we understand ourselves by comparing our attributes with other people's. But the self-concept is social in another sense as well: Others are not just involved in its construction; they are also part of its contents. We describe and think of ourselves not just in terms of our individual characteristics—I *am short, adventurous, somewhat shy*—but also in terms of the groups to which we belong and with which we identify—I *am a French Canadian, Roman Catholic, member of the university marching band*. Self-descriptions that pertain to the person as a separate individual are referred to as **personal identity**, and those that pertain to the social categories or groups to which the person belongs are referred to as **social identity** (Tajfel, 1972). In this section we examine some consequences of viewing ourselves and others in terms of the groups to which we and they belong.



John Tlumacki — The Boston Globe

Social identity can alter self-esteem

When the Boston Red Sox won the baseball World Series, Bostonians, including those who had nothing to do with the victory, felt good about themselves. Our self-esteem is affected by the successes and failures of groups with which we identify, even if we are not actually members of those groups.

Shifting Between Personal and Social Identities

Our self-concepts are relatively consistent from situation to situation, but not rigid (Oakes & others, 1994). We think of ourselves differently at different times, in ways that help us meet the ever-changing challenges of social life. Sometimes, for some purposes, we find it most useful to think of our unique characteristics and motives: our personal identity. At other times, for other purposes, we find it more useful to think of ourselves as components of a larger unit, a group to which we belong: our social identity. Our evolution as a species entailed a continuous balance between the need to assert ourselves as individuals and the need to cooperate with others (Guisinger & Blatt, 1994). To enable us to walk that tightrope, without falling into isolation on the one side or into complete self-sacrifice to the group on the other, natural selection apparently endowed us with the capacity to hold both personal and social identities and to switch between them to meet our needs for survival.

In evolutionary history the groups with which people cooperated included some that were lifelong, such as the family and tribe, and others that were more ephemeral, such as a hunting party organized to track down a particular antelope. Today the relatively permanent groups with which we identify may include our family, ethnic group, co-religionists, and occupational colleagues. The temporary groups include the various teams and coalitions with which we affiliate for particular ends, for periods ranging from minutes to years.

Different cultures in today's world tend to vary in the relative weights they give to personal and social identity. As you might expect, individualist cultures emphasize personal identity, and collectivist cultures emphasize social identity. But the difference is relative. People everywhere think of themselves in terms of both types of identity.

Consequences of Personal and Social Identity for Self-Esteem

Our feelings about ourselves depend not just on our personal achievements but also on the achievements of the groups with which we identify, even when we ourselves play little or no role in those achievements. Sports fans' feelings about themselves rise and fall as "their" team wins and loses (Hirt & others, 1992). Similarly, people feel good about themselves when their town, university, or place of employment achieves high rank or praise.

In some situations, the very same event—high achievement by other members of our group—can temporarily raise or lower our self-esteem, depending on whether our social identity or personal identity is more active. When our social identity predominates, our group-mates are part of us and we experience their success as ours. When our personal identity predominates, our group-mates are the reference group against which we measure our own accomplishments, so their

>>> 20

What value might lie in our flexible capacity to think of ourselves in terms of both personal and social identities?

>>> 21

Why is it that excellent performance by other members of our group can either lower or raise our self-esteem?

success may diminish our view of ourselves. Social psychologists have demonstrated both these effects by priming people to think in terms of either their social or personal identities as they hear of high accomplishments by others in their group (Brewer & Weber, 1994; Gardner & others, 2002).

You read previously that students at highly selective schools think worse of themselves as scholars than do equally able students at less selective schools because of the difference in their relative standing in their reference group. A subsequent study indicated that this is true for those who think primarily in terms of their personal identities but not for those who think primarily in terms of their social identities (McFarland & Buehler, 1995). When the social identity is foremost, self-feelings are elevated, not diminished, by the excellent performance of group-mates. Graceful winners of individual achievement awards know this intuitively, and to promote good feelings and not incur jealousy, they activate their group-mates' social identities by describing the award as belonging properly to the whole group.

Group-Enhancing Biases

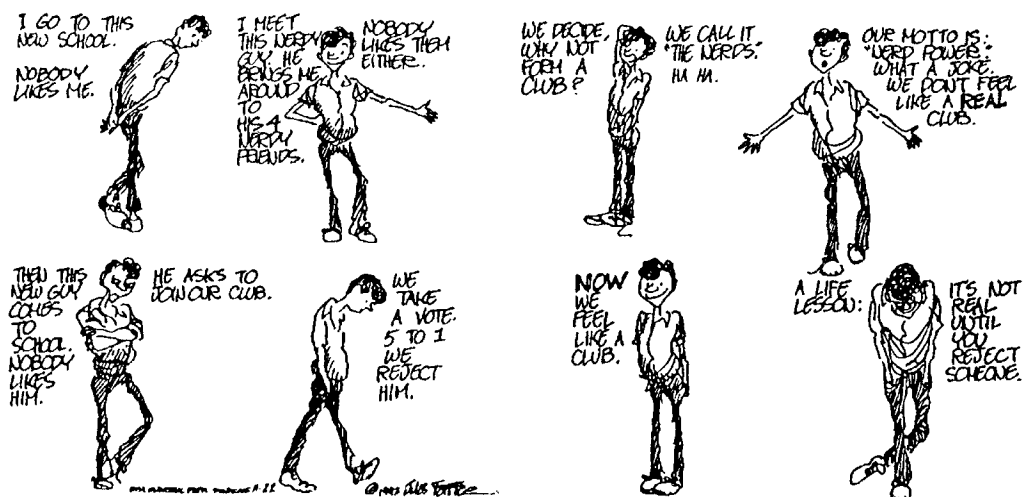
22

According to Tajfel, why do we inflate our view of the groups to which we belong? How does a recent study of members of volleyball teams support Tajfel's theory?

Other studies reveal that our bias to think highly of ourselves applies to our social identity as well as to our personal identity. In some conditions our *group-enhancing biases* are at least as strong as our self-enhancing biases (Bettencourt & others, 2001; Rubin & Hewstone, 1998). In fact, the concept of social identity first became prominent in social psychology when Henri Tajfel (1972, 1982) used it to explain people's strong bias in favor of their own groups over other groups in all sorts of judgments. He argued that we exaggerate the virtues of our own groups to build up the part of our self-esteem that derives from our social identity.

Tajfel and others showed that group-enhancing biases occur even when there is no realistic basis for assuming that one group differs from another. In one laboratory experiment, people who knew they had been assigned to one of two groups by a purely random process—a coin toss—nevertheless rated their own group more positively than they did the other group (Locksley & others, 1980). You don't need a degree in psychology to know that the fans of opposing baseball teams see the same plays differently, in ways that allow them all to leave the game believing that theirs was the better team, regardless of the score. Strike three is attributed by one group to the pitcher's dazzling fastball and by the other to the umpire's unacknowledged need for eyeglasses.

Some research has shown that group-enhancing biases increase when people are primed to think of their social identities and decline when people are primed to think of their personal identities. In one study, volleyball players attributed their team's victories to their team's great skill and their losses to external circumstances such as poor luck or bad officiating. This group-enhancing attributional bias disap-



peared, however, when the players underwent a self-affirmation procedure in which they listed their personal accomplishments in realms other than volleyball (Sherman & Kim, 2005). The self-affirmation presumably raised their self-esteem and primed them to think in terms of their personal rather than social identities, so they no longer felt motivated to think so highly of their team. Other research indicates that, at least in some situations, East Asians show more group-enhancing bias than do North Americans, which is consistent with the greater emphasis that Asian cultures place on social identity (Chen & others, 1998; Lehman & others, 2004).

Stereotypes and Their Automatic Effects on Perception and Behavior

We can switch between personal and social identities in our perceptions of others as well as ourselves. When we view others in terms of their personal identities, we see them as unique individuals. When we view others in terms of their social identities, we gloss over individual differences and see all members of a group as similar to one another. This is particularly true when we view members of out-groups—that is, members of groups to which we do not belong.

The schema, or organized set of knowledge or beliefs, that we carry in our heads about any group of people is referred to as a **stereotype**. The first person to use the term *stereotype* in this way was the journalist Walter Lippmann (1922/1960), who defined it as “the picture in the head” that a person may have of a particular group or category of people. You may have stereotypes for men, women, Asians, African Americans, Californians, Catholics, and college professors. We gain our stereotypes largely from the ways our culture as a whole depicts and describes each social category. A stereotype may accurately portray typical characteristics of a group, or exaggerate those characteristics, or be a complete fabrication based on culture-wide misconceptions.

Stereotypes are useful to the degree that they provide us with some initial, valid information about a person, but they are also sources of prejudice and social injustice. Our stereotypes can lead us to pre-judge others on the basis of the group they belong to, without seeing their qualities as individuals. In the United States, much research on stereotyping has focused on the stereotypes that European Americans have of African Americans, which are seen as a component of the long history of racial bias that began with the institution of slavery.

Distinction Between Implicit and Explicit Stereotypes

Many people in our culture, particularly college students, are sensitized to the harmful effects of stereotypes, especially negative stereotypes about socially oppressed groups, and are reluctant to admit to holding them. Partly for that reason, social psychologists find it useful to distinguish among three levels of stereotypes: public, private, and implicit (Dovidio & others, 1994). The *public* level is what we say to others about a group. The *private* level is what we consciously believe but generally do not say to others. Both public and private stereotypes are referred to as **explicit stereotypes**, because the person consciously uses them in judging other people. Such stereotypes are measured by questionnaires on which people are asked, in various ways, to state their views about a particular group, such as African Americans, women, or old people. Responses on such questionnaires are relatively easy to fake, but the hope is that research subjects who fill them out anonymously will fill them out honestly.



© The New Yorker Collection 1992 Warren Miller from cartoonbank.com. All rights reserved.

“Actually, Lou, I think it was more than just my being in the right place at the right time. I think it was my being the right race, the right religion, the right sex, the right socioeconomic group, having the right accent, the right clothes, going to the right schools . . .”

>>> **23** What is the distinction among public, private, and implicit stereotypes? What are two means by which researchers identify implicit stereotypes?

Implicit stereotypes, in contrast, are sets of mental associations that can guide our judgments and actions toward members of the group in question without our conscious awareness, even if those associations run counter to our conscious beliefs. Most psychological research on stereotyping today has to do with implicit stereotypes. These stereotypes are measured with tests in which the person's attention is focused not on the stereotyped group per se, but on performing quickly and accurately an objective task that makes use of stimuli associated with the stereotype. Two types of tests commonly used in this way are *priming tests* and *implicit association tests*.

Priming as a Means of Assessing Implicit Stereotypes As discussed in Chapter 9, *priming* is a general method for measuring the strengths of mental associations. The premise underlying this method is that any concept presented to a person primes (activates) in the person's mind the entire set of concepts that are closely associated with that concept. Priming the mind with one concept makes the related concepts more quickly retrievable from long-term memory into working memory. In a test of implicit stereotypes about black people and white people, for example, subjects might, on each trial, be presented with either a black face or a white face and then with some adjective to which the person must respond as quickly as possible in a way that demonstrates understanding of the adjective's meaning (Dovidio & others, 1996; Kawakami & Dovidio, 2001). For example, the required response might be to press one button if the adjective could describe a person and another button if it could describe a house but not a person. In some of these experiments the priming stimulus (a black or white face) that precedes the adjective is flashed very quickly on a screen—too quickly to be identified consciously but slowly enough that it can register in the subject's unconscious mind and affect his or her thought processes.

Experiments of this sort have shown that white U.S. college students respond especially quickly to such words as *lazy*, *hostile*, *musical*, and *athletic* when primed with a black face, and especially quickly to such words as *conventional*, *materialistic*, *ambitious*, and *intelligent* when primed with a white face. Although positive and negative traits appear in both stereotypes, the experiments reveal that the implicit stereotypes that white students have of blacks are significantly more negative than the ones they have of whites (Dovidio & others, 1986; Olson & Fazio, 2003).

Implicit Association Tests *Implicit association tests*, developed initially by Anthony Greenwald and his colleagues (1998, 2002), take advantage of the fact that people can classify two concepts together more quickly if they are already strongly associated in their minds than if they are not strongly associated. To get an idea of how such a test works, complete the paper-and-pencil version that is presented in Figure 13.7. You will need a watch or clock that can count seconds to time yourself.

A		B		C	
Male	Female	Male or Violent	Female or Nonviolent	Male or Nonviolent	Female or Violent
☐ John	☐	☐ Henry	☐	☐ Jill	☐
☐ Hank	☐	☐ War	☐	☐ Hug	☐
☐ Joan	☐	☐ Fight	☐	☐ Punch	☐
☐ David	☐	☐ Joan	☐	☐ Bill	☐
☐ Mary	☐	☐ Help	☐	☐ Hank	☐
☐ Jill	☐	☐ Mary	☐	☐ Fight	☐
☐ Susan	☐	☐ David	☐	☐ Susan	☐
☐ Bill	☐	☐ Hank	☐	☐ Help	☐
☐ Henry	☐	☐ Love	☐	☐ Joan	☐
☐ Amy	☐	☐ Jill	☐	☐ Kill	☐
☐ David	☐	☐ Punch	☐	☐ John	☐
☐ Hank	☐	☐ John	☐	☐ David	☐
☐ Joan	☐	☐ Susan	☐	☐ Hit	☐
☐ Jill	☐	☐ Care	☐	☐ Care	☐
☐ John	☐	☐ Hit	☐	☐ War	☐
☐ Mary	☐	☐ Bill	☐	☐ Henry	☐
☐ Amy	☐	☐ Amy	☐	☐ Amy	☐
☐ Bill	☐	☐ Peace	☐	☐ Love	☐
☐ Susan	☐	☐ Kill	☐	☐ Mary	☐

FIGURE 13.7 | An implicit association test How long does it take you to classify the items in each of these columns? Time yourself, in seconds, separately on each column. For column A, place a check to the left of each male name and to the right of each female name. For column B, place a check to the left of each male name or term that has to do with violence, and place a check to the right of each female name or term that has to do with nonviolence. For column C, place a check to the left of each male name or term that has to do with nonviolence, and place a check to the right of each female name or term that has to do with violence. (Modeled after an example in Gladwell, 2005.)

How did you perform? If you are like everyone I know who has taken this test, it took you longer to go through List C, where nonviolent terms are classified with male names and violent terms with female names, than it took you to go through List B, where violent terms go with male names and nonviolent terms with female names. Nearly everyone shares the stereotype that males are relatively violent and females are relatively nonviolent. Implicit association tests used in research are usually presented on a computer. Each individual stimulus appears on the screen, one at a time, and the person must respond by pressing one of two keys on the keyboard, depending on the category of the stimulus. If you wish to take such a test, you might find examples on the Internet at <https://implicit.harvard.edu/implicit> (this site may or may not still be active at the time that you are reading this).

In implicit association tests having to do with race, a typical procedure is to use photographs of white and black faces, and words with “good” meanings (such as *love, happy, truth, terrific*) and words with “bad” meanings (such as *poison, hatred, agony, terrible*) as the stimuli. In one condition, the job is to categorize white faces and “good” words together by pressing one key on the keyboard whenever one of them appears, and to classify black faces and “bad” words together by pressing another key whenever one of those appears. In the other condition, the job is to classify white faces and “bad” words together and black faces and “good” words together. The typical result is that white college students take about 200 milliseconds longer per key press, on average, in the latter condition than in the former (Aberson & others, 2004; Cunningham & others, 2001).

Implicit Stereotypes and Unconscious Discrimination

Implicit stereotypes can lead people who are not consciously prejudiced to behave in prejudicial ways, despite their best intentions. In one series of studies, for example, white students' scores on a test of implicit prejudice toward blacks correlated quite strongly with their subsequent reactions toward black conversation partners with whom they were paired (Dovidio & others, 1997, 2002). Those who scored highest in implicit prejudice spoke friendly words toward their black partners and believed they were behaving in a friendly way, but showed nonverbal signs of negativity or apprehension, which made them come across as unfriendly. They were less likely to make eye contact with their black partners, showed more rapid eye blinking, and were rated by their black partners and other observers as less friendly, despite their words, than were the students whose implicit prejudice scores were lower. None of these effects occurred when the conversation partners were white instead of black. Also, the same studies revealed no significant correlations between any of these measures and the students' *explicit* views of black people, which had been assessed by a questionnaire and were generally quite positive for all of the students.

In other studies, involving fMRI neuroimaging, white subjects showed increased neural activity in the brain's amygdala when they viewed unfamiliar black faces, but not when they viewed unfamiliar white faces. This difference in brain response correlated significantly with measures of the subjects' implicit negative stereotypes of blacks but not with their explicit stereotypes of them (Cunningham & others, 2004; Wheeler & Fiske, 2005). The amygdala, as discussed in Chapter 6, is critically involved in fear and anger.

It is not hard to imagine how such reactions can have real-world consequences. A white employer, who thinks of himself as not prejudiced against black people, interviews several applicants for a job in his company. Some of the applicants are white, others black. Because of the employer's implicit negative stereotype of blacks, his interviews with black candidates come across as strained. He appears unfriendly to them, and they, in consequence, respond in a less friendly way toward him than they otherwise would. The result is that the employer concludes that none of the black candidates has quite the personality required for the job, and he selects one of the white applicants. All this happens despite the employer's sincere intention to be unbiased in his selection.

>>> 24

How have researchers shown that implicit stereotypes can lead to prejudicial behavior even in the absence of conscious prejudice?

Implicit Stereotypes Can Be Deadly

Around midnight, on February 3, 1999, four plain-clothes policemen saw a young black man, Amadou Diallo, standing on the stoop of his apartment building in the South Bronx, in New York City (Gladwell, 2005). He looked suspicious to them, so they stopped their unmarked cruiser and approached him. As they approached, he drew something from his pocket that looked to one of the officers like a gun. Reacting quickly, in self-defense, that officer opened fire on Diallo and killed him instantly. When the officers looked more closely at the dead man they discovered that what he had drawn from his pocket was not a gun but his wallet. Diallo had apparently assumed that the four large, white, non-uniformed men approaching him with guns were intent on robbing him, so, seeing no route for escape, he had taken out his wallet.

25

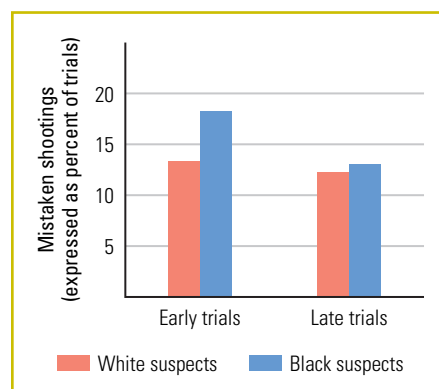
What evidence suggests that implicit prejudice can cause police officers to shoot at black suspects more readily than at white suspects?

<<< Would Diallo have been shot if he had been white? If he had been white, would he have been seen as suspicious and been approached by the police officers, and if he had been approached, would his wallet have been seen as a gun? We can never know the answers as applied to this individual case, but there are grounds for believing that skin color can play a large role in such cases. Many research studies have shown that whites implicitly view unfamiliar blacks as more hostile, violent, and suspicious-looking than unfamiliar whites, and several experiments, inspired by the Diallo case and others like it, reveal that whites implicitly associate black faces more strongly with guns than they do white faces (Correll & others, 2002; Payne, 2001; Plant & Peruche, 2005).

Most such experiments have used college students as subjects, but one was conducted with actual police officers (Plant & Peruche, 2005). In this experiment, which involved computer simulations of self-defense situations, the instructions to each officer included the following words:

Today your task is to determine whether or not to shoot your gun. Pictures of people with objects will appear at various positions on the screen. . . . Some of the pictures will have the face of a person and a gun. These people are the criminals, and you are supposed to shoot at these people. Some of the pictures will have a face of a person and some other object (e.g., a wallet). These people are not criminals and you should not shoot at them. Press the "A" key for "shoot" and press the "L" key on the keyboard for "don't shoot."

FIGURE 13.8 Race of suspect affects decision to shoot in computer simulations During the first 80 trials of a computer simulation game, police officers who had to make split-second decisions "shot" more often at unarmed black suspects than at unarmed white suspects. With practice, however, this bias disappeared. (Based on data in Plant & Peruche, 2005.)



During the experiment, each picture appeared on the screen until the officer responded or until a 630-millisecond time limit had elapsed. As shown in Figure 13.8, during the first 80 trials the officers mistakenly shot unarmed black suspects significantly more often than they shot unarmed white suspects. With practice, during which they received immediate feedback as to whether they had shot an armed or unarmed person, they gradually overcame this bias. They did not show the bias during the final 80 trials of

the 160-trial game. The researchers hope that such training will reduce officers' bias in real-life encounters with criminal suspects.

Defeating Explicit and Implicit Negative Stereotypes

26

What sorts of learning experiences are most effective in reducing (a) explicit and (b) implicit prejudice?

<<< Explicit and implicit stereotypes are psychologically quite different from each other. As you have seen, people often hold implicit stereotypes that do not coincide with their explicit beliefs about the stereotyped group. Whereas explicit stereotypes are products of conscious thought processes, modifiable by deliberate learning and logic, implicit stereotypes appear to be products of more primitive emotional processes, modifiable by such means as classical conditioning. White people who



© Royalty-Free / Corbis

Happy conditioning Experiences such as that depicted here help to counter the development of implicit negative racial stereotypes.

have close black friends exhibit less implicit prejudice than do those without black friends (Aberson & others, 2004), and there is evidence that exposure to admirable black characters in literature, movies, and television programs can reduce implicit prejudice in white people (Rudman, 2004). Apparently, the association of positive feelings with individual members of the stereotyped group helps reduce automatic negative responses toward the group as a whole.

In one long-term study, Laurie Rudman and her colleagues (2001) found that white students who volunteered for a diversity-training course showed, by the end of the course, significant reductions in both explicit and implicit prejudice toward black people, but the two reductions correlated only weakly with each other. Those students who showed the greatest decline in explicit prejudice did not necessarily show much decline in implicit prejudice, and vice versa. Further analysis revealed that those students who felt most enlightened by the verbal information presented in the course, and who reported the greatest conscious desire to overcome their prejudice, showed the greatest declines in explicit prejudice. In contrast, those who made new black friends during the course, or who most liked the African American professor who taught it, showed the greatest declines in implicit prejudice.

SECTION REVIEW

Social identity can be a major determinant of how we perceive ourselves and others.

Personal and Social Self-Identities

- We view ourselves in terms of both personal identity and social identity (the social categories or groups to which we belong). Individualist cultures emphasize personal identity, and collectivist cultures emphasize social identity.
- When social identity is uppermost in our minds, the success of others in our group boosts our self-esteem. When personal identity is uppermost, members of our group become our reference group, and their success can lower our self-esteem.
- Just as we have a self-enhancing bias, we have a group-enhancing bias, especially when our social identity predominates.

Implicit and Explicit Stereotypes

- Stereotypes—the schemas that we have about groups of people—can be explicit (available to consciousness) or implicit (unconscious but able to affect our thoughts, feelings, and actions). Implicit stereotypes are measured through priming and implicit association tests.
- Studies reveal that negative implicit stereotypes can promote prejudiced behavior even without conscious prejudice.
- Implicit prejudices are based on primitive emotional processes, modifiable by classical conditioning. Positive associations with members of the stereotyped group can help to reduce implicit prejudice.

Attitudes: Their Origins and Their Effects on Behavior

Thus far in this chapter we have been discussing the ways in which people evaluate other people and themselves. In doing so, we have been implicitly discussing attitudes. An **attitude** is any belief or opinion that has an evaluative component—a judgment or feeling that something is good or bad, likable or unlikable, moral or immoral, attractive or repulsive. Attitudes tie individuals cognitively and emotionally to their entire social world. We all have attitudes about countless objects, people, events, and ideas, ranging from our feelings about a particular brand of toothpaste to those about democracy or religion. Our most central attitudes, referred to as *values*, help us judge the appropriateness of whole categories of actions. People in our society devote enormous effort to modifying other people's attitudes. Advertising, political campaigning, and the democratic process itself (in which people speak freely in support of their views) are, in essence, attempts to change others' attitudes. Here we shall examine some ideas about the origins of attitudes and the roles they play in our behavior.

Attitudes as Products of Conditioning and Reasoning

Our attitudes are largely products of learning. Through direct experience or from what others convey to us, we learn to like some objects, events, and concepts and to dislike others. Such learning can be automatic, involving no conscious thought, or, at the other extreme, it can be highly controlled, involving deliberate searches for relevant information and rational analysis of that information.

Attitudes Through Classical Conditioning: No Thought

Classical conditioning, a basic learning process discussed in Chapter 4, can be thought of as an automatic attitude-forming system. A new stimulus (the conditioned stimulus) is paired with a stimulus that already elicits a particular reaction (the unconditioned stimulus), and, as a result, the new stimulus comes to elicit, by itself, a reaction similar to that elicited by the original stimulus. Using the language of the present chapter, we can say that Pavlov's dog entered the experiment with a preexisting positive attitude toward meat powder. When Pavlov preceded the meat powder on several occasions with the sound of a bell, the dog acquired a positive attitude toward that sound. The dog now salivated and wagged its tail when the bell rang, and if given a chance, it would have learned to ring the bell itself.

27 How do advertisers use classical conditioning to influence people's attitudes?

<<< Classical conditioning leads us to feel positive about objects and events that have been linked in our experience to pleasant, life-promoting occurrences and to feel negative about those that have been linked to unpleasant, life-threatening occurrences. In today's world, however, where many of the stimulus links we experience are the creations of advertisers and others who want to manipulate us, classical conditioning can have maladaptive consequences. All those ads in which cigarettes, beer, and expensive gas-guzzling cars are paired with beautiful people, happy scenes, and lovely music are designed to exploit our most thoughtless attitude-forming system, classical conditioning. The advertisers want us to salivate, wag our tails, and run out and buy their products.

Controlled experiments have shown that it is relatively easy to condition positive or negative attitudes to such products as a brand of mouthwash by pairing the product with positive or negative scenes or words (Grossman & Till, 1998). In some experiments, such conditioning has been demonstrated even when the subjects are not aware of the pairing of the conditioned and unconditioned stimuli (De Houwer & others, 2001; Olson & Fazio, 2001).

28 What are some examples of decision rules (heuristics) that people use with minimal thought to evaluate messages?

Attitudes Through Heuristics: Superficial Thought

<<< Beyond simple classical conditioning is the more sophisticated but still relatively automatic process of using certain *decision rules*, or *heuristics*, to evaluate informa-

Her royal pleasure John B. Watson, the founder of behaviorism (discussed in Chapters 1 and 4), became an advertising specialist after leaving academia. This ad, created by Watson, helped sell a skin cream by associating it with the wealth and beauty of the queen of Spain. According to the fine print, the queen uses Pond's vanishing and cold creams and "has expressed her royal pleasure in them."

tion and develop attitudes (Chen & Chaiken, 1999). Heuristics provide shortcuts to a full, logical elaboration of the information in a message. Examples of such rules include the following:

1. If there are a lot of numbers and big words in the message, it must be well documented.
2. If the message is phrased in terms of values that I believe in, it is probably right.
3. Famous or successful people are more likely than unknown or unsuccessful people to be correct.
4. If most people believe this message, it is probably true.

We learn to use such rules, presumably, because they often allow us to make useful judgments with minimal expenditures of time and mental energy. The rules become mental habits, which we use implicitly, without awareness that we are using them. Advertisers, of course, exploit these mental habits, just as they exploit the process of classical conditioning. They sprinkle their ads with irrelevant data and high-sounding words such as *integrity*, and they hire celebrities to endorse their products.

Attitudes Through Logical Analysis of the Message: Systematic Thought

Sometimes, of course, we think logically. Generally, we are most likely to do so for issues that really matter to us. In a theory of persuasion called the **elaboration likelihood model**, Richard Petty and John Cacioppo (1986) proposed that a major determinant of whether a message will be processed systematically (through logical analysis of the content) or superficially is the personal relevance of the message. According to Petty and Cacioppo, we tend to be *cognitive misers*; we reserve our elaborative reasoning powers for messages that seem most relevant to us, and we rely on mental shortcuts to evaluate messages that seem less relevant. Much research supports this proposition (Petty & Wegener, 1999; Pierro & others, 2004).

In one experiment on the role of personal relevance in persuasion, Petty and his colleagues (1981) presented college students with messages in favor of requiring students to pass a set of comprehensive examinations in order to graduate. Different groups of students received different messages, which varied in (a) the *strength* of the arguments, (b) the alleged *source* of the arguments, and (c) the *personal relevance* of the message. The weak arguments consisted of slightly relevant quotations, personal opinions, and anecdotal observations; the strong arguments contained well-structured statistical evidence that the proposed policy would improve the reputation of the university and its graduates. In some cases the arguments were said to have been prepared by high school students and in other cases by the Carnegie Commission on Higher Education. Finally, the personal relevance was varied by stating in the high-relevance condition that the proposed policy would take effect the following year, so current students would be subject to it, and in the low-relevance condition that it would begin in 10 years. After hearing the message, students in each condition were asked to rate the extent to which they agreed or disagreed with the proposal.

H·M·VICTORIA EUGENIA Queen of Spain



HER MAJESTY, Victoria Eugenia, Queen of Spain, is the embodiment of queenly beauty. She is granddaughter of Queen Victoria, niece of King Edward VII, and cousin to the reigning King of England. When as a Princess of the British royal household, she married the dark emperous King of Spain, she was "a beauty from the North, with pale golden hair, wild rose complexion and eyes of malachite blue." Today as Victoria Eugenia, Queen of Spain, and mother of six lovely children, she is more beautiful, more regal than ever.

THE gilded silver Tiffany jars below, chased and engraved with the royal crest and monogram and filled with Pond's Gold and Vanishing Creams, were recently sent by the Pond's Extract Company as a gift for Her Majesty's dressing table. Her Majesty, who uses the creams, has expressed her royal pleasure in them.

The silver jar recently sent to Her Majesty is not for general sale.

IF you would like to try these delicate creams, of which Her Majesty has signified her appreciation—Pond's Gold Cream for cleansing and preserving the fine texture of the skin, Pond's Vanishing Cream for lending an exquisite finish, a foundation or powder and complete protection against exposure—send the offer below.

FREE OFFER: Mail coupon for five tubes of Pond's Two-Creams and directions for using. The Pond's Extract Company, Dept. K, 137 Hudson Street, New York City. Please send the five tubes of Pond's Two-Creams.

Name _____
Street _____

A recent portrait of Her Majesty, Victoria Eugenia, Queen of Spain, here reproduced by gracious permission of Her Majesty to the Pond's Extract Company

Courtesy of The J. Walter Thompson Archive, Duke University

>>> 29

How did an experiment support the idea that people tend to reserve systematic thought for messages that are personally relevant to them and to use heuristics for other messages?

Two different sets of norms The couple at the right probably have many attitudes that are quite different from those expressed by the young people sitting near them. This may explain their discomfort and tendency to move away.



© Larry Mangino / The Image Works

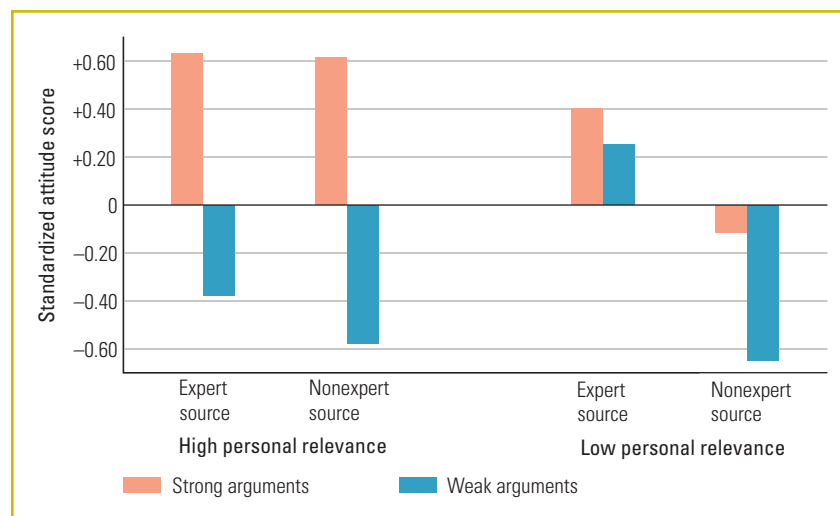
Figure 13.9 shows the results. As you can see, in the high-relevance condition the quality of the arguments was most important. Students in that condition tended to be persuaded by strong arguments and not by weak ones, regardless of the alleged source. Thus, in that condition, students must have listened to and evaluated the arguments. In the low-relevance condition, the quality of the arguments had much less effect, and the source of the arguments had much more. Apparently, when the policy was not going to affect them, students did not attend carefully to the arguments but, instead, relied on the simple decision rule that experts (members of the Carnegie Commission) are more likely to be right than are non-experts (high school students).

There is no surprise in Petty and Cacioppo's theory or in the results supporting it. The idea that people think more logically about issues that directly affect them than about those that don't was a basic premise of philosophers who laid the foundations for democratic forms of government. And, to repeat what is by now a familiar refrain, our mental apparatus evolved to keep us alive and promote our welfare in our social communities; it is no wonder that we use our minds more fully for that than for purposes that have little effect on our well-being.

FIGURE 13.9

Effect of persuasive arguments on attitude under various experimental conditions In this graph, movement above the horizontal axis indicates agreement with the persuasive message, and movement below the axis indicates disagreement. When the issue was of high personal relevance (left half of graph), the strength or weakness of the arguments had more impact than did the source of the arguments; but when the issue was of low personal relevance (right half of graph), the reverse was true. (Adapted from Petty & others, 1981.)

Copyright © 1981 by the American Psychological Association. Adapted by permission.



Attitudes as Social Norms

Values and other attitudes are properties not just of individuals but also of social groups. People who live in the same household or community, or who work at the same occupation, or who grew up in the same generation in the same culture, tend to hold similar attitudes (Cohen, 1996; Rogler, 2002). It is not hard to think of explanations: People experiencing similar conditions and communicating with one another are exposed to similar information, are subjected to the same persuasive messages, may have come together and stayed together partly because of shared attitudes, and may have modified their attitudes to gain greater acceptance in the group.

Research shows that people tend to like and help unfamiliar others more if they believe that those others have attitudes similar to their own than if they believe otherwise (Park & Schaller, 2005). People also tend to feel uncomfortable when they find themselves in a group whose attitudes are quite different from their own. Laboratory studies have shown that, depending on conditions, such discomfort can lead subjects to refrain from expressing their attitudes, or to modify their attitudes in ways that make them more congenial to the group, or to leave the group and join a different one (Matz & Wood, 2005). Much more is said in Chapter 14 about the social pressures that motivate people to conform to group norms.

Attitudes as Rationalizations to Attain Cognitive Consistency

A century ago, Sigmund Freud began developing his controversial theory that human beings are fundamentally irrational. What pass for reasons, according to Freud, are most often rationalizations designed to calm our anxieties and boost our self-esteem. You will read more about Freud's view in Chapter 15. A more moderate view, to be pursued here, is that we are rational but the machinery that makes us so is far from perfect. The same mental machinery that produces logic can produce pseudo-logic.

In the 1950s, Leon Festinger (1957) proposed what he called the **cognitive dissonance theory**, which ever since has been one of social psychology's most central ideas. According to the theory, we have, built into the workings of our mind, a mechanism that creates an uncomfortable feeling of *dissonance*, or lack of harmony, when we sense some inconsistency among the various attitudes, beliefs, and items of knowledge that constitute our mental store. Just as the discomfort of hunger motivates us to seek food, the discomfort of cognitive dissonance motivates us to seek ways to resolve contradictions or inconsistencies among our cognitions.

Such a mechanism could well have evolved to serve adaptive functions related to logic. Inconsistencies imply that we are mistaken about something, and mistakes can lead to danger. Suppose you have a favorable attitude about sunbathing, but you learn that overexposure to the sun's ultraviolet rays is the leading cause of skin cancer. The discrepancy between your preexisting attitude and your new knowledge may create a state of cognitive dissonance. To resolve the dissonance, you might change your attitude about sunbathing from positive to negative or you might bring in a third cognition: "Sunbathing is relatively safe, in moderation, if I use a sunscreen lotion." But the dissonance-reducing drive, like other drives, does not always function adaptively. Just as our hunger can lead us to eat things that aren't good for us, our dissonance-reduction drive can lead us to avoid or reduce dissonance in illogical and maladaptive ways. Those are the effects that particularly intrigued Festinger and many subsequent social psychologists.

Avoiding Dissonant Information

I once heard someone cut off a political discussion with the words, "I'm sorry, but I refuse to listen to something I disagree with." People don't usually come right out and say that, but have you noticed how often they seem to behave that way? Given a choice of books or articles to read, lectures to attend, or documentaries to watch, people generally choose those that they believe will support their existing views.

>>> **30**
How does the cognitive dissonance theory explain people's attraction to some information and avoidance of other information?

That observation is consistent with the cognitive dissonance theory. One way to avoid dissonance is to avoid situations in which we might discover facts or ideas that run counter to our current views. If we avoid listening to or reading about the evidence that ultraviolet rays can cause skin cancer, we can blithely continue to enjoy sunbathing. People don't always avoid dissonant information, but a considerable body of research indicates that they very often do (Frey, 1986; Jonas & others, 2001).

Paul Sweeney and Kathy Gruber (1984) conducted a study during the 1973 Senate Watergate hearings that documented this phenomenon. (The hearings uncovered illegal activities associated with then-President Richard Nixon's reelection campaign against George McGovern.) By interviewing the same voters before, during, and after the hearings, Sweeney and Gruber discovered that (a) Nixon supporters avoided news about the hearings (but not other political news) and were as strongly supportive of Nixon after the hearings as they had been before; (b) McGovern supporters eagerly sought out information about the hearings and were as strongly opposed to Nixon afterward as they had been before; and (c) previously undecided voters paid moderate attention to the hearings and were the only group whose attitude toward Nixon was significantly influenced (in a negative direction) by the hearings. So, consistent with the cognitive dissonance theory, all but the undecideds approached the hearings in a way that seemed designed to protect or strengthen, rather than challenge, their existing views.

Firming Up an Attitude to Be Consistent with an Action

31

How does the cognitive dissonance theory explain why people are more confident of a decision just after they have made it than just before?

<<< We make most of our choices in life with less-than-absolute certainty. We vote for a candidate not knowing for sure if he or she is best, buy one car even though some of the evidence favors another, or choose to major in psychology even though some other fields have their attractions. After we have irrevocably made one choice or another—after we have cast our ballot, made our down payment, or registered for our courses and the deadline for schedule changes is past—any lingering doubts would be discordant with our knowledge of what we have done. So, according to the cognitive dissonance theory, we should be motivated to set those doubts aside.

Dissonance abolished This couple may have agonized long and hard before signing on the dotted line, but once they signed, they set their doubts aside and focused on the positive qualities of their new home. According to the cognitive dissonance theory, people are generally more confident about the correctness of their choices after those choices are made than before.

Many studies have shown that people do tend to set their doubts aside after making an irrevocable decision. Even in the absence of new information, people suddenly become more confident of their choice after acting on it than they were before. For example, in one study, bettors at a horse race were more confident that their horse would win if they were asked immediately after they had placed their bet than if they were asked immediately before (Knox & Inkster, 1968). In another study, voters who were leaving the polling place spoke more positively about their favored candidate than did those who were entering (Frenkel & Doob, 1976). In still another study, photography students who were allowed to choose one photographic print to keep as their own liked their chosen photograph better, several

days later, if they were not allowed to revise their choice than if they were allowed to revise it (Gilbert & Ebert, 2002). In the reversible-choice condition, they could reduce dissonance by saying to themselves that they could exchange the photograph for a different one if they wanted to, so they had less need to bolster their attitude toward the one they had chosen.

Changing an Attitude to Justify an Action: The Insufficient-Justification Effect

Sometimes people behave in ways that run counter to their attitudes and then are faced with the dissonant cognitions, “I believe *this*, but I did *that*.” They can't undo their deed, but they can relieve dissonance by modifying—maybe even reversing—their attitudes. This change



in attitude is called the **insufficient-justification effect**, because it occurs only if the person has no easy way to justify the behavior, given his or her previous attitude.

Example from the Autobiography of Benjamin Franklin More than 200 years ago, the great inventor, statesman, and master of practical psychology Benjamin Franklin recognized what we now call the insufficient-justification effect and used it to his advantage. In his autobiography, Franklin (1818/1949) describes how he changed the attitude of a political opponent who was trying to block his appointment to a high office:

I did not like the opposition of this new member, who was a gentleman of fortune and education with talents that were likely to give him in time great influence. . . . I did not, however, aim at gaining his favour by paying any servile respect to him, but after some time took this other method. Having heard that he had in his library a certain very scarce and curious book, I wrote a note to him expressing my desire of perusing that book and requesting he do me the favour of lending it to me for a few days. He sent it immediately. . . . When we next met in the House, he spoke to me (which he had never done before), and with great civility. And he ever afterwards manifested a readiness to serve me on all occasions, so that we became great friends. . . . This is another instance of the truth of an old maxim I had learned, which says, "He that has once done you a kindness will be more ready to do you another than he whom you yourself have obliged."

How would the cognitive dissonance theory explain why Franklin's former opponent changed his attitude toward Franklin? The opponent received Franklin's request to borrow the book, and for reasons of which he may not have been fully aware, such as the simple habit of courtesy, he did not turn it down. But once he sent the book to Franklin, he was thrown into a state of cognitive dissonance. One thought, *I do not like Ben Franklin*, was discordant with another, *I have just lent Franklin a very valuable book*. The second of these could not be denied, since that was an objective fact, so dissonance could best be relieved by changing the first: *Ben Franklin isn't really a bad sort. At least I know he's honest. If he weren't honest, I certainly wouldn't have lent him that valuable book*. Such thinking reduced or erased the dissonance and set the stage for new, friendlier behaviors toward Franklin in the future.

Conditions That Optimize the Insufficient-Justification Effect: Low Incentive and Free Choice One requirement for the insufficient-justification effect to occur is that there be no obvious, high incentive for performing the counter-attitudinal action. If Franklin had offered to do a large favor for his opponent in exchange for borrowing the book, the opponent could have maintained his dislike for Franklin by saying, "I lent him the book only to get him to do me that favor."

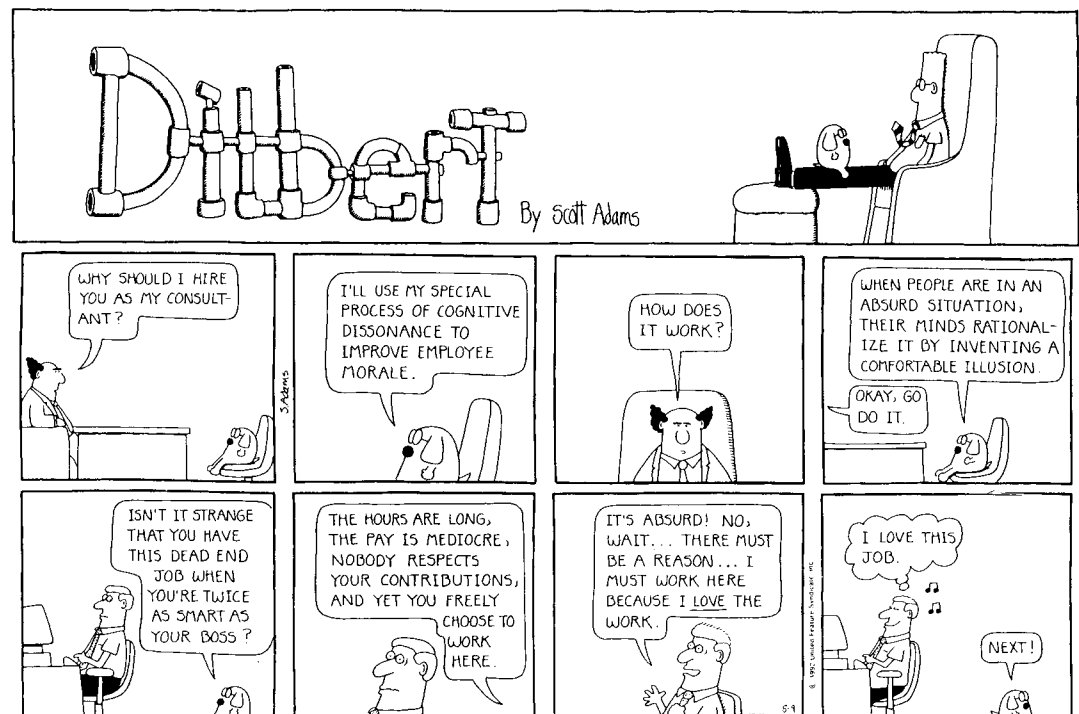
In a classic experiment that demonstrates the low-incentive requirement, Festinger and James Carlsmith (1959) gave college students a boring task (loading spools into trays and turning pegs in a pegboard) and then offered to "hire" them to tell another student that the task was exciting and enjoyable. Some students were offered \$1 for their role in recruiting the other student, and others were offered \$20 (a princely sum at a time when the minimum wage in the United States was \$1 an hour). The result was that those in the \$1 condition changed their attitude toward the task and later recalled it as truly enjoyable, whereas those in the \$20 condition continued to recall it as boring. Presumably, students in the \$1 condition could not justify their lie to the other student on the basis of the little they were promised, so they convinced themselves that they were not lying. Those in the \$20 condition, in contrast, could justify their lie: *I said the task was enjoyable when it was actually boring, but who wouldn't tell such a small lie for \$20?*

Another essential condition for the insufficient-justification effect is that subjects must perceive their action as stemming from their own free choice. Otherwise, they could justify the action—and relieve dissonance—simply by saying, "I was forced to do it."

In an experiment demonstrating the free-choice requirement, students were asked to write essays expressing support for a bill in the state legislature that most students personally opposed (Linder & others, 1967). Students in the *free-choice*

How is the insufficient-justification effect illustrated by Benjamin Franklin's story?

In theory, why should the insufficient-justification effect work best when there is minimal incentive for the action and the action is freely chosen? How was this theory verified by two classic experiments?



Dilbert reprinted by permission of United Features Syndicate.

condition were told clearly that they didn't have to write the essays, but they were encouraged to do so and none refused. Students in the *no-choice condition* were simply told to write the essays, and all complied. After writing the essays, all students were asked to describe their personal attitudes toward the bill. Only those in the free-choice condition showed a significant shift in the direction of favoring the bill; those in the no-choice condition remained as opposed to it as did students who had not been asked to write essays at all. In this case, attitude change in the free-choice condition apparently occurred because the students could justify their choice to write the essays only by deciding that they did, after all, favor the bill. Subsequent experiments, using essentially the same procedure, showed that students in the free-choice condition also exhibited more psychological discomfort and physiological arousal, compared to those in the no-choice condition, as they wrote their essays (Elkin & Leippe, 1986; Elliot & Devine, 1994). This finding is consistent with the view that they were experiencing greater dissonance.

Attitudes as Guides to Behavior

As you have seen, attitudes can help us adapt to our social groups and rationalize our decisions and actions. But common sense tells us that the most basic function of attitudes is to guide our behavior effectively. Presumably, we seek out those objects, people, events, and activities about which we have positive attitudes and avoid those about which we have negative attitudes. Or do we?

Early Findings of Lack of Attitude-Behavior Correlations

34

What early evidence and reasoning led some psychologists to suggest that attitudes have little effect on behavior?

<<< Social psychologists first became interested in attitudes because they conceived of them as mental guides for behavior (Allport, 1935). But early research on the topic was most remarkable for its failure to find reliable correlations between attitudes and actions. In one classic study, for instance, students in a college course filled out a questionnaire aimed at assessing their attitudes toward cheating, and later in the semester they were asked to grade their own true-false tests (Corey, 1937). The tests had already been graded secretly by the instructor, so cheating could be detected. The result was that a great deal of cheating occurred and no cor-

relation at all was found between the attitude assessments and the cheating. A strong correlation was found, however, between cheating and the student's true score on the test: The lower the true score, the more likely the student was to try to raise it by cheating.

Such findings led some psychologists to conclude that attitudes play almost no role in guiding behavior (Wicker, 1969). In addition, theoretical challenges to the idea that attitudes influence behavior were voiced by strict behaviorists, such as B. F. Skinner (1957), who argued that attitudes are simply *verbal habits* that influence what people say but are largely irrelevant to what they do. According to this view, people learn to say things like, "Honesty is the best policy," because they are rewarded for such words by their social group. But when confronted by an inducement to cheat—such as the opportunity to raise a test score—their behavior is controlled by a new set of rewards and punishments that can lead to an action completely contrary to the attitude they profess. Research over the past 40 years, however, has revealed many situations in which attitudes do predict behavior. Today essentially all social psychologists agree that attitudes play a role in people's behavior, and the question is, "When, or under what conditions, do attitudes play that role?"

Implicit Attitudes Automatically Influence Behavior

Earlier in this chapter, in the discussion of stereotypes, I distinguished between implicit and explicit racial attitudes. The same distinction applies to attitudes concerning all sorts of objects (Fazio & Olson, 2003). Implicit attitudes, by definition, are those that automatically, without conscious thought, influence our behavior toward or away from the attitudinal object. Implicit attitudes influence us at the emotional level rather than through conscious thought. In fact, only by conscious thought can we *prevent* them from influencing our behavior. Implicit attitudes are, in general, those that are formed through direct positive or negative experiences with the object, or through repeated experiences in which the object or some representation of it is paired with positive or negative events or images.

As an example, if you have a negative attitude toward bacon because on one or more occasions you ate it and got sick, then the sight and smell of bacon will automatically elicit your negative attitude and you won't eat it. Your attitude is an implicit one. However, if you have a negative attitude toward bacon only because you have read that it is high in nitrates, which are bad for you, your attitude will not be elicited automatically by the sight and smell of bacon. In that case, your attitude is an explicit one and will not automatically lead you to reject bacon.

Explicit Attitudes Must Be Retrieved from Memory to Affect Behavior

If your negative attitude toward bacon is explicit and not implicit, you will probably need some other cue, extrinsic to the bacon, to remind you of your attitude, or you will need to rely on a habit of checking your set of food-related attitudes whenever you eat. Only after repeated rehearsal might your intellectually derived attitude about bacon come automatically to mind whenever you see or smell it. Thus, if every time you see bacon you bring deliberately to mind some image of a disease that high nitrates might cause, then you are establishing the conditions for producing, over time, an implicit negative attitude. You are, in essence, classically conditioning yourself.

A good deal of research has shown that people are most likely to behave in accordance with their explicit attitudes when they are consciously reminded of them just before the behavioral test (Aronson, 1992; Wicklund & Frey, 1980). In the experiment on cheating, the students who cheated may have been immediately overwhelmed by their

>>> 35

How does the distinction between implicit and explicit attitudes contribute to modern understanding of the relation of attitudes to behavior? Under what conditions do explicit attitudes influence behavior?



You promised yourself you'd do something, someday.

Now you can keep your promise. It doesn't take much to help needy children around the world. Just \$24 a month and a lot of love. That's all you need to be a Save the Children sponsor.

Sponsorship is a great way to share your concern and help bring positive, lasting change to disadvantaged children. Your contributions will be combined with those of other sponsors to create self-help programs.

This is not a handout; it's a hand up. Clean water, nutritious food, an education – Save the Children works with each village to develop programs according to their specific needs.

For \$24 a month – just 79¢ a day – you'll help make changes that last a lifetime. Experience the joy of keeping a very important promise.

Call 1-800-614-7979.

Promise kept.

Yes, I'll help needy children through sponsorship.

I prefer to sponsor a _____ boy _____ girl _____, either in the area I've checked below:

☐ Where the need is greatest ☐ Caribbean/Latin America
☐ Middle East ☐ Africa ☐ Asia ☐ United States

Please send my child's photo and personal letter: 55142-M

Name _____ E-mail _____
 Phone _____ Age _____
 Address _____ Sex _____ Zip _____
 City _____

☐ Automatic Credit Card Deductions (Continued)

I authorize \$24 to be paid to Save the Children from my credit card each month. Automatic payments shall begin next month. Please use my credit card to make my initial donation now.

Credit Information: ☐ MasterCard ☐ Visa ☐ Discover ☐ AmEx

Account # _____ Exp. Date _____
 Signature _____

☐ I prefer to pay by check each month. My first monthly sponsorship contribution of \$24 is enclosed.

☐ Instead of becoming a sponsor at this time, I am enclosing a contribution of \$ ____.

☐ Please send me more information.

Information concerning Save the Children Education, Inc. is available through the Internet at www.savethechildren.org.
 505 Wilson Road, Wingham, CT 06095
 ©1995 Save the Children Education, Inc.

Save the Children.

Save the Children

Translating attitudes into actions Ads such as this encourage readers to put their money where their values are.

poor test performance, which reminded them strongly of their negative attitude toward failing a course but did not remind them of their negative attitude toward cheating. Subsequent research suggests that with more time to think about their attitudes, or with more inducement to think about them, fewer students would have cheated and a correlation would have been found between their anticheating attitudes and their behavior.

SECTION REVIEW

An attitude is a belief or opinion that includes an evaluative component.

Sources of Attitudes

- Attitudes can be created or altered through classical conditioning, with no thought required and even without awareness.
- By using heuristics (e.g., “If most people believe this, it is probably true”), people can arrive at attitudes through superficial thought.
- When a message is highly relevant to us, we tend to base our attitudes on logical analysis of the content.
- People may adopt attitudes that form the social norms of their group. It feels uncomfortable to hold attitudes that are different from those of others in one’s group.

Attitudes and Cognitive Dissonance

- We are motivated to reduce cognitive dissonance—a discomforting lack of accord among our beliefs, attitudes, and knowledge.
- The desire to prevent or reduce cognitive dissonance often leads people to avoid dissonant information and to set aside doubts about a decision once it has been made.
- When we freely and with little incentive do something contrary to an attitude, we may alter the attitude to better fit the action; this is called the insufficient-justification effect.

Attitudes and Behavior

- Early research suggested that attitudes do not affect behavior; more recently, we see that attitudes affect behavior under some conditions and not others.
- Implicit attitudes—those formed through direct experience or repeated associations—influence behavior automatically.
- Explicit attitudes must be brought to mind somehow before they can affect behavior.

Concluding Thoughts

To organize your thoughts about this chapter, you might find it useful to reflect on four themes that ran through it: the themes of bias, function, cultural differences, and implicit versus explicit mental processes.

1. Biases in social perceptions and attitudes can lead people to make judgments that are objectively untrue and unfair.

You might think about each bias discussed in the chapter with the following questions in mind: What evidence was presented for the existence of the bias? In what contexts does the bias seem to occur or not occur? What, if any, functions might the bias serve for the person manifesting it? What harm might result from this bias, either to the person who manifests it or to the objects of the biased perception or thought? You will not find the answers to all these questions for every bias described in the chapter, but in most cases you will at least find hints. Apply such questions to the person bias, actor–observer discrepancy, biasing effects of first impressions, attractiveness bias, baby-face bias, biasing effects of others’ appraisals on self-concept, big-fish-in-little-pond effect, self-enhancing biases (including the self-serving attributional bias), group-enhancing biases, biasing effects of stereotypes, and insufficient-justification effect.

2. Social perceptions and attitudes serve life-promoting functions for the individual. From the functionalist perspective, the first question asked about any consistent human ten-

dency is: How might it serve the needs of the individual who manifests that tendency? Some of the ideas in this chapter were presented quite explicitly from that perspective, and others were not. As you review each of the phenomena described in the chapter, think about whether and how it serves the individual. Concerning the biases in person perception, you might well conclude that some do serve useful functions for the perceiver and that others may be non-adaptive side effects of the ways that our minds work. Concerning self-perception, think about the potential survival-promoting values of the looking-glass self, self-esteem (according to the sociometer theory), the use of social comparison to understand oneself, the self-enhancing biases, and the capacity to shift between personal identity and social identity. Think also about the functions of attitudes for social acceptance as well as for guiding behavior.

3. Culture plays a powerful role in shaping our social perceptions and attitudes. The human mind evolved to meet the social needs of our hunter-gatherer ancestors, but each individual person’s mind is honed, through development, to meet the requirements of life within that person’s culture. In this chapter you have read about some consistent differences between Westerners and East Asians, including differences in susceptibility to self-enhancement and differences in the balance between personal identity and social identity. How might these differences be explained in terms of the different philosophical traditions of the West and the East?

4. Social perceptions and attitudes may be implicit and automatic or explicit and controlled. The distinction between implicit and explicit mental processes was most directly discussed in the sections on stereotypes and on the relationship between attitudes and behavior. You saw how implicit stereotypes and attitudes affect judgments and actions automatically and unconsciously, while explicit stereotypes and attitudes operate through conscious, deliberate means. The implicit-explicit distinction is relevant to all the topics of this chapter. All the biases affecting the impressions we form of other individuals—including the person bias in attributions, the attrac-

tiveness bias, and the baby-face bias—operate implicitly; we are usually unaware of their effects on our judgments and actions. You also read of implicit and explicit processes in the section dealing with the roles of classical conditioning, heuristics, and rational thought in attitude formation. As you review each of the social perceptual and attitudinal phenomena discussed in this chapter, think about the degree to which it may arise from implicit or explicit mental processes. How might we bring explicit processes to bear in reducing or countering the irrational and unfair implicit judgments that we make?

Further Reading

MALCOLM GLADWELL (2005). *Blink: The power of thinking without thinking*. Boston: Little, Brown.

This is a beautifully written brief book about the beneficial and harmful effects of our capacity to make quick, implicit, relatively unconscious judgments about people and other objects. Gladwell is a professional writer who has thoroughly familiarized himself with contemporary social psychological research. Here he relates research on implicit judgments to such practical issues as speed dating, market research, advertising, and racial prejudice.

MARK LEARY (2004). *The curse of the self: Self-awareness, egotism, and the quality of human life*. Oxford: Oxford University Press.

Our capacity for self-reflection is a large part of what makes us human. It allows us to think about and learn from our past successes and failures, to see ourselves somewhat as others see us, and to contemplate and plan our futures. However, that same capacity also contributes to human misery. Overemphasis on the self can lead to fruitless rumination, crippling self-conscious anxiety, egocen-

tric selfishness and narcissism, and life-threatening depression when the bubble of self-inflation breaks or fails to materialize. Leary is the social psychologist who developed the sociometer theory of self-esteem. In this book, for the general reader, he writes clearly and persuasively about the dark side of too much focus on the self and too little focus on the world outside ourselves.

MELINDA JONES (2002). *Social psychology of prejudice*. Upper Saddle River, NJ: Prentice Hall.

This brief, easy-to-read paperback textbook summarizes much of the social-psychological research on stereotyping and prejudice. It includes chapters on racism, sexism, and antigay prejudice; the relationship between values and prejudice; the cognitive foundations of categorization and stereotyping; intergroup comparison and conflict; the experiences of victims of prejudice; and methods of reducing prejudice.

LEWIS R. AIKEN (2002). *Attitudes and related psychosocial constructs: Theories, assessment, and research*. Thousand Oaks, CA: Sage.

The first three chapters of this brief, clear textbook deal with general theoretical and practical issues concerning attitudes, their measurement, and their formation. The remaining six chapters describe research and theories pertinent to attitudes in the following specific social domains: social prejudice and discrimination; living conditions and health; natural and unnatural death; schooling and students; work, leisure, and retirement; and politics, religion, and morality.

ROY BAUMEISTER (2005). *The cultural animal: Human nature, meaning, and social life*. Oxford: Oxford University Press.

In this intellectually stimulating book, Baumeister, an eminent contemporary social psychologist, develops the thesis that we have been cultural animals for a long time and that evolution has shaped us for culture. The book brings together much of what social scientists (economists, sociologists, and anthropologists as well as psychologists) have learned about human motivation, emotion, thought, and behavior.



Social Influences on Behavior

EFFECTS OF BEING OBSERVED AND EVALUATED

- Facilitating and Interfering Effects of an Audience
- Choking Under Pressure:
The Working-Memory Explanation
- Impression Management:
Behavior as Performance

EFFECTS OF OTHERS' EXAMPLES AND OPINIONS

- Asch's Classic Conformity Experiments
- Norms as Forces for Helpful and Harmful Actions
- Social Pressure in Group Discussions

EFFECTS OF OTHERS' REQUESTS

- Sales Pressure: Some Principles of Compliance
- Conditions That Promote Obedience:
Milgram's Experiments

TO COOPERATE OR NOT: THE DILEMMA OF SOCIAL LIFE

- The Tragedy of the Commons:
A Social-Dilemma Allegory
- The Logic of Social Dilemmas Exemplified in Games
- Conditions That Promote Cooperation
- Group Against Group: Lessons from Robbers Cave

EMOTIONAL FOUNDATIONS OF OUR SOCIAL NATURE

- Emotions Help Keep Us Together and Coordinate Our Actions
- Our Self-Conscious Emotions Make Us Socially Acceptable

A central theme of social psychology is that human behavior is influenced powerfully by the social situation in which it occurs. Think of the men who hijacked planes and smashed them into the World Trade Center towers in 2001; of the firefighters and others who rushed into those flaming towers and saved many lives; of the countless soldiers who have risked or given their lives on battlefields; of the soldiers who tortured and humiliated defenseless Iraqi prisoners at Abu Ghraib. Most people, at least most Westerners, tend to attribute such actions primarily to the inner character, or personality, of the people involved: Evil men flew into the towers or master-minded those flights, heroes rushed in to save people, heroes acted valiantly in battle, and "a few bad apples" tortured prisoners. When social psychologists think about such events, however, they think primarily of the social settings that surround the actors and prompt them to behave as they did (Fiske & others, 2004). They think of the social norms, the examples set by peers, the explicit or implicit orders or expectations of those in authority, the social identities of the actors, and the emotions that prompt people to behave as they do. How do such forces operate to cause normal people to behave in such a wide variety of ways?

This chapter is about the ways by which the social environment influences human behavior. It deals with such topics as evaluation anxiety, conformity, obedience, cooperation, interpersonal and intergroup conflict, and the roles of emotions in social behavior. A very general concept that runs through the chapter is that of **social pressure**, which can be defined as the entire set of psychological forces that are exerted on us by others' examples, judgments, expectations, and demands, whether real or imagined. At any given moment, we are most strongly influenced by those people who are physically or psychologically closest to us. If we believe that others are observing and judging us as we perform a task, we feel pressure to perform well, and this pressure can either improve or worsen our performance. If our peers signal through their actions or words that one way of behaving is normal and another is not, we feel pressured to behave in the way deemed normal. If close others or authority figures make requests or demands, we feel pressure to comply or obey.

Social pressure arises from the ways we interpret and respond emotionally to the social situations around us. It is a normal part of being human. Such pressure is useful because it promotes our social acceptance by others and helps create order and predictability in social interactions. But it can also lead us, in some situations, to behave in ways that are objectively foolish or morally repugnant.

Effects of Being Observed and Evaluated

Let us begin by considering what might seem to be the minimal form of social pressure—the mere presence of other people who can observe us perform. We do not behave in exactly the same way when others can see us as we do when we are alone. As thinking, social beings, we are concerned about the impressions we make on other people, and that concern influences our behavior.

Facilitating and Interfering Effects of an Audience

In some of the earliest experiments in social psychology, people performed various tasks better when one or more observers were present than they did when alone. In one experiment, for example, college students who had achieved skill at a task involving eye–hand coordination (moving a hand-held pointer to follow a moving target) subsequently performed it more accurately when observed by a group of graduate students than when tested alone (Travis, 1925). The enhancing effect of an audience on task performance was soon accepted as a general law of behavior and was given a name: **social facilitation**.

Other early experiments, however, demonstrated an opposite effect: **social interference** (also called *social inhibition*), a decline in performance when observers are present. For example, students who were asked to develop arguments opposing the views of certain classical philosophers developed better arguments when they worked alone than when they worked in the presence of observers (Allport, 1920). The presence of observers also reduced performance in solving math problems (Moore, 1917), learning a finger maze (Husband, 1931), and memorizing lists of nonsense syllables (Pressin, 1933).

Increased Drive as an Explanation of Social Facilitation and Interference

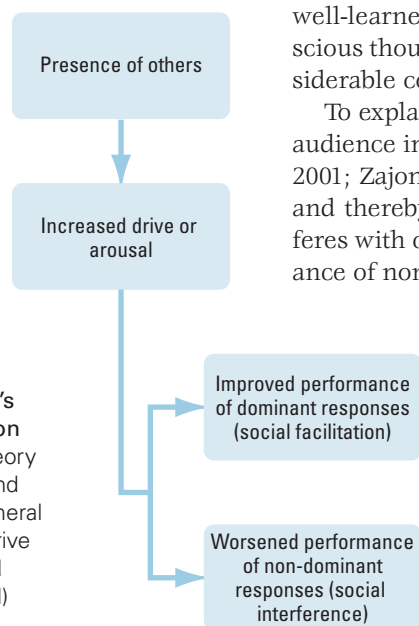
1 How does Zajonc's theory explain both social facilitation and social interference? What evidence supports the theory and suggests, further, that evaluation anxiety is involved?

Why does an audience sometimes improve a person's performance and other times worsen it? In reviewing the experiments, Robert Zajonc [pronounced **Zai-yons**] (1965) noticed that social facilitation usually occurred with relatively simple or well-learned tasks and that social interference usually occurred with complex tasks or tasks that involved new learning. From this observation, Zajonc proposed the following generalization: *The presence of others facilitates performance of dominant actions and interferes with performance of non-dominant actions*. In this statement, the term *dominant actions* refers to actions that are so simple, or instinctive, or well-learned that they can be produced automatically, with little conscious thought; and *non-dominant actions* refers to actions that require considerable conscious thought.

To explain both effects, Zajonc further proposed that the presence of an audience increases a person's level of drive and arousal (Aiello & Douthitt, 2001; Zajonc, 1965). The heightening of drive and arousal improves effort and thereby facilitates the performance of dominant actions, but it interferes with controlled, conscious thought and thereby worsens the performance of non-dominant actions (see Figure 14.1).

Much subsequent research has supported Zajonc's theory. The presence of observers does increase drive and arousal, as measured by self-reports and by physiological indices such as increased heart rate and muscle tension (Cacioppo & others, 1990; Zajonc, 1980). Other studies showed, just as Zajonc had predicted, that either facilitation or interference can occur in the very same task, depending on the performers' skill. In one experiment, for example, expert pool players performed better when they were watched conspicuously by a group of four observers than when they thought they were not

FIGURE 14.1 Zajonc's theory of social facilitation and interference This theory relates social facilitation and interference to a more general effect of high arousal or drive on dominant (habitual) and non-dominant (nonhabitual) responses.



being observed, and the opposite occurred for novice pool players (Michaels & others, 1982).

Zajonc did not explain why the presence of observers increases drive and arousal, but subsequent research showed, as you might expect, that the effect depends at least partly on the performers' concern about being evaluated. Social facilitation and interference decline when the audience is blindfolded or inattentive to the performer (Cottrell & others, 1968) and increase when the audience is high in status or expertise and is present explicitly to evaluate (Geen, 1980, 1991).

Choking Under Pressure: The Working-Memory Explanation

Social interference can be thought of as a subcategory of a more general phenomenon commonly referred to as *choking under pressure*. The highly aroused mental state produced by any strong form of pressure to perform well can, ironically, cause performance to worsen.

Using the terminology of modern cognitive psychology, “choking” is especially likely to occur with tasks that make strong demands on working memory. Working memory, as described in Chapter 9, is the part of the mind that controls conscious attention and holds, in consciousness, those items of information that are needed to solve a problem. The kinds of tasks that Zajonc referred to as involving non-dominant responses are, in general, tasks that make heavy demands on working memory. Pressure can worsen performance of such tasks by creating distracting thoughts—thoughts about being evaluated, about the difficulty of the task, about the consequences of failing, and so on—which usurp much of the limited capacity of working memory and thereby interfere with the person's ability to concentrate on the problem to be solved.

Choking on Academic Tests

Students who suffer from severe test anxiety report that distracting and disturbing thoughts flood their minds and interfere with their performance on important tests (Zeidner, 1998). With sufficient pressure, such choking can occur even in students who normally do not suffer from test anxiety, and researchers have found that it occurs specifically with test items that make the highest demands on working memory.

In one series of experiments, for example, students were given math problems that varied in difficulty and in the degree to which students had an opportunity to practice them in advance (Beilock & others, 2004). Pressure to perform well was manipulated by telling some students that they were part of a team and that if they failed to perform above a certain criterion neither they nor their teammates—who had already performed above the criterion—would win a certain prize. They were also told that they would be videotaped as they worked on the problems, so that math teachers and professors could examine the problem-solving process. Students in the low-pressure group, in contrast, were simply asked to solve the problems as best they could. The result was that the high-pressure group performed significantly worse than the low-pressure group on the unpracticed difficult problems, but not on the easy or thoroughly practiced problems, which were less taxing on working memory.

Stereotype Threat as a Special Cause of Choking

Stereotype threat, a particularly potent cause of choking on academic tests, was identified several years ago by Claude Steele (1997; Steele & others, 2002). In a series of experiments, Steele found that African American college students, but not white college students, performed worse on various tests if they were referred to as



Duomo / Corbis

Social facilitation When the performer is an expert, like Tiger Woods, the presence of an audience improves performance. Such an audience would worsen performance for a novice.

>>> 2
What evidence supports the view that choking on tests occurs because distracting thoughts interfere with working memory?

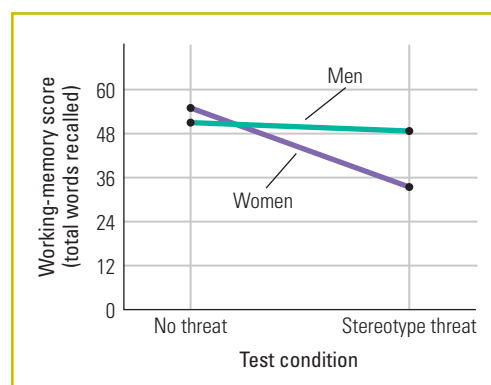
>>> 3
How can the activation of a stereotype influence test performance? What evidence suggests that this effect, like other forms of choking, involves interference with working memory?

intelligence tests than if the same tests were referred to by other labels. He found, further, that this drop in performance became even greater if the students were deliberately reminded of their race just before taking the test. As in the case of other examples of choking, stereotype threat seems to produce its effects by increasing anxiety and mental distraction (Cadinu & others, 2005; Wheeler & Petty, 2001). African Americans are painfully aware of the stereotype that they are less intelligent than whites, and activation of this stereotype during a test is a powerful source of anxiety and distraction.

Stereotype threat has since been demonstrated in other stigmatized groups as well. For example, women consistently perform worse on problem-solving tests that are referred to as math tests than on the same tests when given other names, and this effect is magnified if attention is drawn to the stereotype that women have lower aptitude for math than do men (Johns & others, 2005; Wheeler & Petty, 2001). Stereotype threat is one cause of what, in Chapter 13, were referred to as self-fulfilling prophecies. The stereotypes that blacks have low intelligence and that women are bad at math can, in themselves, reduce scores on intelligence tests for blacks and on math tests for women.

In one recent experiment, a particularly large effect of stereotype threat was demonstrated in women using a standard test of working-memory capacity (Schmader & Johns, 2003). Each trial of this test requires subjects to hold a list of words in mind while solving simple arithmetic problems and then to recall and report as many of the words as possible. In the experiment, men and women took this test under one of two conditions. In the *non-threat condition*, the test was described as a test of memory and no mention was made of gender differences. In the *stereotype-threat condition*, the test was described as a mathematics test and the students were informed that their gender might play a role in their performance. The results are shown in Figure 14.2. In the non-threat condition women and men performed equally well, as measured by the number of words recalled correctly, but in the stereotype-threat condition women performed far worse than men. The large

FIGURE 14.2 | Stereotype threat reduces working-memory score in women When the test of working memory was described as a math test and attention was drawn to the stereotype that women do poorly in math (stereotype-threat condition), women scored much lower than men. In the non-threat condition, women and men scored equally well. (Data from Schmader & Johns, 2003. The maximum score on the test was 72.)



stereotype-threat effect found with this test supports the theory that stereotype threat, like other causes of choking, entails a reduction in available working-memory capacity.

In case you yourself have suffered from stereotype threat, you may be interested to know that simple awareness of the stereotype-threat phenomenon helps people overcome it (Johns & others, 2005). Apparently, such awareness leads test takers to attribute their incipient anxiety to stereotype threat rather

than to the difficulty of the problems or to some inadequacy in themselves, and this helps them to concentrate on the problems rather than on their own fears. Other experiments suggest that thoughts about positive role models—people whose achievements counter the stereotype—can reduce or abolish stereotype threat (Marx & Roman, 2002; McIntyre & others, 2003).

Impression Management: Behavior as Performance

Besides influencing our abilities to perform specific tasks, social pressures influence our choices of what to say and do in front of other people. Because we care what others think of us, we strive to influence their thoughts. To that end, we behave differently when witnesses are present than when we are alone, and differently in front of some witnesses than in front of others. Social psychologists use the term *impression management* to refer to the entire set of ways by which

people consciously and unconsciously modify their behavior to influence others' impressions of them (Schlenker, 1980).

Humans as Actors and as Politicians

Poets and philosophers have always been aware of, and have frequently ridiculed, the human concern about appearances. As Shakespeare put it, "All the world's a stage, and all the men and women merely players." The sociologist Erving Goffman developed an entire approach to thinking about human behavior based on this metaphor. In a classic book titled *The Presentation of Self in Everyday Life*, Goffman (1959) portrayed us as actors, playing at different times on different stages to different audiences, always trying to convince our current audience that we are whom we are playing. In Goffman's view, we are not necessarily aware that we are performing. There need not be a division in our minds between the image we try to project and our sincere beliefs about ourselves (Schlenker & Pontari, 2000). At any given moment we may simply be trying to exhibit our best self, or those aspects of our self that seem most appropriate and useful to meet the moment's needs, which change with our audience.

An alternative metaphor, which captures long-range purposes of our performances, is that we are *intuitive politicians* (suggested by Tetlock, 1991, 2002). We perform in front of others not just to tell a good story or portray a character but to achieve real-life ends that may be selfish or noble, or to some degree both. To do what we want to do in life, we need the approval and cooperation of other people—their votes, as it were—and to secure those votes, we perform and compromise in various ways. We are *intuitive politicians* in that we campaign for ourselves and our interests quite naturally, often without consciousness of our political ingenuity and strategies.

Depending on our needs, our capacities, and our audience, we may at any given time portray ourselves as pitiful, enraged, stern, or even irrational and unpredictable. These can be effective strategies for certain ends. For some people these strategies may even become regular ploys. But most of us, most of the time, try to make ourselves look *good* to other people. We want to come across as attractive, friendly, competent, rational, trustworthy, and moral, because we know that others will be more inclined to collaborate with us if they see those qualities in us than if they don't. We also want to look *modest*, so people will think we are understating, not overstating, our virtues. And we want to look *sincere*, not as if we are putting on a show or trying to ingratiate ourselves. We may or may not be conscious of our delicate balancing act between showing off and appearing modest, or between sincerity and ingratiation, but the act requires effort nevertheless (Vohs & others, 2005).

Impressing Acquaintances More Than Close Friends

In general, we are more concerned with impression management with new acquaintances than with familiar friends and companions (Leary & Kowalski, 1995; Tice & others, 1995). That concern makes sense in light of our understanding that first impressions can have long-lasting effects (discussed in Chapter 13). We have less need to manage impressions with close friends because they already know us well. A slip will not so seriously harm our reputation with friends as it will with strangers, and friends may see through our act no matter how clever the performance.

Other research shows, not surprisingly, that dating partners are much more concerned with making good impressions on each other than are married partners. Dating partners rate themselves as most intimate and secure in their relationship when they have a highly favorable impression of each other, but married partners rate themselves as most intimate and secure when they have a "true" impression of each other—that is, when each sees the other as the other sees himself or herself (Swann & others, 1994). To feel comfortable in marriage, the two partners must feel that they can "be themselves" and not be rejected for it.

How do the theatrical and political metaphors portray the nature of impression management? Why do we, as intuitive politicians, want to look "good" to other people?



FIVE DISTINGUISHED PROFESSORS, EACH TRYING TO LOOK LIKE A MORE DISTINGUISHED PROFESSOR THAN THE OTHER FOUR

© Sidney Harris

SECTION REVIEW

Social pressure affects our performance and leads us to try to control how others see us.

Effects of Having an Audience

- The presence of an audience can produce either social facilitation (improvement in performance) or social interference (decline in performance).
- Zajonc theorized that the presence of others leads to heightened drive and arousal, which facilitates dominant tasks and interferes with non-dominant tasks. Much evidence supports this view.

Choking Under Pressure

- Pressure to perform well, on an academic test for example, can cause a decline in performance (choking).
- Choking occurs because pressure produces distracting thoughts that compete with the task itself for limited-capacity working memory.
- Stereotype threat is a powerful form of choking that occurs when members of a stigmatized group are reminded of stereotypes about their group before performing a relevant task.

Impression Management

- Because of social pressure, we consciously or unconsciously modify our behavior in order to influence others' perceptions of us.
- This tendency to manage impressions has led social scientists to characterize us as actors, playing roles, or as politicians, promoting ourselves and our agendas.

Effects of Others' Examples and Opinions

Other people influence our behavior not just through their roles as observers and evaluators but also through the examples they set. There are two general reasons why we tend to conform to others' examples.

5 What are two classes of reasons why people tend to conform to examples set by others?

<<< One reason has to do with information and pragmatics. If other people cross bridge A and avoid bridge B, they may know something about the bridges that we don't know. To be safe, we had better stick with bridge A too. If other people say rhubarb leaves are poisonous, then to be safe, in the absence of better information, we shouldn't eat them, and we should probably tell our children they are poisonous. One of the great advantages of social life lies in the sharing of information. We don't all have to learn everything from scratch, by trial and error. Rather, we can follow the examples of others and profit from trials and errors that may have occurred generations ago. Social influence that works through providing clues about the objective nature of an event or situation is referred to as **informational influence**.

The other general reason for conforming is to promote group cohesion and acceptance by the group. Social groups can exist only if some degree of behavioral coordination exists among the group members. Conformity allows a group to act as a coordinated unit rather than as a set of separate individuals. We tend to adopt the ideas, myths, and habits of our group because doing so generates a sense of closeness with others, promotes our acceptance by them, and enables the group to function as a unit. We all cross bridge A because we are the bridge A people, and proud of it! If you cross bridge B, you may look like you don't want to be one of us or you may look strange to us. Social influence that works through the person's desire to be part of a group or to be approved of by others is called **normative influence**.

Asch's Classic Conformity Experiments

Under some conditions, conformity can lead people to say or do things that are objectively ridiculous, as demonstrated in a famous series of experiments conducted by Solomon Asch in the 1950s. Asch's original purpose was to demonstrate the limits of conformity (Asch, 1952). Previous research had shown that people conform to others' judgments when the objective evidence is ambiguous (Sherif, 1936), and Asch expected to demonstrate that they would not conform when the evidence is clear-cut. But his results surprised him and changed the direction of his research.



© Hilary B. Price. Reprinted with permission of King Features Syndicate.

Basic Procedure and Finding

Asch's (1956) procedure was as follows: A college-student volunteer was brought into the lab and seated with six to eight other students, and the group was told that their task was to judge the lengths of lines. On each trial they were shown one standard line and three comparison lines and were asked to judge which comparison line was identical in length to the standard (see Figure 14.3). As a perceptual task, this was absurdly easy; in previous tests, subjects performing the task alone almost never made mistakes. But, of course, this was not really a perceptual task; it was a test of conformity. Unbeknownst to the real subject, the others in the group were confederates of the experimenter and had been instructed to give a specific wrong answer on certain prearranged "critical" trials. Choices were stated out loud by the group members, one at a time in the order of seating, and seating had been arranged so that the real subject was always the next to last to respond (see Figure 14.4). The question of interest was this: On the critical trials, would subjects be swayed by the confederates' wrong answers?

Of more than 100 subjects tested, 75 percent were swayed by the confederates on at least one of the 12 critical trials in the experiment. Some of the subjects conformed on every trial, others on only one or two. On average, subjects conformed on 37 percent of the critical trials. That is, on more than one-third of the trials on which the confederates gave a wrong answer, the subject also gave a wrong answer, usually the same wrong answer as the confederates had given. Asch's experiment has since been replicated dozens of times, in at least 17 different countries (Bond & Smith, 1996). The results reveal some decline in conformity in North America after the 1950s and some variation across cultures, but they still reveal a considerable amount of conformity whenever and wherever the experiment is conducted.

How did Asch demonstrate that a tendency to conform can lead people to disclaim the evidence of their own eyes?

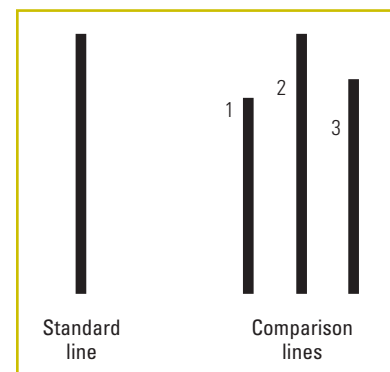


FIGURE 14.3 | Sample stimuli used by Asch to study conformity
The task on each trial was to select the comparison line that was identical in length to the standard. On critical trials, the confederates unanimously made a specific wrong choice (either 1 or 3, in this example). (Adapted from Asch, 1956.)



William Vandivert / Scientific American

FIGURE 14.4 | A perplexed subject in Asch's experiment It is not hard to tell who the real subject is in this photograph taken during a critical trial in one of Asch's experiments.

Was the Influence Informational or Normative?

7

What evidence led Asch to conclude that conformity in his experiments was caused more by normative than by informational influences?

<<<

Did Asch's subjects conform as a result of *informational* or *normative* influence? That is, did the subjects use the majority response as evidence regarding the objective lengths of the lines, or did they conform out of fear of looking different (non-normative) to the others present? When Asch (1956) questioned the subjects after the experiment, very few said that they had actually seen the lines as the confederates had seemed to see them, but many said that they had been led to doubt their own perceptual ability. They made such comments as: "I thought that maybe because I wore glasses there was some defect"; "At first I thought I had the wrong instructions, then that something was wrong with my eyes and my head"; and "There's a greater probability of eight being right [than one]." Such statements suggest that to some degree the subjects did yield because of informational influence; they may have believed that the majority was right.

But maybe these comments were, to some degree, rationalizations. Maybe the main reasons for conformity had more to do with a desire to be liked or accepted by the others (normative influence) than with a desire to be right. To test this possibility, Asch (1956) repeated the experiment under conditions in which the confederates responded out loud as before, but the subjects responded privately in writing. To accomplish this, Asch arranged to have the real subjects arrive "late" to the experiment and be told that although no more subjects were needed, they might participate in a different way by listening to the others and then writing down, rather than voicing aloud, the answer they believed to be correct. In this condition, the amount of conformity dropped to about one-third of that in the earlier experiments. This indicates that the social influence on Asch's original subjects was partly informational but mostly normative. When subjects did not have to respond publicly, so that there was no fear of appearing odd to the other subjects, their degree of conformity dropped sharply.

Experiments following Asch's indicate that conformity is most often motivated by normative influences when the perceptual task is easy (as in Asch's original experiment) and by informational influences when the task is difficult or ambiguous (Campbell & Fairey, 1989; Campbell & others, 1986). With an apparently easy task, people

expect everyone to give the same response, so failure to side with the majority could be judged by others as a sign of abnormality or perverseness. With a difficult or ambiguous task, people expect responses to vary, so there is less stigma attached to being different; in that case the primary motive for conformity stems from the assumption that the others may be right.

The Liberating and Thought-Provoking Influence of a Nonconformist

8

What valuable effects can a single nonconformist have on others in the group?

<<<

When Asch (1956) changed his procedure so that a single confederate gave a different answer from the others, the amount of conformity on the line-judging task dropped dramatically—to about one-fourth of that in the unanimous condition. This effect occurred regardless of how many other confederates there were (from 2 to 14) and regardless of whether the dissenter gave the right answer or a different wrong answer from the others. Any response that differed from the majority encouraged the real subject to resist the majority's influence and to give the correct answer.

Since Asch's time, other experiments, using more difficult tasks or problems than Asch's, have shown that a single dissension from the majority can have beneficial effects not just through reducing normative pressure to conform but also through informational means, by shaking people out of their complacent view that the majority must be right (Nemeth, 1986). When people hear a dissenting opinion, they become motivated to examine the evidence more closely, and that can lead to a better solution.



Norms as Forces for Helpful and Harmful Actions

In Asch's experiments, the social context was provided by the artificial situation of a group of people misjudging the length of a line. In everyday life, the social context consists not just of what we hear others say or see them do, but also of the various telltale signs that inform us implicitly about which behaviors are normal and which are not in the setting in which we find ourselves. The norms established by such signs can have serious consequences for all of us.

The “Broken Windows” Theory of Crime

During the 1990s, New York City saw a precipitous drop in crime. The rates of murders and felonies dropped, respectively, to one-third and one-half of what they had been in the 1980s. A number of social forces contributed to the change, no doubt, but according to some analyses, at least some of the credit goes to the adoption of new law-enforcement policies recommended by George Kelling, a criminologist who was hired by the city as a consultant (Gladwell, 2000).

Kelling was already known as a developer of the *broken windows theory of crime* (Kelling & Coles, 1996). According to this theory, crime is encouraged by physical evidence of chaos and lack of care. Broken windows, litter, graffiti, and the like send signals that disrespect for law, order, and the rights of residents is normal. People regularly exposed to such an environment can develop a “law of the jungle” mentality, which leads not just to more petty crime but to thefts and murders as well. By physically cleaning up the subways, streets, and vacant buildings, and by cracking down on petty crime, New York City authorities helped create a social environment that signaled that law and order are normal and lawbreaking is abnormal. That signal may have helped to reduce crimes of all sorts, ranging from littering to murder. People are motivated to behave in ways they see as normal.

>>> **9**
According to the “broken windows” theory of crime, how do normative influences alter the crime rate?

Effects of Implicit Norms in Public-Service Messages

Many public-service messages include statements about the large number of people who engage in some undesirable behavior, such as smoking, drunk driving, or littering. According to Robert Cialdini, an expert on persuasion, such messages may undermine themselves. At the same time that they are urging people not to behave in a certain way, they are sending the implicit message that behaving in that way is *normal*—many people *do* behave in that way. Cialdini suggests that public-service messages would be more effective if they emphasized that the majority of people behave in the desired way, not the undesired way, and implicitly portrayed the undesired behavior as abnormal.

As a demonstration of this principle, Cialdini (2003) developed a public-service message designed to increase household recycling, and aired it on local radio and TV stations in four Arizona communities. The message depicted a group of people all recycling and speaking disapprovingly of a lone person who did not recycle. The result was a 25 percent increase in recycling in those communities—a far bigger effect than is usually achieved by public-service messages.

In another study, Cialdini (2003) and his colleagues created two signs aimed at decreasing the pilfering of petrified wood from Petrified Forest National Park. One sign read, “Many past visitors have removed petrified wood from the Park, changing the natural state of the Petrified Forest,” and depicted three visitors taking petrified wood. The other sign read, “Please do not remove the petrified wood from the Park, in order to preserve the natural state of the Petrified Forest,” and depicted a single visitor taking a piece of wood, with a red circle-and-bar symbol superimposed over his hand. On alternate weekends, one or the other of these signs was placed near the beginning of each path in the park. To measure theft, marked pieces of petrified wood were placed along each path. The dramatic result was that 7.92 percent of visitors tried to steal marked pieces on the days when the first sign was present, compared with only 1.67 percent on the days when the second sign was present. Previous research had shown that, with neither of these signs present

>>> **10**
According to Cialdini, how can public-service messages best capitalize on normative influences? What evidence does he provide for this idea?

but only the usual park injunctions against stealing, approximately 3 percent of visitors stole petrified wood from the park. Apparently, by emphasizing that many people steal, the first sign *increased* the amount of stealing sharply above the baseline rate; and the second sign, by implying that stealing is rare as well as wrong, decreased it to well below the baseline rate.

Conformity as a Basis for Failure to Help

A man lies ill on the sidewalk in full view of hundreds of passersby, all of whom fail to stop and ask if he needs help. A woman is brutally beaten in front of witnesses who fail to come to her aid or even to call the police. How can such incidents occur?

In many experiments, social psychologists have found that a person is much more likely to help in an emergency if he or she is the only witness than if one or more other witnesses are also present (Latané & Nida, 1981). In one experiment, for example, college students filling out a questionnaire were interrupted by the sound of the researcher, behind a screen, falling and crying out, “Oh . . . my foot . . . I . . . can’t move it; oh . . . my ankle . . . I can’t get this thing off me” (Latané & Rodin, 1969). In some cases the student was alone, and in other cases two students sat together filling out questionnaires. The remarkable result was that 70 percent of those who were alone went to the aid of the researcher, but only 20 percent of those who were in pairs did so. Apparently an accident victim is better off with just one potential helper present than with two! Why? Part of the answer probably has to do with diffusion of responsibility. The more people present, the less any one person feels it is his or her responsibility to help (Schwartz & Gottlieb, 1980). But conformity also seems to contribute.

11

How can the failure of multiple bystanders to help a person in need be explained in terms of informational and normative influences?

◀◀ If you are the only witness to an incident, you decide whether it is an emergency or not, and whether you can help or not, on the basis of your assessment of the victim’s situation. But if other witnesses are present, you look also at them. You wait just a bit to see what they are going to do, and chances are you find that they do nothing (because they are waiting to see what you are going to do). Their inaction is a source of information that may lead you to question your initial judgment: Maybe this is not an emergency, or if it is, maybe nothing can be done. Their inaction also establishes an implicit social norm. If you spring into action, you might look foolish to the others, who seem so complacent. Thus, each person’s inaction can promote inaction in others through both informational and normative influences.

These interpretations are supported by experiments showing that the inhibiting effect of others’ presence is reduced or abolished by circumstances that alter the informational and normative influences of the other bystanders. If the bystanders indicate, by voice or facial expressions, that they *do* interpret the situation as an emergency, then their presence has a much smaller or no inhibiting effect on the target person’s likelihood of helping (Bickman, 1972). If the bystanders know one another well—and thus have less need to manage their impressions in front of one another, or know that each shares a norm of helping—they are more likely to help than if they don’t know one another well (Rutkowski & others, 1983; Schwartz & Gottlieb, 1980).



David McNew / Liaison / Newsmakers

Passing by The presence of other witnesses decreases each witness’s likelihood of helping.

social pressure. Such pressure can occur whenever one person expresses an opinion or takes a position on an issue in front of another: Are you with me or against me? It feels good to be with, uncomfortable to be against. There is unstated pressure to agree.

Social Pressure in Group Discussions

When people get together to discuss an idea or make a decision, their explicit goal usually is to share information. But whether they want to or not, group members also influence one another through normative



Jim Bourg / Reuters / Landov

Polarized groups When people talk primarily with others who agree with them, they come to hold their original opinions more strongly and fervently than before. As a result, different groups, who start out with different opinions, can become so polarized in their views that they begin to see each other as enemies. Here, in front of the Massachusetts State House, we see such group polarization centered on the question of gay marriage.

Group Discussion Can Make Attitudes More Extreme

When a group is evenly split on an issue, the result is often a compromise (Burnstein & Vinokur, 1977). Each side partially convinces the other, so the majority leave the room with a more moderate view on the issue than they had when they entered. However, if the group is not evenly split—if all or a large majority of the members argue on the same side of the issue—discussion typically pushes that majority toward a more extreme view in the same direction as their initial view. This phenomenon is called **group polarization**.

Group polarization has been demonstrated in many experiments, with a wide variety of problems or issues for discussion. In one experiment, mock juries evaluated traffic-violation cases that had been constructed to produce either high or low initial judgments of guilt. After group discussion the jurors rated the high-guilt cases as indicating even higher levels of guilt, and the low-guilt cases as indicating even lower levels of guilt, than they had before the discussion (Myers & Kaplan, 1976). In other experiments, researchers divided people into groups on the basis of their initial views on a controversial issue and found that discussions held separately by each group widened the gaps between the groups (see Figure 14.5).

>>> **12**
How has group polarization been demonstrated?

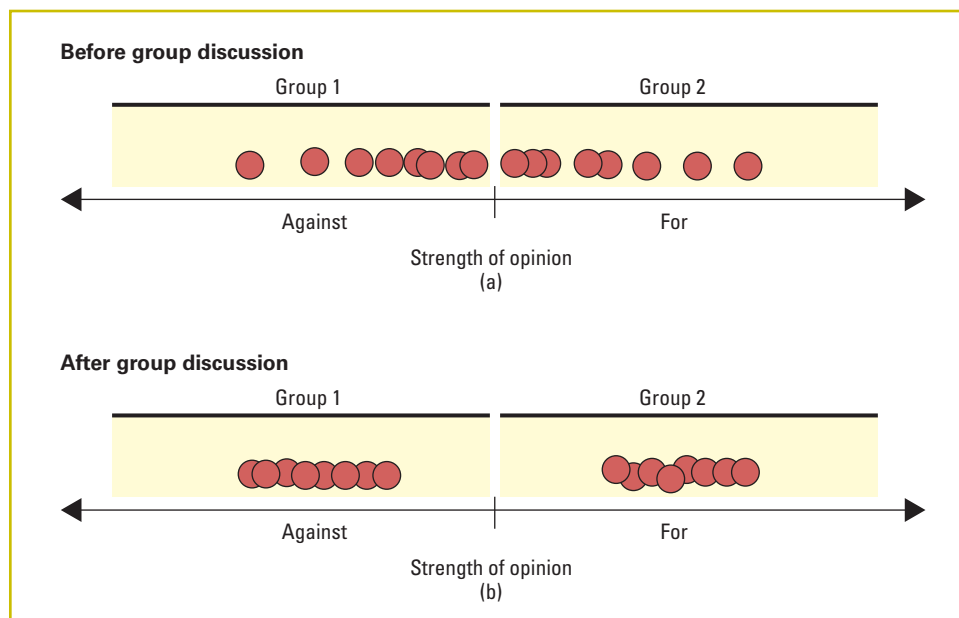


FIGURE 14.5 Schematic illustration of group polarization Each circle represents the opinion of one individual. When the individuals are divided into two groups on the basis of the direction of their initial position (a) and then discuss the issue with other members of their group, most shift toward a more extreme position than they held before (b).

In one experiment, for example, discussion caused groups favoring a strengthening of the military to favor it even more strongly and groups favoring a paring down of the military to favor that more strongly (Minix, 1976; Semmel, 1976).

Group polarization can have socially important consequences. When students whose political views are barely to the right of center get together to form a Young Conservatives club, their views are likely to shift further toward the right; similarly, a Young Liberals club is likely to shift to the left. Prisoners who enter jail with little respect for the law and spend their time talking with other prisoners who share that view are likely to leave prison with even less respect for the law than they had before. Systematic studies of naturally occurring social groups suggest that such shifts are indeed quite common (Sunstein, 2003).

What Causes Group Polarization?

13

How might group polarization be explained in terms of (a) informational and (b) normative influences?

<<< By now it should be no surprise to you that social psychologists have proposed two classes of explanations for group polarization: informational and normative. *Informational* explanations focus on the pooling of arguments that occurs during group discussion. People vigorously put forth arguments favoring the side toward which they lean and tend to withhold arguments that favor the other side. As a result, the group members hear a disproportionate number of arguments on the side of their initial leaning, which may persuade them to lean further in that direction (Kaplan, 1987; Vinokur & Burnstein, 1974). Moreover, simply hearing others repeat one's own arguments in the course of discussion can have a validating effect. People become more convinced of the soundness of their own logic and the truth of the "facts" they know when they hear the logic and facts repeated by another person (Brauer & others, 1995).

Normative explanations attribute group polarization to people's concerns about being approved of by other group members. One might expect normative influences to cause opinions within a group to become more similar to one another, but not more extreme. In fact, as illustrated in Figure 14.5, they do become more similar as well as, on average, more extreme. Several normative hypotheses have been offered to explain why opinions within a like-minded group become more extreme. According to one, which we can call the *one-upmanship hypothesis*, group members vie with one another to become the most vigorous supporter of the position that most people favor, and this competition pushes the group as a whole toward an increasingly extreme position (Levinger & Schneider, 1969; Myers, 1982). According to another hypothesis, which we can call the *group differentiation hypothesis*, people who see themselves as a group often exaggerate their shared group opinions as a way of clearly distinguishing themselves from other groups (Hogg & others, 1990; Keltner & Robinson, 1996). The Young Liberals become more liberal to show clearly that they are not like those hard-hearted conservatives in the other club, and the Young Conservatives become more conservative to show that they are not like those soft-headed liberals.

As is so often the case in psychology, the hypotheses developed by different researchers are all probably valid, to differing degrees, depending on conditions. Group polarization in most instances probably results from a combination of normative and informational influences.

Conditions That Lead to Good or Bad Group Decisions

Decisions made by groups are sometimes better and sometimes worse than decisions made by individuals working alone. To the degree that the group decision arises from the sharing of the best available evidence and logic, it is likely to be better (Surowiecki, 2004b). To the degree that it arises from shared misinformation, selective withholding of arguments on the less-favored side, and participants' attempts to please or impress one another rather than to arrive at the best decision, the group decision is likely to be worse than the decision most group members would have made alone.

In a now-classic book titled *Groupthink*, Irving Janis (1982) analyzed some of the most misguided policy decisions made in the U.S. White House. Among them were the decisions to sponsor the Bay of Pigs invasion of Cuba in 1961 (an invasion that failed disastrously); to escalate the Vietnam War during the late 1960s; and to cover up the Watergate burglary in the early 1970s. Janis contends that each of these decisions came about because a tightly knit clique of presidential advisers, whose principal concerns were upholding group unity and pleasing their leader (Presidents Kennedy, Johnson, and Nixon, respectively), failed to examine critically the choice that their leader seemed to favor and instead devoted their energy to defending that choice and suppressing criticisms of it.

To refer to such processes, Janis coined the term **groupthink**, which he defined as “a mode of thinking that people engage in when they are deeply involved in a cohesive ingroup, when the members’ striving for unanimity overrides their motivation to realistically appraise alternative courses of action.” More recently, many other ill-advised decisions, including the decision by NASA to launch the U.S. space shuttle *Challenger* in below-freezing weather (Moorhead & others, 1991) and the endorsements by corporate boards of shady accounting practices, which led to the recent downfalls of Enron and several other large corporations, have been attributed to groupthink (Postmes & others, 2001; Surowiecki, 2004a).

A number of experiments conducted both before and after Janis published his groupthink model have affirmed it. In one, actual bomber crews—each consisting of a pilot, a navigator, and a gunner who were used to working together—were given logic problems to solve individually or as a group. As individuals the navigators were best at solving the problems, but working in a group the navigators were relatively ineffective because both they and the gunners deferred to the pilots, who held the highest status in the crew (Torrance, 1954). Other experiments have shown that the ability of groups to solve problems and make effective decisions is improved if (a) the leaders refrain from advocating a view themselves and instead encourage group members to present their own views and challenge one another (Leana, 1985; Neck & Moorhead, 1995) and (b) the groups focus on the problem to be solved rather than on developing group cohesion (Mullen & others, 1994; Quinn & Schlenker, 2002).

14 How did Janis explain some White House policy blunders with his groupthink theory? How can the tendency toward groupthink be countered?



A. Tannenbaum / Sigma

The *Challenger* explosion The flawed decision-making that Irving Janis called *groupthink* has been implicated in the 1986 explosion of the U.S. space shuttle *Challenger*. In striving for unanimity in the decision to launch, managers ignored engineers’ warnings about the dangers of launching after a night of freezing temperatures.

SECTION REVIEW

The opinions and examples of others can have strong effects on what we say and do.

Conformity Experiments

- Asch found that subjects frequently conformed to others’ stated judgments about the lengths of lines, even if it meant contradicting the evidence of their own eyes.
- In such experiments, the cause of conformity is usually normative (motivated by a desire to be accepted) when the task is easy and informational (motivated by a desire to be correct) when the task is difficult.
- Even a single dissenting opinion can sharply reduce the tendency of others to conform.

Helpful and Harmful Norms

- According to the broken windows theory of crime, cues suggesting that disrespect for the law is normal can lead to an escalation in crime.
- Public-service messages can be made more effective if they portray the undesirable behavior (e.g., smoking or pilfering) as abnormal—that is, non-normative.
- The more bystanders present at an emergency, the less likely any of them are to help. This may result from both informational and normative influences.

Social Pressure Within Groups

- When most of or all the people in a group initially agree, group discussion typically moves them toward a more extreme version of their initial view—a phenomenon called group polarization.
- Both informational and normative influences may contribute to group polarization.
- Groupthink occurs when group members are more concerned with group cohesion and unanimity than with genuine appraisal of various approaches to a problem. When this occurs at high levels of government or business, it can lead to disastrous decisions.

Effects of Others' Requests

One of the least subtle yet most potent forms of social influence is the direct request. If the request is small and made politely, we tend to comply automatically (Langer & others, 1978), and this tendency increases when our attention is otherwise occupied (Cialdini, 1993). But even if the request is onerous or offensive, people often find it hard to look a requester in the eye and say no. The tendency to comply usually serves us well. Most requests are reasonable, and we know that in the long run doing things for others pays off, as others in turn do things for us. But there are people who—out of selfishness, or because their jobs demand it, or because they are working for causes in which they sincerely believe—will exploit our tendency to comply. It is useful to know their techniques so that, instead of succumbing to pressure, we give when we want to give and buy what we want to buy.

Sales Pressure: Some Principles of Compliance

Robert Cialdini (1987, 2001) is a social psychologist who has devoted more than lip service to the idea of combining real-world observations with laboratory studies. To learn about compliance from the real-world experts, he took training in how to sell encyclopedias, automobiles, and insurance; infiltrated advertising agencies and fund-raising organizations; and interviewed recruiters, public-relations specialists, and political lobbyists. He learned their techniques, extracted what seemed to be basic principles, and tested those principles in controlled experiments. The following paragraphs describe a sample of compliance principles taken largely from Cialdini's work but also much studied by other social psychologists.

Cognitive Dissonance as a Force for Compliance

Chapter 13 contains an extensive discussion of the theory of cognitive dissonance. The basic idea of the theory is that people are made uncomfortable by contradictions among their beliefs, or between their beliefs and their actions, and their discomfort motivates them to change their beliefs or actions to maintain consistency. According to Cialdini's analysis, a number of standard sales tricks make use of cognitive dissonance to elicit compliance. Here are two of them.

15

How can the low-ball sales technique be explained in terms of cognitive dissonance? What evidence supports this explanation?

➤➤➤ **Throwing the Low Ball: Increasing the Price After Commitment to Buy** One of the most underhanded sales tricks is the **low-ball technique**. The essence of this technique is that the customer first agrees to buy a product at a low price and then, after a delay, the salesperson “discovers” that the low price isn't possible and the product must be sold for more. Experiments conducted by Cialdini and others (1978) suggest that the trick works because customers, after agreeing to the initial deal, are motivated to reduce cognitive dissonance by setting aside any lingering doubts they may have about the product. During the delay between the low-ball offer and the real offer, they mentally exaggerate the product's value; they set their heart on the house, or car, or ice cream treat that they had agreed to purchase. Having done this, they are now primed to pay more than they would have initially.

Consistent with this interpretation, researchers have found that the low-ball technique works only if the customer makes a verbal commitment to the original, low-ball deal (Burger & Cornelius, 2003). If the initial offer is withdrawn and a worse offer is proposed *before* the person agrees verbally to the initial offer, the result is a *reduction* in compliance rather than an increase. With no verbal commitment, the customer has no need to reduce dissonance, no need to exaggerate mentally the value of the purchase, and in that case the low-ball offer just makes the final, real offer look bad by comparison.

Putting a Foot in the Door: Making a Small Request to Prepare the Ground for a Large One With some chagrin, I can introduce this topic with a true story in which I was outwitted by a clever gang of driveway sealers. While I was raking leaves in front of my house, these men pulled up in their truck and asked if they

could have a drink of water. I, of course, said yes; how could I say no to a request like that? Then they got out of the truck and one said, “Oh, if you have some lemonade or soda, that would be even better; we’d really appreciate that.” Well, all right, I did have some lemonade. As I brought it to them, one of the men pointed to the cracks in my driveway and commented that they had just enough sealing material and time to do my driveway that afternoon, and they could give me a special deal. Normally, I would never have agreed to a bargain like that on the spot; but I found myself unable to say no. I ended up paying far more than I should have, and they did a very poor job. I had been taken in by what I now see to be a novel twist on the foot-in-the-door sales technique.

The basis of the **foot-in-the-door technique** is that people are more likely to agree to a large request if they have already agreed to a small one. The driveway sealers got me twice on that: Their request for water primed me to agree to their request for lemonade, and their request for lemonade primed me to agree to their offer to seal my driveway. Cialdini (1987) has argued that the foot-in-the-door technique works largely through the principle of cognitive dissonance. Having agreed, apparently of my own free will, to give the men lemonade, I must have justified that action to myself by thinking, *These are a pretty good bunch of guys*, and that thought was dissonant with any temptation I might have had a few moments later, when they proposed the driveway deal, to think, *These people may be overcharging me*. So I pushed the latter thought out of my mind before it fully registered.

In situations like my encounter with the driveway sealers, the foot-in-the-door technique may work because compliance with the first request induces a sense of trust, commitment, or compassion toward the person making that request. In other situations it may work by inducing a sense of commitment toward a particular product or cause (Burger, 1999). The technique has proved to be especially effective in soliciting donations for political causes and charities. People who first agree to make a small gesture of support, such as by signing a petition or giving a few minutes of their time, are subsequently more willing than they otherwise would be to make a much larger contribution (Bell & others, 1994; Cialdini, 2001; Freedman & Fraser, 1966). Apparently the small donation leads the person to develop a firmer sense of support for the cause—“I contributed to it, so I must believe in it”—which in turn promotes willingness to make a larger donation.

The Reciprocity Norm as a Force for Compliance

Anthropologists and sociologists have found that people all over the world abide by a **reciprocity norm** (Gouldner, 1960; Whatley & others, 1999). That is, people everywhere feel obliged to return favors. This norm is so ingrained that people may even feel driven to reciprocate favors that they didn’t want in the first place. Cialdini (1993) suggests that this is why the technique known as *pregiving*—such as pinning a flower on the lapel of an unwary stranger before asking for a donation or giving a free bottle of furniture polish to a potential vacuum-cleaner customer—is effective. Having received the gift, the victim finds it hard to turn away without giving something in return.

Notice that pregiving works through a means that is opposite to that proposed for the foot-in-the-door technique. The foot-in-the-door target is first led to make a small contribution, which induces a sense of commitment and thereafter a larger contribution. The reciprocity target, in contrast, is first presented with a gift, which leads to a felt need to give something back. If the two procedures were combined, they should tend to cancel each other out. The contribution would be seen as payment for the gift, reducing any further need to reciprocate, and the gift would be seen as justification for the contribution, reducing cognitive dissonance and thereby reducing the psychological drive to become more committed to the cause.

>>> **16**
How can the foot-in-the-door sales technique be explained in terms of cognitive dissonance?



Bordán Hrynewych / Stock, Boston

A foot in the door If the homeowner here agrees to the first request, to sign a petition, she may then find it hard to turn down a second, larger request, to donate time or money to the cause.

>>> **17**
How can the effectiveness of pregiving be explained in terms of the reciprocity norm?

>>> **18**
Logically, why should the pregiving and foot-in-the-door techniques be ineffective if combined? What evidence supports this logic?

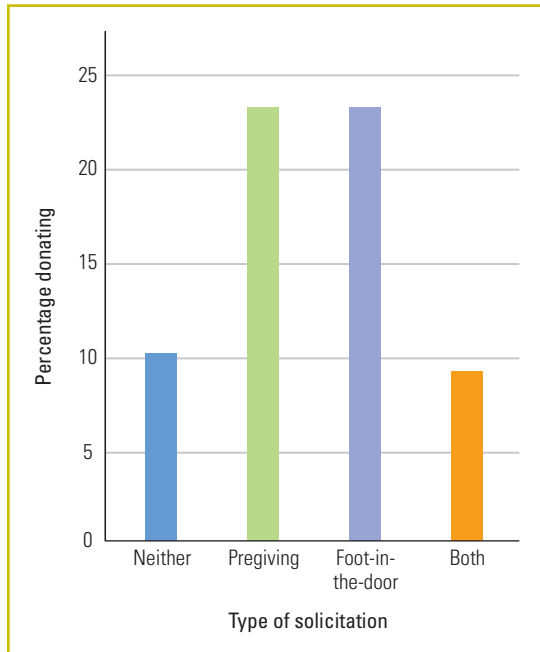


FIGURE 14.6 Evidence that the pregiving and foot-in-the-door techniques work in ways that oppose each other. In this field experiment, involving actual door-to-door solicitation to support an AIDS foundation, the pregiving and foot-in-the-door techniques were effective when each was used alone, but the two together were ineffective. (Data from Bell & others, 1994.)



ChevWilkinson / Getty Images

The reciprocity norm at Christmas The felt need to reciprocate can make the Christmas season quite burdensome. When we receive Christmas cards from others, we feel compelled to send cards back to them, so each year we send more and more cards. For many people, the same principle also applies to the exchange of Christmas gifts.

An experiment involving actual door-to-door solicitation of funds for a local AIDS foundation showed that in fact the two techniques do cancel each other out (Bell & others, 1994). Each person who was solicited first heard the same standard spiel about the foundation's good work. Some were then immediately asked for a donation (*control condition*). Others were given an attractive brochure containing "life-saving information about AIDS" (*pregiving condition*), or were asked to sign a petition supporting AIDS education (*foot-in-the-door condition*), or both (*combined condition*), before they were asked for a donation. As shown in Figure 14.6, the pregiving and foot-in-the-door techniques markedly increased the frequency of giving when either was used alone but had no effect beyond that of the control condition when the two were combined. This finding not only provides practical information for fund-raisers but also supports the proposed theories about the mechanisms of the two effects. If the two techniques operated simply by increasing the amount of interaction between the solicitor and the person being solicited, then the combined condition should have been most effective.

Shared Identity, or Friendship, as a Force for Compliance

19

What evidence suggests that even trivial connections between a sales person and potential customers can increase sales?

<<< Great salespeople are skilled at identifying quickly the things they have in common with potential customers and at developing a sense of friendship or connectedness: "What a coincidence, my aunt Millie lives in the same town where you were born." We automatically tend to like and trust people who have something in common with us or with whom we have enjoyed some friendly conversation, and we are inclined to do business with those people. In one series of experiments, compliance to various requests increased greatly when the targets of the requests were led to believe that the requester had the same birthday as they or the same first name or a similar-looking fingerprint (Burger & others, 2004). In other experiments, people were much more likely to comply with a request if they had spent a few minutes sitting quietly in the same room with the requester than if they had never seen that person before (Burger & others, 2001). Apparently, even a brief period of silent exposure makes the person seem familiar and therefore trustworthy.

Conditions That Promote Obedience: Milgram's Experiments

Obedience refers to those cases of compliance in which the requester is perceived as an authority figure or leader and the request is perceived as an order. Obedience is often a good thing. Obedience to parents and teachers is part of nearly everyone's social training. Running an army, an orchestra, a hospital, or any enterprise involving large numbers of people would be almost impossible if people did not routinely carry out the instructions given to them by their leaders or bosses. But obedience also has a dark side. Most tragic are the cases in which people obey a leader who is malevolent, unreasonable, or sadly mistaken. Cases in which people, in response to others' orders, carry out unethical or illegal actions have been referred to as *crimes of obedience* (Kellman & Hamilton, 1989).

Sometimes crimes of obedience occur because the order is backed up by threats; the subordinate's job or life may be at stake. In other cases, they occur because the subordinate accepts the authority's cause and doesn't interpret the action as wrong. In still other cases, however, people follow orders that they believe are wrong, even when there would be no punishment for disobeying. Those are the cases that interest us here, because they are the cases that must be understood in terms of psychological pressures. When social psychologists think of such cases, they relate them to a series of experiments performed by Stanley Milgram at Yale University in the early 1960s, which rank among the most famous of all experiments in social psychology. Milgram's goal was to identify some of the psychological pressures that underlie a person's willingness to follow a malevolent order. As you will see, these pressures include the social norm of obedience to authority figures, the apparent willingness of the authority figure to accept full responsibility, the difficulty of saying no to an authority who is immediately present, the absence of a model demonstrating that disobedience is legitimate, and the incremental nature of the objectionable demands that are made.

Milgram's Basic Procedure and Finding

To understand Milgram's experiments emotionally as well as intellectually, it is useful to imagine yourself as one of his subjects. You enter the laboratory and meet the experimenter and another person, who is introduced to you as a volunteer subject like yourself. The experimenter, a stern and expressionless man, explains that this is a study of the effects of punishment on learning and that one of you will serve as teacher and the other as learner. You draw slips of paper to see who will play which role and find that your slip says "teacher." The other subject, a pleasant middle-aged man, will be the learner. You watch while his arms are strapped to his chair and electrodes are taped to his wrist (see Figure 14.7). The experimenter explains that the straps will prevent excessive movement while the learner is shocked and that the electrode paste on the skin has been applied "to avoid blisters and burns." While he is being strapped in, the learner expresses some apprehension, saying that he is concerned because he has a heart condition.

After observing this part of the procedure, you—the teacher—are taken to an adjoining room, from which you can communicate with the learner through an intercom. Your job is to read off the questions on a test of verbal memory and to give the learner an electric shock whenever he gives a wrong answer. The shock generator in front of you has 30 switches, labeled with voltage designations from 15 to 450 volts. Additional labels next to



© Bart Nagel

He exposed real crimes of obedience Seymour Hersh, shown during an interview at the University of California at Berkeley, is widely known as the journalist who exposed the My Lai massacre and its cover-up during the Vietnam War. In April 2004, he was the first to write about U.S. soldiers' documentation of their own role in the torture and humiliation of Iraqi prisoners at Abu Ghraib Prison.

>>> 20

How did Milgram demonstrate that a remarkably high percentage of people would follow a series of orders to hurt another person?

From the film *Obedience* distributed by New York University Film Library

FIGURE 14.7 | The "learner" in Milgram's obedience experiments While being strapped into a chair and fitted with electrodes, this pleasant man—the "learner"—mentioned that he had a heart condition.

the switches describe the shocks as ranging from “Slight shock” to “Danger, severe shock,” followed by two switches labeled “XXX.”

As the experiment progresses, the learner makes frequent mistakes, and at each mistake you give him a stronger shock than you gave before. The learner receives the early shocks silently, but when you get to 75 volts, he responds with an audible “unghh,” and at stronger shocks his protests become more vehement. At 150 volts he cries out, “Experimenter, get me out of here! I won’t be in the experiment any more! I refuse to go on!” At 180 volts he hollers, “I can’t stand the pain!” By 270 volts his response to each shock is an agonized scream, and at 300 volts he shouts in desperation that he will no longer provide answers in the memory test. The experimenter instructs you to continue anyway and to treat each nonresponse as a wrong answer. At 315 and 330 volts the learner screams violently, and then, most frightening of all, from 345 volts on, the learner makes no sound at all. He does not respond to your questions, and he does not react to the shock.

At various points you look to the experimenter and ask if he should check on the learner or if the experiment should be terminated. You might even plead with the experimenter to let you quit giving shocks. At each of these junctures, the experimenter responds firmly with well-rehearsed prompts. First, he says, “Please continue.” If you still protest, he responds, “The experiment requires that you continue.” This is followed, if necessary, by “It is absolutely essential that you continue” and “You have no other choice; you must go on.” These prompts are always used in the sequence given. If you still refuse to go on after the last prompt, the experiment is discontinued.

In reality—as you, serenely reading this book, have probably figured out—the learner receives no shocks. He is a confederate of the experimenter, trained to play his role. But you, as a subject in the experiment, do not know that. You believe that the learner is suffering, and at some point you begin to think that his life may be in danger. What do you do? If you are like the majority of people, you will go on with the experiment to the very end and eventually give the learner the strongest shock on the board—450 volts, “XXX.” In a typical rendition of this experiment, 65 percent (26 out of 40) of the subjects continued to the very end of the series. They did not find this easy to do. Many pleaded with the experimenter to let them stop, and almost all of them showed signs of great tension, such as sweating and nervous tics, yet they went on.

Why didn’t they quit? There was no reason to fear retribution for halting the experiment. The experimenter, although stern, did not look physically aggressive. He did not make any threats. The \$5 pay for participating was so small as to be irrelevant; and all subjects had been told that the \$5 was theirs just for showing up. So why didn’t they quit?

Explaining the Finding

21

Why does Milgram’s finding call for an explanation in terms of the social situation rather than in terms of unique characteristics of the subjects?

◀◀ Upon first hearing about the results of Milgram’s experiment, people are tempted to suggest that the volunteers must have been in some way abnormal to give painful, perhaps deadly, shocks to a middle-aged man with a heart condition. But that explanation doesn’t hold up. The experiment was replicated dozens of times, using many different groups of subjects, and yielded essentially the same results each time. Milgram (1974) himself found the same results for women as for men and the same results for college students, professionals, and workers of a wide range of ages and backgrounds. Others repeated the experiment outside the United States, and the consistency from group to group was far more striking than the differences (A. G. Miller, 1986). No category of person has been found to be immune from the tendency to obey at a high rate in the Milgram experiment.

Another temptation is to interpret the results as evidence that people in general are sadistic. But nobody who has seen Milgram’s film of subjects actually giving the shocks would conclude that. The subjects showed no pleasure in what they were doing, and they were obviously upset by their belief that the learner was in pain. How, then, can the results be explained? By varying the conditions of the experi-

ment, Milgram (1974) and other social psychologists (A. G. Miller, 1986) identified a number of factors that contributed, in these experiments, to the psychological pressure to obey:

- **The norm of obedience to legitimate authorities** The volunteer comes to the laboratory as a product of a social world that effectively, and usually for beneficent reasons, trains people to obey legitimate authorities and to play by the rules. Social psychologists refer to this as the *norm of obedience* (Cialdini & Goldstein, 2004). An experimenter, especially one at such a reputable institution as Yale University, must surely be a legitimate authority in the context of the laboratory, a context that the subject respects but doesn't fully understand. Thus, the person enters the laboratory highly prepared to do what the experimenter asks. Consistent with the idea that the perceived legitimacy of the authority contributed to this obedience, Milgram found that when he moved the experiment from Yale to a downtown office building, under the auspices of a fictitious organization, Research Associates of Bridgeport, the percentage who were fully obedient dropped somewhat—from 65 to 48 percent. Presumably, it was easier to doubt the legitimacy of a researcher at this unknown office than that of a social scientist at Yale.
- **The experimenter's self-assurance and acceptance of responsibility** Obedience is predicated on the assumption that the person giving orders is in control and responsible and that your role is essentially that of a cog in a machine. The preexisting beliefs mentioned earlier helped prepare subjects to accept the cog's role, but the experimenter's unruffled self-confidence during what seemed to be a time of crisis no doubt helped subjects to continue accepting that role as the experiment progressed. To reassure themselves, they often asked the experimenter questions like "Who is responsible if that man is hurt?" and the experimenter routinely answered that he was responsible for anything that might happen. The importance of attributing responsibility to the experimenter was shown directly in an experiment conducted by another researcher (Tilker, 1970), patterned after Milgram's. Obedience dropped sharply when subjects were told beforehand that they, the subjects, were responsible for the learner's well-being.
- **The proximity of the experimenter and the distance of the learner** We can picture Milgram's subjects as caught between two conflicting social forces. On one side was the experimenter demanding that the experiment be continued, and on the other was the learner pleading that it be stopped. Not only did the experimenter have the greater initial authority, but he was also physically closer and perceptually more prominent. He was standing in the same room with the subject, while the learner was in another room, out of sight. To test the importance of physical closeness, Milgram (1974) varied the placement of the experimenter or the learner.

In one variation, the experimenter left the room when the experiment began and communicated with the subject by telephone, using the same verbal prompts as in the original study; in this case, only 23 percent obeyed to the end, compared with 65 percent in the original condition. In another variation, the experimenter remained in the room with the subject, but the learner was also brought into that room; in this case, 40 percent obeyed to the end. In still another variation, the subject was required to hold the learner's arm on the shock plate while the shock was administered (see Figure 14.8), with the result that 30 percent obeyed to the end. Thus, any change that moved the experimenter farther away from the subject, or the learner closer to the subject, tended to tip the balance away from obedience.
- **The absence of an alternative model of how to behave** Milgram's subjects were in a novel situation. Unlike the subjects in Asch's experiments, those in most variations of Milgram's experiment saw no other subjects who were in the same situation as they, so there were no examples of how

>>> **22**

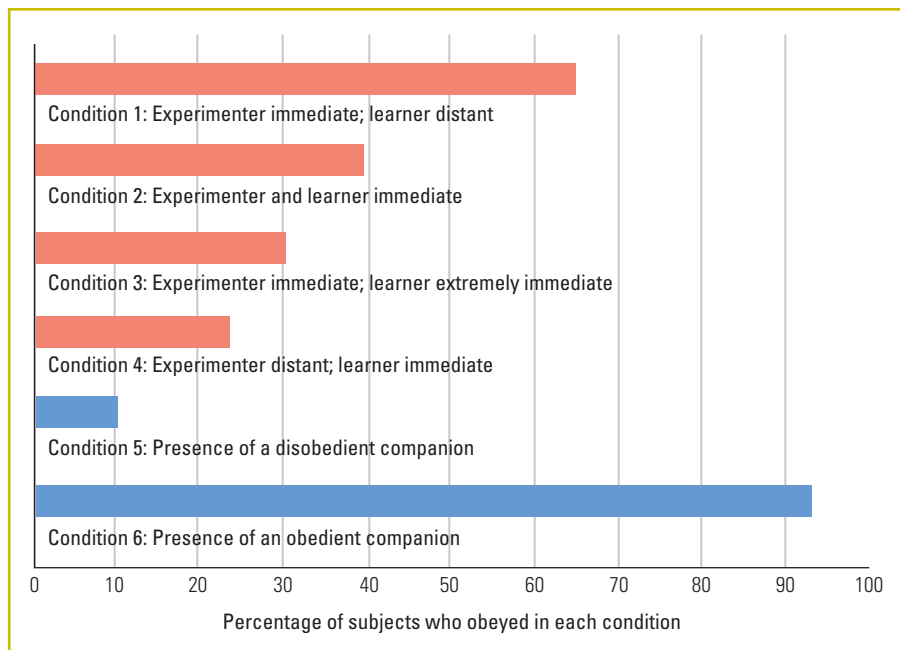
How might the high rate of obedience in Milgram's experiments be explained in terms of (a) the norm of obedience, (b) the experimenter's acceptance of responsibility, (c) the proximity of the experimenter, (d) the lack of a model for rebellion, and (e) the incremental nature of the requests?

FIGURE 14.8 Giving a shock while in close proximity to the learner In one of Milgram's experiments, subjects were required to hold the learner's arm on the shock plate each time a shock was given. Fewer obeyed in this condition than when the learner received shocks in another room, out of sight of the subjects.



From the film *Obedience* distributed by New York University Film Library

FIGURE 14.9 | Results of Milgram's obedience experiments In the first four conditions of the experiment, the greatest degree of obedience occurred when the experimenter and the subject were in the same room and the learner was out of sight in another room. Obedience dropped with decreased proximity of the experimenter and increased proximity of the learner. In conditions 5 and 6, the presence of a disobedient or obedient confederate, masquerading as another subject, dramatically altered the tendency to obey. In each condition, obedience is defined as continuing to give shocks, up through the highest shock level.



to respond to the experimenter's orders. In two variations, however, a model was provided in the form of another ostensible subject, actually a confederate of the experimenter, who shared with the real subject the task of giving shocks (Milgram, 1974). When the confederate refused to continue at a specific point and the experimenter asked the real subject to take over the whole job, only 10 percent of the real subjects obeyed to the end. When the confederate continued to the end, 93 percent of the real subjects did too. In an unfamiliar and stressful situation, having a model to follow has a potent effect. (For a summary of the results of the variations in Milgram's experiments, see Figure 14.9.)

- **The incremental nature of the requests** At the very beginning of the experiment, Milgram's subjects had no compelling reason to quit. After all, the first few shocks were very weak, and subjects had no way of knowing how many errors the learner would make or how strong the shocks would become before the experiment ended. Although Milgram did not use the term, we might think of his method as a very effective version of the foot-in-the-door technique. Having complied with earlier, smaller requests (giving weaker shocks), subjects found it hard to refuse new, larger requests (giving stronger shocks). The technique was especially effective in this case because each shock was only a little stronger than the previous one. At no point were subjects asked to do something radically different from what they had already done. To refuse to give the next shock would be to admit that it was probably also wrong to have given the previous shocks—a thought that would be dissonant with subjects' knowledge that they indeed had given those shocks.

Critiques of Milgram's Experiments

23

How has Milgram's research been criticized on grounds of ethics and real-world validity, and how has the research been defended?

<<< Because of their dramatic results, Milgram's experiments immediately attracted much attention and criticism from psychologists and other scholars.

The Ethical Critique Some of Milgram's critics focused on ethics (for example, Baumrind, 1964). They were disturbed by such statements as this one made in Milgram's (1963) initial report: "I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse." Was the study of sufficient scientific merit to warrant inflicting such stress on subjects, leading some to believe that they might have killed a man?

Milgram took great care to protect his subjects from psychological harm. Before leaving the lab, they were fully informed of the real nature and purpose of the experiment; they were informed that most people in this situation obey the orders to the end; they were reminded of how reluctant they had been to give shocks; and they were reintroduced to the learner, who offered further reassurance that he was fine and felt well disposed toward them. In a survey made a year after their participation, 84 percent of Milgram's subjects said they were glad to have participated, and fewer than 2 percent said they were sorry (Milgram, 1964). Psychiatric interviews of 40 of the former subjects revealed no evidence of harm (Errera, 1972). Still, because of concern for possible harm, such research would not be approved today by the ethics review boards at most research institutions.

The Question of Generalizability to Real-World Crimes of Obedience Other critics have suggested that Milgram's results may be unique to the artificial conditions of the laboratory and have little to tell us about crimes of obedience in the real world, such as the Nazi Holocaust (Miller, 2004). Some of these argued that Milgram's subjects must have known, at some level of their consciousness, that they could not really be hurting the learner, because no sane experimenter would allow that to happen (Orne & Holland, 1968). From that perspective, the subjects' real conflict may have been between the belief that they weren't hurting the learner and the possibility that they were. Unlike subjects in Milgram's experiments, Nazis who were gassing people could have had no doubt about the effects of their actions. Another difference is that Milgram's subjects had no opportunity, outside the stressful situation in which the orders were given, to reflect on what they were doing. In contrast, Nazis who were murdering people would go home at night and then return to the gas chambers and kill more people the next day. Historians have pointed to motives for obedience on the part of Hitler's followers—including rampant anti-Semitism and nationalism (Goldhagen, 1996; Miller, 2004)—that are unlike any motives that Milgram's subjects had.

Most social psychologists would agree that Milgram's findings do not provide full explanations of real-world crimes of obedience, but do shed light on some general principles that apply, in varying degrees, to such crimes. Preexisting beliefs about the legitimacy of the endeavor, the authority's confident manner, the immediacy of authority figures, the lack of alternative models of how to behave, and the incremental nature of the requests or orders may contribute to real crimes of obedience in much the same way that they contributed to obedience in Milgram's studies, even when the motives are very different. This may be true for crimes ranging from acts of genocide on down to the illegal “cooking of books” by lower-level corporate executives responding to orders from higher-ups.

SECTION REVIEW

Our tendency to comply can be exploited by others.

Sales Pressure and Compliance

- Salespersons can manipulate us into making purchases through the low-ball and foot-in-the-door techniques, both of which depend on our tendency to reduce cognitive dissonance.
- Pregiving—such as handing someone a free calendar before requesting a donation—takes advantage of the reciprocity norm, the universal tendency to return favors.
- A sense of having something in common with the salesperson—even something quite trivial—can make us more likely to comply with the sales pitch.

Conditions That Promote Obedience

- In his studies of obedience, Milgram found that most subjects obeyed commands to harm another person, even though it distressed them to do so and there was no real threat or reward involved.
- A number of factors help to explain the high rate of obedience in Milgram's studies. These include the prevailing norm of obedience to legitimate authorities, the authority's acceptance of responsibility, the proximity of the authority figure, and the lack of a model for disobeying.
- Though Milgram took precautions to protect his subjects from psychological harm, some critics object to the studies on ethical grounds. Others question whether the results can be generalized to real crimes of obedience, such as the Nazi Holocaust.

To Cooperate or Not: The Dilemma of Social Life

From an evolutionary perspective, the fundamental purpose of sociability is cooperation. Individual ants, wolves, chimpanzees, and humans live and work with others of their kind because they can fulfill certain life-sustaining needs better through cooperation than they could alone. At the same time that they are teammates working for common ends, however, the members of a group are also individuals with self-interests that can run counter to those of the group as a whole. The tension between acting for the good of the group (*cooperation*) and acting for one's own selfish good at the expense of the others (*defection*) is epitomized in social dilemmas. A **social dilemma** exists whenever a particular course of action or inaction will benefit the individual but harm the others in the group and cause more harm than good to everyone if everyone takes that course.

The Tragedy of the Commons: A Social-Dilemma Allegory

24

How does “the tragedy of the commons” illustrate the critical importance of social dilemmas to human survival? What are some examples of real-world social dilemmas?

<<< The significance of social dilemmas for human survival was dramatically illustrated by the ecologist Garrett Hardin (1968) with an allegory that he called *the tragedy of the commons*. Hardin compared our whole planet with the common grazing land that used to lie at the center of New England towns. When the number of cattle grazing the pasture began to reach the limit that the pasture could support, each farmer was faced with a dilemma: “Should I add another cow to my herd? One more cow will only slightly hurt the pasture and my neighbors, and it will significantly increase my profits. But if everyone adds a cow, the pasture will fail and all the cattle will die.” The dilemma becomes a tragedy if all the farmers reason: “It is not my increase in cattle, but everyone else’s, that will determine the fate of the commons. I will lose if everyone increases his herd, but I will lose even more if everyone except me increases his herd.” So they all add a cow, the pasture gives out, the cattle die, and the townspeople all suffer the loss.

We are all constantly involved in social dilemmas, some so grand in scale as to encompass all members of our species as a single group and others much smaller in scale. Here’s a grand one: Sound logic tells me that the pollution I personally add to the Earth’s atmosphere by driving a gasoline-burning automobile does not seriously damage the air. It is the pollution of the millions of cars driven by others that causes the serious damage. So I keep driving, everyone else does too, and the pollution keeps getting worse.

Here’s a social dilemma of smaller scale, more typical of the kind that most of us actually experience as a dilemma: If you are part of a team of students working on

a project to which everyone is supposed to contribute and for which everyone will receive the same grade, you might benefit by slacking off and letting others do the work. That way you could spend your time on other courses, where your grade depends only on your own effort. But if everyone in your group reasoned that way, the group project would not get done and you, along with the others in your group, would fail.

Another example, intermediate in scale, is that of public television, which depends on voluntary contributions from its viewers. Any individual viewer might reason that his or her contribution or lack of it won’t make or break the station and might decide, therefore, to leave the contributing to others. If everyone reasoned that way, the station might disappear.

Every project that depends on group effort or voluntary contributions poses a social dilemma. In each case, *social working*, or *contributing*, is the individual’s cooperative solution, and so-

A modern tragedy of the commons

The oceans are common fishing grounds where thousands of people make their living. When too many fish are caught, the supply diminishes and valued species may even become extinct. Each fisherman can reason, logically, that his catch contributes very little to the problem. The diminished supply of fish is caused by the thousands of other fishermen.



Darren McColester / Newsmakers

cial loafing, or *free riding*, is the noncooperative solution. We are all involved in social dilemmas every day. What are the factors that lead us to cooperate or not in any given instance? Let's first look more closely at the logic of social dilemmas, as exemplified in laboratory games, and then at some conditions that promote cooperation in real-life social dilemmas.

The Logic of Social Dilemmas Exemplified in Games

To study the logic of social dilemmas, stripped of some of their real-world complexity, social scientists have invented a variety of social-dilemma games. Going further—stripping away also the emotions, values, and norms that human subjects carry with them—some social scientists use computer programs rather than humans as the players. Such studies help researchers understand the underlying logic of the dilemmas and develop hypotheses about real-life conditions that would tip the balance, logically, toward or away from cooperation.

The One-Trial Prisoner's Dilemma Game

The most common social-dilemma games used by researchers are two-person games called **prisoner's dilemma games**. They are called that because of their historical origin in a hypothetical dilemma in which each of two prisoners must choose between remaining silent and confessing. If both remain silent, both will get a short prison sentence based on other charges. If both confess, they will both get a moderately long sentence. If only one confesses, that one will be granted immunity and get no sentence but the partner will get a very long sentence. They can neither communicate nor learn the other's choice until both have chosen.

In prisoner's dilemma games played in the psychology laboratory, the consequence for one choice or the other is not a reduced or increased prison sentence but a lower or greater monetary reward. Figure 14.10 shows a typical payoff matrix for a two-person game. On each trial, each player can choose either to cooperate for the common good or to defect. Neither learns of the other's choice until both have responded, and the payoff to each player depends on the combination of their two responses. As in all prisoner's dilemma games, the payoff matrix has the following characteristics:

- The *highest individual payoff* goes to the player who defects while the other cooperates (\$5 in the example in Figure 14.10).
- The *lowest individual payoff* goes to the player who cooperates while the other defects (\$0 in the example).
- The *highest total payoff* to the two players combined occurs if both cooperate (\$3 + \$3 = \$6 in the example).
- The *lowest total payoff* occurs if they both defect (\$1 + \$1 = \$2 in the example).

Both players are informed of the payoff matrix and have adequate time to think about their choice. The game is a social dilemma because the highest individual payoff to either player comes from defecting, but the highest total payoff to the two combined comes from cooperating. If the other player defects, you will get more for defecting (\$1) than for cooperating (\$0); if the other cooperates, you will still get more for defecting (\$5) than for cooperating (\$3); but if you both defect, you will each get less (\$1) than you would if you had both cooperated (\$3).

In the *one-trial* version of the game, each player plays only once with a given other player. When the players are anonymous to each other, are not allowed to discuss their choices, and see their task as winning as much for themselves as they can, they usually defect. A person who always defects on one-trial prisoner's dilemma games will necessarily win more money than a person who sometimes or always cooperates. Logic tells both players to defect, so they both do and get \$1 each. What a pity they are so logical! If they were illogical, they might have both cooperated and received \$3 each.

25 What are the defining features of prisoner's dilemma games, and how do they put each player into a social dilemma?

		Player 2	
		Cooperates	Defects
Player 1	Cooperates	\$3 (blue), \$3 (green)	\$0 (blue), \$5 (green)
	Defects	\$5 (blue), \$0 (green)	\$1 (blue), \$1 (green)

FIGURE 14.10 | Sample payoff matrix for a prisoner's dilemma game On each trial, each player must decide whether to cooperate or defect, without knowing in advance what the other player will do. The payoff to each player depends on the combination of the two players' decisions. In this example, the possible payoffs to player 1 are shown in the blue portions of the matrix, and the possible payoffs to player 2 are shown in the green portions.

The Iterative Prisoner's Dilemma Game: The Power of Reciprocity

26

Why are players more likely to cooperate in an iterative (repeated) prisoner's dilemma game than in a one-trial game?

◀◀◀ If two players play the same prisoner's dilemma game repeatedly (iteratively) with each other for a series of trials rather than just once, the logic changes. Cooperation becomes a reasonable choice even when the only goal is to maximize one's own profits. Each player might now reason: "If I cooperate on this trial, maybe that will convince the other player to cooperate on the next. We'll both do a lot better over time if we cooperate and get \$3 each per trial than if we defect and get \$1 each per trial." In other words, logic and selfishness, which lead players to defect in the one-trial game, can lead them to cooperate in the iterative game.

In order to identify the best strategy in an iterative prisoner's dilemma game (the strategy that maximizes one's own earnings), Robert Axelrod (1984) held two round-robin tournaments played by computer programs rather than by human subjects. The programs were submitted by various experts on game strategy, including social scientists, mathematicians, and teenage computer experts. Each program was made to play a series of trials with each other program, using the payoff system shown in Figure 14.10. The objective was to see which program would accumulate the greatest number of points in the tournament. Some of the programs were very complicated, able to remember all the previous plays of the other player and take them into account in deciding on the next play. But in the end, the simplest program won both tournaments.

27

What computer program won Axelrod's prisoner's dilemma tournaments, and why was it so successful?

◀◀◀ The winning program, sent in by psychologist Anatol Rapoport, was called *Tit-for-Tat* (or *TFT*). It consisted of just two rules: (1) The first time you meet a new program, cooperate with it. (2) After that, do on each trial what the other program did on its most recent previous trial with you.

Notice that TFT is incapable of "beating" any other program in head-to-head competition. It never cooperates less often than the other program does, and therefore it can never win more points in play with another program than that program wins in play with it. The TFT program won the tournament not by beating other programs in individual encounters but by getting them to cooperate with it. Other programs earned at least as many points as TFT in their games with TFT, but they did not do so well in their games with each other, and that is why TFT earned the most points in the end.

Why is TFT so effective in eliciting cooperation? According to Axelrod's analysis, there are four reasons:

1. *TFT is nice.* By cooperating from the start, it encourages the other player to cooperate.
2. *TFT is not exploitable.* By reciprocating every failure to cooperate with its own failure to cooperate on the next turn, it discourages the other player from defecting.
3. *TFT is forgiving.* It resumes cooperating with any program as soon as that program begins to cooperate with it.
4. *TFT is transparent.* It is so simple that other programs quickly figure out its strategy and learn that they are best off cooperating with it.

Studies since Axelrod's have shown that TFT is highly effective in eliciting cooperation not just from other computer programs but also from human subjects in the laboratory (Komorita & Parks, 1999; Sheldon, 1999). People, like computer programs, figure out rather quickly that their best strategy when playing with TFT is to cooperate.

Decline in Cooperation as the Number of Players Increases

28

Why might we logically expect small groups to cooperate more than large ones? What evidence indicates that indeed they do?

◀◀◀ Prisoner's dilemma games involve just two players, but other social-dilemma games—including those called *public-goods games*—can involve any number. In a typical public-goods game, each player is given a sum of money and then, under conditions of anonymity, must choose whether to keep the money or contribute it

to a common pool (the public good). Then, if and only if at least a certain percentage of players (say, 75 percent) have contributed, all players, including those who haven't contributed, receive a reward that is substantially greater than the amount each person was asked to contribute.

In such games, each individual player's choice to contribute or not is more significant (because it represents a larger portion of the total contributions needed) if the number of players is small than if it is large, so the temptation to refrain from contributing becomes greater as the number of players increases. With few players, each one might logically reason: "I'd better contribute or we won't reach the needed percentage to win the reward." But with many players, each might logically reason: "My contribution will have little effect on the total pool of contributions, so my best strategy is to keep my money and hope enough others will contribute to produce the reward."

The result is that far fewer rewards are won in such games when group size is large than when it is small (Glance & Huberman, 1994; Komorita & Parks, 1995). The same thing happens in experiments in which groups of people are asked to exert effort for a common goal. People work harder in small groups than in large ones (Karau & Williams, 1995, 2001). You are more likely to contribute to a course project for a common grade if you are one member on a team of 3 than if you are one on a team of 30. The larger the group, the greater is the diffusion of responsibility, and the less is the likelihood that a given individual will contribute.



Michael Grecco / Stock, Boston

A social dilemma in dining out

When diners agree before a meal to split the bill equally, a temptation may be created in each one to "cheat" by ordering a more expensive meal than he or she normally would. Since the added expense created by any one person's expensive meal is shared by the whole group, the temptation to order an expensive meal increases as the number of diners increases. But if everyone orders an expensive meal, then everyone pays dearly in the end.

Conditions That Promote Cooperation

In real life, and even in laboratory games, people cooperate more in social dilemmas than would be expected if their choices were based solely on immediate self-interest (Fehr & Fischbacher, 2003). Many people work hard on group projects even in large groups. Many people contribute to public television. Some people (though not nearly enough) even choose to ride a bicycle or use public transportation, rather than drive a car, to reduce pollution and help save the planet. What are the forces that lead us to cooperate in such "illogical" ways?

Evolution, cultural history, and our own individual experiences have combined to produce in us decision-making mechanisms that are not confined to an immediate cost-benefit analysis. Consciously or unconsciously, thoughtfully or automatically, we take into account factors that have to do with not just our short-term interests but also our long-term interests, which often reside in maintaining good relationships with other people. Many aspects of our social nature can be thought of as adaptations for cooperating in social dilemmas.

Accountability, Reputation, and Reciprocity as Forces for Cooperation

In the iterative prisoner's dilemma game, people and computer programs using the TFT strategy do well by cooperating because cooperation encourages other players to cooperate with them in the future. By cooperating, players greatly increase their long-term earnings at the expense of slightly reduced short-term earnings on any given play. The strategy works only because the other players can identify who did or did not cooperate, can remember that information from one trial to the next, and are inclined to respond to each player in accordance with that player's previous action toward them. In human terms, we would say that TFT is successful because each player is *accountable* for his or her actions. Through that accountability, TFT establishes a *reputation* as one who helps others and *reciprocates* help given by others but who won't be exploited by those who fail to reciprocate.

29

Why is TFT especially successful in situations where players can choose their partners? How do real-life social dilemmas differ from laboratory games with respect to accountability, reputation, and reciprocity?

<<< When the players of laboratory social-dilemma games are free to choose the partners with whom they play, they favor those who have already developed a reputation for cooperation (Sheldon & others, 2000; Wedekind & Milinski, 2000). The TFT strategy does especially well in this situation, because it attracts partners who seek to cooperate and repels potential partners who seek to compete (Van Lange & Visser, 1999).

In laboratory games, the factors that foster cooperation—accountability, reputation, and reciprocity—are neatly confined to the specific actions allowed by the game. But in real life they are not confined; they spill out everywhere. If I help a person today, that person and others who hear of my help may be disposed to help me in the future, in ways that I cannot even guess at today. Stories of long-range, unanticipated reciprocity are easy to find in everyone's autobiography. Research studies, in various cultures, including hunter-gatherer cultures, suggest that people everywhere tend to keep track of the degree to which others are helpful and to offer the greatest help to those who have themselves been most helpful in the past (Fehr, 2004; Gurven, 2004; Price, 2003). Because there is no way to predict how or when reciprocity may occur, it hardly pays to think about reciprocity when we have an opportunity to help another without too much sacrifice to ourselves. It may be more efficient in such cases to help automatically, and that indeed describes fairly well what most people do in the absence of factors that inhibit helping (Cialdini, 2001).

Norms of Fairness and Punishment of Cheaters as Forces for Cooperation

People everywhere seem to have a strong sense of fairness, which goes beyond immediate self-interest. In many situations, people would rather gain nothing than enter into an unfair agreement by which they gain a little and another person gains a lot (Fehr & Fischbacher, 2003, 2004). This sense of fairness has been illustrated using laboratory games called *ultimatum games*.

30

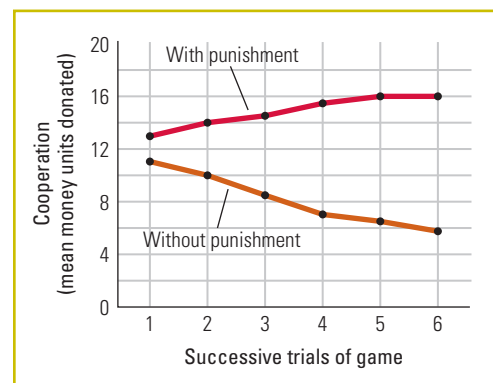
How have laboratory games demonstrated the human sense of justice and willingness to punish even at a personal cost? How does such behavior promote long-term cooperation?

<<< In a typical ultimatum game, two players are told that a certain amount of money, say \$100, will be divided between them. One player, the *proposer*, is allowed to propose how to divide the money, and the other player, the *responder*, has the choice of accepting or rejecting the offer. If the offer is accepted, the two players keep the amounts of money proposed by the proposer; if the offer is rejected, nobody gets anything. When this game is played in a one-shot way (there will be no future games) under conditions of complete anonymity (neither player will ever know who the other player is), it is irrational, from an economic point of view, for the responder to reject any offer above \$0. Yet, the repeated result, wherever this game is played, is that responders usually reject offers that are considerably less than half of the total money. This is true even when the amount of money to be divided is sufficiently substantial that the smaller (rejected) portion would be enough to pay one's living expenses for several weeks (Cameron, 1999).

Other experiments show that in social-dilemma games with several players, people are willing to give up some of their own earnings in order to punish a

player who has contributed substantially less than his or her share to the public good (Fehr & Fischbacher, 2004; Shinada & others, 2004). The punishment, in such cases, involves removing some of the winnings that the "cheater" has garnered. As Figure 14.11 shows, when the opportunity to punish cheaters is introduced into the rules of such a game, most people stop cheating, so the total amount of cooperation—and thus, the total amount earned by all play-

FIGURE 14.11 | Altruistic punishment increases cooperation In each trial of this social-dilemma game, each player was given 20 money units and could contribute any amount of that to a common pool. The money contributed was then multiplied by 1.4 and redistributed evenly among the four players. In one condition, players could punish low contributors by giving up one of their own money units to have three units taken away from the punished player. With the punishment option, cooperation increased from trial to trial. Without that option, it decreased from trial to trial. At the end of all trials, the players could exchange the money units they had accumulated for real money. (Data from Fehr & Gächter, 2002.)



ers combined—increases dramatically. Punishment of this sort is referred to as **altruistic punishment**, because the whole group benefits from the improvement in cooperation, while only the punisher and the person being punished pay a cost.

The fact that people are willing to go out of their way and pay a cost to punish cheaters may be a fact of human nature—and a norm resulting from cultural training—that helps maintain cooperation in human societies. Such behavior is mediated by emotions. Neuroimaging studies of people in ultimatum and public-goods games have shown that when people are cheated, brain areas associated with anger become active (Sanfey & others, 2003), and when people punish cheaters, brain areas associated with pleasure become active (de Quervain & others, 2004). Apparently, we are emotionally wired to get mad at cheaters and to enjoy revenge, even when that revenge entails an economic cost to ourselves.

Social Identity Promotes Cooperation Within Groups and Competition Across Groups

As discussed in Chapter 13, people everywhere have two different ways of thinking about themselves, which serve different functions. One is *personal identity*, which entails thought of oneself as an independent person with self-interests distinct from those of other people. The other is *social identity*, which entails thought of oneself as a more or less interchangeable member of a larger entity, the group, whose interests are shared by all members. Evolutionarily, the two modes of self-thought may have arisen from our need to survive both as individuals and as groups (Guisinger & Blatt, 1994). If I save myself but destroy the group on which I depend, I will, in the long run, destroy myself. We don't logically think the issue through each time but, rather, tend automatically to cooperate more when we think of the others as members of our group than when we don't.

Many experiments have shown that people in all types of social-dilemma games cooperate much more when they think of the other players as group-mates than when they think of them as separate individuals or as members of other groups (Dawes & Messick, 2000; Van Vugt & Hart, 2004). In one study, for example, players differing in age cooperated much more if they were introduced to one another as citizens of the same town than if they were introduced as representatives of different age groups (Kramer & Brewer, 1984). In another study, simply referring to a set of unacquainted players as a “group,” rather than as a set of individuals, increased their cooperation (Wit & Wilke, 1992).

Identification with a group increases people's willingness to help members of their own group but *decreases* their willingness to help members of another group. Groups playing social-dilemma games with other groups are far more likely to defect than are individuals playing with other individuals, even when the payoffs to individuals for cooperating or defecting are identical in the two conditions (Insko & others, 1994, 2005; Wildschut & others, 2003). People are much less likely to trust others and more likely to cheat others when they view those others as part of another group than when they view them as individuals. In real life, as in the laboratory, interactions between groups are typically more hostile than interactions between individuals (Hoyle & others, 1989). All too often the hostility becomes extreme.

Group Against Group: Lessons from Robbers Cave

The most vicious and tragic side of our nature seems to emerge when we see ourselves as part of a group united against some other group. Perhaps because of an evolutionary history of intertribal warfare, we can be easily provoked into thinking of other groups as enemies and as inferior beings, unworthy of respectful treatment. The history of humankind can be read as a sad, continuing tale of intergroup strife that often becomes intergroup atrocity.

There are limits to the degree to which researchers can, ethically, create the conditions that bring out intergroup hostility for the purpose of study. Some years



Lorne Resnick / Stone

Should I drive or cycle to work? For some people it is easy to resolve this dilemma in the cooperative direction. Bicycle commuting not only helps preserve the earth's atmosphere but is also good regular exercise.

31 What is some evidence that social identity can lead to helping group-mates and hurting those who are not group-mates?

ago, Muzafer Sherif and his colleagues approached those limits in a now-famous study with 11- and 12-year-old boys at a 3-week camping program in Oklahoma's Robbers Cave Park (so named because it was once used as a hideout by the famous outlaw Jesse James). The researchers were interested in understanding how hostility between groups develops and how it can be resolved.

The Escalation of Conflict

To establish two groups of campers, Sherif and his colleagues divided the boys into two separate cabins and assigned separate tasks to each group, such as setting up camping equipment and improving the swimming area. Within a few days, with little adult intervention, each cabin of boys acquired the characteristics of a distinct social group. Each group had its own leaders, its own rules and norms of behavior, and its own name—the Eagles and the Rattlers.

When the groups were well established, the researchers proposed a series of competitions, and the boys eagerly accepted the suggestion. They would compete for valued prizes in such games as baseball, touch football, and tug-of-war. As Sherif had predicted from previous research, the competitions promoted three changes in the relationships among the boys within and between groups:

32

What changes occurred within and between two groups of boys as a result of intergroup competitions at a summer camp?



1. **Within-group solidarity** As the boys worked on plans to defeat the other group, they set aside their internal squabbles and differences, and their loyalty to their own group became even stronger than it was before.
2. **Negative stereotyping of the other group** Even though the boys had all come from the same background (white, Protestant, and middle class) and had been assigned to the groups on a purely random basis, they began to see members of the other group as very different from themselves and as very similar to one another in negative ways. For example, the Eagles began to see the Rattlers as dirty and rough, and in order to distinguish themselves from that group they adopted a “goodness” norm and a “holier-than-thou” attitude.
3. **Hostile between-group interactions** Initial good sportsmanship collapsed. The boys began to call their rivals names, accuse them of cheating, and cheat in retaliation. After being defeated in one game, the Eagles burned one of the Rattlers' banners, which led to an escalating series of raids and other hostilities. What at first was a peaceful camping experience turned gradually into something verging on intertribal warfare.

Rattlers versus Eagles What at first were friendly competitions, such as the tug-of-war shown here, degenerated into serious hostility and aggression between the two groups of boys in Sherif's field study at Robbers Cave.

Resolving the Conflict by Creating Common Goals

In the final phase of their study, Sherif and his colleagues tried to reduce hostility between the two groups, a more difficult task than provoking it had been. In two previous studies similar to the one at Robbers Cave, Sherif had tried a number of procedures to reduce hostility, all of which had failed. Peace meetings between leaders failed because those who agreed to meet lost status within their own groups for conceding to the enemy. Individual competitions (similar to the Olympic Games) failed because the boys turned them into group competitions by tallying the total victories for each group (just as we find with the international Olympics). Sermons on brotherly love and forgiveness failed because, while claiming to agree with the messages, the boys simply did not apply them to their own actions.

At Robbers Cave, the researchers tried two new strategies. The first involved joint participation in pleasant activities. Hoping that mutual enjoyment of noncompetitive activities would lead the boys to forget their hostility, the researchers brought the two groups together for such activities as meals, movies, and shooting firecrackers. This



From *Social Psychology* by Muzafer Sherif and Carolyn W. Sherif, Fig. 11.18a (p. 258). Copyright © 1969 by Muzafer Sherif and Carolyn Sherif, by permission of HarperCollins, Inc.

didn't work either. It merely provided opportunities for further hostilities. Meals were transformed into what the boys called "garbage wars."

The second new strategy, however, was successful. This involved the establishment of **superordinate goals**, defined as goals that were desired by both groups and could be achieved best through cooperation between the groups. The researchers created one such goal by staging a breakdown in the camp's water supply. In response to this crisis, boys in both groups volunteered to explore the mile-long water line to find the break, and together they worked out a strategy to divide their efforts in doing so. Two other staged events similarly elicited cooperation. By the end of this series of cooperative adventures, hostilities had nearly ceased, and the two groups were arranging friendly encounters on their own initiative, including a campfire meeting at which they took turns presenting skits and singing songs. On their way home, one group treated the other to milkshakes with money left from its prizes.

Research since Sherif's suggests that the intergroup harmony brought on by superordinate goals involves the fading of group boundaries (Bodenhausen, 1991; Gaertner & others, 1990). The groups merge into one, and each person's social identity expands to encompass those who were formerly excluded. The boys at Robbers Cave might say, "I am an Eagle [or a Rattler], but I am also a member of the larger camp group of Eagles plus Rattlers—the group that found the leak in the camp's water supply."

If there is hope for a better human future, one not fraught with wars, it may lie in an increased understanding of the common needs of people everywhere and the establishment of superordinate goals. Such goals might include those of stopping the pollution of our shared atmosphere and oceans, the international drug trade, diseases such as AIDS that spread worldwide, and the famines that strike periodically and disrupt the world. Is it possible that we can conceive of all humanity as one group, spinning together on a small and fragile planet, dependent on the cooperation of all to find and stop the leaks?

How did Sherif and his colleagues succeed in promoting peace between the two groups of boys?

A superordinate task brings people together To protect their town in western Washington, these volunteers are piling sandbags to hold back the flooding waters of the Skagit River. None of them could succeed at this task alone; they need one another. Social psychological theory predicts that this experience will create positive bonds among the volunteers and promote their future cooperation.



AP / Wide World

SECTION REVIEW

Social dilemmas require individuals to decide whether or not to act for the common good.

The Tragedy of the Commons

- In a social dilemma, the choice to behave in a certain way produces personal benefit at the expense of the group and leads to greater harm than good for all if everyone chooses that option.
- In the illustration called the tragedy of the commons, each individual puts one extra cow on the common pasture, thinking his or her cow will make little difference; but the collective effect is disastrous to all.

Social-Dilemma Games

- Social-dilemma games, including prisoner's dilemma games, are used to study the conditions that promote cooperation.
- Players tend to defect in one-shot prisoner's dilemma games. When players can play each other repeatedly, however, a tit-for-tat strategy can be used to promote cooperation.
- In social-dilemma games with multiple players, cooperation declines as the number of players increases.

Roles of Accountability and Social Identity

- Cooperation increases when players are accountable for their actions and can develop reputations as cooperators or cheaters.
- The tendency for people to reject unfair offers and to punish cheaters, even at their own expense, is a force for cooperation.
- Shared social identity among group members increases cooperation within the group but decreases cooperation with other groups.

Group Against Group

- In Sherif's Robbers Cave experiment, competition between the two groups of boys led to solidarity within groups, negative stereotyping of the other group, and hostile interactions between groups.
- Hostility was greatly reduced by superordinate goals that required the two groups to cooperate.

Emotional Foundations of Our Social Nature

All the social influences that you have been reading about in this chapter involve emotions. Our social situations affect our emotions, and our emotions affect our behavior. As social beings, we are endowed with emotions that help us to keep connected with others, to coordinate our behavior with that of others, to gain and retain social acceptance, and, at the same time, to avoid being exploited by others.

Emotions Help Keep Us Together and Coordinate Our Actions

The Pain of Exclusion and Social Loss

When we are excluded by others, we say that our “feelings are hurt.” When we lose, through whatever means, a lover, dear friend, or family member, we say our “heart is broken.” In many different languages, people describe social loss and exclusion using the same terms as those used for physical pain (Eisenberger & Lieberman, 2004). Psychologists use the term **social pain** to refer to the discomfort that people feel when they are socially rejected or when they lose a valued companion.

34

What is known about the neural basis of social pain? What, presumably, is the evolutionary function of social pain?

<<< Social pain and physical pain overlap in many ways (Eisenberger & Liberman, 2004; MacDonald & Leary, 2005). Social pain magnifies physical pain, and vice versa. People who are experiencing a social loss suffer more from physical pain than do people who are not experiencing such a loss. People suffering from physical pain find relief in social companionship. Drugs that reduce physical pain also reduce social pain.

Neuroimaging research indicates that social pain involves some of the same brain areas that are involved in physical pain. As is discussed in Chapter 7, the discomforting, aversive quality of physical pain, as opposed to its sensory quality, derives from activity in specific portions of the limbic system, notably the anterior cingulate cortex and the anterior insula. Nancy Eisenberger and her colleagues (2003) found that these same areas of the brain became unusually active when subjects in an fMRI scanner were led to believe that other subjects in the experiment had excluded them from a game. Moreover, the degree of activity in one of these areas, the anterior cingulate cortex, correlated positively with the amount of distress that the subjects reported feeling when they learned that they had been excluded. This, along with other evidence, suggests that the social pain system uses parts of the same neural system that evolved originally to serve the function of physical pain (MacDonald & Leary, 2005).

The purpose of social pain, most likely, is to motivate us to avoid social rejection and to maintain close social ties with other people. As social beings, the loss of people with whom we can cooperate is as detrimental to our well-being and survival as is serious damage to our bodily tissues.

Emotional Contagion as a Force for Group Cohesion

A social group is not just a collection of individuals. It is a *unified* collection. People in a social group behave, in many ways, more like one another than the same people would if they were not in a group. They tend automatically to mimic one another's postures, mannerisms, and styles of speech, and this imitation contributes to their sense of rapport (Lakin & Chartrand, 2003; van Baaren & others, 2004). They also tend to take on the same emotions.

35

What is the value, for group life, of the spread of sadness, anger, fear, and laughter from person to person? How might emotional contagion figure into the rise of a group leader?

<<< As noted in Chapter 3, people everywhere express emotions in relatively similar ways. Such expressions help members of a group know how to interact with one another. By seeing others' emotional expressions, group members know who needs help, who should be avoided, and who is most approachable at the moment for help. In addition, people tend automatically to adopt the emotions that they perceive in those around them, and this helps the group to function as a unit (Hatfield & others, 1994; Wild & others, 2001). Sadness in one person tends to induce sad-

ness in others nearby, and that is part of the mechanism of empathy by which others become motivated to help the one in distress (discussed in Chapter 12). Anger expressed in the presence of potential allies may lead to shared anger that helps in the recruitment of allies into a common cause. Likewise, fear in one person tends to induce fear in others nearby, placing them all in a state of heightened vigilance and thereby adding a measure of protection for the whole group.

One of the most contagious of all emotional signals is laughter, as evidenced by the regularity of mutual laughter in everyday interactions and the effectiveness of laugh tracks (often added to recorded comedy) in inducing laughter in audiences (Provine, 1996, 2004). The easy spread of laughter apparently helps put a group into a shared mood of playfulness, which reduces the chance that one person will be offended by the remarks or actions of another. In addition to coordinating people's actions, shared emotions seem to promote bonds of affection or attachment among those who share the moods, which contribute to the long-term stability of social relationships and groups.

Political leaders often achieve their status at least partly through their ability to manipulate others' emotions, and their own emotional expressions are part of that process. Former U.S. presidents Ronald Reagan and Bill Clinton were both known as "great communicators," and much of their communication occurred through their persuasive facial expressions. In a research study conducted shortly after Reagan was elected president, university students watched film clips of Reagan expressing happiness, anger, or fear as he spoke to the American public about events that faced the nation (McHugo & others, 1985). In some cases the sound track was kept on, in others it was turned off, and in all cases the students' own emotional reactions were recorded by measuring their heart rates, their sweating, and the movements of particular facial muscles. Regardless of whether they claimed to be his supporters or opponents, and regardless of whether they could or could not hear what he was saying, the students' bodily changes indicated that they were responding to Reagan's performance with emotions similar to those he was displaying.

More recently, researchers have found that the spread of emotions can occur completely unconsciously. In one study, a face expressing happiness or anger was flashed on a screen too quickly for subjects to see it consciously, yet subjects responded with the same facial expression that had been flashed (Dimberg & others, 2000).

Our Self-Conscious Emotions Make Us Socially Acceptable

Our emotions not only help us coordinate our actions with those of others, they also help to make us socially acceptable. This is particularly true of *guilt*, *shame*, *embarrassment*, and *pride*, when they function as they have evolved to function. These four emotions are called the **self-conscious emotions**, because they are linked to thoughts about the self or one's own actions (Tangney, 1999). The feelings of guilt, shame, and embarrassment are painful, and the pain may motivate us to make amends when we have offended others or been ineffective in some way. Conversely, the feeling of pride rewards us for behaving effectively and gaining others' approval.



Vittoriano Rastelli / Corbis

A poker face is hard to keep We tend automatically to signal our emotions when in the company of others. Only through conscious effort can we refrain from giving away our hand.

Emotional contagion Many American citizens, regardless of their political views, found it hard not to feel good when they saw Ronald Reagan looking like this.



UPI / Corbis-Bettmann

36

What differing functions are proposed for (a) guilt, (b) shame, and (c) embarrassment, and what evidence supports each function?



The self-conscious emotions have been studied primarily by two means. One is to observe them in very young children, who are less adept than older children or adults at hiding them and whose emotions can be manipulated relatively easily in the laboratory. The other means is to ask adults to describe one or more of their previous experiences with a particular emotion: What brought it on? What effects did it have on your behavior and that of others? How was the emotion resolved?

Guilt as a Motivator of Relationship Repair

When adults are asked to describe episodes in their lives that led them to experience guilt, they most often tell about hurting or disappointing someone they care about (Smith & others, 2002; Tangney, 1995). Typical guilt stories involve neglecting, offending, or being disloyal to a friend or partner.

In one study, Roy Baumeister and his colleagues (1995) asked university students to write out two episodes in which they caused someone else to become angry at them, one in which they responded with guilt and one in which they didn't. Compared with the non-guilt stories, the guilt stories were far more likely to involve selfish behavior toward someone whom they respected and with whom they had a mutually giving relationship. The guilt stories were also far more likely than the non-guilt stories to end with an apology or with a behavioral change designed to correct or prevent recurrence of the offense. Another study revealed that people who are most able to perceive events from another person's perspective are also most prone to experience guilt after an interpersonal conflict (Leith & Baumeister, 1998). When people perceive that they have hurt a valued partner, they experience guilt, which motivates them to behave in ways that bring the relationship back into balance.

Shame as a Motivator of Social Withdrawal

While guilt focuses our attention on another person's feelings and our role in causing those feelings, shame focuses our attention on some real or imagined flaw in ourselves—a flaw in appearance, ability, or moral character (Lindsay-Hartz & others, 1995; Smith & others, 2002). In one research study, 3-year-olds manifested more evidence of shame in response to failing an easy task than to failing a difficult one, already showing a link between shame and perceived incompetence (Lewis, 1991). In another study, university students were asked to describe one episode in which they had experienced shame and another in which they had experienced guilt (Tangney, 1995). Compared with the guilt stories, the shame stories expressed more deeply felt pain and were more difficult for the students to tell. In the guilt episodes students saw themselves as relatively powerful—capable of either hurting or helping another person and motivated now to help. In the shame stories they saw themselves as blemished and powerless, motivated not to help but to disappear or hide.

Stories of shame suggest that it most often occurs when the person's failings have been publicly exposed (Smith & others, 2002), when it may truly be in the person's best interest to hide from negative judgment. Expressions of shame may also help communicate the person's feelings of self-derogation and powerlessness, which may induce compassion, perhaps even guilt, in those who would judge or punish the person.

Embarrassment as a Rectifier of Awkward Situations

When undergraduates were asked to write separate essays describing their experiences of embarrassment, shame, and guilt, the embarrassment essays far more often than the others focused on trivial or humorous incidents, not serious misconduct or deeply negative self-evaluations (Tangney & others, 1996). Embarrassment most often follows the inadvertent violation of a social norm or the receipt of unexpected or undesired attention from others. A house guest walks into the bathroom and finds it occupied; an after-dinner speaker notices that he has gravy on his

necktie; a young girl is asked by her grandmother to sing in front of strangers; a student is complimented effusively by a professor in front of classmates.

Unlike guilt, embarrassment most often involves acquaintances or strangers, whose reactions are not predictable, rather than close friends or family. Even toddlers as young as 15 months old have been observed to show signs of embarrassment when made the center of attention (Lewis, 1995).

The function of embarrassment appears to lie primarily in its communicative value. The blush and the sheepish grin, which crest just after the head and eyes have turned aside, are clear and immediate signals that what just happened is the result of circumstance or blunder or someone else's doing and is not a deliberate attempt by the self to disrupt the others' activities or become the center of attention (Miller, 1995). The sheepish or silly grin of embarrassment seems to say, non-verbally, "How silly of me" or "What a foolish thing this is that happened." The human display of embarrassment is similar in form and function to the appeasement displays shown by other primates (Keltner & Anderson, 2000). As noted in Chapter 3, apes and monkeys grin and turn their heads away or down as a nonverbal way of saying, "No offense was meant; please don't attack me."

Some people try to hide their embarrassment, but others—who know its power—may cultivate it and use it more or less deliberately for good effect. Several studies have shown that people are more attracted to those who manifest embarrassment in awkward situations than to those who don't (Keltner & Buswell, 1997). Embarrassment indicates concern for others' reactions and lack of haughtiness. In one experiment, subjects saw a video depicting a shopper bumping into a grocery display and then either expressing embarrassment or not as he began to pick up what he had knocked down (Semin & Manstead, 1982). The viewers who saw the embarrassment version rated the shopper as more likable than those who saw the calm, non-embarrassment version.

Pride as an Index of Social Acceptability

In the dance of social life, we experience guilt at ignoring our partner too long, shame at our chronic lack of grace, and embarrassment when we accidentally bump a person we don't know very well. These reactions are useful; they help keep others from throwing us out despite our flaws. But sometimes we seem to do everything right, and then we experience *pride*. We throw our shoulders back, hold our heads high, speak in richer tones, and strut a little as the dance goes on, inviting others to admire us. As long as those others do indeed admire us, this drawing of attention to ourselves is useful; it adds to the cohort who may wish to make alliances with us for mutual benefit in the future.

Pride seems to be the direct opposite of shame. While shame is the emotional component of low self-esteem, pride is the emotional component of high self-esteem (Brown & Marshall, 2001). While shame focuses our attention on our own flaws and failures, pride focuses it on our own strengths and successes. While shame is felt and manifested as a desire to shrivel up and disappear, pride is felt and manifested as a desire to swell up and be seen by as many others as possible (Mascolo & Fischer, 1995; Tracy & Robins, 2004). Just as 2-year-olds react with shame when they fail at a task or are criticized, they react with pride when they succeed or are praised (Stipek, 1995). The good feeling

Pride The expression of pride is easy to recognize. When it is too extreme or occurs in inappropriate situations, it draws criticism, such as "She's got her nose in the air." When used more moderately, in situations where others are likely to admire us, however, it can draw positive attention to our accomplishments and confidence.



Andrea Barbe/Corbis

Embarrassment Blushing, grinning sheepishly, looking down to one side, and touching one's own face (as if symbolically hiding behind the hand) are common components of the human display of embarrassment.

37 When is it safe to express pride? What differing functions may be served by the feeling and expression of pride?



© Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science*, 15, 194–197.

of pride is a self-reward for virtuous, diligent, or skillful behavior and an inducement to continue such behavior in the future. The expression of pride is perhaps best understood functionally as an advertisement of self-confidence that invites others to feel confident in us also. One must be careful about expressing pride, however. If shown too often, or in the wrong contexts, or with the wrong people, or in amounts that are disproportionate to the actual accomplishment, it is perceived as arrogance and can lead to derision and rejection.

SECTION REVIEW

Social influences on our behavior are mediated by our emotions.

Emotions Can Promote Group Cohesion

- Social pain—the discomfort we feel when we are socially rejected or lose a valued relationship—involves some of the same brain areas as physical pain. It motivates us to avoid situations that could lead to social isolation.
- Emotional contagion acts to unite the members of a group, coordinating their actions and promoting bonds of attachment. Successful leaders are often particularly good at expressing emotions in ways that lead others to share those emotions.

Functions of the Self-Conscious Emotions

- The self-conscious emotions—guilt, shame, embarrassment, and pride—serve to make us more socially acceptable to others.
- Guilt is typically felt when we have hurt someone we value. It motivates us to behave so as to bring the relationship back into balance.
- Shame—the emotional response to a real or imagined flaw in ourselves—often produces deeper pain than guilt, including a sense of powerlessness. It motivates us to withdraw, and it may evoke compassion if others witness it.
- Embarrassment is generally felt after we inadvertently breach a social norm or receive unwanted or unexpected attention. Embarrassment reveals our concern for others' feelings and can make us more likable.
- Pride is the emotional component of high self-esteem. In the right circumstances it can focus attention on our strengths and successes and lead others to join us in the confidence we feel in ourselves.

Concluding Thoughts

1. The desire to be accepted by others underlies much of social influence In surveying the body of research and theory on social influence, one cannot help being struck by the frequent recurrence of a single, simple idea: Human beings have a remarkably strong desire to be approved of by other human beings nearby.

Why do people become aroused—leading to social interference or facilitation—when they know that their performance is being evaluated, even if the evaluator is a stranger and the evaluation doesn't count for anything? Why are people so concerned with impression management? Why did subjects in Asch's experiments deny the clear evidence of their own two eyes when it ran counter to what others were saying? Why are people so motivated to abide by social norms? Why do group polarization and groupthink occur? Why do people find it hard to refuse a direct request? Why do people find it hard not to reciprocate a favor, even one that they did not want in the first place? Why did Milgram's subjects find it hard to tell the experimenter that he was asking them to do a terrible thing and that they would not go on? Why do subjects in prisoner's dilemma games find it much easier to take a competitive stance if they have colleagues (a group) on their side than if they play alone? Why do group polarization and groupthink occur? Why is social rejection painful? Why do we tend to adopt the emotions of

those around us? Why do people respond with guilt, shame, or embarrassment when they feel they might have hurt or offended someone, even slightly?

I don't want to oversimplify. The desire to be accepted is surely not the *whole* answer to these questions, but it seems to be an important part of it. As you review each of the phenomena and experiments described in the chapter, you might ask yourself: To what extent (if at all) can this be explained by the desire for acceptance, and what additional explanatory principles seem to be needed?

2. Conformity derives from both normative and informational influences Normative social influences reflect people's desire to behave in accordance with group norms so as to be accepted or liked by the group. Informational social influences reflect people's use of others' responses as information to be included in solving a problem, so as to arrive at the objectively best solution. The difference between these two types of social influence was discussed explicitly in relation to (a) Asch's conformity experiments, (b) the effect that other bystanders have on any one bystander's willingness to help a person in distress, and (c) the polarizing effect that group discussion has on people's opinions or attitudes. As you review the chapter, you might think about how these two types of influence may apply

to other phenomena as well. For example, how might both sorts of influence be involved in obedience, as observed by Milgram's experiments, and in cooperation, as studied in social-dilemma games?

3. Human adaptations for group living As social beings, we are endowed with characteristics that draw us to other members of our species and help us to function effectively in groups. We feel lonely when alone, pained when rejected, and satisfied or proud when accepted. Our self-conscious emotions help to

prevent others from rejecting us. We tend to adopt the attitudes, behavioral styles, and emotions of others in our group, and that conformity allows the group to function as a unit. We also have characteristics that keep us from being exploited by others in the group, as demonstrated by our concern for fairness and our tendency to either punish or avoid those who treat us unfairly. As you review each social-influence phenomenon discussed in this chapter, you might think about the aspects of human nature that underlie it and help to make group living possible and beneficial.

Further Reading

ROBERT CIALDINI (2001). *Influence: Science and practice* (4th ed.). Boston: Allyn & Bacon.

This fun-to-read paperback, full of anecdotes as well as references to research, has chapters on reciprocity, cognitive consistency, conformity, friendliness, and authority as means of influencing people and gaining compliance. After reading it, you will not be as susceptible to sales pressure as you were before.

JAMES SUROWIECKI (2004). *The wisdom of crowds*. New York: Doubleday.

Traditionally, social scientists have emphasized the irrationality of group thinking. This book, in contrast, emphasizes its rationality. Surowiecki, a professional writer, has delved into a diverse literature—in psychology, history, economics, politics, and business management—to support his thesis that groups, when they function well, make better decisions than even the most informed individuals in the group would make alone. To function well, however, groups must avoid the pitfalls of group polarization and groupthink, which are also discussed and exemplified in this delightful book.

MALCOLM GLADWELL (2000). *The tipping point: How little things can make a big difference*. Boston: Little, Brown.

This is a brief, fun-to-read book about social epidemics—the rapid spread of new fashions, ideas, and ways of behaving. Why do some innovations take off while others do not? To address this question, Gladwell, a professional writer, draws from social-psychological research and uses examples from history and contemporary popular culture. Along the way, he presents a convincing explanation of why Paul Revere's famous midnight ride was so successful, why crime in New York City began to plummet in the early 1990s, and why the book *Divine Secrets of the Ya-Ya Sisterhood* sold millions of copies.

STANLEY MILGRAM (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

This is a fascinating, firsthand account of one of the most famous series of experiments in social psychology. Milgram describes his reasons for initiating the research, his findings in many variations of the basic experiment, and his interpretations of and reactions to the findings.

JAMES WALLER (2002). *Becoming evil: How ordinary people commit genocide and mass killing*. Oxford: Oxford University Press.

Millions of people have played active roles in the massacres of millions of oth-

ers. How can we explain the willingness, even eagerness, of ordinary people to partake in such genocides as those committed by European colonists against Native Americans, Turks against Armenians, Christians against Jews, and various tribes in Africa, the Near East, Eastern Europe, and Asia against closely related other tribes? In this book, a social psychologist brings together various explanations of those aspects of human nature and cultural conditioning that make such atrocities possible.

ROBERT AXELROD (1984). *The evolution of cooperation*. New York: Basic Books.

The first half of this brief, nontechnical book tells the story of Axelrod's computer tournaments of a prisoner's dilemma game and presents his analysis of the success of the tit-for-tat strategy. The second half presents examples of the reciprocity principle in evolution and modern life. One chapter tells how a "live-and-let-live" reciprocity norm emerged between opposing lines in trench warfare in World War I and operated against officers' commands to shoot.



PERSONALITY AND DISORDERS

We do not all approach life in the same way. We vary in our emotions, motives, and styles of thinking and behaving, and these differences give each of us a unique personality. Most of these differences are healthy and add spice to life, but some create problems and are classed as mental disorders. This final unit has three chapters. Chapter 15 is about ways of describing and explaining normal personality differences. Chapter 16 is about identifying mental disorders and understanding their origins. And Chapter 17 is about methods to help people overcome or cope with their mental problems and disorders.

part 8

Personality

PERSONALITY AS BEHAVIORAL DISPOSITIONS, OR TRAITS

Trait Theories: Efficient Systems for Describing Personalities
Questions About the Predictive Value of Personality Traits
Genetic Foundations of Personality Traits

PERSONALITY AS ADAPTATION TO LIFE CONDITIONS

Advantages of Being Different from One Another
Adapting to the Family Environment
Adapting to One's Gender

PERSONALITY AS MENTAL PROCESSES I: PSYCHODYNAMIC AND HUMANISTIC VIEWS

Elements of the Psychodynamic Perspective
Contemporary Research on Defense Mechanisms
The Humanistic Perspective: The Self and Life's Meanings

PERSONALITY AS MENTAL PROCESSES II: SOCIAL-COGNITIVE VIEWS

Habitual Ways of Thinking as Personality Traits
The Idea of Situation-Specific Personality Traits
Cross-Cultural Differences in Personality

Personality refers to a person's general style of interacting with the world, especially with other people—whether one is withdrawn or outgoing, excitable or placid, conscientious or careless, kind or stern. A basic assumption of the personality concept is that people do differ from one another in their styles of behavior in ways that are fairly consistent across time and place. Most chapters of this book emphasize the ways in which we are similar to one another, but in this chapter we turn explicitly to differences among us.

Most people are fascinated by human differences. Such fascination is natural and useful. In everyday life we take for granted those aspects of a person that are common to all people, and we focus, instead, on aspects that distinguish one person from another. Attention to differences helps us decide whom we want for partners and friends and how to deal with the different people that we know. *Personality psychologists* make a scientific study of such differences. Using questionnaires and other assessment tools, they conduct research to measure personality differences and explain their origins. They try to relate personality to the varying roles and habitats that people occupy in the social world, and they try to describe objectively the mental processes that underlie the differences. Here we shall examine these endeavors.

The chapter is divided into four main sections. The first section is concerned with the basic concept of personality traits and with questions about their validity, stability, and biological bases. The second is concerned with the adaptive functions of personality: How might individual differences prepare people for life within different niches of the social environment? The third and fourth sections are about the unconscious and conscious mental processes that may underlie and help explain behavioral differences among individuals. In those last two sections you will read of psychodynamic, humanistic, and social-cognitive theories of personality.

chapter 15

Personality as Behavioral Dispositions, or Traits

What are the first three adjectives that come to mind concerning your own personality? Do those adjectives apply to you in all settings, or only in some? How clearly do they distinguish you from other people you know? Do you have any idea why you have those characteristics?

1 What is a trait, and how do traits differ from states? What does it mean to say that a trait is a *dimension* rather than an all-or-none characteristic?

<<< The most central concept in personality psychology is the **trait**, which can be defined as a relatively stable predisposition to behave in a certain way. Traits are considered to be part of the person, not part of the environment. People carry their traits with them from one environment to another, although the actual manifestation of a trait in the form of behavior usually requires some perceived cue or trigger in the environment. For example, the trait of aggressiveness might be defined as an inner predisposition to argue or fight. That predisposition is presumed to stay

with the person in all environments, but actual arguing or fighting is unlikely to occur unless the person perceives provocations in the environment. Aggressiveness or kindness or any other personality trait is, in that sense, analogous to the physical trait of “meltability” in margarine. Margarine melts only when subjected to heat (a characteristic of the environment); but some brands need less heat to melt than others do, and that difference lies in the margarine, not in the environment.

It is useful to distinguish *traits* from *states*. States of motivation and emotion (discussed in Chapter 6) are, like traits, defined as inner entities that can be inferred from observed behavior. But while traits are enduring, states are temporary. In fact, a trait might be defined as an enduring attribute that describes one's likelihood of entering temporarily into a particular state. The personality trait of aggressiveness describes the likeli-

A state is not a trait This man is clearly in a state of anger. However, before we judged him as having an aggressive or hostile personality we would have to know how easily he is provoked into anger.

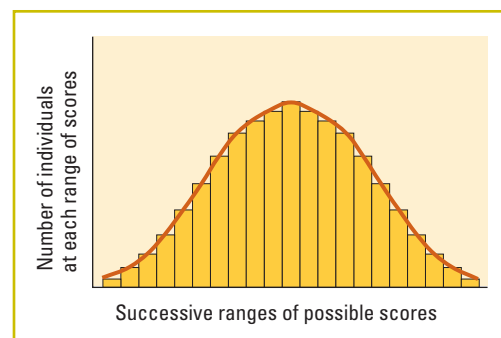
2 Why is a trait considered to be a description rather than an explanation of behavior?

<<< Traits describe how people differ from others on particular dimensions, but they are not themselves explanations of those differences. To say that a person is high in aggressiveness simply means that the person tends to argue or fight a lot, in situations that would not provoke such behavior in most people. The trait is inferred from the behavior. It would be meaningless, then, for me to say that Harry argues and fights a lot *because* he is highly aggressive. That would be essentially the same as saying, “Harry argues and fights a lot because he argues and fights a lot.”



© Anthony Redpath / Corbis

FIGURE 15.1 Typical distribution of scores measuring a personality trait When many individuals are tested for a personality trait—such as boldness versus shyness—the majority usually fall in the middle of the range and the frequency tapers off toward zero at the extremes, in a manner that approximates the bell-shaped *normal distribution* depicted here. (For a more complete description of normal distributions, see the Statistical Appendix at the end of the book.)



Trait Theories: Efficient Systems for Describing Personalities

In everyday life we use an enormous number of personality descriptors. Gordon Allport (1937), one of the pioneers of personality psychology, identified 18,000 such terms in a standard English dictionary. Most of them have varying meanings, many of which overlap with the meanings of other terms. Consider, for example, the various connotations and overlapping meanings of *affable*, *agreeable*, *amiable*, *amicable*, *companionable*, *congenial*, *convivial*, *cordial*, *friendly*, *genial*, *gracious*, *hospitable*, *kind*, *sociable*, *warmhearted*, and *welcoming*. Personality psychologists have long been interested in devising a more efficient vocabulary for describing personality. The goal of any **trait theory** of personality is to specify a manageable set of distinct personality dimensions that can be used to summarize the fundamental psychological differences among individuals.

Factor Analysis as a Tool for Identifying an Efficient Set of Traits

In order to distill all the trait terms of everyday language down to a manageable number of meaningful, different dimensions of personality, trait theorists use a statistical technique called **factor analysis**. Factor analysis is a method of analyzing patterns of correlations in order to extract mathematically defined *factors*, which underlie and help make sense of those patterns. I will illustrate the technique here, with hypothetical data.

The first step in a factor-analytic study of personality is to collect data in the form of a set of personality measures taken across a large sampling of people. For example, the researcher might present a group of people with a set of adjectives and ask each person to indicate, on a scale of 1 to 5, the degree to which each adjective describes himself or herself. For the sake of simplicity, let us imagine a study in which just seven adjectives are used—*carefree*, *compliant*, *dependable*, *hard-working*, *kind*, *rude*, and *trusting*. To get a sense of what it is like to be a subject in such a study, you might rate yourself on each of these traits. For each term write 1 if you think the term is a very poor descriptor of you, 2 if you think it is a somewhat poor descriptor of you, 3 if you think it is neither a poor nor a good descriptor of you, 4 if you think it is a somewhat good descriptor of you, and 5 if you think it is a very good descriptor of you. Be honest in your ratings; you are doing this anonymously (as are the subjects of actual personality studies).

Once the data are collected, the researcher statistically correlates the scores for each adjective with those for each of the other adjectives, using the method of correlation described in Chapter 2. The result is a matrix of correlation coefficients, showing the correlation for every possible pair of scores (illustrated in Table 15.1). As you may recall from Chapter 2, correlation coefficients can be positive or negative. A positive correlation between two measures means that people who score high on one measure also tend to score high on the other. A negative correlation

>>> **3**
How is factor analysis used to identify trait dimensions that are not redundant with one another?

TABLE 15.1

Hypothetical matrix of correlations among adjectives used as personality self-descriptions

Self-description	Compliant	Dependable	Hard-working	Kind	Rude	Trusting
Carefree	−.11	−.60	−.43	+.01	+.21	+.17
Compliant	—	+.29	+.13	+.39	−.70	+.53
Dependable		—	+.49	+.19	−.13	+.17
Hard-working			—	+.08	−.11	+.09
Kind				—	−.65	+.55
Rude					—	−.48
Trusting						—

means that people who score high on one measure tend to score low on the other. The strength of either a positive or negative correlation is indicated by the absolute value of the correlation coefficient. The closer that value is to 1 (or the farther it is from 0), the stronger is the correlation. Table 15.1 shows hypothetical correlation coefficients that might be obtained for the seven self-rating items of our imaginary study.

The next step in a factor analysis is a mathematically complex one called *factor extraction*. Through statistical means, the researcher identifies mathematical factors, which best account for the observed pattern of correlations. Roughly speaking, each factor corresponds to a cluster of measures that correlate relatively strongly (either positively or negatively) with one another and only weakly with items not in that cluster. For the data depicted in Table 15.1, such an analysis would identify two rather clear factors. One factor corresponds most closely with the adjectives *care-free*, *dependable*, and *hard-working*; and the other factor corresponds most closely with the adjectives *compliant*, *kind*, *rude*, and *trusting*. If you examine the table, you will see that within each of these two clusters of adjectives the correlations are relatively strong (the absolute values range from .39 to .70), and between clusters the correlations are relatively weak (the absolute values range from .01 to .29).

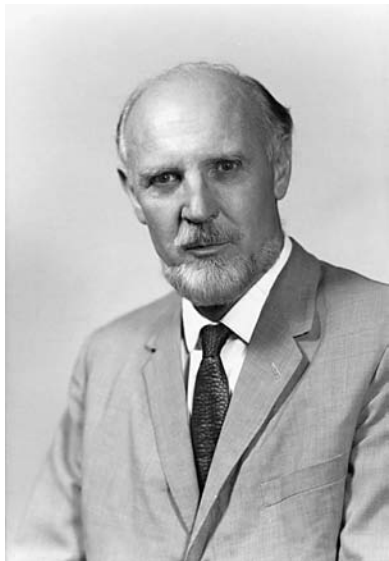
The final step is that of labeling the factors. This is a subjective process, based on the researcher's conception of the combined meanings of the cluster of items that contribute most heavily to the factor being named. In our hypothetical example, the factor that corresponds with *carefree*, *dependable*, and *hard-working* might be referred to as the *conscientiousness* dimension; and the factor that corresponds with *compliant*, *kind*, *rude*, and *trusting* might be referred to as the *agreeableness* dimension. Notice that each of these factor terms refers to a *dimension* of personality. The conscientiousness dimension runs from very low to very high conscientiousness, and the agreeableness dimension runs from very low to very high agreeableness. Items that contribute heavily to one of these trait factors may be at either the high end or the low end of the dimension. For example, *dependable* and *hard-working* represent the high end of the conscientiousness dimension, and *carefree*, which correlates negatively with the other two, represents the low end of that dimension.

What the factor analysis tells us is that these two dimensions of personality are relatively independent of each other. People who are high in conscientiousness are about equally likely to be either high or low in agreeableness, and vice versa. Thus, conscientiousness and agreeableness are useful, efficient trait dimensions because they are not redundant with each other and because each captures at least part of the essence of a set of more specific trait terms.

Cattell's Pioneering Use of Factor Analysis to Develop a Trait Theory

The first well-known trait theory was developed by Raymond Cattell (1950). Cattell's undergraduate degree was in chemistry, and his goal in psychology was to develop a sort of chemistry of personality. Cattell argued that just as an infinite number of different molecules can be formed from a finite number of atoms, an infinite number of different personalities can be formed from a finite number of traits. To pursue this idea, he needed to identify the elements of personality (trait dimensions) and develop a way to measure them.

<<< Cattell began his research by condensing Allport's 18,000 English adjectives describing personality down to about 170 that he took to be logically different from one another. He had large numbers of people rate themselves on each of these characteristics and then used factor analysis to account for patterns of correlation among the 170 different trait terms—a monumental task in the 1940s, before computers. Through this means, Cattell identified a preliminary set of traits and gave each a name. Then he developed various questionnaires aimed at assessing those traits in people and factor-analyzed the questionnaire results.



Courtesy of the University of Illinois Archives

Raymond Cattell Cattell (1905–1998) was a bold innovator in psychology who applied factor analysis to develop a two-factor theory of intelligence (discussed in Chapter 10) and a 16-factor theory of personality.

4

How did Cattell develop his trait theory?

The upshot of his research, which spanned many years, was the identification of 16 basic trait dimensions, listed in Table 15.2, and a questionnaire called the *16 PF Questionnaire* to measure them (Cattell, 1950, 1973). (*PF* stands for *personality factors*.) The questionnaire consists of nearly 200 statements about specific aspects of behavior, such as “I like to go to parties.” For each statement the respondent must select one of three possible answers—*yes*, *occasionally*, or *no*. The 16 PF Questionnaire is still sometimes used today, both as a clinical tool for assessing personality characteristics of clients in psychotherapy and as a research tool in studies of personality.

The Five-Factor Model of Personality

Many trait researchers find Cattell’s 16-factor theory to be overly complex, with redundant factors. Indeed, even as Cattell was finalizing his theory, D. W. Fiske (1949) reported that many of Cattell’s personality traits correlated rather strongly with one another. Fiske’s reanalysis of Cattell’s data resulted in a five-factor solution. Although Fiske’s report lay in relative obscurity for many years, it was the first bit of evidence supporting a five-factor model of personality.

Today, with modern computers, it is much easier to perform large-scale factor-analytic studies than it was at the time of Cattell’s early research. Over the past 35 years or so, hundreds of such studies have been done, with many different measures of personality and many different groups of subjects. Some studies have used self-ratings, based on adjectives or on responses to questions about behavior. Others have used ratings that people make of the personality of someone they know well rather than of themselves. Such studies have been conducted with children as well as with adults, and with people from many different cultures, using questionnaires translated into many different languages (McCrae & Allik, 2002; McCrae & Terracciano, 2005). Overall, the results have been fairly consistent in supporting what has become known as the **five-factor model**, or the “*Big Five*” theory, of personality.

The five-factor model, in its most commonly accepted form, is summarized in Table 15.3. According to the model, a person’s personality is most efficiently described in terms of his or her score on each of five relatively independent global trait dimensions: *neuroticism* (vulnerability to emotional upset), *extraversion* (tendency to be outgoing), *openness to experience*, *agreeableness*, and *conscientiousness*. Nearly all of the thousands of adjectives commonly used to describe personalities correlate at least to some degree with one or another of these five traits. The model also posits that each global trait dimension encompasses six subordinate trait dimensions referred to as *facets* of that trait (Costa & McCrae, 1992; Paunonen & Ashton, 2001). The facets within any given trait dimension correlate with one another, but the correlations are far from perfect. Thus, a detailed description of

>>> **5**
Why is the five-factor model of personality generally preferred over Cattell’s trait theory today?

TABLE 15.2 |
Cattell’s 16 source traits, or personality factors

1. Sociable–unsociable
2. Intelligent–unintelligent
3. Emotionally stable–unstable
4. Dominant–submissive
5. Cheerful–brooding
6. Conscientious–undependable
7. Bold–timid
8. Sensitive–insensitive
9. Suspicious–trusting
10. Imaginative–practical
11. Shrewd–naïve
12. Guilt proclivity–guilt rejection
13. Radicalism–conservatism
14. Self-sufficiency–group adherence
15. Self-disciplined–uncontrolled will
16. Tense–relaxed

Note: In this table, descriptive terms have been substituted for the technical terms that Cattell coined for each trait.

Source: Adapted from *Personality and mood by questionnaire* (pp. 53–54) by R. B. Cattell, 1973, San Francisco: Jossey-Bass.

TABLE 15.3 |
The Big Five personality factors and their 30 facets

- **Neuroticism–stability (N):** Anxious–calm; Angry–placid; Depressed–not depressed; Self-conscious–not self-conscious; Impulsive–controlled; Vulnerable–secure.
- **Extraversion–introversion (E):** Warm–detached; Gregarious–withdrawn; Assertive–unassertive; Active–contemplative; Excitement-seeking–tranquility-seeking; Positive emotions–modulated emotions.
- **Openness to experience–non-openness (O):** Here the six facets refer to openness versus non-openness to experience in each of six realms: Fantasy, Aesthetics, Feelings, Actions, Ideas, and Values.
- **Agreeableness–antagonism (A):** Trusting–suspicious; Straightforward–conniving; Altruistic–selfish; Compliant–noncompliant; Modest–self-aggrandizing; Tender-minded–hard-headed.
- **Conscientiousness–undirectedness (C):** Competent–incompetent; Ordered–disordered; Dutiful–neglectful; Achievement-striving–not achievement-striving; Self-disciplined–not self-disciplined; Deliberative–careless.

Note: In this table, the five major traits and the six facets of each trait are indicated as dimensions, using antonyms to indicate the two ends of each dimension. Usually the traits and facets are referred to using just the first term of each of the antonym pairs shown here.

someone's personality would include not just a score for each of the five global traits but also a score for each of the 30 facets. The facet terms shown in Table 15.3 can help you identify the characteristics that each of the five major trait dimensions includes.

Measurement of the Big Five Traits and Their Facets

6

How are the Big Five traits and their facets measured in individuals?

<<< Over the years, trait theorists have developed many different questionnaires to assess personality. The questionnaire most often used to measure the Big Five traits and their facets is the *NEO Personality Inventory* (where *N*, *E*, and *O* stand for three of the five major traits), developed by Paul Costa and Robert McCrae (1992). This questionnaire is now in its third revision, referred to as the NEO-PI-3 (McCrae & others, 2005). In its full form, the person being tested rates 240 statements on a five-point scale ranging from “strongly disagree” to “strongly agree.” Each statement is designed to assess one facet of one of the five major traits. For example, here are two statements from the NEO-PI-3 designed to assess the sixth facet (values) of openness to experience:

118. *Our ideas of right and wrong may not be right for everyone in the world.*

238. *People should honor traditional values and not question them.*

Agreement with the first statement and disagreement with the second would count toward a high score on the values facet of openness. In the development of such tests, trait theorists submit the results of trial tests to factor analysis and use the results to eliminate items whose scores do not correlate strongly (positively or negatively) with the scores of other items designed to measure the same facet.

At the time that I am writing this, several shortened versions of the NEO-PI are available on the Internet for self-testing. You can probably find one or another of these by using the search term “Big Five personality test” with Google or any other search engine. For one example of research findings using the NEO-PI, see Figure 15.2.

One problem with almost all personality questionnaires, which you will discover if you take such a test, is that the questions are quite transparent. A person who wants to present himself as a particular kind of person can easily do so. The usefulness of the questionnaires depends on the honesty and insight of the respondent about his or her own behavior and attitudes. In some research studies, and for some clinical purposes, personality inventories are filled out both by the person being evaluated and by others who know that person well. Agreement among the different ratings of the same person adds to the likelihood that the ratings are accurate.

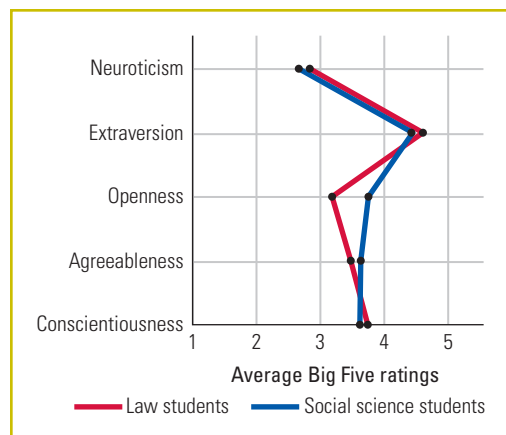


FIGURE 15.2 Personality profiles of law students and social science students at one university
Shown here are the average Big Five personality ratings for students studying law and for students studying social sciences at Tel Aviv University, in Israel. The questionnaire used was a Hebrew version of the NEO Personality Inventory. The individual ratings could range from 1 to 5. The differences between the two groups on openness to experience and agreeableness were statistically significant. (Data from Rubinstein, 2005.)



“When you lie about yourself, is it to appear closer to or farther away from the middle of the bell curve?”

© The New Yorker Collection 2003 William Haefeli from cartoonbank.com. All rights reserved.

Questions About the Predictive Value of Personality Traits

You now have some understanding of how personality psychologists identify and measure traits, and you know that traits describe but do not explain psychological differences. Personality tests are often used to assess a person's suitability for one or another line of work, or as part of a clinical assessment for a person undergoing psychotherapy. How useful are they? Do scores on a personality test allow one to predict what the person will be like tomorrow or 30 years from now? Do they allow one to predict, with any degree of accuracy, how the person meets various types of challenges in the real world?

The Stability of Personality Measures over Time

If you ever have the opportunity to attend the twenty-fifth reunion of your high school class, don't miss it; it is bound to be a remarkable experience. Before you stands a person who claims to be your old friend Marty, whom you haven't seen for 25 years. The last time you saw Marty, he was a skinny kid with lots of hair, wearing floppy sneakers and a sweatshirt. What you notice first are the differences: This Marty has a pot belly and very little hair, and he is wearing a business suit. But after talking with him for a few minutes, you have the almost eerie knowledge that this is the same person you used to play basketball with behind the school. The voice is the same, the sparkle in the eyes, the quiet sense of humor, the way of walking. And when it is Marty's turn to stand up and speak to the group, he, who always was the most nervous about speaking before the class, is still most reluctant to do so. There's no doubt about it—this Marty, who now has two kids older than he was when you last saw him, is the same Marty you always knew.

But that is all impression. Is he really the same? Perhaps Marty is in some sense *your* construction, and it is your construction that has not changed over the years. Or maybe it's just the situation. This, after all, is a high school reunion held in your old school gymnasium, and maybe you've all been transported back in your minds and are coming across much more like your old selves than you normally would. Maybe you're all trying to be the same kids you were 25 years ago, if only so that your former classmates will recognize you. Clearly, if we really want to answer the question of how consistent personality is over long periods, we've got to be more scientific.

Many studies have been conducted in which people fill out personality questionnaires, or are rated on personality characteristics by family members or friends, at widely separated times in their lives. The results indicate a rather high stability of

>>> **7**
What is the evidence that personality is relatively stable throughout adulthood?



© Mary Kate Denny / PhotoEdit, Inc. All Rights Reserved.

Twenty-fifth reunion of the class of 1958 After 25 years, these women may not have recognized each other at first. But once they began talking and gesturing, so that their personalities came through, they most likely began to see each other as the same person they had known as an 18-year-old classmate.

personality throughout adulthood (Caspi & others, 2005). Correlation coefficients on repeated measures of major traits (such as the Big Five) during adulthood typically range from .50 to .70, even with intervals of 30 or 40 years between the first and second tests (Costa & McCrae, 1994). Such stability apparently cannot be dismissed as resulting from a consistent bias in how individuals fill out personality questionnaires, as similar consistency is found even when the people rating the participants' personalities are not the same in the second test as in the first (Mussen & others, 1980).

Such studies also indicate that personality becomes increasingly stable with increasing age up to about age 50, and it remains at a relatively constant level of stability after age 50 (Caspi & others, 2005). One analysis of many studies, for example, revealed that the average test-retest correlation of personality measures across 7-year periods was .31 in childhood, .54 during young adulthood, .64 at around age 30, and .74 between ages 50 and 70 (Roberts & DelVecchio, 2000). Apparently, the older one gets, up to about age 50, the less one's personality is likely to change. But some degree of change can occur at any age.

Patterns of Change in Personality with Age

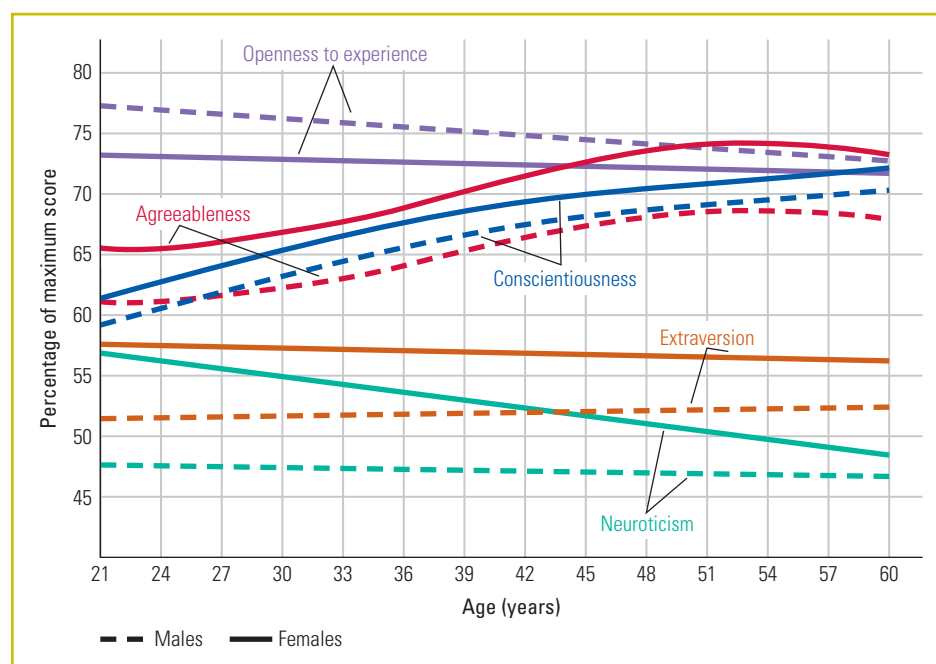
8

In what ways does personality tend to change with age?

<<< Some of the changes in personality that occur with age are relatively consistent across samplings of individuals and constitute what is commonly summarized as increased maturity. Studies in many different cultures on measures of the Big Five indicate that, over the adult years, neuroticism and openness to experience tend to decline somewhat, and conscientiousness and agreeableness tend to increase somewhat, at least up to age 50 (Caspi & others, 2005; Srivastava & others, 2003). Data from one such study are shown in Figure 15.3.

In addition to the general age trends, other research indicates that an individual's personality can change, at least to some degree, in any direction, at any age, in response to a major life change, such as a new career, an altered marital status, or the onset of a chronic illness (Caspi & Roberts, 1999; Helson & Stewart, 1994). Other studies suggest that people who have a particular personality characteristic often make life choices that alter their personality even further in the preexisting direction (Caspi & others, 2005; Roberts & Robins, 2004). For example, a highly extraverted person may choose a career that involves a lot of social activity, which may cause the person to become even more extraverted than before.

FIGURE 15.3 Average Big Five scores vary somewhat with age and sex This graph depicts the mean score for men and women on each of the Big Five traits as a function of age. The results are derived from a sample of 132,515 adults, aged 21–60, who completed a version of the Big Five Personality Inventory that was posted on the Internet. Other studies, using more systematically chosen samples, have shown similar results. (Adapted from Srivastava & others, 2003, p. 1047.)



The Relationship of Personality Measures to People's Actual Behavior

Do scores on personality tests predict anything useful about people's behavior in the real world? It is theoretically possible (as discussed in Chapter 2) for a test to be highly *reliable* (consistent from one test to the next) but not at all *valid*. A test is valid to the degree that its scores are true measures of the characteristics they were meant to measure. A personality test is valid to the degree that scores on each of its trait measures correlate with aspects of the person's real-world behavior that are relevant to that trait. Countless studies have shown that personality tests such as the NEO-PI are, to at least some degree, valid. Here are some examples of correlations—between personality measures and actual behaviors—that demonstrate the validity of the personality measures:

- People who score high on *neuroticism*, compared to those who score lower, have been found to (a) pay more attention to, and exhibit better recall of, threats and unpleasant information in their environment (Matthews & others, 2000); (b) manifest more distress when given a surprise math test (Schneider, 2004); and (c) be much more susceptible to mental disorders, especially depression and anxiety disorders (Hettema & others, 2004).
- People who score high on *extraversion*, compared to those who score as introverts, have been found to (a) attend more parties and be rated as more popular (Paunonen, 2003); (b) be more often seen as leaders in their fraternity, sorority, or dormitory groups (Anderson & others, 2001); (c) live with and work with more people (Diener & others, 1992); and (d) be less disturbed by sudden loud sounds or other intense stimuli (Eysenck, 1990; Geen, 1984; Stelmack, 1990).
- People who score high on *openness to experience*, when compared to those who score lower, have been found to (a) be more likely to enroll in liberal arts programs rather than professional training programs in college (Paunonen, 2003); (b) change careers more often in middle adulthood (McCrae & Costa, 1985); (c) perform better in job training programs (Goodstein & Lanyon, 1999); (d) be more likely to play a musical instrument (Paunonen, 2003); and (e) exhibit less racial prejudice (Flynn, 2005).
- People who score high on *agreeableness*, compared to those who score lower, have been found to (a) be more willing to lend money (Paunonen & Ashton, 2001); (b) have higher school grades and fewer behavior problems during childhood (Laursen & others, 2002); and (c) manifest less alcoholism or arrests in adulthood (Laursen & others, 2002).
- People who score high on *conscientiousness*, compared to those who score lower, have been found to (a) be more sexually faithful to their spouses (Buss, 1996); (b) receive higher ratings for job performance (Goodstein & Lanyon, 1999); and (c) smoke less, drink less, drive more safely, follow more healthful diets, and live longer (Bogg & Roberts, 2004; Friedman & others, 1995).

All such findings demonstrate that people's answers to personality test questions reflect, relatively accurately, the ways they actually behave and respond to challenges in the real world.

Personality differences do not reveal themselves equally well in all settings. When you watch people in familiar roles and settings, conforming to learned, socially agreed-upon behavioral norms—at their jobs, in the classroom, or at formal functions such as weddings and funerals—the common influence of the situation may override individual personality styles. Avshalom Caspi and Terrie Moffitt

How do researchers assess the validity of personality tests? What are some sample findings that show that measures of the Big Five personality traits are, to at least some degree, valid?



Bob Daemrich / The Image Works

A stressful situation The first day on campus for first-year students is the kind of event that brings out personality differences.

Why might personality traits be most apparent in novel situations or life transitions?

(1991, 1993) have suggested, with some evidence, that personality differences are most clearly revealed when people are in novel, ambiguous, stressful situations and in life transitions, where there are no or few cues as to what actions are appropriate. According to Caspi and Moffitt (1993), in the absence of cues as to how to behave, “the reticent become withdrawn, the irritable become aggressive, and the capable take charge.”

Genetic Foundations of Personality Traits

Where do traits come from? To some degree, at least, global traits such as the Big Five appear to derive from inherited physiological qualities of the nervous system.

The Heritability of Traits

11

How have researchers assessed the heritability of personality traits? What are the general results of such studies?

<<< How heritable are personality traits? As discussed in Chapter 10, *heritability* refers to the degree to which individual differences derive from differences in genes rather than from differences in environmental experiences. Numerous research studies—using the behavior-genetics methods described in Chapter 10—have shown that the traits identified by trait theories are at least moderately heritable. The most common approach in these studies has been to administer standard per-

sonality questionnaires to pairs of identical twins and fraternal twins (who are no more similar genetically than are ordinary siblings). The usual finding is that identical twins are much more similar than are fraternal twins on every personality dimension measured, similar enough to lead to an average heritability estimate of roughly 0.40 to 0.55 for most traits, including all the Big Five (Bouchard, 2004; Loehlin, 1992). As explained in Chapter 10, a heritability of 0.50 means that about 50 percent of the variability among individuals results from genetic differences.

Such findings have been criticized on the grounds that parents and others may treat identical twins more similarly than fraternal twins, and similar treatment may lead to similar personality. To get around that possibility, researchers at the University of Minnesota, led by David Lykken, gave personality tests to twins who had been separated in infancy and raised in different homes, as well as to twins raised in the same home (Bouchard, 1991, 1994; Tellegen & others, 1988). Their results were consistent with the previous studies: The identical twins were more similar to each other than were the fraternal twins on essentially every measure, whether they had been raised in the same home or in different homes, again leading to heritability scores averaging close to 0.50. Subsequent studies, by other researchers in several different countries, have produced similar results (Plomin & Caspi, 1999).



“I was a good boy, grandpa was a good boy, his father was a good boy. In fact, since the dawn of history, there have only been good boys in this family. That’s why you have to be a good boy.”

12

What evidence suggests that being raised in the same family does not promote similarity in personality?

<<< In the past it was common for psychologists to attribute personality characteristics largely to the examples and training that people gain from their mothers and (to a lesser degree) their fathers (Harris, 1998). A common assumption was that people raised in the same family have similar personalities not just because of their shared genes but also because of their shared family environment. Perhaps the most surprising finding in the Minnesota study, confirmed since by other studies, is that this assumption is not true. Being raised in the same family has an almost negligible effect on measures of personality (Bouchard, 2004; McGue & Bouchard, 1998; Turkheimer & Waldron, 2000). Twin pairs who had been raised in different families were, on average, as similar to—and as different from—each other as were twin pairs who had been raised in the same family.

The contradiction between this finding and long-standing beliefs about the influence of parents is dramatically illustrated by some of the explanations that

twins gave for their own behavioral traits (Neubauer & Neubauer, 1996). When one young man was asked to explain his almost pathological need to keep his environment neat and clean, he responded, "My mother. When I was growing up she always kept the house perfectly ordered. . . . I learned from her. What else could I do?" When his identical twin brother, who had grown up in a different home in a different country but was also compulsively neat and clean, was asked the same question, he said, "The reason is quite simple. I'm reacting to my mother, who was a complete slob!" In this case the similarity of the twins in their compulsive tendency was almost certainly the result of their shared genetic makeup, yet each one blamed this tendency on his adoptive mother.

The results of such studies of twins do not mean that the environment is unimportant in personality differentiation, or even that the family environment is unimportant. The results do imply, however, that those aspects of the family environment that contribute to personality differentiation are typically as different for people raised in the same family as they are for people raised in different families. Two children raised in the same family may experience that environment very differently from each other. Sandra Scarr and her colleagues (1981) earlier came to a similar conclusion through a different route. They compared non-twin, adopted children with both their biological and their adoptive siblings and found that biological siblings were far more similar to each other than were adoptive siblings. In fact, for most personality measures, they found that the children were no more similar to their adoptive siblings than any two randomly compared children were to each other. Later in the chapter we will examine some possible causes of the large individual differences in personality found among children raised in the same family.



Aaron Haupt / Stock, Boston

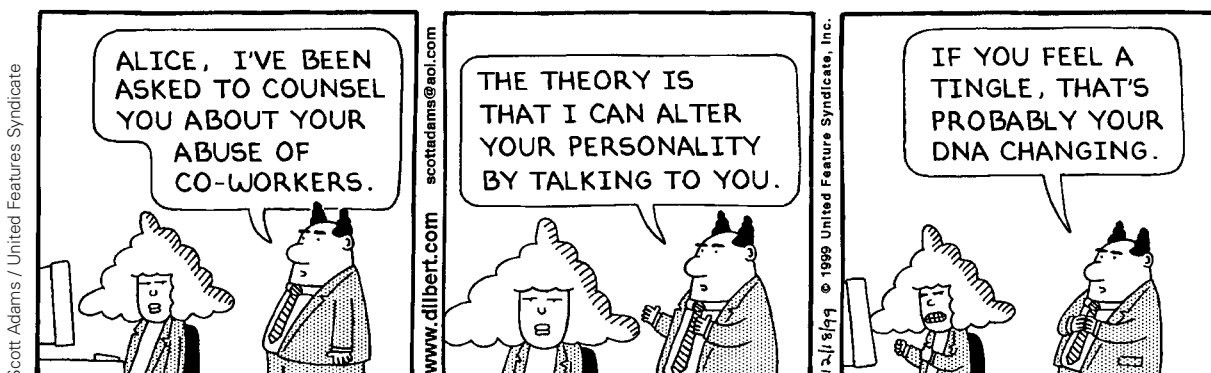
Identical twins To study the heritability of personality, researchers at the University of Minnesota located identical twins who had been adopted at an early age and raised in different homes, often in different parts of the country. They paid the twins to come to Minnesota for study, and, as a side effect of the study, reunited twins who had long been separated.

Single Genes and the Physiology of Traits

If genes affect personality, they presumably do so primarily by influencing the physiological characteristics of the nervous system in one way or another. One likely route by which they might exert such effects is through their influence on neurotransmission in the brain. In recent years, several laboratories have reported small but significant correlations between specific personality characteristics and specific genes that alter neurotransmission.

For example, some researchers have found a significant relationship between neuroticism and a gene allele that increases the action of the neurotransmitter serotonin (Lesch, 2003; Lesch & others, 1996). Others have found a significant relationship between the trait of novelty seeking—which includes elements of impulsiveness, excitability, and extravagance—and an allele that decreases the action of the neurotransmitter dopamine (Benjamin & others, 1996, 1998; Ebstein & others, 2003). Attempts to replicate such findings, however, have thus far produced inconsistent results (Ebstein & others, 2003). It seems inevitable that, with time, researchers will find direct relationships between the brain effects of individual genes and personality variations, but the picture is not yet clear.

>>> **13**
How might variation in single genes influence personality?



SECTION REVIEW

The trait—a relatively stable behavioral predisposition—is a key concept in personality.

Trait Theories

- Factor analysis provides a mathematical means to identify an efficient, non-redundant set of personality traits based on correlations among larger sets of such traits. Each trait is a continuous dimension: A person can score high, low, or anywhere in between on any given trait.
- Cattell pioneered this approach, producing a theory with 16 basic traits.
- The most widely accepted trait theory today posits five major traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), each with six subordinate traits called facets.
- Questionnaires designed to measure individuals on the Big Five traits or other traits all require honesty and insight from the respondent to yield accurate results.

Predictive Value of Traits

- Studies show that adult personality is relatively stable. Correlation coefficients for repeated tests, even many years apart, range from .50 to .70.
- Personality does change, however. Increased age is typically accompanied by increased conscientiousness and agreeableness and decreased neuroticism and openness to experience.
- An individual's personality can change to some extent in any direction at any age in response to life changes.
- All of the Big Five traits have been shown to predict behavior at better than chance levels, which helps to establish the validity of the personality measures.

Genetic Basis of Traits

- Studies comparing pairs of identical and fraternal twins yield heritability estimates for personality averaging about 0.50.
- The personalities of biological relatives raised in the same family are generally no more similar than those of equally related people raised apart.
- Researchers are searching for specific gene alleles that contribute to particular traits, but results so far have been inconsistent.

14

How does an ultimate explanation of personality variability differ from a proximate one?

Personality as Adaptation to Life Conditions

Why are people different from one another in personality? As you may recall from Chapter 3, psychologists and biologists distinguish between two different types of answers to *why* questions. One type, referred to as *proximate explanation*, focuses on causal mechanisms that operate in the lifetime of the individual to produce the phenomenon in question. Proximate explanations of personality differences focus on ways by which differing genes and experiences work to make us different. The other type of answer, referred to as *ultimate explanation*, focuses on function, or evolutionary survival value, rather than mechanisms. How might personality differences help individuals survive longer and produce more offspring than they would if all individuals were identical in personality? Why were genetic, developmental, and learning mechanisms that ensure diversity favored by natural selection over mechanisms that would have made us more uniform? These are the questions to which we turn now.

Advantages of Being Different from One Another

It is conceivable that variability in personality has no adaptive advantage. Perhaps the kinds of differences that personality theorists study simply represent a degree of randomness that could not be weeded out by natural selection. Perhaps; but I doubt it.

As noted in Chapter 3, sexual reproduction itself seems to be an adaptation that ensures the diversity of offspring. Mixing half of one parent's genes in a random manner with half of the other parent's genes leads to an essentially infinite number of possible new gene combinations in the offspring. From a purely biomechanical point of view, cloning is a far simpler and less costly form of reproduction than is sexual reproduction, so it is hard to imagine why sexual reproduction would have evolved and persisted if there were no advantage in diversity. It seems equally unlikely that learning mechanisms and other developmental mechanisms

that lead us to become different in response to different environmental conditions would have evolved if they produced no survival advantage. Research on dozens of species of other animals, ranging from ants to octopuses to chimpanzees, has revealed, in every species tested, individual differences in behavioral styles that can be described in terms similar to those that are used to describe human personality differences (Gosling, 2001). Personality appears to be a basic, biological aspect of animal existence.

Diversifying One's Investment in Offspring

One way to think about the value of genetic diversity and personality differences is through an analogy between producing offspring and investing money (Miller, 1997). Investors who put all their money into one company risk going broke if that company suddenly collapses. Smart investors diversify: They invest in a variety of stocks, bonds, and other financial instruments, which are likely to respond differently to economic forces, so that when one goes down, another goes up or at least doesn't go down as rapidly as the first. Diversified investment greatly reduces the potential for dramatic loss while maintaining the potential for substantial gains over the long run.

From the perspective of evolution by natural selection, producing offspring is an investment, the goal of which is to send multiple copies of one's genes into future generations. Since conditions of survival and reproduction vary in unpredictable ways over time, the chance that an individual's genes will die out can be reduced if the offspring differ from one another in some of their characteristics. Therefore, over the course of evolution, mechanisms that ensure diversity of offspring—even the random diversity that results from genetic mixing in sexual reproduction—would be favored by natural selection.

Studies of the Bold-Cautious Dimension in Fish

The most commonly studied dimension of personality in animals is that of boldness versus cautiousness (Gosling, 2001). This dimension has been especially well documented and studied in various fish species, including pumpkinseed sunfish (Wilson, 1998), perch (Magnhagen & Staffan, 2005), and three-spined sticklebacks (Ward & others, 2004).

In one series of studies, David Wilson and his colleagues (1993; Wilson, 1998) categorized pumpkinseed sunfish as bold or cautious on the basis of their tendency to approach (and get caught in) or to avoid wire traps placed in their habitat in a small experimental pond. Those caught by the trap were considered bold and those that avoided the traps (and were later caught by seining) were considered cautious. Subsequent study of the two groups of fish in laboratory tanks demonstrated the consistency and generality of the trait difference. Compared with the cautious fish, the bold fish (a) adapted more readily to the tank, as shown by a quicker return to normal feeding, (b) more frequently swam off on their own rather than staying close to the other fish, and (c) were more likely to approach rather than avoid a human observer. In addition, analysis of the stomach contents of freshly caught bold and cautious fish revealed a striking difference in their diets. The bold fish had eaten far more copepods (tiny crustaceans), which are most plentiful in the shoreline areas where vegetation is least dense and the risk of predation by fish and birds is greatest.

One can well imagine the advantages and disadvantages of being either cautious or bold, among pumpkinseeds or any other vertebrate species. Caution reduces the risk of sudden death through predation or accident, but it also confines one's life to a relatively narrow portion of the available environment. Boldness allows one to explore new areas and locate foods that cautious individuals would not find, though at increased risk. Consistent with this view is the finding, with stickleback fish, that bold individuals eat more and grow more rapidly than do cautious individuals (Ward & others, 2004). Boldness may be especially valuable when the narrow niche

>>> 15

How does an analogy to financial investment explain the value of producing offspring who differ from one another in personality?

>>> 16

How has the bold-cautious dimension been studied in fish? What findings support the view that there is an evolutionary advantage in diversity?

© Galen Rowell / Corbis



A bold bear Animals of many species vary on a bold-cautious dimension. In places where wild animals are protected, such as Yosemite National Park, the bold individuals do especially well, resulting in a rather rapid evolution toward increased boldness. One result is a great increase in the number of cars that are broken into by bears looking for food.

occupied by cautious individuals is nearly filled, so the risks entailed in exploring new objects, areas, and life strategies are offset by the potential for finding new, needed resources. Parents who produce offspring that differ from one another in their degree of caution or boldness are hedging their bets, much like investors who invest in a combination of conservative and high-risk bonds and stocks.

There is also evidence that individual fish can become either bolder or more cautious, depending on the number of other bold or cautious fish in their environment. In one experiment, with perch, cautious fish became bolder when they were grouped, in large aquaria, only with other cautious fish; and bold fish became more cautious when grouped only with other bold fish (Magnhagen & Staffan, 2005). Such context-driven changes in personality are apparently adaptive: If the “cautious niche” is filled by many cautious individuals, then it is better to become bolder and explore new realms. Conversely, if the “bold niche” is filled by many bold individuals, who are spending little of their time in the relatively safe areas, so food is still available in the safe areas, then it is better to be cautious and remain where it is safe.

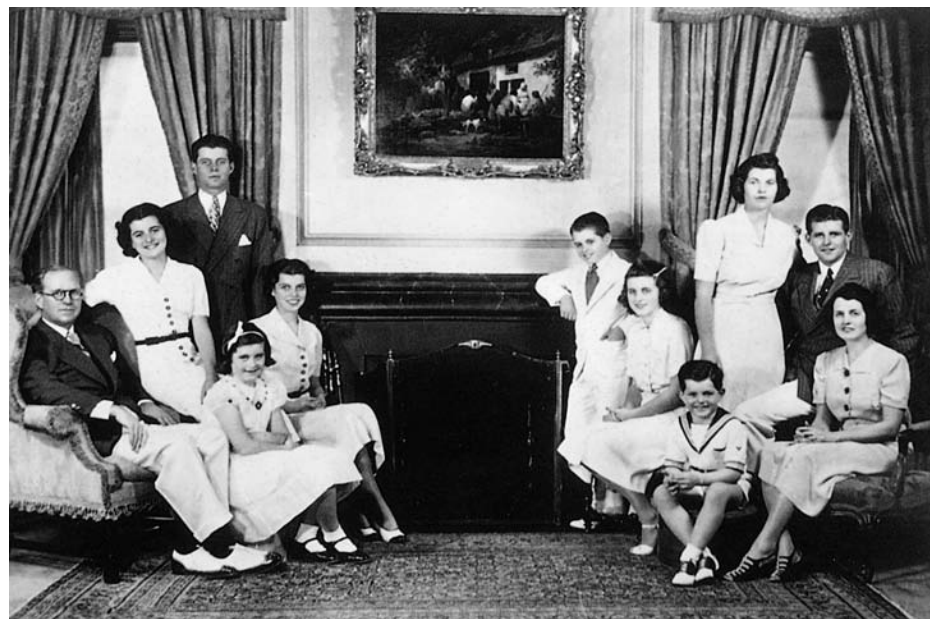
Personality Traits as Alternative Problem-Solving Strategies

17

How might both heritable and non-heritable variations in the Big Five dimensions be explained in terms of alternative life strategies?

<<< From an evolutionary perspective, personality traits in humans can be thought of as alternative general strategies for solving problems related to survival and reproduction. Consistent with this view, David Buss (1996) has found significant correlations between many measures of the Big Five trait dimensions and mating strategies. For instance, variation in conscientiousness correlates with one's tendency to be sexually faithful or unfaithful to one's mate, and variation in agreeable-

A family famous for boldness The members of the Kennedy clan are well known for their willingness to take risks in their political and personal lives. This photograph, taken in 1938, shows Joseph Kennedy (far left) with his wife Rose (far right) and their children (left to right): Patricia, John (who became the thirty-fifth president of the United States), Jean, Eunice, Robert, Kathleen, Edward, Rosemary, and Joseph. The subsequent histories of these people and of their own children offer repeated examples of triumphs and tragedies resulting from risky actions.



© Topham / The Image Works

ness correlates with one's tendency to become aggressive, or not, toward sexual rivals. As an exercise, you might turn back to Table 15.3 (p. 541) and think of ways in which each of the Big Five traits may influence one's approach to a wide range of problems that can enhance or reduce one's ability to survive and reproduce, depending on conditions that are not always predictable. From an evolutionary perspective, what might be the advantages and disadvantages of being near either end of each trait dimension? Your speculations will no doubt generate many hypotheses, and you might think further about how such hypotheses could be tested through research.

The finding that individual differences on trait dimensions are partly heritable and partly the product of environmental experience is consistent with the evolutionary perspective. In an environment that varies unpredictably, the chances that at least some offspring will survive are enhanced if the offspring are genetically inclined toward different life strategies. Genetic inclinations need to be flexible, however, so that individuals can move along a personality dimension in a direction compatible with their life situation. Perch that are genetically inclined to be cautious and people who are genetically inclined toward introversion can become more bold and extraverted if their condition of life promotes or requires such adaptation. (If I had to make my living selling books rather than writing them, then even I, a prototypical introvert, might become a bit less introverted.)

Adapting to the Family Environment

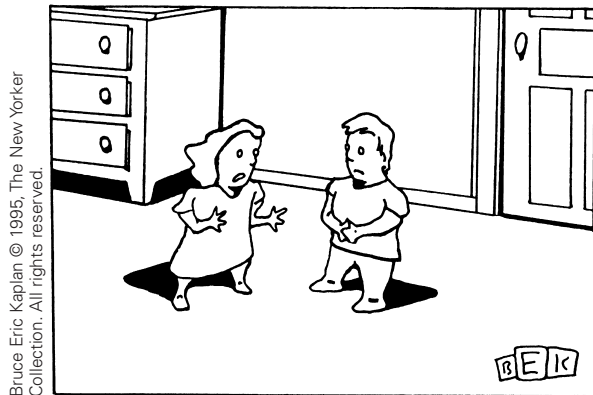
The first social environment to which most new human beings must adapt is that of the family into which they are born. The first view of other people, the first attempts at communication, the first emotional attachments, the first strivings for acceptance and approval, and the first competitions with others for resources occur most often within the context of the family. Children come into the world with nervous systems predisposed for behaving in certain ways, but those ways of behaving are first exercised and built upon within the family. Not surprisingly, personality psychologists from Sigmund Freud on have emphasized the role of the early family environment in the development of each individual's unique personality.

Some psychologists contend that the family may not play as big a role in personality development as was previously believed (Harris, 1998; Rowe, 1994). That contention stems primarily from the research indicating that people who were raised in the same home typically score, on average, just about as differently from one another on personality tests as do equally related people who were raised in different homes. To the degree that siblings are similar to one another in global personality traits, that similarity seems to come mainly from their shared genes, not from their shared home environment.

One response to such findings has been to suggest that siblings raised in the same home, by the same parents, might nevertheless experience quite different environments (Dunn & Plomin, 1990; Vernon & others, 1997). Some differences may stem from chance events: One child has an accident that leads to extensive care from the parents which is not given to the other, or one child is assigned to a warm and caring teacher at school and another is assigned to a harsh one. Other differences may stem from choices made by the siblings themselves. Preexisting differences in personality may lead them to choose very different sets of friends and activities, which may lead them to diverge even further in personality. Preexisting personality differences can also lead them to interpret the same objective events quite differently, with different consequences for further personality development: A parent's consistent demand that chores be done might be understood by one child as a caring act designed to instill good work habits and by another as harsh discipline. Judy Dunn and Robert Plomin (1990) have summarized evidence that siblings indeed do have different experiences for all these kinds of reasons.

>>> 18

What are some ways by which the personality-forming experiences of children raised in the same family may differ?



Bruce Eric Kaplan © 1995, The New Yorker Collection. All rights reserved.

"And then, as soon as I had carved out my niche, they went and had another kid."

Sibling Contrast: Carving Out a Unique Niche Within the Family

Differences between siblings growing up in the same family may become exaggerated because siblings tend to define themselves as different from one another and to accentuate their differences through their behavioral choices (Lalumière & others, 1996). When people are asked to describe their brother or sister, they commonly begin by saying, in effect, "Oh, he (or she) is entirely different from me," and then proceed to point out how (Dunn & Plomin, 1990). Parents likewise tend to focus more on differences than on similarities when they describe two or more of their children (Schachter, 1982). This within-family emphasis on the differences between siblings is referred to as *sibling contrast*. Possibly related to sibling contrast is a strong tendency for siblings with two parents to identify more strongly with different parents; this is referred to as *split-parent identification*. If the first child identifies more strongly with the mother, the second typically identifies more strongly with the father, and vice versa. Sibling contrast and split-parent identification have been documented as highly reliable phenomena in questionnaires filled out by parents as well as in those filled out by the siblings themselves (Schachter, 1982).

19

How might sibling contrast and split-parent identification be useful in reducing sibling rivalry and diversifying parental investment?

Why do family members accentuate the differences rather than the similarities between siblings? A possible answer, proposed by Frances Schachter (1982), is that sibling contrast and split-parent identification are devices by which parents and children consciously or unconsciously strive to reduce sibling rivalry, which can be highly disruptive to family functioning. If siblings are seen by themselves and their parents as having very different abilities, needs, and dispositions, then the siblings are less likely to compete with one another and more likely to be valued and rewarded separately for their unique characteristics. If Joan is understood to be reserved and scholarly and her sister Mary is understood to be outgoing and athletic, then Joan and Mary can be appreciated by their parents and by each other for their separate traits rather than viewed as competitors on the same dimension—a competition that one of them would have to lose. From an evolutionary perspective, such differentiation may promote the survival of the two siblings and other members of their family not only by reducing rivalry but also by diversifying the parental investment (Lalumière & others, 1996; Sulloway, 1996). To the degree that Joan and Mary move into different life niches, they will compete less with each other for limited resources both within the family and in

the larger environment outside the family, and they may develop separate skills through which they can help each other and the family as a whole.

In support of the view that sibling contrast and split-parent identification serve to reduce sibling rivalry, Schachter (1982) found that both phenomena are much stronger for adjacent pairs of siblings than for pairs who are separated in birth order (see Figure 15.4) and stronger for same-sex pairs of siblings than for opposite-sex pairs. Other researchers have also found

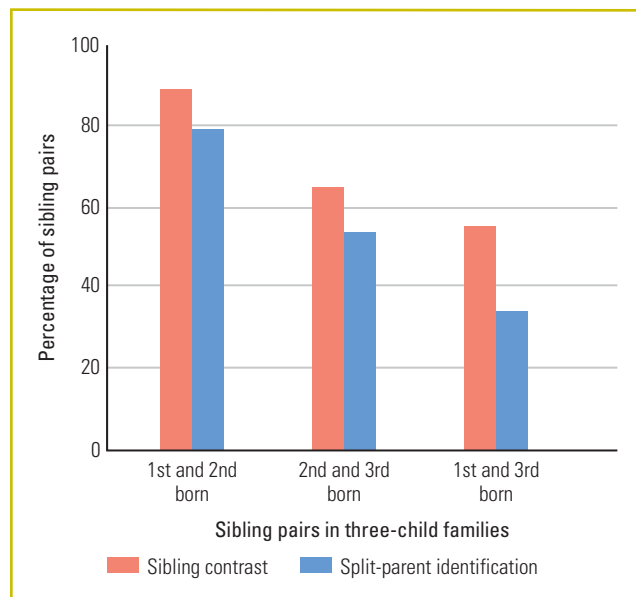


FIGURE 15.4 Sibling contrast and split-parent identification in three-child families Mothers were asked to compare pairs of their children by stating whether the two were similar or different on various personality dimensions and by stating which parent each child identified with more strongly. *Sibling contrast* (orange bars) refers to the percentage of times that the mothers said "different" rather than "alike" in response to personality questions, and *split-parent identification* (blue bars) refers to the percentage of times two siblings were said to identify with different parents rather than with the same parent. (Data from Schachter, 1982, pp. 128 and 141.)

greater sibling contrast for siblings who are close in age than for those who are more distant in age (Feinberg & Hetherington, 2000). It seems reasonable that siblings who are the same sex, adjacent to each other in birth order, and close in age would be most subject to implicit comparisons and possible rivalries and would therefore have the greatest need to reduce rivalry through contrast and split-parent identification. The observation (shown in Figure 15.4) that sibling contrast and split-parent identification are strongest for the first two children in a family is likewise consistent with the rivalry-reduction hypothesis. The first two children are likely to experience the greatest degree of sibling rivalry because, for a period of time, they are the only two children being compared. When third and subsequent children come along, multiple comparisons can be made, so the intensity of comparisons between any two, and the potential for rivalry between them, may be somewhat muted.

Influence of Birth Order on Personality

With the exception of twins, children within a family all differ from one another in one seemingly important way: birth order. Birth order certainly affects children's lives within the family. Earlier-born children, simply by virtue of being older, are, at least for most of their childhoods, bigger, stronger, more knowledgeable, and more competent than their later-born siblings. Because of this, earlier-born children are likely to dominate later-borns and may often be called upon to help care for them. Firstborns in particular occupy a special position. They are *first*: first in competitions, first in receiving new privileges and responsibilities, first to grow up. Do these differences in childhood experiences have lasting consequences for personality? One person who thinks they do is Frank Sulloway, a historian of science who became interested in this topic as a result of his interest in scientific revolutions.

Sulloway wondered why, at any given point in history, some people support and others oppose the new, revolutionary scientific theories of their time—theories such as Copernicus's idea that the earth revolves around the sun, Darwin's idea of evolution by natural selection, and Einstein's concept of special relativity. With the help of other historians of science, he identified hundreds of scientists and other intellectuals who, over the past 500 years of Western history, were on record as either supporting or opposing a particular revolutionary or innovative scientific idea of their time. He then classified these people according to birth order and found, overall, that later-borns were far more likely to support such ideas than were firstborns. For example, the odds that a later-born person relative to a firstborn supported the new scientific idea were 5.4 to 1 for Copernicus's theory and 4.6 to 1 for Darwin's theory. In a subsequent study, Sulloway also found that firstborns were disproportionately likely to support conservative political causes and later-borns were disproportionately likely to support liberal political causes.

Largely on the basis of these historical studies, Sulloway (1996) developed the theory that children's experiences within the family have long-lasting effects, such that firstborns tend to be conservative and traditional throughout their lives and later-borns tend to be more open to new ideas and more likely to rebel against established ways. Because firstborns are, for a while, the only child in their family, they identify strongly with their parents, and in doing so they develop strong respect for authority and for conventional ways of

A brilliant, rebellious lastborn Voltaire, the eighteenth-century satirist and leading critic of the Catholic Church—whose literary work sometimes led to periods of imprisonment or exile—was, according to Sulloway, reacting against his sanctimonious, pious eldest brother, his father's favorite son.

>>> **20**
What birth-order differences in personality did Sulloway find in his analyses of historical data? How did he explain these differences?



Archivo Iconografico, S. A. / Corbis

doing things. Later-born children enter a family that already has at least one older child. They cannot compete with the firstborn child at the activities that are most favored by their parents, so they are motivated to explore new activities and new ways of thinking. The status quo does not favor them, so they look for ways to shake things up. They also tend to identify with the underdog; they know what it is like to be dominated by their more powerful older siblings. Such forces, according to Sulloway, lead later-borns to be much more supportive than firstborns of new ideas and liberal political causes.

21

How have studies correlating personality with birth order cast some doubt on Sulloway's theory?

<<< Over the past several decades, psychologists have conducted dozens of research studies on the relationship between birth order and scores on standard personality tests. In some of these studies, consistent with Sulloway's theory, firstborn children were found, on average, to be higher in conscientiousness and lower on openness to experience than later-born children (Paulhus & others, 1999; Sulloway, 1996). But, compared to the very large effects found by Sulloway in his historical studies, these effects are quite small and inconsistent. Indeed, many studies have shown no effects at all of birth order on personality (Ernst & Angst, 1983; Steelman & others, 2002).

At present it is not clear how to reconcile Sulloway's findings and theory with the relative lack of birth-order effects found in other studies. Perhaps the effects found by Sulloway depend on family practices that vary from place to place and from time to time in history. Sulloway's historical study was primarily of famous men who grew up in upper-class homes a century or more ago. It is entirely possible that family practices that applied to those men—such as primogeniture (the practice of leaving most or all of the parents' wealth to the eldest son)—affected personality in ways that practices in most families today do not. It is also possible that people who eventually become famous are different from the rest of us, in ways that include a greater susceptibility to birth-order effects on personality. For example, they may be people who are strongly driven to high achievement, so birth-order effects on potential ways of achieving—and ways of thinking—may act more strongly on them than on the rest of us.

Adapting to One's Gender

Human beings, like other sexually reproducing animals, come in two varieties, female and male. If you have read the previous chapters of this book, you have already read a good deal about sex differences. For instance, in Chapter 3 you read of the different problems that females and males must solve in mating and of their possible different strategies for solving them. In Chapter 12 you read of differences in how the two sexes are treated, how they are expected to behave, and how they typically do behave at each life stage. As is pointed out in Chapter 12, psychologists generally use the word *sex* to refer to the clear-cut biological differences between males and females and the word *gender* to refer to the differences that may be heavily influenced by cultural forces and expectations.

One way to think about gender is in terms of niches to which we must adapt. The female and male niches are defined partly by biological sex: Females play one role in sexual reproduction, males another. The gender niches are further defined and elaborated upon socially in ways that vary at least somewhat from culture to culture. Females and males are viewed by the culture as having different attributes and being best suited for different social roles. We choose neither our sex nor our gender. Pure chance normally decides sex—X, you're a female; Y, you're a male. Then, throughout life, most of us live with that assignment. What consequences might all this have for personality?

Some Gender Differences in Personality

22

What differences have researchers found between women and men in personality traits?

<<< Standard personality tests do reveal relatively consistent differences, small to moderate in size, between men and women in average scores on many personality traits (Costa & others, 2001; Feingold, 1994; McCrae & Terracciano, 2005). One of the largest and most consistent of such differences, found in essentially every



© Janine Wiedel / Photolibrary / Alamy



© George Disario / Corbis

culture where tests have been conducted, is that women score higher than men in agreeableness. This result is in line with the results of many other studies, using a variety of measures, showing that women are, on average, more concerned than men with developing and maintaining positive social relationships (Kashima & others, 1995; Taylor & others, 2000).

Relatively consistent gender differences have also been observed in neuroticism, primarily because women report higher levels of anxiety and feelings of vulnerability than do men. Women also generally score slightly but significantly higher on conscientiousness. Results on the other two Big Five trait dimensions are more varied. Women tend to score considerably higher than men on the warmth and gregariousness facets of extraversion, but considerably lower on the excitement-seeking facet of that trait (McCrae & Terracciano, 2005). Women also tend to score considerably higher than men on the feelings and aesthetics facets of openness to experience, but not on the other four facets (fantasy, actions, ideas, and values) of that trait (McCrae & Terracciano, 2005).

Gender not only influences the kind of personality one develops but also affects the relationship of personality to life satisfaction (Rothbart & Bates, 1998). For example, researchers have found that shyness or behavioral inhibition correlates positively with feelings of emotional distress and unhappiness in young men but not in young women (Gest, 1997). That difference is not present in early childhood but seems to emerge in adolescence and reach a peak in young adulthood. A possible explanation lies in cultural expectations that make life more difficult for shy or inhibited young men than for similarly shy or inhibited young women. In our culture and all others that have been studied, men are expected to initiate romantic and sexual relationships and to be more assertive or dominant in social interactions of all types, and shyness or lack of assertion is considered more attractive in women than in men. In other respects, too, personality dispositions that run counter to stereotypes can have emotional costs. For instance, women who have a relatively competitive orientation toward others, rather than an orientation emphasizing similarity and agreeableness, typically score lower on measures of self-esteem than do other women, while the opposite is true for men (Josephs & others, 1992).

As noted earlier, personality measures are merely descriptions, not explanations, of psychological differences among people. What accounts for the just-described average differences between women and men on personality measures?

A gender difference in personality

Closeness, caring, and intimacy come more easily to girls and women than to boys and men. Such gender differences may be predisposed by biological sex differences and magnified by cultural learning.

>>> 23

How might gender alter the relationship between personality and life satisfaction?



"You've had to be kind to survive, and I've had to be nasty."

© The New Yorker Collection 2005 Victoria Roberts from cartoonbank.com. All rights reserved.

In addressing that question, some psychologists focus on evolutionary history and others focus on the present-day pressures and expectations of the cultures in which we develop.

Evolutionary Foundations of Gender Differences

24

How might gender differences in personality be understood in terms of natural selection? What evidence suggests that hormones may provide a basis for such gender differences?

<<< Personality theorists who favor evolutionary explanations of gender differences (such as Larsen & Buss, 2005) point to the universality of certain gender differences and to the long history of evolution in which males and females were subject, generation after generation, to different reproductive challenges. Females' greater role in child care, and perhaps a need for cooperative relationships with other adults in relation to child care, may have led to selection for personality qualities promoting nurturance, cooperation, and caution; and males' greater need to compete in order to reproduce may have led to selection for competitiveness, aggressiveness, and risk taking (as described in Chapter 3).

Consistent with this evolutionary perspective, Shelley Taylor and her colleagues (2000) have amassed considerable evidence that male and female mammals in general, including humans, tend to respond differently to stressful situations. Whereas males tend to respond to stress by becoming more aggressive, females tend to respond by becoming more nurturant and more motivated to strengthen social connections. According to Taylor, females are more likely than males to attempt to placate their rivals rather than intimidate them and are more likely than males to look to friends for comfort and support. This difference, of course, is not all-or-none, but a matter of degree, and there is considerable overlap between the sexes. Taylor also summarizes evidence that sex differences in hormones contribute to these differences in personality. Oxytocin, which is at higher levels in females than in males, tends to promote affiliation; and testosterone, which is at higher levels in males, tends to promote aggression.

Cultural Foundations of Gender Differences

Cultural theorists, in contrast, point to the different experiences, expectations, role models, and opportunities provided by the culture for girls and boys—all the differences discussed in Chapter 12. From the cultural perspective, the most relevant niches to think about are not those existing in past generations of humans and prehumans but are those in existence right now (Bussey & Bandura, 1999; Wood & Eagly, 2002). The immediate causes of gender differences in personality are social forces that encourage girls to develop the nurturant, agreeable, and conscientious aspects of their nature and boys to develop their competitive, aggressive, and risk-taking aspects.

25

What evidence supports the view that gender differences in personality are influenced by cultural variation?

<<< Consistent with the cultural explanation is evidence that some gender differences in personality have changed, over historical time, in keeping with changing social roles and expectations. In particular, a systematic analysis of scores on various tests of *assertiveness*, given to men and women in the United States between 1931 and 1993, revealed that gender differences in this trait have changed over time in keeping with changes in the culture (Twenge, 2001). During the Great Depression and World War II, when women were generally expected to be self-sufficient, personality tests revealed relatively small gender differences in assertiveness. After the war, and through the 1950s and early 1960s, however, women were generally expected to be passive and domestic, and during those years women's scores on assertiveness tests declined considerably while men's scores remained relatively constant. Since the mid-1960s, however, women have entered the workforce in ever-greater numbers and have taken on roles previously considered to be masculine, and their assertiveness scores have increased sharply. In fact, some studies suggest that women's average assertiveness scores are now as high as men's (Twenge, 2001).

Few psychologists today would attribute gender differences in personality entirely to evolution or entirely to culture. Even the most ardent cultural theorists ac-



Taxi / Getty Images

Becoming assertive In the 30 years from 1970 to 2000, girls' and women's participation in formal sports increased by nearly tenfold (Twenge, 2001). Perhaps such cultural changes help to explain the increased assertiveness of women, as measured by personality tests, that occurred over that same period.

knowledge that evolved, biological differences between the sexes may account for consistencies, from culture to culture, in the kinds of cultural pressures exerted. For example, Wendy Wood and Alice Eagly (2002) suggest that cultures everywhere tend to see women as warm and nurturing because of their biological role in child-birth and nursing, and to see men as aggressive defenders of the group because of their larger size and strength. Conversely, even the most ardent evolutionary theorists acknowledge that cultural roles and expectations can alter evolved gender differences, either diminishing them or exaggerating them.

SECTION REVIEW

Personality differences may represent alternative adaptations to life's variable conditions.

Advantages of Being Different

- Just as diversified investments help protect one's financial future in a world of unpredictable change, diverse personalities, resulting from sexual reproduction, may protect one's genetic investment.
- Fish (including pumpkinseeds and perch) can be bold or cautious; each tendency has benefits and risks. Such variation is affected by the environment (including the number of other bold or cautious fish present) as well as by heredity.
- Human variations in the Big Five traits can, likewise, be viewed as alternative strategies for survival and reproduction.

Adapting to the Family Environment

- Siblings raised together may experience quite different environments, for reasons that include chance events, their own different choices, and differences in how they interpret the same occurrences.
- The tendency to exaggerate differences between siblings (sibling contrast) and the tendency for siblings to identify with different parents (split-parent identification) may reduce sibling rivalry and diversify parental investment.
- A historical analysis suggested that firstborns are more conservative and traditional, while later-borns are more open to new ideas and more likely to rebel against established ways. Contemporary studies of birth order have produced less dramatic, less consistent results.

Adapting to One's Gender

- On average, women score slightly to moderately higher than men on agreeableness, neuroticism, and conscientiousness.
- Apparently because of cultural pressures, some personality characteristics that run counter to gender stereotypes correlate with unhappiness. For example, shy young men are generally less happy than shy young women.
- Both evolutionary and cultural forces may help to account for gender differences in personality.



Sigmund Freud The founder of psychoanalysis viewed himself as a detective whose task was to use cues in people's behavior to uncover the secrets of their unconscious minds.

Personality as Mental Processes I: Psychodynamic and Humanistic Views

Trait theories, such as the five-factor model, are useful as general schemes for describing human psychological differences and for thinking about possible functions of those differences. Such theories have little to say, however, about the internal mental processes that lead people to behave in particular ways. The rest of this chapter is devoted to ideas about the mental underpinnings of personality.

We begin by examining two very general, classic theoretical perspectives on personality: the psychodynamic and humanistic perspectives. These two perspectives derive largely from the observations and speculations of clinical psychologists, who were attempting to make sense of the symptoms and statements of their clients or patients. These two perspectives also provide much of the basis for the general public's understanding of psychology.

Elements of the Psychodynamic Perspective

The pioneer of clinical psychology, and of the clinical approach to understanding personality, was Sigmund Freud (1856–1939). Freud was a physician who specialized in neurology and, from 1886 on, worked with patients at his private office in Vienna. He found that many people who came to him had no detectable medical problems; their suffering seemed to derive from their memories, especially their memories of disturbing events in their childhoods. In many cases, they could not recall such memories consciously, but clues in their behavior led Freud to believe that disturbing memories were present nevertheless, buried in what he referred to as the *unconscious mind*. From this insight, Freud developed a method of treatment in which his patients would talk freely about themselves and he would analyze what they said in order to uncover buried memories and hidden emotions and motives. The goal was to make the patient conscious of his or her unconscious memories, motives, and emotions, so that the patient's conscious mind could work out ways of dealing with them.

26 What characteristics of the mind underlie personality differences, according to the psychodynamic perspective?

<<< Freud coined the term **psychoanalysis** to refer both to his method of treatment (discussed in Chapter 17) and to his theory of personality. Freud's psychoanalytic theory was the first of what today are called **psychodynamic theories**—personality theories that emphasize the interplay of mental forces (the word *dynamic* refers to energy or force and *psycho* refers to mind). Two guiding premises of psychodynamic theories are that (a) people are often unconscious of their motives and (b) processes called *defense mechanisms* work within the mind to keep unacceptable or anxiety-producing motives and thoughts out of consciousness. According to psychodynamic theories, personality differences lie in variations in people's unconscious motives, in how those motives are manifested, and in the ways that people defend themselves from anxiety.

The Concept of Unconscious Motivation

Freud proposed that the main causes of behavior lie deeply buried in the unconscious mind—that is, in the part of the mind that affects the individual's conscious thought and action but is not itself open to conscious inspection. The reasons people give to explain their behavior often are not the true causes. To illustrate this idea, Freud (1912/1932) drew an analogy between everyday behavior and the phenomenon of posthypnotic suggestion.

27 How is the concept of unconscious motivation illustrated by posthypnotic suggestion?

<<< In a demonstration of posthypnotic suggestion, a person is hypnotized and given an instruction such as "When you awake, you will not remember what happened during hypnosis. However, when the clock chimes, you will walk across the room, pick up the umbrella lying there, and open it." When awakened, the subject appears to behave in a perfectly normal, self-directed way until the clock chimes. At this signal the subject consciously senses an irresistible impulse to perform the

commanded action, and consciously performs it, but has no conscious memory of the origin of the impulse (the hypnotist's command). If asked why he or she is opening the umbrella, the subject may come up with a plausible though clearly false reason, such as "I thought I should test it because it may rain later." According to Freud, the real reasons behind our everyday actions are likewise hidden in our unconscious minds, and our conscious reasons are cover-ups, plausible but false rationalizations that we believe to be true.

Freud believed that to understand his patients' actions, problems, and personalities, he had to learn about the contents of their unconscious minds. But how could he do that when the unconscious, by definition, consists only of information that the patient cannot talk about? He claimed that he could do it by *analyzing* certain aspects of their speech and other observable behavior to draw inferences about their unconscious motives. This is where the term *psychoanalysis* comes from. His technique was to sift the patient's behavior for clues to the unconscious. Like a detective, he collected clues and tried to piece them together into a coherent story about the unconscious causes of the person's conscious thoughts and actions.

Because the conscious mind always attempts to act in ways that are consistent with conventional logic, Freud reasoned that the elements of thought and behavior that are *least* logical would provide the best clues to the unconscious. They would represent elements of the unconscious mind that leaked out relatively unmodified by consciousness. Freud therefore paid particular attention to his patients' slips of the tongue and other mistakes as clues to the unconscious. He also asked them to describe their dreams and to report in uncensored fashion whatever thoughts came to mind in response to particular words or phrases. These methods are described more fully, with examples, in Chapter 17.

Sex and Aggression as Motivating Forces in Freud's Theory

Unlike most modern psychologists, Freud considered drives to be analogous to physical forms of energy that build up over time and must somehow be released. To live peaceably in society (especially in the Victorian society that Freud grew up in), one must often control one's direct expressions of the sexual and aggressive drives, so these are the drives that are most likely to build up and exert themselves in indirect ways. Freud concluded from his observations that much of human behavior consists of disguised manifestations of sex and aggression and that personality differences lie in the different ways that people disguise and channel these drives. Over time, Freud (1933/1964) came to define these drives increasingly broadly. He considered the sex drive to be the main pleasure-seeking and life-seeking drive and the aggressive drive to be the force that lies behind all sorts of destructive actions, including actions that harm oneself.

Social Drives as Motives in Other Psychodynamic Theories

Freud viewed people as basically asocial, forced to live in societies more by necessity than by desire, and whose social interactions derived primarily from sex, aggression, and displaced forms of these drives. In contrast, most psychodynamic theorists since Freud's time have viewed people as inherently social beings whose motives for interacting with others extend well beyond sex and aggression.

For example, Karen Horney developed a psychodynamic theory that focused on *security* as an inborn human need that can be filled only by other people. The most fundamental emotion in her theory is *basic anxiety*, defined as "the feeling a child has of being isolated and helpless in a potentially hostile world" (Horney, 1945). Parents influence a child's lifelong personality through the ways in which they succeed or fail in relieving the child's basic anxiety and helping the child feel secure. Closely related to Horney's theory are more recent *object relations theories* of personality (Ingram & Lerner, 1992; Kernberg, 1976), which derive partly from John Bowlby's ideas concerning the emotional attachments that infants make with their caregivers (discussed in Chapter 12). The term *object relations*, in these theories, refers to the interactions between a child and the objects (people) to whom the

>>> **28**
How did Freud draw inferences about the content of his clients' unconscious minds?

>>> **29**
Why did Freud consider sex and aggression to be especially significant drives in personality formation?

>>> **30**
What drives or human needs provide bases for personality differences in Horney's and Adler's theories of personality?



**"Of course I'm feeling uncomfortable.
I'm not allowed on the couch."**

child feels attached. According to these theories, the way people respond to others throughout life depends on the nature of their early attachments. For instance, inconsistent approval from attachment objects (typically parents) early in childhood—veering from fawning admiration to rejection—may result in an excessive need for others' approval later in life.

As another example, Alfred Adler developed a psychodynamic theory that centers on people's drive to feel competent. Adler (1930) contended that everyone begins life with a feeling of inferiority, which stems from the helpless and dependent nature of early childhood, and that the manner in which people learn to cope with or to overcome this feeling provides the basis for their lifelong personalities. While Horney emphasized emotional attachments as means of overcoming early anxieties, Adler emphasized the role of personal achievements. People who become overwhelmed by a sense of inferiority will develop either an *inferiority complex*, and go through life acting incompetent and dependent, or a *superiority complex*, and go through life trying to prove they are better than others as a means of masking their inferiority.

In all psychodynamic theories, including Freud's, the first few years of life are especially crucial in forming the personality. One's earliest attempts to satisfy drives result in positive or negative responses from others that, taken together, have lifelong effects on how those drives are subsequently manifested.

The Idea That the Mind Defends Itself Against Anxiety

A central idea in psychodynamic theories is that mental processes of self-deception, referred to as ***defense mechanisms***, operate to reduce one's consciousness of wishes, memories, and other thoughts that would threaten one's self-esteem or in other ways provoke a strong sense of insecurity, or anxiety. The theory of defense mechanisms was most thoroughly developed by Anna Freud (1936/1946), Sigmund's daughter, who was also a prominent psychoanalyst. Some examples of defense mechanisms are repression, displacement, reaction formation, projection, and rationalization.

31

How do repression, displacement, reaction formation, projection, and rationalization each serve to defend against anxiety?

Feeling good about feeling bad

Blues music, sung here by BB King, may appeal to us because it turns our unconscious as well as conscious pain into something beautiful. In that sense, it may be a form of sublimation.



S. Gaboury / DMI / RexFeatures

< Repression (discussed more fully in the next section) is the process by which anxiety-producing thoughts are pushed out or kept out of the conscious mind. In Freud's theory, repression provides the basis for most of the other defense mechanisms. Freud visualized repression as a damming up of mental energy. Just as water will leak through any crack in a dam, repressed wishes and memories will leak through the barriers separating the unconscious mind from the conscious mind. When such thoughts leak through, however, the mind can still defend itself by distorting the ideas in ways that make them less threatening. The other defense mechanisms are means to such distortion.

Displacement occurs when an unconscious wish or drive that would be unacceptable to the conscious mind is redirected toward a more acceptable alternative. For example, a child long past infancy may still have a desire to suck at the mother's breast—a desire that is now threatening and repressed, because it violates

the child's conscious assessment of what is proper and possible. The desire might be displaced toward sucking on a lollipop, an action that is symbolically equivalent to the original desire but more acceptable and realistic. In some cases, displacement may direct one's energies toward activities that are particularly valued by society, such as artistic, scientific, or humanitarian endeavors. In these cases displacement is referred to as **sublimation**. A highly aggressive person might perform valuable service in a competitive profession, such as being a trial lawyer, as sublimation of the drive to beat others physically. As another example, Freud (1910/1947) suggested that Leonardo da Vinci's fascination with painting Madonnas was a sublimation of his desire for his mother, which had been frustrated by his separation from her.

Reaction formation is the conversion of a frightening wish into its safer opposite. For example, a young woman who unconsciously hates her mother and wishes her dead may consciously experience these feelings as intense love for her mother and strong concern for her safety. Psychodynamic theorists have long speculated that *homophobia*—the irrational fear and hatred of homosexuals—may often stem from reaction formation. People who have a tendency toward homosexuality, but fear it, may protect themselves from recognizing it by vigorously separating themselves from homosexuals (West, 1977). In a study supporting that explanation, men who scored as highly homophobic on a questionnaire were subsequently found, through direct physical measurement, to show more penile erection while watching a male homosexual video than did other men, even though they denied experiencing any sexual arousal while watching the video (Adams & others, 1996).

Projection occurs when a person consciously experiences an unconscious drive or wish as though it were someone else's. A person with intense, unconscious anger may project that anger onto her friend—she may feel that it is her friend, not she, who is angry. In an early study of projection conducted at college fraternity houses, Robert Sears (1936) found that men who were rated by their fraternity brothers as extreme on a particular characteristic, such as stinginess, but who denied that trait in themselves tended to rate others as particularly high on that same characteristic.

Rationalization is the use of conscious reasoning to explain away anxiety-provoking thoughts or feelings. A man who cannot face his own sadistic tendencies may rationalize the beatings he gives his children by convincing himself that children need to be beaten and that he is only carrying out his fatherly duty. Psychodynamic theories encourage us to be wary of conscious logic, since it often serves to mask true feelings and motives. (For discussion of the left-hemisphere interpreter, which seems to be the source of rationalization, look back at Chapter 5, p. 162.)

Contemporary Research on Defense Mechanisms

The concept of psychological defenses is one of the most widely accepted of Freud's ideas today; it provides the implicit foundation for many modern ideas about social cognition (Cramer, 2000). Examples discussed in Chapter 13 include the self-serving biases, which lead people to think that they are more competent and ethical than they actually are, and the ways in which people rationalize discrepancies between their consciously held attitudes and their actual behavior. Moreover, controlled research studies have shown that at least some of the classic defense mechanisms posited by Sigmund and Anna Freud do exist and do help to reduce the conscious experience of anxiety (Baumeister & others, 1998; Cramer, 2000).

Studies of Repression

The Controversy Concerning Repression of Traumatic Memories The most dramatic examples of repression, in Freud's work and in that of other psychoanalysts, involve people who seemed to have forgotten traumatic incidents that occurred in their childhood, but who recovered memories of those incidents in the course of their psychotherapy. A problem with such clinical observations is that the original experiences are rarely corroborated by outside observers. Without corroboration, there is no way to know if a person's memories, reconstructed in therapy, are accurate or false constructions. As discussed in Chapter 9, memories—especially memories of childhood events—are very susceptible to the power of suggestion.

Contemporary research involving people who have suffered from known (corroborated) traumatic experiences indicates that such experiences are usually remembered very well (McNally, 2003). Survivors of such experiences generally report that they wish they could forget, but cannot. Indeed, in general, the more

Why is Freud's idea that people repress memories of traumatic experiences still controversial?

32

traumatic the experience, the more likely it is that people will remember it (Goodman & others, 2003). Any forgetting that does occur seems to follow the same rules as does forgetting in general (discussed in Chapter 9). People who experience a traumatic event, such as sexual abuse, before the age of 7 are much more likely to forget it than are people who experience it at a later age; and people who spend little time consciously thinking about or talking about the event are more likely to forget it than are others (Goodman & others, 2003; McNally, 2003).

It is possible that repression of memories for traumatic experiences exists primarily in cases where nobody except the perpetrator is aware of the abuse. Indeed, it makes some sense that repression would be most effective in those cases. When people don't know of the abuse, they cannot talk about it and remind the victim of it. Unfortunately, in such cases there is usually no scientific way of corroborating the abuse and thereby demonstrating the existence of repression. For this reason, the question of whether, or to what degree, repression of traumatic experiences actually occurs has been hotly debated, and unresolved, from Freud's time to the present.

33

What evidence exists that some people regularly repress anxious feelings? In general, how do such “repressors” differ from others in their reactions to stressful situations?

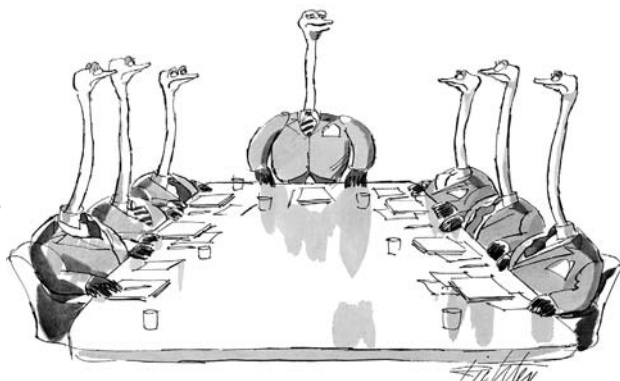


Repression of Emotional Feelings In contrast, there is considerable evidence for repression of a different sort, a sort that is also consistent with Freud's theory. Some people, apparently, regularly repress the conscious emotional *feelings* that accompany disturbing events. Using standardized questionnaires, researchers have identified individuals—referred to as *repressors*—who report that they experience relatively little anxiety and who answer other questions in a way that seems highly defensive (Weinberger, 1990). Their questionnaire responses indicate that these people have an especially strong need to view themselves in a very favorable light; they do not admit to the foibles that characterize essentially all normal people.

Many research studies have compared repressors, identified in this way, with non-repressors. When asked to complete sentences that contained sexual or aggressive themes, repressors claimed not to feel stressed, yet by physiological indices (heart rate, muscle tension, and sweating) they exhibited more stress than did other subjects, who consciously felt stressed (Weinberger, 1990; Weinberger & others, 1979). In other studies, repressors reported less anxiety or other unpleasant emotions in daily diaries, recalled fewer childhood emotional experiences, and were less likely to notice consciously or remember emotion-arousing words or phrases in messages presented during an experiment than was the case for non-repressors (Bonanno & others, 1991; Cutler & others, 1996; Davis & Schwartz, 1987).

The research strongly suggests that repressors are truly unconscious of their own negative emotions and are not just pretending to lack them, even though, by physiological measures, they react even more strongly to emotion-arousing events than do other people (Furnham & others, 2002). There is also evidence that repressors have more health problems and complain of more chronic pain than do non-repressors (Burns, 2000; Schwartz, 1990). This is consistent with the idea that they experience stress physically rather than as conscious emotion.

© The New Yorker Collection 1986 Mischia Richter from cartoonbank.com. All Rights Reserved.



“The motion has been made and seconded that we stick our heads in the sand.”

Distinction Between Mature and Immature Defensive Styles

In the late 1930s, an ambitious longitudinal study was begun of male sophomores at Harvard University. Each year until 1955, and less often after that, the men filled out an extensive questionnaire concerning such issues as their work, ambitions, social relationships, emotions, and health. Nearly 30 years after the study began, when the men were in their late forties, a research team led by George Vaillant (1995) interviewed in depth 95 of these men, randomly selected. By systematically analyzing the content and the style of their responses in the interview



LOOKALIKE PRODUCTIONS

A very healthy defense Alfred Adler claimed that people who have a handicap often work hard to compensate, sometimes to the point that they excel at the very activity that would otherwise be most difficult for them. Emmanuel Osuff Yeboah of Ghana, shown here on his bicycle, may be an example. He was born with only one leg, but has become an outstanding bicyclist. In Ghana people born with a handicap are greatly stigmatized; they are expected to become beggars. Yeboah has been campaigning against that attitude by cycling through every part of his country to demonstrate what people with one leg can do.

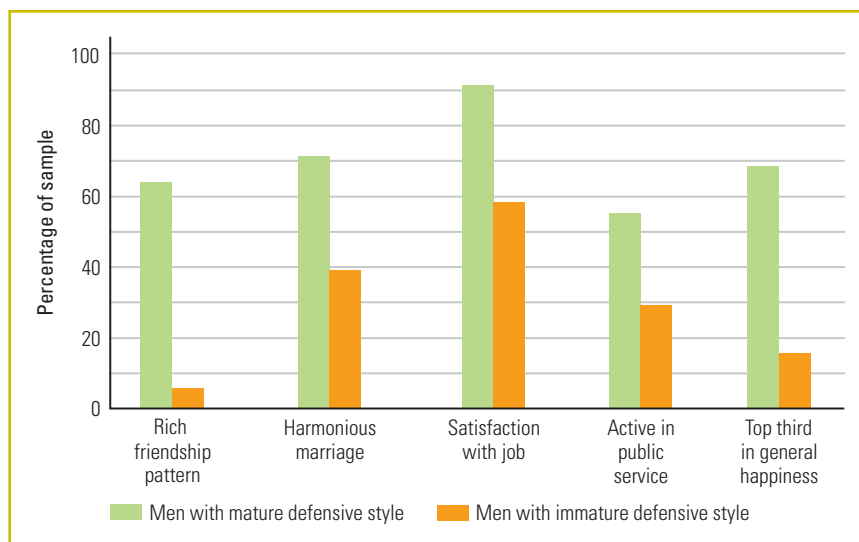
and on the previous questionnaires, the researchers rated the extent to which each man used specific defense mechanisms.

Vaillant divided the various defense mechanisms into categories according to his judgment of the degree to which they would promote either ineffective or effective behavior. *Immature defenses* were those presumed to distort reality the most and to lead to the most ineffective behavior. Projection was included in this category. *Intermediate defenses* (referred to by Vaillant as “neurotic defenses”), including repression and reaction formation, were presumed to involve less distortion of reality and to lead to somewhat more effective coping. *Mature defenses* were presumed to involve the least distortion of reality and to lead to the most adaptive behaviors. One of the most common of the mature defenses was *suppression*, which involves the conscious avoidance of negative thinking. Suppression differs from repression in that the negative information remains available to the conscious mind and can be thought about constructively when the person chooses to do so. Another defense in the mature category was humor, which, according to Freud and other psychodynamic theorists, reduces fear by making fun of feared ideas (for a comparable view, see the discussion of the evolution of laughter in Chapter 3).

Consistent with Vaillant's expectations, the men who used the most mature defenses were the most successful on all measures of ability to love and work (Freud's criteria for mature adulthood). They were also, by their own reports, the happiest. Figure 15.5 depicts some of the comparisons between the men who used mainly mature defenses and those who used mainly immature defenses. Subsequently, Vaillant found similar results in two other longitudinal studies: one of men from a working-class background and the other of women who had been identified in childhood as unusually bright (Vaillant & Vaillant, 1992). Such correlations do not prove that mature defenses help cause successful coping, but they do show that mature defenses and such success tend to go together. Vaillant (1995) also found, not surprisingly, that as the Harvard men matured—from age 19 into their forties—the average maturity of their defenses increased. Immature defenses such as projection declined, and mature defenses such as suppression and humor increased.

>>> 34
What relationships did Vaillant find between defensive styles and measures of life satisfaction?

FIGURE 15.5 Love, work, and happiness in men with mature and immature defensive styles Harvard alumni who were classified as using primarily mature defenses were more frequently rated as having rich friendships, a good marriage, satisfaction with their work, active involvement in public service, and a high degree of happiness than were those who were classified as using primarily immature defenses. (Data from Vaillant, 1977.)



The Humanistic Perspective: The Self and Life's Meanings

35

How, in general, do humanistic theories differ from psychodynamic theories?

<<< **Humanistic theories** of personality arose, in the mid-twentieth century, partly in reaction to the then-dominant psychodynamic theories. While psychodynamic theories emphasize unconscious motivation and defenses, humanistic theories emphasize people's conscious understanding of themselves and their capacity to choose their own paths to fulfillment. They are called *humanistic* because they center on an aspect of human nature that seems to distinguish us clearly from other animals—our tendency to create belief systems, to develop meaningful stories about ourselves and our world, and to govern our lives in accordance with those stories.

Phenomenology is the study of conscious perceptions and understandings, and humanistic theorists use the term **phenomenological reality** to refer to each person's conscious understanding of his or her world. Humanistic theorists commonly claim that one's phenomenological reality is one's real world; it provides the basis for the person's contentment or lack of contentment and for the meaning that he or she finds in life. In the words of one of the leaders of humanistic psychology, Carl Rogers (1980): "The only reality you can possibly know is the world as you perceive and experience it. . . . And the only certainty is that those perceived realities are different. There are as many 'real worlds' as there are people."



Courtesy of Natalie Rogers / © Bob Van Doren

Carl Rogers Rogers (1902–1987) developed a humanistic theory of personality that centers on the self-concept and the ways in which it can be distorted by socially imposed conditions of worth.

Being One's Self; Making One's Own Decisions

36

What human drive, or goal, characterizes healthy development in Rogers's self theory?

<<< According to most humanistic theories, a central aspect of one's phenomenological reality is the *self-concept*, the person's understanding of who he or she is. Rogers referred to his own version of humanistic theory as *self theory*. He claimed that at first he avoided the construct of self because it seemed unscientific, but was forced to consider it through listening to his clients in therapy sessions. Person after person would say, in effect, "I feel I am not being my real self"; "I wouldn't want anyone to know the real me"; or "I wonder who I am." From such statements, Rogers (1959) gradually came to believe that a concept of self is a crucial part of a person's phenomenological reality and that a common goal of people is to "discover their real selves" and "become their real selves." According to Rogers, people often are diverted from becoming themselves by the demands and judgments placed upon them by other people, particularly by authority figures such as parents and teachers. To be oneself, according to Rogers, is to live life according to one's own wishes rather than someone else's. Thus, an important dimension of individual difference, in Rogers's theory, has to do with the degree to which a person feels in charge of his or her own life.

37

What research findings on self-determination appear to support Rogers's theory?

<<< As would be predicted by Rogers's theory, researchers have found that people who feel as if a decision is fully their own, made freely by themselves, are more likely to follow through and act on it effectively than are people who feel that they are doing something because of social pressure or because an authority figure thinks they should do it. Those who think and talk about their intention to lose weight, stop smoking, exercise more, or take prescribed medicines on schedule as "following the doctor's orders" are less likely to succeed at such goals than are those who think and talk about the decision as their own (Ryan & Deci, 2000; Williams & others, 1995, 1996; Wilson & others, 2003). Other research—conducted in such diverse countries as Canada, Brazil, Russia, South Korea, Turkey, and the United States—has shown that people who feel that they are following their own desires are happier and more productive than those who feel that they are responding to social pressures (Chirkov & others, 2003, 2005). In different cultures, people tend to have

different values and to choose different activities, but in each culture those who see those choices as their own claim to be most satisfied with their lives.

Self-Actualization and Maslow's Hierarchy of Human Needs

Humanistic theorists use the term **self-actualization** to refer to the process of becoming one's full self—that is, of realizing one's dreams and capabilities. The specific route to self-actualization will vary from person to person and from time to time within a person's lifetime, but for each individual the route must be self-chosen.

Rogers (1963, 1977) often compared self-actualization in humans to physical growth in plants. A tree growing on a cliff by the sea must battle against the wind and saltwater and does not grow as well as it would in a more favorable setting, yet its inner potential continues to operate and it grows as well as it can under the circumstances. Nobody can tell the tree how to grow; its growth potential lies within itself. Humanistic theorists hold that full growth, full actualization, requires a fertile environment, but that the direction of actualization and the ways of using the environment must come from within the organism. In the course of evolution, organisms have acquired the capacity to use the environment in ways that maximize growth. In humans, the capacity to make free, conscious choices that promote positive psychological growth is the actualizing tendency. To grow best, individuals must be permitted to make those choices and must trust themselves to do so.

Another pioneer of humanistic psychology, Abraham Maslow (1970), suggested that to self-actualize, one must satisfy five sets of needs that can be arranged in a hierarchy (see Figure 15.6). From bottom to top, they are (1) physiological needs (the minimal essentials for life, such as food and water); (2) safety needs (protection from dangers in the environment); (3) attachment needs (belongingness and love); (4) esteem needs (competence, respect from others and the self); and (5) self-actualization needs. In Maslow's view, the self-actualization needs encompass the needs for self-expression, creativity, and “a sense of connectedness with the broader universe.” Maslow argued that a person can focus on higher needs only if lower ones, which are more immediately linked to survival, are sufficiently satisfied so that they do not claim the person's full attention and energy.

Maslow's needs hierarchy makes some sense from an evolutionary perspective. The physiological and safety needs are most basic in that they are most immediately linked to survival. If one is starving or dehydrated, or if a tiger is charging,



Courtesy of the Sudbury Valley School

Self-actualization Humanistic theorists draw an analogy between the self-actualization process in humans and the inner growth potential of all living things. This beech tree has long been using its environment to promote its own growth, and these children have for a much shorter time been doing the same.

What is Maslow's theory about the relationship among various human needs? How might the theory be reconciled with an evolutionary perspective?



FIGURE 15.6 Maslow's hierarchy of human needs According to Maslow, needs at the lower portion of the hierarchy must be satisfied at least to some degree before people can focus on needs higher up. The most psychologically healthy people are those whose deficiency needs are sufficiently satisfied to free their energies for self-actualization.

then survival depends immediately upon devoting one's full resources to solving that problem. The social needs for belongingness, love, and esteem are also linked to survival, though not in quite as direct and immediate a fashion. We need to maintain good social relationships with others to ensure their future cooperation in meeting our physiological and safety needs and in helping us reproduce. The self-actualization needs are best construed, from an evolutionary perspective, as self-educative needs. Playing, exploring, and creating can lead to the acquisition of skills and knowledge that help one later in such endeavors as obtaining food, fending off predators, attracting mates, and securing the goodwill and protection of the community. From this perspective, self-actualization is not ultimately higher than the other needs but is part of the long-term human system for satisfying those needs.

The Life Story or Personal Myth as a Basis for Personality

39

How is the life-story conception of personality consistent with the humanistic theorists' focus on phenomenology?

<<< People strive to make sense of their lives as a whole, and they may establish goals and make choices in accordance with that larger conception. One way to learn about individual people, which is gaining increasing acceptance among psychologists, is to ask them to tell their life stories. On the basis of his analysis of many such stories, Dan McAdams (1993, 2001) contends that the formats of the stories generally share certain characteristics, though the contents vary greatly. Each story contains themes and subthemes, morals, subordinate characters who influence the main character (the self), conflicts that get resolved, and a relatively consistent narrative tone (such as serious or lighthearted). The stories are not simply chronological chains of events but, rather, are integrated wholes in which events are related to one another with themes pertaining to larger life purposes or a sense of personal destiny. The integrating themes are creations of the storyteller.

McAdams refers to the self-told story as a *personal myth*, because it seems to provide for the individual what religious myths provide for whole cultures. It gives a sense of direction and meaning to life. People understand and explain individual events in their lives in terms of how those events fit into their larger life stories, and they make new decisions that are consistent with their stories. In their research, McAdams and his colleagues have found positive correlations between the coherence of people's life stories and measures of their mental health (Baerger & McAdams, 1999; Bauer & McAdams, 2004).

The personal myth changes with time and experience, usually gradually but sometimes dramatically and quickly. As circumstances change, the story may change to accommodate the new situation (Baumeister, 1994; McAdams, 1994). For instance, a couple's first meeting may be understood and described in romantic terms as part of their life destiny when the couple is happily married, but years later, when the couple is divorced, the same meeting might be described as an interfering accident, a digression from the meaningful life course (Vaughan, 1986). From a phenomenological perspective, neither understanding is right or wrong; each is a way of making sense of one's current and ongoing life.

40

How does the life-story approach to personality differ from the trait approach in its conclusions about the modifiability of personality in adulthood?

<<< In a study of personal transformation, William Miller and Janet C'deBaca (1994) advertised for volunteers "who have been transformed in a relatively short period of time—who have had a deep shift in core values, feelings, attitudes, or actions." They had no trouble finding people who claimed to have had such an experience at some time in their lives, and they interviewed extensively 55 such people. Each story of change was different, but some common themes emerged. In general, these people saw themselves as acquiring over a short period of time, often less than 24 hours, a dramatic and lasting new insight, a new way of understanding themselves and their purposes in life. Usually they described themselves as emotionally distressed before the new insight and relieved and happy after it. Usually they described the change as liberating and involving a shift in values. Most often the shift was from worldly values—such as wealth, achievement, adventure, fitting in, and attractiveness—to social and spiritual values—such as honesty, generosity,

personal peace, and obedience to God's will. Such changes occurred for different people at very different ages and in response to a wide variety of life events.

If personality is defined phenomenologically as one's self-told story, then personality can apparently change dramatically at any time in adulthood. Traits such as the Big Five, measured by standard personality tests, may not change much as people grow older and move from one life experience to another. But through their personal myths people can apparently acquire new ways of understanding and employing those traits at any time in life—new meanings, new self-definitions, and new life goals.

SECTION REVIEW

Psychodynamic and humanistic personality theories focus on mental processes.

The Psychodynamic Perspective

- Freud, whose psychoanalytic views originated this perspective, believed that the real causes of behavior lie in the unconscious mind, with sexual and aggressive motives being especially important.
- Other psychodynamic theorists emphasized unconscious effects of other drives. Horney and Adler emphasized, respectively, the drives for security and competence.
- Defense mechanisms serve to reduce conscious awareness of unacceptable or emotionally threatening thoughts, wishes, and feelings.

Studies of Defense Mechanisms

- Among people who have experienced known trauma, repression of that trauma seems to be the exception rather than the rule.
- People classified as repressors routinely repress disturbing emotional feelings. Though they claim to feel little anxiety, their bodies react strongly to stressful situations.
- In a longitudinal study of men, Vaillant found that defensive styles that involved less distortion of reality and led to more effective behavior were correlated with greater success in all areas of life.

The Humanistic Perspective

- Humanistic theories emphasize phenomenological reality (the self and world as perceived by the individual).
- Rogers proposed that a person must move past social demands and judgments to become his or her real self. Self-determination does correlate with greater life satisfaction.
- In Maslow's hierarchy of needs, self-actualization (becoming one's full self, living one's dreams) is addressed only when more basic needs are adequately met.
- In the life-story approach to personality, a person's self-told story serves as a personal myth providing meaning and direction.

Personality as Mental Processes II: Social-Cognitive Views

Social-cognitive theories of personality, sometimes called *social-learning* or *social-cognitive-learning theories*, draw both from clinical psychologists' experiences with their clients and from academic psychologists' research on learning, cognition, and social influence. In place of the instinctive, unconscious motives posited by psychodynamic theories as the prime shapers of personality, social-cognitive theories emphasize the roles of beliefs and habits of thought that are acquired through one's unique experiences in the social environment. These habitual ways of thinking may be conscious, but they may also become so ingrained and automatic that they exert their influence without the person's conscious awareness. Thus, to social-cognitive theorists, the term *unconscious* generally refers to automatic mental processes, in the same sense as discussed in earlier chapters in this book, not to thoughts that are actively barred from consciousness by defense mechanisms. As you will see, social-cognitive ideas overlap considerably with humanistic ideas about personality, but they are based more on laboratory research and have more to do with predicting people's behavior in specific situations, and less to do with global life choices, than is the case for humanistic theories.

>>> 41

How, in general, do social-cognitive theories differ from psychodynamic theories?

Habitual Ways of Thinking as Personality Traits

Much of the personality research within the social-cognitive tradition has focused on learned beliefs and habits of thinking that increase or decrease people's tendency to take control of their lives and accomplish the tasks they would like to accomplish (Bandura, 2001).

42

What, in Rotter's research, predicted people's improvement or failure to improve at laboratory tasks? How did this lead to the concept of locus of control?



Beliefs About the Locus of Control over Desired Effects

If there is a principal founder of the social-cognitive perspective on personality, it is Julian Rotter, who wrote the first book explicitly describing a social-cognitive approach to personality (Rotter, 1954/1980). In his own early research, Rotter found that people behaved differently at various tasks or games in the laboratory, depending on whether they believed that success depended on skill or luck (Rotter & others, 1961). To the degree that they believed that success depended on skill (which it did), they worked hard and improved. To the degree that they believed that it depended on luck, they did not work hard and did not improve. Partly on the basis of these observations, Rotter argued that people's behavior depends not just on the objective relationship between their responses and rewards but also on their subjective beliefs about that relationship.



Paul Cézanne, *The Card Players*, 1892. Paris, Musée d'Orsay. Photo: Erich Lessing / Art Resource.

Skill or luck? People approach an activity—such as a game of cards—very differently depending on whether they believe its potential rewards are controlled by skill or luck. This insight lay behind Rotter's concept of locus of control.

Generalized Locus of Control In many life situations it is not clear to what degree we have control over rewards. For example, it is not completely clear that studying hard will lead to a good grade on an exam or that diet and exercise will prevent us from having a heart attack. Rotter (1966) suggested that in such situations people tend to behave according to a generalized disposition (a personality trait), acquired from past experience, to believe that rewards either are or are not usually controllable by their own efforts. He referred to this disposition as **locus** (location) **of control** and developed a questionnaire designed to measure it. Table 15.4 shows sample test items from Rotter's locus-of-control questionnaire. People whose answers reflect a belief that individuals control their own rewards (and, by extension, their own fate) are said to have an *internal* locus of control. People

whose answers reflect a belief that rewards (and fate) are controlled by factors outside themselves are said to have an *external* locus of control.

43

What sorts of behaviors correlate with an internal locus of control? Why have tests been developed to assess domain-specific locus of control, not just generalized locus of control?



Since its development, hundreds of studies have shown consistent, though usually not very high, correlations between scores on Rotter's locus-of-control scale and actual behavior in various situations. People who score toward the internal-locus end of the scale are, on average, more likely than those who score toward the external-locus end to try to control their own fate. They are more likely to take pre-

TABLE 15.4

Sample test items from Rotter's locus-of-control scale

The task on each item is to decide which alternative (*a* or *b*) seems more true. The actual test consists of 23 items similar to those shown here.

- Item: a. In the long run, people get the respect they deserve in this world.
b. Unfortunately, an individual's work often passes unrecognized, no matter how hard he or she tries.
- Item: a. I have often found that what is going to happen will happen.
b. Trusting to fate has never turned out as well for me as making a decision to take a definite course of action.
- Item: a. In the case of the well-prepared student, there is rarely if ever such a thing as an unfair test.
b. Many times exam questions tend to be so unrelated to course work that studying is really useless.

Note: For the items shown here, *internal* locus of control is indicated by choosing *a* for the first and third items and *b* for the second item.

Source: "Generalized expectancies for internal versus external locus of control of reinforcement" by J. B. Rotter, 1966, *Psychological Monographs: General and Applied*, 80 (Whole no. 609), p. 11.

ventive health-care measures (Reich & others, 1997); to succeed in weight-loss programs (Adolfsson & others, 2005); to seek information on how to protect themselves during a tornado warning (Sims & Baumann, 1972); to resist group pressure in laboratory tests of conformity (Crowne & Liverant, 1963); and to prefer games of skill over games of chance (Schneider, 1972). Other research has indicated that people who score toward the internal end of the scale are, on average, less anxious and more content with life than those who score toward the external end (Phares, 1978, 1984). Apparently, the sense of control helps calm people's fears about potential mishaps and dangers.

Domain-Specific Locus of Control In recent years, focus has shifted from the idea of a generalized locus of control to the idea that people may vary in their locus-of-control beliefs for different realms of activity. For example, a person who has a highly internal locus of control concerning academic performance may have an external locus of control concerning maintenance of health, or vice versa. In keeping with this idea, separate locus-of-control questionnaires have been developed for separate realms. Scores on these tests, in general, show quite strong correlations with the relevant behaviors. For example, people who score as having an internal locus of control for work and careers perform much better at their work than do people who score as external on this measure (Hattrup & others, 2005). There is even a locus-of-control questionnaire related to driving. Research with this instrument shows that people who think they have strong control over whether or not they will have a traffic accident drive much more safely than do people who are more external on this measure (Ozkan & Lajunen, 2005).

Of course, as with all correlational research, we cannot be sure what is cause and what is effect. Does a sense of control promote hard work, responsible behavior, and success? Or do hard work, responsible behavior, and success promote a sense of control? Most social-cognitive theorists would contend that both of these causal hypotheses are correct to some degree. People who succeed develop a stronger sense of internal control in the realm of their success, which, in turn, leads to further success.

Beliefs About One's Own Ability to Perform Specific Tasks

Another pioneer of the social-cognitive perspective on personality is Albert Bandura, who, like Rotter, earned a degree in clinical psychology and then went on to a career of laboratory research. Bandura's early research on observational learning is described in Chapter 4. Much of Bandura's more recent research centers on people's beliefs about their own abilities to perform specific tasks, which he refers to as **self-efficacy**. People who expect that they can perform a certain task are said to have high self-efficacy about the task, and people who expect the opposite are said to have low self-efficacy about it.

Self-efficacy may seem similar to locus of control, but Bandura (1997) considers the two to be distinct. Self-efficacy refers to the person's sense of his or her own ability, while locus of control refers to the person's sense of whether or not that ability will produce rewards. Although self-efficacy and an internal locus of control usually go together, they do not always. If you believe, for example, that you are skilled at math but that the skill is worthless (perhaps because it is unrecognized by your math professor or others in society), then you have high self-efficacy but an external locus of control in that area. Conversely, if you believe that skill at math would bring rewards but that you don't have the skill, then you have low self-efficacy and an internal locus of control in that area. As is the case for locus of control, self-efficacy may be quite specific to a very narrow range of tasks or quite general over a broad range of tasks (Cervone, 1997; Welch & West, 1995).

Bandura and his colleagues have repeatedly demonstrated that improved self-efficacy for a task predicts improvement in actual performance of the task. In one study, for example, various treatments were used to help people overcome their fear of snakes. The subjects who claimed after treatment that they now expected to

>>> **44**
How does self-efficacy differ from locus of control?

>>> **45**
What evidence supports the theory that high self-efficacy (a) predicts high performance and (b) may help cause high performance?

© Larry Williams / Corbis



Self-efficacy According to Albert Bandura, the first step in acquiring a skill like this—or in doing anything that is difficult—is to believe you can do it.

46

What evidence supports the value of optimism? Through what mechanisms might optimism produce its beneficial effects?



A number of psychologists have developed questionnaires designed to assess people's general tendency to think either positively or negatively. C. Rick Snyder (1994) and his colleagues developed a questionnaire to assess *hope*, which they construe as a belief in one's ability to solve solvable problems (generalized self-efficacy) combined with a belief that most problems in life are solvable. Martin Seligman (1990) and his colleagues developed a questionnaire to assess the degree to which people explain negative events in their lives in a pessimistic or optimistic manner (discussed in relation to theories of depression in Chapter 16). Michael Scheier and Charles Carver (1993) developed a questionnaire to assess dispositional optimism, the tendency to believe in a rosy future. On the questionnaire, people indicate the degree to which they agree or disagree with such statements as "In uncertain times, I usually expect the best."

Correlational studies using all these questionnaires have shown that, in general, people with an optimistic style of thought tend to cope more effectively with life's stressors than do people who have a pessimistic style (Peterson, 2000; Schneider, 2001; Snyder, 1994; Taylor & others, 2003). In one such study, Scheier and his colleagues (1989) used their questionnaire to assess dispositional optimism in middle-aged men who were about to undergo coronary artery bypass surgery. They found that those who scored high on optimism before the surgery made quicker recoveries than did those who scored low, even when the medical conditions that led to surgery were equivalent. The optimists were quicker to sit up on their own, to walk, to resume vigorous exercise, and to get back to work full-time than were the pessimists.

A. Ramey / Stock, Boston



Optimism The belief that one can overcome adversity may become a self-fulfilling prophecy if that belief leads to effective actions. This woman, the victim of a drunk driver, works hard at strengthening her legs so she can walk again.

The most likely explanation for this and other positive correlations with optimism is that optimistic thinking leads people to devote attention and energy to solving their problems or recovering from their disabilities, which in turn leads to positive results. Pessimists are relatively more likely to say, "It won't work out anyway, so why try?" A more controversial idea is that optimistic thinking and psychological equanimity in general may facilitate physical health and recovery through direct effects on body chemistry (discussed in Chapter 16).

Adaptive and Maladaptive Optimism and Pessimism

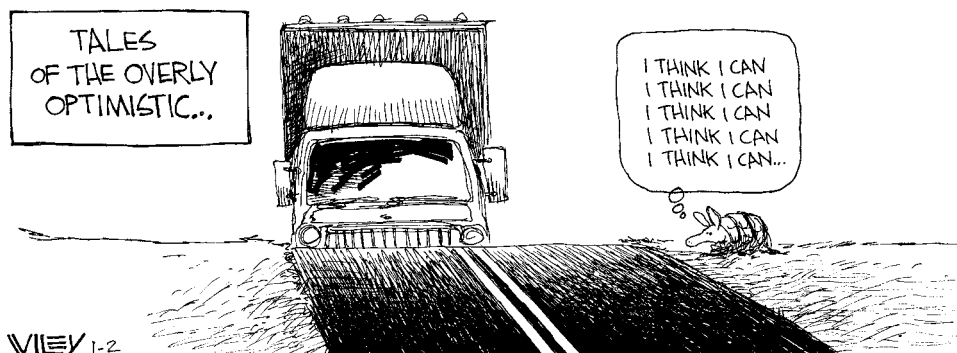
Before concluding that optimism is always best, before rushing out to trade the clear lenses on our glasses for rose-tinted ones, we should consider a potential hazard of optimism. Health psychologists have long pointed out the danger of unrealistic, self-delusional forms of optimism. Many people, especially adolescents and young adults, optimistically believe that they are invulnerable to such catastrophes as AIDS, lung cancer, drug addiction, and automobile accidents and fail to take precautions to avoid such dangers (Schwarzer, 1994; Weinstein, 1980, 1982). Similarly, an optimistic, inflated belief in their academic or career abilities blinds some people to their own shortcomings and prevents them from taking steps to improve (Dunning & others, 2004). Optimism of this sort, which in the psychodynamic tradition is called *defensive optimism*, may reduce anxiety by diverting thoughts away from fearful possibilities, but it may also lead to serious harm. The optimistic belief that you can control your fate through active self-care and self-improvement usually leads to constructive behaviors, but the optimistic belief that fate will protect you without your participation can lead to dangerously imprudent behaviors.

Just as optimism can be adaptive or maladaptive, depending on whether or not it translates into constructive action, so can pessimism. In research on the cognitive underpinnings of success in college, Julie Norem and Shaun Illingworth (1993, 2004) found students who use apparently opposite mental strategies to perform well academically. Some students use an adaptive form of optimism. They believe they will do well, and that belief, coupled with their thoughts about the positive consequences of doing well, motivates them to work hard and actually do well. Other students, however, use an adaptive form of pessimism. They believe they will not do well, despite having done well in the past, and that belief, along with thoughts about the negative consequences of failure, motivates them to work hard to avoid failure. As a result, and apparently to their surprise, they not only pass but achieve high grades. Still, the optimists are probably better off in the long run than the pessimists. One study of adjustment to college life revealed that the pessimists' constant anxiety about failure led them to focus too narrowly on grades and lose the intrinsic pleasure of academic work (Cantor & Harlow, 1994).

The Idea of Situation-Specific Personality Traits

Social-cognitive theorists have long contended that global traits—such as those specified by the five-factor model—tell only part of the story of personality. To understand a person, according to these theorists, one must not only know that a person tends to be extraverted or introverted, for example, but must also know the *contexts* in which the person typically manifests those tendencies. One person might be shy (introverted) at parties but outspoken (extraverted) at formal meetings, while the opposite might be true of another person. One person who scores as

>>> 47
What seems to differentiate adaptive from maladaptive optimism and adaptive from maladaptive pessimism?



© 1999 Wiley Miller / distributed by the Washington Post Writers Group

disagreeable on the NEO-PI might be most disagreeable to subordinates while another, with the same score, might be most disagreeable to authority figures. The leading advocate for this contextual view of personality is social-cognitive theorist Walter Mischel (1984, 2004).

48

What evidence supports Mischel's concept of situation-specific dispositions?

<<< In one study, Mischel and Phillip Peake (1982) assessed repeatedly, by direct observation, 19 different forms of behavior presumed to be related to the trait of conscientiousness in a group of students at Carleton College in Minnesota. Included were measures of regularity of class attendance, promptness in completing assignments, bed neatness, and neatness of class notes. They found high consistency within any one of these measures but relatively low consistency across measures. For instance, students who kept neat notes for one class were very likely to keep neat notes for another class but only slightly more likely than average to keep their beds neat.

In another study, Mischel and his colleagues found that children with social adjustment problems at a summer camp were not well described by such global traits

as “aggressive” or “withdrawn” but were quite well described and differentiated from one another by terms that referred to the social situations that prompted them to act aggressively or to withdraw (Mischel & Shoda, 1995; Shoda & others, 1994). Figure 15.7 shows sample results, from that study, for verbal aggressiveness for two children. As shown in the figure, both children were somewhat more verbally aggressive than the average child at the camp, but the two were very different from each other with regard to the situations in which they exhibited aggression. Child 28 was highly aggressive to peers who approached him in a friendly manner but not particularly aggressive in other situations, and child 9 was highly aggressive to adults but not to peers. Knowledge of that difference is essential to any clinically useful understanding of the children, but that knowledge would be obscured in a global rating of aggressiveness or disagreeableness for the two children.

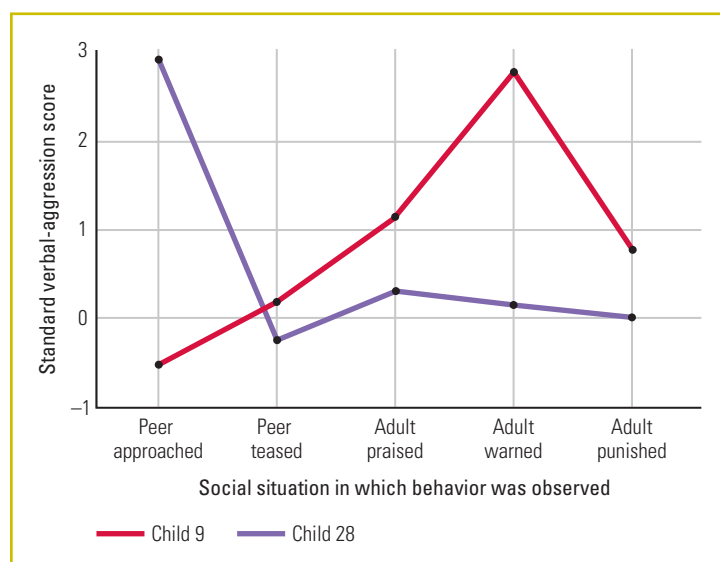


FIGURE 15.7 Situation-specific profiles of verbal aggression for two children Shoda, Mischel, and Wright (1994) recorded various categories of behaviors among emotionally disturbed children in various social situations at a summer camp. Shown here are results concerning verbal aggressiveness for two children. Zero on the y axis represents the average aggressiveness for all the children observed. In overall verbal aggressiveness, these two children were similar, but they were very different with respect to the situations that elicited their aggression.

Journal of Personality and Social Psychology, 67, 678 (Figure 1). Reproduced by permission.

Cross-Cultural Differences in Personality

If personality is, in part, ingrained beliefs and habits of thought that affect behavior, as social-cognitive theorists contend, then we should expect personality to vary across cultures. People growing up in different cultures are exposed to different values, philosophies, economic conditions, and models of how to behave.

Allocentrism-Ideocentrism as a Personality Dimension

As noted in Chapter 13, cultures vary in the degree to which they have a *collectivist* versus an *individualist* orientation. Collectivist cultures are those that emphasize the interdependence of people and the duties that people have to other members of their family and community. Individualist cultures, in contrast, place relatively more emphasis on personal freedom and rights, and relatively less emphasis on responsibilities to others. The dominant cultures of North America, Australia, and Western Europe generally fall on the individualist side of this dimension, and the dominant cultures of East Asia, Africa, and South America generally fall on the collectivist side. Many studies indicate that personalities of people in collectivist and individualist cultures differ from each other in predictable ways.

49

In general, how do personalities in collectivist cultures differ from those in individualist cultures?

<<< The collectivist and individualist manners of thinking and behaving, when taken as personality traits, are referred to respectively as **allocentrism** and **ideocentrism** (Triandis & Suh, 2002). Allocentric individuals are highly concerned with personal relationships and promoting the interests of the groups to which they belong. Ideocentric individuals, in contrast, focus more on their own interests and



B. Daemmrich / The Image Works



B. Daemmrich / The Image Works

abilities. While allocentrics emphasize the similarities between themselves and other group members, ideocentrics emphasize their own uniqueness. While allocentrics see themselves as responding primarily to the conditions of their social environment, ideocentrics see themselves as motivated by their own inner needs and aspirations. The humanistic concept of self-actualization is a quintessentially ideocentric concept.

As predicted, personality tests that are aimed at assessing the allocentrism-ideocentrism dimension show that people in collectivist cultures are, on average, more allocentric and less ideocentric than are people in individualist cultures. The differences, however, are by no means all or none. According to data summarized by Harry Triandis and his colleagues, roughly 40 percent of people in collectivist cultures score on the ideocentric side of this personality dimension, and roughly 40 percent of people in individualist cultures score on the allocentric side (Triandis & Suh, 2002). Allocentrics in individualist cultures tend to be loyal members of groups and avoid the “rat race” of individual competition; ideocentrics in collectivist cultures often report feeling oppressed and held back by their culture (Triandis & Suh, 2002).

Cultural differences in personality training

In the American art class (left-hand photo), a major goal is to allow children to express their individuality. In the Chinese art class (right-hand photo), a major goal is to impart traditional Chinese artistic methods so that the children’s paintings will resemble the model. Such differences in expectations and training lead children in China and the United States to develop different personality styles.

Cultural Differences in Conceptions of Personality

People in different cultures tend to differ not only in their average scores on various personality measures but also in their views about the significance of personality and the relative importance of particular traits. As discussed in Chapter 13, people in collectivist cultures, especially East Asian cultures, place less emphasis on personality than do people in individualist cultures. They are more likely to attribute individual differences in behavior to differences in the environmental situation than to inner personality differences. Moreover, when East Asians do talk about personality, they tend to emphasize trait dimensions that are somewhat different from those emphasized in individualist cultures.

For example, in China much emphasis is placed on such traits as *harmony* (inner peace of mind and a harmonious way of interacting with others), *face* (a concern with maintaining one’s dignity or reputation in relationships with others), and *ren qing* (a relationship orientation that emphasizes the mutual exchange of favors) (Lin & Church, 2004). None of these traits quite matches any of the facets of Western psychologists’ five-factor model (the model depicted in Table 15.3). When psychologists in China developed their own indigenous trait theory, using Chinese terms rather than translated versions of Western terms, their factor analysis produced factors that were in some ways quite different from those of Western psychologists’ five-factor model. The most clearly different factor was one that they labeled *interpersonal relatedness*, which includes elements of harmony, concern for reciprocity, and concern for traditional Chinese ways of relating to others (Cheung, 2004). My own reading of the findings of Asian psychologists suggests that Asians

>>> 50

In what sense do Westerners give more weight to the concept of personality than do East Asians?

>>> 51

What sorts of trait dimensions are emphasized in China more than in Western cultures?

tend to view personality in terms of styles of self-control, while Westerners tend to view personality in terms of spontaneous manifestations of drives and emotions, which vary from person to person.

Even today, most personality tests used in non-Western cultures are translated versions of personality inventories developed in the West. As many cultural psychologists have pointed out, the results of such tests may often be misleading, because people in other cultures may interpret the questions differently than do people in the West (Cheung, 2004). Only recently have psychologists in non-Western cultures begun to develop their own personality models and tests, using their cultures' own terms and concepts. Such research promises to enrich our understanding of the potential ways that human beings can differ from one another and of the value of such differences.

SECTION REVIEW

Social-cognitive theorists stress beliefs and habitual thoughts learned in the social context.

Habitual Ways of Thinking as Personality Traits

- People have an internal or external locus of control, depending on whether they do or do not believe that rewards are controlled by their own efforts.
- People have high or low self-efficacy, depending on whether they do or do not believe they can accomplish the relevant tasks.
- People with an internal locus of control and high self-efficacy tend to apply themselves more and to be more successful.
- In general, people with optimistic styles of thought cope better than others with life's demands. However, defensive optimism can cause harm, and some people use pessimism adaptively.

Domain-Specific and Situation-Specific Traits

- Locus of control and self-efficacy beliefs can be general, applying to many tasks, or domain-specific, applying to particular types of tasks. Domain-specific measures of these beliefs have the greatest predictive value.
- Social-cognitive theorists have also shown that traits such as conscientiousness and aggressiveness can vary across contexts, with the pattern of variation depending on the individual. They contend that situation-specific measures of traits have more predictive value than do global trait measures.

Cross-Cultural Personality Differences

- Because the social environment differs from one culture to another, social-cognitive theorists expect beliefs and habitual ways of thinking to differ cross-culturally.
- In collectivist cultures, such as those of East Asia, most people have allocentric personalities, which emphasize personal relationships and the interests of the group. In individualist cultures, most people are ideocentric, concerned more with their own interests and abilities.
- In non-Western cultures, the traits that are most useful in characterizing personality may not closely match the five-factor model.

Concluding Thoughts

In this chapter you read of ideas about how personality can be described, why personalities differ across individuals, and how such differences might be understood in terms of mental processes. Two general ideas might help you organize your thinking as you review the particular ideas of the chapter.

1. The varying purposes of personality theories Different personality theories have been developed to serve different purposes. Trait theorists try to distill the essential personality dimensions common to all people, while clinical theorists try to discover the mental processes and beliefs that help or hinder people in coping with life's demands.

Trait theories, such as the five-factor model, are attempts to describe the diversity of human personality objectively and efficiently, by identifying sets of non-redundant global traits and facets and ways to measure them. The trait measures, usually

made with questionnaires, are used in studies comparing one group of people with another, such as firstborns with laterborns, men with women, or people in one career with those in another. They are also used clinically (as explained in Chapter 17) as a first approximation to understanding individuals who seek help. Trait theories do not explain personality; they only describe its elements.

Psychodynamic, humanistic, and social-cognitive theories, in contrast, were designed to explain the particular behaviors, emotions, and thoughts of individual people, especially of people undergoing psychotherapy. Psychodynamic theories explain personality in terms of unconscious motives and defenses against anxiety. Humanistic theories explain personality in terms of people's subjective understanding of their world and themselves and their strivings for self-actualization. Social-cognitive theories also attempt to explain the behavior of indi-

viduals in terms of their beliefs but take a less holistic approach than do either humanistic or psychodynamic theories. Social-cognitive theorists are more often academic research psychologists than clinicians, and their interest tends to center more on a specific mental construct (such as locus of control) than on individuals as whole entities.

2. Adaptive functions of individual differences Because of the close tie between personality research and clinical research, personality theories have often been concerned with distinctions between healthy (or adaptive) and unhealthy (or maladaptive) personality styles. This concern is reflected in such distinctions as that between mature and immature defenses or between adaptive and maladaptive forms of optimism. An alternative

way to think about personality differences, however, is to view them as adaptations to different niches or as different strategies for solving life's problems. In this way of thinking, two quite different styles might be equally healthy or adaptive. In this chapter you saw examples of this idea reflected especially in research and theories concerning animal personality and concerning sibling and gender differences in personality.

As you review each of the dimensions of personality differences discussed in the chapter—including differences in defensive style, locus of control, optimism versus pessimism, and allocentrism versus ideocentrism, as well as the Big Five traits—think about ways in which variation in either direction could be either adaptive or maladaptive, depending on one's life circumstances.

Further Reading

DAVID C. FUNDER & DANIEL J. OZER (Eds.) (2004). *Pieces of the personality puzzle: Readings in theory and research* (3rd ed.). New York: Norton.

This is a collection of articles that have helped to shape contemporary personality psychology. The authors include such well-known psychologists as Sigmund Freud and Carl Jung (on the psychoanalytic perspective); Abraham Maslow and Carl Rogers (on the humanistic perspective); Walter Mischel and Albert Bandura (on the social-cognitive perspective); and Gordon Allport, Robert McCrae, and Paul Costa (on the trait perspective).

ANNELI RUFUS (2003). *Party of one: The loners' manifesto*. New York: Marlow.

You have read repeatedly, in this textbook, about the social nature of human beings. But the truth is, some of us humans are much less social than others. In this beautifully written, often humorous book, Anneli Rufus, herself a happy, well-adjusted loner, writes about the struggles that loners often face in a world of people who try to convince them that there is something weird or sad about preferring to be alone. If you

are by nature a loner, you will find comfort and good advice in this book; if you are not a loner, you may enjoy reading about the motives and feelings of people who are different from you.

SHELLEY E. TAYLOR (2002). *The tending instinct: How nurturing is essential for who we are and how we live*. New York: Holt.

Taylor is a prominent social psychologist who combines an understanding of social forces with a biological, evolutionary perspective. Here, for the general reader, she presents her thesis that human beings—and other mammals as well—respond to stress not just in terms of the classic “fight or flight” response, but also with a response that she calls “tend and befriend,” which involves turning toward others to give and receive help and comfort. Both men and women manifest, to varying degrees, both of these response patterns, but, overall, men show more of the first and women more of the second. These differences, primed by differences in sex hormones, provide a foundation for male–female personality differences, according to Taylor.

SIGMUND FREUD (1901; reprinted 1960). *The psychopathology of everyday life* (J. Strachey, Ed.; A. Tyson, Trans.). New York: Norton.

This is one of Freud's most popular and fun-to-read books. It is full of anecdotes having to do with forgetting, slips of the tongue, and bungled actions. In each anecdote, Freud argues that an apparent mistake was really an expression of an unconscious wish.

EDWARD C. CHANG (Ed.) (2001). *Optimism and pessimism: Implications for theory, research, and practice*. Washington, DC: American Psychological Association.

Many popular books have been written on the value of optimistic thinking. This book, in contrast, is a scholarly but readable summary of research on optimism and pessimism. Each chapter is by a different expert or set of experts. One emerging theme is that the typical and perhaps most adaptive personality includes a balance of optimism and pessimism, each of which, in its own way, can help the person to prepare for and meet life's challenges.



Mental Disorders

THE CONCEPT OF MENTAL DISORDER AND ITS RELATION TO CULTURE

- Categorizing and Diagnosing Mental Disorders
- Cultural Variations in Disorders and Diagnoses

CAUSES OF MENTAL DISORDERS

ANXIETY DISORDERS

- Generalized Anxiety Disorder
- Phobias
- Obsessive-Compulsive Disorder
- Panic Disorder
- Posttraumatic Stress Disorder

MOOD DISORDERS

- Depression
- Bipolar Disorders

PSYCHOLOGICAL INFLUENCES ON PHYSICAL SYMPTOMS AND DISEASES

- Somatoform Disorders
- Psychological Factors Affecting Medical Condition

SCHIZOPHRENIA

- Diagnostic Characteristics of Schizophrenia
- Underlying Cognitive and Neural Deficits in Schizophrenia
- Genetic and Environmental Causes of Schizophrenia

A theme running through this book is that psychological processes are usually adaptive; that is, they usually promote survival and well-being. Drives and emotions, including those experienced as uncomfortable or painful, typically motivate survival-enhancing actions; perceptions provide useful information to guide such actions; and thoughts produce effective plans for actions. But sometimes these processes break down and become maladaptive: Drives become too strong, too weak, or misdirected; emotions become overwhelming; perceptions become distorted; thoughts become confused; and behavior becomes ineffective. All of us experience such disturbances occasionally, to some degree, and accept them as a normal part of life. But sometimes such disturbances are so strong, prolonged, or recurrent that they seriously interfere with a person's ability to live a satisfying life. Then the person is said to have a mental disorder.

This chapter begins with general discussions of the nature and causes of mental disorders and then moves on to separate sections devoted to some of the most common classes of disorders—including anxiety disorders, mood disorders, and schizophrenia.

The Concept of Mental Disorder and Its Relation to Culture

Mental disorder is a fuzzy concept, impossible to define in a precise way. This should come as no surprise. Most concepts, if you think about them, are fuzzy. Try to define precisely, for example, such an everyday concept as a house. Some things are clearly houses, and others, like the shell that that a turtle carries on its back, are only "sort of" houses. Every attempt to define *mental disorder* contains ambiguities and therefore raises controversy.

1 How is the concept of mental disorder defined by the American Psychiatric Association? What ambiguities lie in that definition?

<<< The most frequently used definition of mental disorder today is that which appears in the current (fourth, text-revised) edition of the *Diagnostic and Statistical Manual of Mental Disorders*, referred to as **DSM-IV**, the American Psychiatric Association's official guide for diagnosing mental disorders. This definition regards mental disorders as analogous to medical diseases and borrows from medicine the terms *symptom* and *syndrome*. A **symptom** is any characteristic of a person's actions, thoughts, or feelings that could be a potential indicator of a mental disorder, and a **syndrome** is a constellation of interrelated symptoms manifested by a given individual. According to *DSM-IV* (American Psychiatric Association, 2000), a syndrome is considered to be a mental disorder if, and only if, it satisfies the following three criteria:



Cleo Freelance / Jeroboam, Inc.

Mental distress or mental disorder? Feelings of sadness, pessimism, and low self-esteem are evident here, but is the source of the distress the situation or something inside the person? This question is central to defining the concept *mental disorder*.

- *It involves a clinically significant detriment.* The syndrome must involve *distress* (painful feelings) or *impairment of functioning* (interference with ability to work, play, or get along with people) or both, and it must be *clinically significant*, meaning that the distress or impairment must be serious enough to warrant professional treatment.
- *It derives from an internal source.* The source of distress or impairment must be located within the person—that is, in the person's biology, mental structure (ways of perceiving, thinking, or feeling), or learned habits—and not in the person's immediate environment. The distress or impairment may have been prepared by past environmental circumstances and may be aggravated or triggered by present circumstances, but it cannot be simply an expectable reaction to present circumstances, such as despondency brought on by poverty or a period of grief brought on by the death of a loved one.
- *It is not subject to voluntary control.* The syndrome cannot be the result of a deliberate, voluntary decision to act in a manner contrary to the norms of society. Thus, a person who voluntarily undergoes starvation to protest a government policy, or who deliberately behaves in bizarre ways in order to amuse or shock others, is not considered to have a mental disorder.

Although these criteria are useful guidelines for thinking about and identifying mental disorders, they are by nature ambiguous and in some cases seem to be contradicted by the descriptions of particular disorders in *DSM-IV* (Widiger, 2005; Widiger & Sankis, 2000). Just how “distressing” or “impairing” must a syndrome be to be considered “clinically significant”? As all behavior involves an interaction between the person and the environment, how can we tell whether the impairment is really within the person, rather than just in the environment? For example, in the case of someone living in poverty or experiencing discrimination, how can we tell if the person's actions are normal responses to those conditions or represent something more? When people claim that they are deliberately choosing to behave in a way that violates social norms and could behave normally if they wanted to, how do we know when to believe them? A person who starves to protest a government policy may not be mentally disordered, but what about a person who starves to protest the “fact” that Martians have seized control of the universe? Who has the right to decide whether a person is or is not mentally disordered: the person, the person's family, a psychiatrist or psychologist, a court of law, a health insurance administrator who must approve or not approve payment for therapy?

These are tough questions that can never be answered strictly scientifically. The answers always represent human judgments, and they are always tinged by the social values and pragmatic concerns of those doing the judging.

Categorizing and Diagnosing Mental Disorders

We humans are inveterate categorizers. Rarely do we see an unfamiliar object or event as completely new and unique; instead, we see it as a member of some familiar category—a house, a fruit, a party. Our world is more predictable and describable when we can relate each new object and event to a familiar category, even if the categories are fuzzy and the boundaries arbitrary. This same response applies in the realm of mental disorders. Beginning long before the era of modern psychology and psychiatry, people everywhere have had systems for categorizing and labeling their psychological miseries. In keeping with the common Western practice of likening mental disorders to physical diseases, the process of assigning a label to a person's mental disorder is referred to as *diagnosis*.

Classification and diagnosis are essential to the scientific study of mental disorders. Without a system to identify people whose disorders are similar, the accumulation of knowledge about causes, effective treatments, and eventual outcomes would be impossible. But such a system is useful only to the degree that it is reliable and valid.

The Quest for Reliability: Development of *DSM-IV*

The *reliability* of a diagnostic system refers to the extent to which different diagnosticians, all trained in the use of the system, reach the same conclusion when they independently diagnose the same individuals. If you have ever gone to two different doctors with a physical complaint and been given two different, non-synonymous labels for your disease, you know that, even in the realm of physical disorders, diagnosis is by no means completely reliable. Until relatively recently, psychiatrists and clinical psychologists had no reliable, agreed-upon method for diagnosing mental disorders. As a consequence, the same troubled person might be given a diagnosis of schizophrenia by one clinician, depression by another, and dementia by a third.

As a first step toward remedying this sort of problem, the American Psychiatric Association published the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* as a standard system for labeling and diagnosing mental disorders. The first two editions, which appeared in 1952 and 1968, were quite slim; diagnostic categories were defined briefly, in general terms, and much was left to the judgment of individual clinicians. The result was some degree of standardization in labeling, but still relatively little diagnostic reliability (Matarazzo, 1983; Spitzer & Fleiss, 1974). The third edition of the manual, referred to as *DSM-III*, which appeared in 1980, was much thicker and more useful.

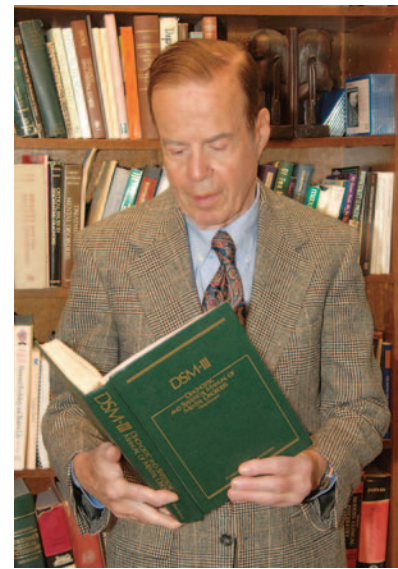
The goal of the hundreds of psychologists, clinical psychologists, and other mental health professionals who produced *DSM-III* was to remove as much guesswork as possible from the task of diagnosis and thereby increase reliability. To do this, they strove to define mental disorders as objectively as possible, in terms of symptoms that clinicians could observe or could assess by asking relatively simple, straightforward questions of the person being diagnosed or of those who knew that person well. To test alternative ways of diagnosing each disorder, they conducted field studies in which people who might have a particular disorder were diagnosed independently by a number of clinicians or researchers using each of several alternative diagnostic systems. In general, the systems that produced the greatest reliability—that is, the greatest agreement among the diagnosticians as to who had or did not have a particular disorder—were retained. As a result, the diagnostic reliability of *DSM-III* was quite high—comparable in many cases to the levels achieved in diagnosing physical disorders (Matarazzo, 1983; Segal & Coolidge, 2001). More recently, further revisions have made the manual even more reliable (Widiger, 2005). The most recent edition is the text-revised version of *DSM-IV*, published in 2000. (This version is sometimes referred to as *DSM-IV-TR*. The text revision involved an improvement in wording but little change in content from the previous version of *DSM-IV*, published in 1994.)

As an example of diagnostic criteria specified in *DSM-IV*, consider those for *anorexia nervosa*, an eating disorder that can result in self-starvation. The person must (a) refuse to maintain body weight at or above a minimally normal weight for age and height; (b) express an intense fear of gaining weight or becoming fat; (c) manifest a disturbance in the experience of her or his own body weight or shape, show an undue influence of body weight or shape on self-evaluation, or deny the seriousness of the current low body weight; and (d) if a postpubertal female, have missed at least three successive menstrual periods (a condition brought on by a lack of body fat). If any one of these criteria is not met, a diagnosis of *anorexia nervosa* would not be made. Notice that all these criteria are based on observable characteristics or self-descriptions by the person being diagnosed; none rely on inferences about underlying causes or unconscious symptoms that could easily result in disagreement among diagnosticians who have different perspectives.

The Question of Validity

The *validity* of a diagnostic system is an index of the extent to which the categories it identifies are clinically meaningful. (See Chapter 2, pp. 40–41, for a more general discussion of both validity and reliability.) In theory, a diagnostic system could be highly reliable without being valid. For example, a system that reliably categorizes

What does reliability mean with reference to a diagnostic system, and how did the developers of the *DSM* strive to increase reliability?



Leader of a revolution As head of the task force that created the third edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)*, Robert Spitzer, shown here, led a revolution in the manner by which mental disorders are diagnosed. He and those who worked with him created detailed checklists of objective symptoms for each disorder described in the manual. Spitzer views this and subsequent editions of the manual not as finished products but as steps in an ongoing process of creating more accurate labels and diagnostic guidelines. He refers to each edition of the manual as “a set of hypotheses to be tested.”

How does validity differ from reliability? How can the validity of the *DSM* be improved through further research and revisions?

a group of people as suffering from Disorder X, on the basis of certain superficial characteristics, would not be valid if further work failed to reveal any clinical usefulness in that diagnosis. Do people with the same diagnosis truly suffer in similar ways? Does their suffering stem from similar causes? Does the label help predict the future course of the disorder and help in deciding on a beneficial treatment? To the degree that questions like these can be answered in the affirmative, a diagnostic system is valid.

The question of validity is much more complicated than that of reliability and must be based on extensive research. The developers of *DSM-III* and *DSM-IV* maintain that reliability is a prerequisite for validity (Spiegel, 2005; Wilson, 1993). In order to conduct the research needed to determine whether or not a diagnosis is valid, by the criteria listed earlier, one must first form a tentative, reliable diagnostic system. To determine whether people suffering from anorexia nervosa, for example, have similar histories, experience similar outcomes, and benefit from similar treatment, one must first have a consistent way of deciding who falls into the population to be studied. The results of such studies can lead to new means of diagnosis or to new subcategories that describe variations within the original category, leading to increased validity.

The main categories of disorders identified in *DSM-IV* are listed in Table 16.1.

Possible Dangers in Labeling

4

What are some negative consequences of labeling a person as mentally disordered? What is recommended as a partial solution?

<<< Diagnosing and labeling may be essential for the scientific study of mental disorders, but labels can be harmful. A label implying mental disorder can blind clinicians and others to qualities of the person that are not captured by the label, can reduce the esteem accorded to the person by others, can reduce the labeled person's self-esteem, and, through these means, can interfere with the person's ability to cope well with his or her environment.

To reduce, at least somewhat, the likelihood of such effects, the American Psychiatric Association (2000) recommends that clinicians apply diagnostic labels only to people's disorders, not to people themselves. For example, a client or patient might be referred to as *a person who has schizophrenia* or *who suffers from alcoholism* but should not be referred to as *a schizophrenic* or *an alcoholic*. The distinction might at first seem subtle, but if you think about it, you may agree that it is not so subtle in psychological impact. If we say, "John has schizophrenia," we tend to be reminded that John is first and foremost a person, with qualities like those of other people, and that his having schizophrenia is just one of many things we could say about him. In contrast, the statement "John is a schizophrenic" tends to imply that everything about him is summed up by that label.

As I talk about specific disorders in the remainder of this chapter and the next, I will attempt to follow this advice even though it produces some awkward wording at times. I also urge you to add, in your mind, yet another step of linguistic complexity. When I refer to "a person who has schizophrenia," you should read this statement as "a person *who has been diagnosed by someone as having* schizophrenia," keeping in mind that diagnostic systems are never completely reliable.

Medical Students' Disease

The ability of labels to cause psychological harm demonstrates the power of suggestion, which also underlies *medical students' disease*. This disease, which could also be called *introductory psychology students' disease*, is characterized by a strong tendency to relate personally to, and to find in oneself, the symptoms of any disease or disorder described in a textbook. Medical students' disease was described by the nineteenth-century humorist Jerome K. Jerome (1889/1982) in an essay about his own discomfort upon reading a textbook of medical diagnoses. After explaining his discovery that he must have typhoid fever, St. Vitus's dance, and a multitude of



"You made that diagnosis just to be mean."

TABLE 16.1
Summary of *DSM-IV* categories of mental disorders

Anxiety disorders* Disorders in which fear or anxiety is a prominent symptom. Examples: <u>generalized anxiety disorder</u>, <u>phobias</u>, <u>obsessive-compulsive disorder</u>, <u>panic disorder</u>, and <u>posttraumatic stress disorder</u>. Mood disorders* Disorders marked by depression or mania. Examples: <u>major depression</u>, <u>dysthymia</u>, <u>bipolar I disorder</u>, and <u>bipolar II disorder</u>. Somatoform disorders* Disorders involving physical (somatic) symptoms arising from unconscious psychological processes. Examples: <u>somatization disorder</u> and <u>conversion disorder</u>. Schizophrenia and other psychotic disorders* <u>Schizophrenia</u> is marked by disorganized thought and speech, delusions, hallucinations, disorganized behavior, and flattened or inappropriate affect. Another psychotic disorder is <u>delusional disorder</u>, which involves persistent delusions (usually of persecution) not accompanied by other disruptions of thought or mood that would lead to a diagnosis of schizophrenia. Dissociative disorders Disorders in which a part of one's experience is separated off (dissociated) from one's conscious memory or identity. Examples: <u>psychogenic amnesia</u>, <u>fugue states</u>, and <u>dissociative identity disorder</u>. Disorders usually first diagnosed in infancy, childhood, or adolescence A diverse group of disorders that always or almost always first appear before adulthood. Included are <u>Down syndrome</u> and other forms of mental retardation, <u>autism</u>, and <u>ADHD</u> (<u>attention deficit/hyperactivity disorder</u>). Delirium, dementia, amnesia, and other cognitive disorders A diverse group of disorders of perception, memory, and thought that stem from known damage to the brain. Included are disorders due to strokes, physical trauma to the brain, and degenerative brain diseases such as <u>Alzheimer's disease</u>. Eating disorders Disorders marked by extreme undereating, overeating, or purging or by excessive concern about gaining weight. Examples: <u>anorexia nervosa</u> and <u>bulimia nervosa</u>.	Substance-related disorders Disorders brought on by drugs such as alcohol, cocaine, or opiates. Included are <u>dependence</u> (the intense craving for the substance), <u>abuse</u> (use of the substance in ways that harm the self or others), and effects of brain damage caused by prolonged use of the substance. Sexual and gender identity disorders <u>Sexual disorders</u> are those of sexual functioning; they include <u>paraphilias</u> (in which bizarre imagery or acts are necessary for sexual excitement, such as <u>fetishism</u>, <u>exhibitionism</u>, and <u>sexual sadism</u>) and <u>psychosexual dysfunctions</u> (inability to become sexually aroused or to complete intercourse in a satisfying way). <u>Gender identity disorders</u> involve strong and persistent desires to be, or appear to be, a member of the other gender. Impulse control disorders not elsewhere specified Disorders characterized by impulsive behaviors that are harmful to the self or others. Examples: <u>intermittent explosive disorder</u> (outbursts of aggression resulting in assault or property destruction), <u>kleptomania</u> (impulsive stealing), <u>pyromania</u> (impulsive setting of fires), and <u>pathological gambling</u>. Sleep disorders Disorders include <u>insomnia</u> (too little sleep), <u>hypersomnia</u> (too much sleep), <u>sleep-wake disorder</u> (inability to establish a sleep-wake cycle corresponding with the 24-hour day), and disorders involving sleepwalking, fear of sleep, or fear of nightmares. Adjustment disorder Maladaptive, excessive emotional reaction to an identified stressful event that occurred within the previous 6 months. Personality disorders† Disorders involving inflexible, maladaptive personality traits. Examples: <u>antisocial personality disorder</u> (a history of antisocial acts and violation of others' rights, with no sense of guilt), <u>histrionic personality disorder</u> (excessively emotional, overly dramatic attention seeking), and <u>narcissistic personality disorder</u> (unwarranted sense of self-importance and demand for constant attention or admiration).
--	--

*The first four categories in the list correspond with major sections of this chapter. Throughout the table, all disorders that are referred to or described in the chapter are underlined.
†Because personality disorders involve a person's long-standing style of thinking and acting rather than a change in the person, they are categorized on a separate dimension, or axis (Axis II), from the other categories (which constitute Axis I).
Source: *Diagnostic and statistical manual of mental disorders* (4th, text-revised ed.). Washington, D.C.: American Psychiatric Association, 2000.

diseases he had never heard of before, he wrote, "The only malady I concluded I had *not* got was housemaid's knee. . . . I had walked into that reading-room a happy, healthy man. I crawled out a decrepit wreck." As you read about specific disorders later in this chapter, brace yourself against medical students' disease. Everyone has at least some of the symptoms, to some degree, of essentially every disorder that can be found in this chapter, in *DSM-IV*, or in any other compendium.

Cultural Variations in Disorders and Diagnoses

Mental disorder is, to a considerable degree, a cultural construct. The kinds of distress that people experience, the ways in which they express that distress, and the ways in which other people respond to the distressed person vary greatly from culture to culture and over time in any given culture. Moreover, cultural beliefs and values help determine whether particular syndromes are considered to be disorders or variations of normal behavior.

Culture-Bound Syndromes

The most striking evidence of cross-cultural variation in mental disorders can be found in *culture-bound syndromes*—expressions of mental distress that are almost completely limited to specific cultural groups. In some cases, such syndromes represent exaggerated forms of behaviors that, in more moderate forms, are admired by the culture.

Cultural foundation for *taijin kyofusho*

Modesty and concern not to offend others are much valued in Japan. Such modesty and concern, when taken to the extreme, may provide the foundation for the mental disorder *taijin kyofusho*.



Charles Gupton / Getty Images

5

How do *taijin kyofusho* and eating disorders illustrate the role of culture in the mental disorders that people develop?

<<< An example of such a syndrome is *taijin kyofusho*, which is one of the most commonly diagnosed disorders in Japan but is almost unknown in other parts of the world (Dinnel & others, 2002; Tarumi & others, 2004). This disorder, which is diagnosed in men more often than in women, involves an incapacitating fear of offending or harming others through one's own awkward social behavior or imagined physical defect. In particular, the person may believe that he offends others through blushing, or through eye contact, or because of some imagined physical defect or a foul odor that he emits. Unlike the case for social phobias in the West, the conscious fear in *taijin kyofusho* centers on the possibility of harming others, not on possible harm to the self from others (such as being rejected or ridiculed). Japanese psychiatrists consider the disorder to be a pathological exaggeration of the modesty and sensitive regard for others that, at lower levels, are considered proper in Japan (Kirmayer, 1991).

Examples of culture-related syndromes closer to home are the eating disorders anorexia nervosa and bulimia nervosa. *Anorexia nervosa*, as noted earlier, is characterized by an extraordinary preoccupation with thinness and a refusal to eat, sometimes to the point of death by starvation. *Bulimia nervosa* is characterized by periods of extreme binge eating followed by self-induced vomiting, misuse of laxatives or other drugs, or other means to undo the effects of a binge. It is probably not a coincidence that these disorders began to appear with some frequency in the 1970s in North America and Western Europe, primarily among adolescent girls and young women of the middle and upper classes, and that their prevalence increased through the remainder of the twentieth century (Gordon, 1990; Hoek, 2002). During that period, Western culture became increasingly obsessed with dieting and an ideal of female thinness while, at the same time, weight control became more difficult because of the increased availability of high-fat foods. At present, somewhere between 1 and 2 percent of women in Western industrialized cultures meet the *DSM-IV* criteria for anorexia nervosa or bulimia nervosa (Porzelius & others, 2001; Powers & Bannon, 2004). In non-Western cultures that have not been heavily influenced by Western habits and values, anorexia nervosa is much less prevalent and bulimia nervosa is essentially unknown (Keel & Klump, 2003).



Joseph Neumayer / Design Conceptions

Western values and abnormality Anorexia nervosa and bulimia nervosa are eating disorders found largely in Western countries. Many clinicians believe that these are culture-bound disorders caused in part by Western society's overemphasis on thinness as the aesthetic ideal for women, a preoccupation on display in Western ads, magazines, movies, and the like.

Role of Cultural Values in Determining What Is a Disorder

Culture does not affect just the types of behaviors and syndromes that people manifest; it also affects clinicians' decisions about what to label as disorders. A prime example has to do with homosexuality. Until 1973, homosexuality was officially—according to the American Psychiatric Association—a mental disorder in the United States; in that year, the association voted to drop homosexuality from its list of disorders (Minton, 2002). The vote was based partly on research showing that the suffering and impairment associated with homosexuality derived not from the condition itself but from social prejudice directed against homosexuals. The vote was also prompted by an increasingly vocal gay and lesbian community that objected to their sexual preferences being referred to as a disorder, and by gradual changes in attitudes among many in the straight community, who were beginning to accept the normality of homosexuality.

>>> **6**
How does the example of homosexuality illustrate the role of culture in determining what is or is not a “disorder”?

Cultural Values and the Diagnosis of ADHD

The American Psychiatric Association has added many more disorders to *DSM* over the last three or four decades than it has subtracted. The additions have come partly from increased understanding and partly from a general cultural shift toward seeing mental disorder where people previously saw normal human variation. An example that illustrates both of these trends is *attention deficit/hyperactivity disorder*, more commonly known by its acronym **ADHD**, which is currently the single most frequently diagnosed disorder among children in the United States (Bradley & Golden, 2001).

Most diagnoses of ADHD originate because of difficulties in school. Indeed, many of the *DSM-IV* symptoms of the disorder refer specifically to schoolwork (American Psychiatric Association, 2000). The manual describes three varieties of ADHD: The *predominantly inattentive type* is characterized by lack of attention to instructions, failure to concentrate on schoolwork or other such tasks, and carelessness in completing assignments. The *predominantly hyperactive-impulsive type* is characterized by such behaviors as fidgeting, leaving one's seat without permission, talking excessively, interrupting others, and blurting out answers before the question is completed. The *combined type* is characterized by both sets of symptoms.

>>> **7**
How is ADHD identified, explained, and treated? How do critics of the high rates of diagnosis of ADHD in the United States explain those high rates?

A prominent theory of the neural basis of ADHD is that it involves deficits in the prefrontal lobes of the cortex—a portion of the brain responsible for focusing attention on tasks and inhibiting spontaneous activities (Barkley, 2001). Consistent with this theory, researchers have found that the right prefrontal cortex is, on average, slightly smaller in children diagnosed with ADHD than in other children (Banaschewski & others, 2005). The most common treatment, by far, is the drug methylphenidate, sold in various short- and long-acting forms under such trade names as Ritalin and Concerta. Methylphenidate increases the activity of the neurotransmitters dopamine and norepinephrine in the brain, and its effectiveness may derive from its ability to boost neural activity in the prefrontal cortex (Keating & others, 2001). This drug reduces the immediate symptoms of ADHD in most diagnosed children, but, to date, there have been no long-term studies showing that the drug improves children's lives overall (LeFever & others, 2003; Pehlam & others, 2005).

Normal playfulness or ADHD?

There is no clear dividing line between normal rambunctious playfulness in children and ADHD. As a result, ADHD is diagnosed at much higher levels in some school districts than in others.

No other country in the world diagnoses ADHD at anything close to the rate seen in the United States, which consumes 90 percent of the world's production of methylphenidate (Diller, 1999). During the 1990s, the rate of diagnosis and treatment of ADHD in the United States increased eightfold, and currently



somewhere between 4 and 8 percent of U.S. schoolchildren are diagnosed with the disorder (LeFever & others, 2003). Since boys are diagnosed at rates that are 3 to 10 times those of girls (Rafalovich, 2004), this means that the prevalence of ADHD diagnosis among boys is somewhere between 6 and 15 percent. The rate varies tremendously, however, depending on location. By actual count, in two middle-class school districts in Virginia, 33 percent of white boys were diagnosed with ADHD (LeFever & others, 2003). Whites, in general, are diagnosed with the disorder more frequently than are African Americans, not because they are more disruptive in school but because they are more often referred to clinicians for diagnosis and treatment (LeFever & others, 2003).

Some psychologists and sociologists have argued that the explosion in diagnosis of ADHD in the United States derives at least partly from the nation's current obsession with school performance (Rafalovich, 2004). Children, especially boys, everywhere get into trouble in school because of their need for vigorous activity, their impulsiveness, their carelessness about schoolwork, and their willingness to defy teachers and other authority figures. These characteristics vary in a continuous manner, showing the typical bell-shaped distribution of any normal personality dimension (Nigg & others, 2002). Even defenders of the high rate of diagnosis and treatment of ADHD acknowledge that the characteristics of this disorder exist to some degree in all children, more so in boys than in girls, and that children classed as having ADHD are simply those who have these characteristics to a greater extent than do other children (Maddux & others, 2005; Stevenson & others, 2005). Also widely acknowledged is the fact that most (although not all) diagnosed children eventually outgrow their ADHD symptoms (Rowland & others, 2002).

Some critics of the high rate of ADHD diagnosis—including a prominent neuroscientist who has conducted research on brain systems involved in this condition (Panksepp, 1998, 2002)—contend that American society has made it increasingly difficult for children to engage in the sorts of rough-and-tumble free play needed by all young mammals, especially young males, for normal development. As a culture, they contend, we have chosen to treat many children with strong drugs, whose long-term consequences are still unknown, rather than to develop systems of schooling that can accommodate a wider range of personalities.

SECTION REVIEW

Mental disorder presents numerous conceptual, diagnostic, and social challenges.

Categorizing and Diagnosing Mental Disorders

- To be considered a mental disorder by *DSM-IV* standards, a syndrome (set of interrelated symptoms) must involve a clinically significant detriment, derive from an internal source, and not be subject to voluntary control. Though these guidelines are useful, “mental disorder” is still a fuzzy concept.
- Classification and diagnosis (assigning a label to a person's mental disorder) are essential for clinical purposes and for scientific study of mental disorders.
- *DSM-III* and *DSM-IV* substantially increased reliability (the probability that independent diagnosticians would agree about a person's diagnosis) by using objective symptoms. Validity is a more complex issue.
- Because labeling a person can have negative consequences (e.g., lowering self-esteem or the esteem of others), it is advised that labels be applied only to the disorder, not to the person.
- Beware of medical students' disease—the tendency to relate personally (and falsely) to the disorders described in a textbook.

Cultural Variations in Disorders and Diagnoses

- Culture-bound syndromes are expressions of mental distress limited to specific cultural groups. Examples are *taijin kyofusho* (in Japan) and bulimia nervosa (in cultures influenced by modern Western values).
- Culture also affects the types of behaviors or characteristics thought to warrant a diagnosis of mental disorder. Until 1973, homosexuality was officially classed as a mental disorder in the United States.
- The great increase in diagnosis of ADHD (attention deficit/hyperactivity disorder) in the United States may result not just from increased understanding but also from reduced opportunity for and tolerance of rough-and-tumble play and an increased emphasis on school performance.

Causes of Mental Disorders

All thoughts, emotions, and actions, whether they are adaptive or maladaptive, are products of the brain. Differences among people in predisposition to any type of mental disorder are differences that reside in the brain. All the factors that contribute to the causation of mental disorders do so by acting, in one way or another, on the brain. These include genes that influence brain development; environmental assaults on the brain, such as those produced by a blow to the head, oxygen deprivation, viruses, or bacteria; and, more subtly, the effects of learning, which are consolidated in pathways in the brain.

Irreversible Mental Disorders Derive from Irreversible Brain Damage

The role of the brain is most obvious in certain chronic mental disorders—that is, in certain disorders that stay with a person for life once they appear. In these cases the brain deficits are irreversible and affect processes that are crucial for daily functioning in all normal environments. Three quite prevalent examples of such disorders are Down syndrome, autism, and Alzheimer's disease.

Down syndrome is a congenital (present at birth) disorder that appears in about 1 out of every 800 newborn babies (American Psychiatric Association, 2000). It is caused by an error in meiosis, which results in an extra chromosome 21 in the egg cell or (less often) the sperm cell (the numbering of chromosomes is shown in Figure 3.3, p. 52). The extra chromosome is retained in the zygote and in all cells of the newly developing individual. Through means that are not well understood, it causes damage to many regions the developing brain, such that the person goes through life with moderate to severe mental retardation and with difficulties in physical coordination.

Autism, or *autistic disorder*, is a congenital disorder that is characterized primarily by a severe deficit in the social instincts. Infants with autism fail to smile at smiling human faces, respond negatively or not at all to cuddling or other signs of affection, fail to pay special attention to people and to human speech, and, in general, seem to treat people as they would any other objects in their environment (Baron-Cohen & Belmonte, 2005; Dawson & others, 2004). Apparently because of their inattention to people or speech, they usually do not develop language except through special training, and their language is never as natural and spontaneous as is that of other people. They fail to develop an ability to empathize with others' emotions or to understand what is going on in other people's minds (Baron-Cohen & Belmonte, 2005). This disorder, which occurs in approximately 1 person out of

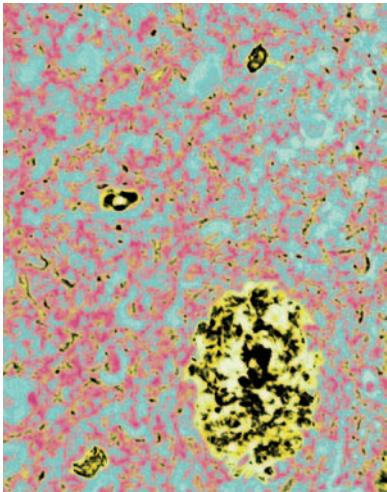
>>> **8**
How are Down syndrome, autism, and Alzheimer's disease characterized as brain diseases?



© Rosalie Winard

A voice of autism Temple Grandin, despite having autism, developed language through special training and has gone on to a successful career in agricultural science. She claims that her unique perspective on the world, in which she has had to learn about human beings as though she were “an anthropologist on Mars,” has given her a better understanding of non-human animals. Her autobiography is one of the Further Readings listed at the end of this chapter.

Simon Fraser / Photo Researchers, Inc.



Alzheimer's brain This slice of brain tissue (magnified X50) from a recently deceased person who suffered from Alzheimer's disease reveals a large amyloid plaque—the yellow and black structure in the lower right-hand corner.

500, appears to be caused in some cases primarily by genes and in other cases primarily by prenatal toxins or birth complications that disrupt normal brain development (Kabot & others, 2003).

Alzheimer's disease, found primarily in the elderly, is becoming increasingly prevalent as ever more people live into old age. It occurs in about 1 percent of people who are in their sixties, 3 percent of those in their seventies, 12 percent of those in their eighties, and 40 percent of those in their nineties (Nussbaum & Ellis, 2003). The disorder is characterized psychologically by a progressive deterioration, over the final years of the person's life, in all cognitive abilities—including memory, reasoning, spatial perception, and language—followed by deterioration in the brain's control of bodily functions. Neurologically, the disorder is characterized by certain physical disruptions in the brain, including the presence of *amyloid plaques*. The plaques are deposits of a particular protein, called *beta amyloid*, which form in the spaces between neurons and may disrupt neural communication (Nicoll & others, 2004). The disorder appears to be caused by a combination of genetic predisposition and general debilitating effects of old age. Among the genes that may contribute are those that affect the rate of production and breakdown of beta amyloid. Age may contribute partly through the deterioration of blood vessels, which become ever less effective in carrying excess beta amyloid out of the brain (Nicoll & others, 2004).

Role of the Brain in Episodic Mental Disorders

Many disorders, including all the disorders discussed in the remaining sections of this chapter, are *episodic*, meaning that they are reversible. They may come and go, in episodes. Episodes of a disorder may be brought on by stressful environmental experiences, but the predisposition for the disorder nevertheless resides in one way or another in the brain.

Most mental disorders, including those that are episodic, are to some degree heritable. The more closely related two people are, genetically, the more likely it is that they share the same mental disorder or disorders, regardless of whether or not they were raised in the same home (Howland, 2005; Rutter, 2002). In most cases it is not known just which genes are involved or how they influence the likelihood of developing the disorder, but it is reasonable to assume that such effects occur primarily through the genes' roles in altering the biology of the brain. Environmental assaults to the brain and the effects of learning can also contribute to the predisposition for episodic disorders.

A Framework for Thinking About Multiple Causes of Mental Disorders

Most mental disorders derive from more than one cause. Most of them are not present at birth, but manifest themselves at some point later in life, often in early adulthood, and the subsequent course of most disorders is affected by experiences that one has after the disorder begins. It is useful, therefore, to distinguish among three categories of causes of mental disorders: predisposing, precipitating, and perpetuating causes—the *three Ps*.

Predisposing causes of mental disorders are those that were in place well before the onset of the disorder and make the person susceptible to the disorder. Genetically inherited characteristics that affect the brain are most often mentioned in this category. Predispositions for mental disorders can also arise from damaging environmental effects on the brain, including effects that occur before or during birth. Such environmental assaults as poisons (including alcohol or other drugs consumed by the mother during pregnancy), birth difficulties (such as oxygen deprivation during birth), and viruses or bacteria that attack the brain can predispose a child for the subsequent development of one or more mental disorders.

Prolonged psychologically distressing situations—such as living with abusive parents or an abusive spouse or having to take responsibility for events that cannot

9

How can the causes of mental disorders be categorized into three types—"the three Ps"?



be controlled—can also predispose a person for one or another mental disorder. Other predisposing causes include certain types of learned beliefs and maladaptive patterns of reacting to or thinking about stressful situations. A young woman reared in upper-class Western society is more likely to acquire beliefs and values that predispose her to anorexia nervosa than is a young woman from a rural community in China. Highly pessimistic habits of thought, in which one regularly anticipates the worst and fails to think about reasons for hope, predispose people for mood disorders (particularly depression) and anxiety disorders.

Precipitating causes are the immediate events in a person's life that bring on the disorder. Any loss, such as the death of a loved one or the loss of a job; any real or perceived threat to one's well-being, such as physical disease; any new responsibilities, such as might occur as a result of marriage or job promotion; or any large change in the day-to-day course of life can, in the sufficiently predisposed person, bring on the mood or behavioral change that leads to diagnosis of a mental disorder. Precipitating causes are often talked about under the rubric of *stress*, a term that sometimes refers to the life event itself and sometimes to the worry, anxiety, hopelessness, or other negative experiences that accompany the life event (Lazarus, 1993). When the predisposition is very high, an event that seems trivial to others can be sufficiently stressful to bring on a mental disorder. When the predisposition is very low, even an extraordinarily high degree of loss, threat, or change may fail to bring on a mental disorder. Figure 16.1 depicts this inverse relationship.

Perpetuating causes are those consequences of a disorder that help keep it going once it begins. In some cases, a person may gain rewards, such as extra attention, which help perpetuate the maladaptive behavior. More often, the negative consequences of the disorder help perpetuate it. For example, a sufferer of depression may withdraw from friends, and lack of friends can perpetuate the depression. Behavioral changes brought on by a disorder, such as poor diet, irregular sleep, and lack of exercise, may also, through physiological means, help to prolong the disorder. Expectations associated with a particular disorder may also play a perpetuating role. In a culture that regards a particular disorder as incurable, a person diagnosed with that disorder may simply give up trying to change for the better.

Possible Causes of Sex Differences in the Prevalence of Specific Disorders

Little difference occurs between men and women in the prevalence of mental disorder when all disorders are combined, but large differences are found for specific disorders (American Psychiatric Association, 2000). Women are diagnosed with *anxiety disorders* and *depression* at rates that are nearly twice as great as those for men. Men are diagnosed with *intermittent explosive disorder* (characterized by relatively unprovoked violent outbursts of anger) and with *antisocial personality disorder* (characterized by a history of antisocial acts with no sense of guilt) at rates that are three or four times those for women. Men are also diagnosed with *substance-use disorders* (including alcohol dependence and other drug dependence) at rates that are nearly twice as great as those for women. These differences may arise from a number of causes, including the following:

- **Differences in reporting or suppressing psychological distress** Diagnoses of anxiety disorders and of depression necessarily depend to a great extent on self-reporting. Men, who are supposed to be the “stronger” sex, may be less inclined than women to admit to anxiety and despondency in interviews or questionnaires. Supporting this view, experiments have shown that when men and women are subjected to the same stressful situation, such as a school examination, men report less anxiety than do women even though they show physiological signs of distress that are as great as, or greater than, those shown by women (Polefrone & Manuck, 1987). Anger is the one negative emotion that is generally more acceptable in men than in women, so a reporting bias might operate in the other direction for intermittent explosive disorder and antisocial personality disorder.

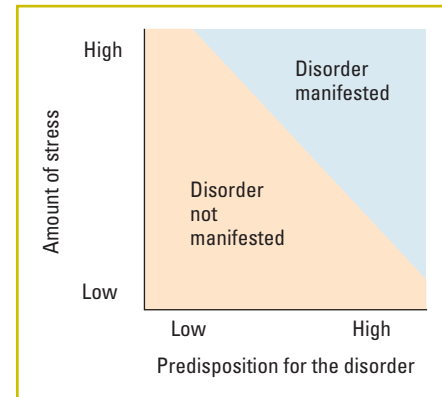


FIGURE 16.1 Relationship between predisposition and stress needed to trigger a mental disorder The amount of stress needed to bring on a mental disorder decreases as the predisposition for the disorder increases.

>>> **10** What are four possible ways of explaining sex differences in the prevalence of specific mental disorders?

Hysterical or angry? Even today the word “hysterical” may come to mind more quickly when we view an angry woman than when we view an angry man. That same bias may contribute to the more frequent diagnosis of histrionic personality in women than in men.



SuperStock



FPG

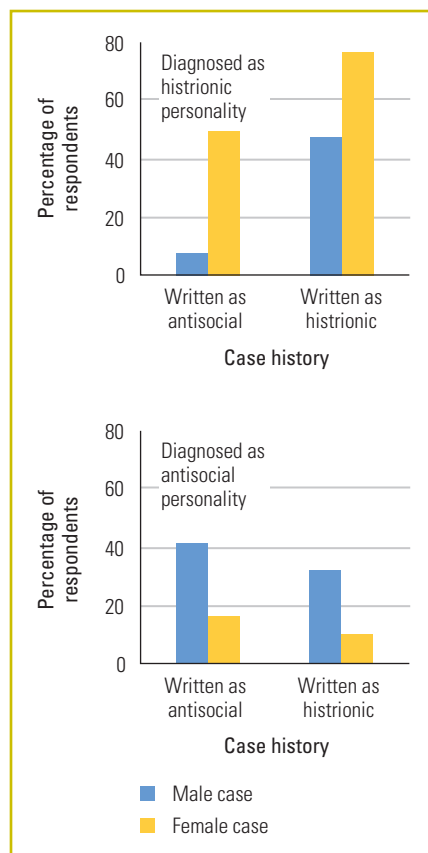


FIGURE 16.2 Evidence of a gender bias in diagnosis In this study, case histories were more likely to be diagnosed as antisocial personality if they described a fictitious male patient and as histrionic personality if they described a fictitious female patient, regardless of which disorder the case history was designed to portray. (Data from Ford & Widiger, 1989.)

- Clinicians' expectations** Diagnosticians may, to some degree, find a disorder more often in one sex than in the other because they *expect* to find it there. In an experiment demonstrating such an expectancy bias, several hundred clinical psychologists in the United States were asked to make diagnoses on the bases of written case histories that were mailed to them (Ford & Widiger, 1989). For some, the case history was constructed to resemble the *DSM-III* criteria for *antisocial personality disorder*, which is diagnosed more often in men than in women. For others it was constructed to resemble the *DSM-III* criteria for *histrionic personality disorder* (characterized by dramatic expressions of emotion to seek attention), which is diagnosed more often in women. Each of these case histories was written in duplicate forms, for different clinicians, differing only in the sex of the person being described. As you can see in Figure 16.2, the diagnoses were strongly affected by the patient's sex. Given the exact same case histories, the man was far more likely than the woman to receive a diagnosis of antisocial personality, and the woman was far more likely than the man to receive a diagnosis of histrionic personality.
- Differences in social roles and experiences** A number of well-controlled studies indicate that at least some sex differences in the prevalence of disorders are real—they cannot be explained by sex differences in reporting or by biased diagnoses (Klose & Jacobi, 2004). One way to explain such actual differences is to search for differences in the social experiences of men and women. Women, throughout the world, are more likely than men to live in poverty, to experience discrimination, to have been sexually or physically abused in childhood, and to be physically abused by their spouses—all of which can contribute to depression, anxiety, and various other disorders that occur more often in women than in men (Koss, 1990; Nolen-Hoeksema, 2001). There is also evidence that the typical roles that women assume in the family, such as caring for children, are more conducive to anxiety and depression than are the typical roles that men assume (Almeida & Kessler, 1998; Barnett & Baruch, 1987).
- Differences in ways of responding to stressful situations** The two sexes not only tend to experience different sorts of stressful situations, but also tend to respond differently to objectively similar situations (Hankin & Abramson, 2001; Nolen-Hoeksema, 2001). Women tend to “internalize” their discomfort; they dwell mentally (ruminate) on their distress and seek causes within themselves. This manner of responding, in either sex, tends to promote both anxiety and depression (Nolen-Hoeksema, 2001). Men, in contrast, more often “externalize” their discomfort; they tend to look for causes outside of themselves and to try to control those causes, sometimes in ways that involve aggression or violence. It is not clear what causes these differences in ways of responding, but it is reason-

able to suppose that they are, in part, biologically predisposed. Such differences are observed throughout the world, and there is evidence that male and female hormones promote the typically male and female ways of reacting to stressful situations (Taylor & others, 2000).

SECTION REVIEW

Though mental disorders have many possible causes, all exert their effects via the brain.

The Brain's Role in Chronic and Episodic Mental Disorders

- Chronic mental disorders such as Down syndrome, autism, and Alzheimer's disease arise from irreversible brain deficits.
- Down syndrome derives from an extra chromosome 21 that causes widespread brain damage and mental retardation.
- In autism, which involves a severe deficit in social instincts, the brain damage may stem from genes, prenatal toxins, or birth complications.
- Alzheimer's disease, which involves progressive decline in all cognitive functions, usually in old age, may result from disruptive effects of amyloid plaques in the brain.
- All the other disorders discussed in the chapter are episodic, meaning that their active manifestations come and go. Causes may include hereditary influences on the brain's biology, environmental assaults on the brain, and effects of learning.

A Framework for Thinking About Multiple Causes

- Predisposing causes, which make a person susceptible to a mental disorder, include genes, early environmental effects on the brain, and learned beliefs.
- Precipitating causes, which bring on episodes of a disorder, are often stressful life changes or losses.
- Perpetuating causes, which decrease the chance of recovery, include poor self-care, social withdrawal, and negative reactions from others.

Causes of Sex Differences in Prevalence of Specific Disorders

- Some diagnoses (e.g., anxiety disorder) are much more prevalent in women, and some (e.g., antisocial personality disorder) are much more prevalent in men.
- Such differences may derive from sex differences in (a) the tendency to report or suppress psychological distress; (b) clinicians' expectations of seeing certain disorders more often in one sex than in the other; (c) sex differences in stress associated with differing social roles; or (d) sex differences in ways of responding to stress.

Anxiety Disorders

The rabbit crouches, frozen in the grass, minimizing its chance of being detected. Its sensory systems are at maximum alertness, all attuned to the approaching dog. With muscles tense, the rabbit is poised to leap away at the first sign that it has been detected; its heart races, and its adrenal glands secrete hormones that help prepare the body for extended flight if necessary. Here we see fear operating adaptively, as it was designed to operate by natural selection.

We humans differ from the rabbit on two counts: Our biological evolution added a massive, thinking cerebral cortex atop the more primitive structures that organize fear, and our cultural evolution led us to develop a habitat far different from that of our ancestors. Our pattern of fear is not unlike the rabbit's, but it can be triggered by an enormously greater variety of both real and imagined stimuli, and in many cases the fear is not adaptive. Fear is not adaptive when it causes a job candidate to freeze during an interview, or overwhelms a student's mind during an examination, or constricts the life of a person who can vividly imagine the worst possible consequence of every action.

Anxiety disorders are those in which fear or anxiety is the most prominent symptom. Although *fear* and *anxiety* can be used as synonyms, *fear* is more often used when the feared

The adaptive value of fear Fear causes this rabbit to remain motionless but highly alert to the nearby predator. The rabbit is camouflaged but ready to flee at the first indication of being detected.



S. J. Krasemann / Peter Arnold, Inc.

stimulus or event is specific and present, and *anxiety* is more often used when the stimulus or event is vague, not identifiable, or in the future. Cognitively, anxious people worry excessively, especially about their own lack of ability to cope with unpredictable future dangers (Barlow, 2002).

Among the classes of anxiety disorders recognized by *DSM-IV* are generalized anxiety disorder, phobias, obsessive-compulsive disorder, panic disorder, and post-traumatic stress disorder. Genetic differences play a considerable role in the predisposition for all these disorders. Research with twins indicates that roughly 30 to 50 percent of the individual variability in risk to develop any given anxiety disorder derives from genetic variability (Gordon & Hen, 2004). In what follows, I will describe the main diagnostic characteristics and precipitating causes of each of these common types of anxiety disorders.

Generalized Anxiety Disorder

Generalized anxiety is called *generalized* because it is not focused on any one specific threat; instead, it attaches itself to various threats, real or imagined. It manifests itself primarily as *worry*. Sufferers of **generalized anxiety disorder** worry more or less continuously, about multiple issues, and they experience muscle tension, irritability, difficulty in sleeping, and sometimes gastrointestinal upset due to overactivity of the autonomic nervous system. They worry about the same kinds of issues that most of us worry about—family members, money, work, illness, daily hassles—but to a far greater extent and with much less provocation (Becker & others, 2003).

To receive a *DSM-IV* diagnosis of generalized anxiety disorder, such life-disrupting worry must occur on more days than not for at least 6 months and occur independently of other diagnosable mental disorders (American Psychiatric Association, 2000). By *DSM-IV* criteria, roughly 6 percent of people in North America could be diagnosed as having this disorder at some time in their lives (Kessler & others, 2005).

Unpredictable traumatic experiences in childhood, as well as genetic disposition, can apparently predispose a person to this disorder (Torgersen, 1986). In predisposed people, the disorder often first appears following a major life change in adulthood, such as getting a new job or having a baby, or after a disturbing event, such as an accident or illness (Blazer & others, 1987).

Laboratory research shows that people diagnosed with generalized anxiety disorder are particularly attuned to threatening stimuli. For example, they notice threatening words (such as “cancer” or “collapse”), but not unthreatening words, more quickly, reliably, and automatically than do other people (Mogg & Bradley, 2005). Such automatic attention to potential threat is referred to as *hypervigilance*, and there is some evidence that such vigilance characterizes people who are prone to generalized anxiety disorder even before they develop the disorder (Eysenck, 1992). Hypervigilance may provide a link between childhood trauma and generalized anxiety. Early trauma may teach the person to look constantly for cues of impending danger and to see such cues even where they don’t exist. Cognitive therapy (discussed in Chapter 17) for generalized anxiety is often aimed at helping the person overcome the automatic tendency to see signs of threat everywhere while failing to see signs of safety.

11

How can the onset of generalized anxiety disorder be explained in terms of predisposing and precipitating causes?



Peanuts © United Features Syndicate, Inc.

Surveys taken at various times, using comparable measures, indicate that the average level of generalized anxiety in Western cultures has increased sharply since the middle of the twentieth century (Twenge, 2000). From a cultural perspective, that increase may be attributable to the reduced stability of the typical person's life. In a world of rapid technological change, we can't be sure that the job skills needed today will still be useful tomorrow. In a world of frequent divorce and high mobility, we can't be sure that the people we are attached to now will be with us in the future. In a world of rapidly changing values and expectations, we have difficulty judging right from wrong or safe from unsafe. Such threats may be felt only dimly and lead not to conscious articulation of specific fears but to generalized anxiety. And, unlike the predator that scares the rabbit one minute and is gone the next, these threats are with us always.

How might the high rate of generalized anxiety in modern Western cultures be explained?

Phobias

In contrast to generalized anxiety, a **phobia** is an intense, irrational fear that is very clearly related to a particular category of object or event. In *DSM-IV*, phobias are divided into two main types: social and specific. In a **social phobia**, the fear is of being scrutinized or evaluated by other people; included here are fears of public speaking, of eating in public places, and of meeting new people. In a **specific phobia**, the fear is of some specific, nonsocial category of object or situation. It may be of a particular type of animal (such as snakes), substance (such as blood), or situation (such as heights or being closed in). For a diagnosis to be given, the fear must be long-standing and sufficiently strong to disrupt everyday life in some way—such as causing one to leave a job or to refrain from leaving home in order to avoid encountering the feared object. Both categories of phobias are quite prevalent; surveys suggest that they are diagnosable, at some time in life, in somewhere between 7 percent and 13 percent of people in Western societies (Furmark, 2002; Kessler & others, 2005).

Usually a phobia sufferer is aware that his or her fear is irrational but still cannot control it. The person knows well that the neighbor's kitten won't claw anyone to death, that common garter snakes aren't venomous, or that the crowd of 10-year-olds at the corner won't attack. People with phobias suffer doubly—from the fear itself and from knowing how irrational they are to have such a fear (Williams, 2005).

Probably everyone has some irrational fears, and, as in all other anxiety disorders, the difference between the normal condition and the disorder is one of degree, not kind. Specific phobias are usually of things that many people fear to some extent, such as snakes, spiders, blood, darkness, or heights (Marks, 1987); and social phobias

What evidence links phobias to normal fears?



© George D. Lepp / Corbis

A spider for the hypervigilant
People who suffer from a specific phobia tend to see the object of their fear even when it is so well camouflaged that most people would miss it. A sight such as this might cause someone with a spider phobia to avoid sunflowers forever after.

14

How might phobias be explained in terms of learning and evolution?

may simply be extreme forms of shyness (Rapee & Spence, 2004). Also consistent with the idea that phobias lie on a continuum with normal fears is the observation that specific phobias are much more often diagnosed in women than in men, whereas social phobias are diagnosed about equally often in the two sexes (American Psychiatric Association, 2000). In the population as a whole, men and boys are much less likely than women and girls to report fears of such things as spiders and darkness but are about equally likely to report shyness. The sex difference in specific phobias could stem from the fact that boys are more strongly encouraged than are girls to overcome or to hide their childhood fears (Fodor, 1982).

<<< Relatively little is known about how phobias usually arise. Behaviorists, from John B. Watson (1924) on, have argued that specific phobias are acquired by classical conditioning, through experiences in which the now-feared stimulus had been paired with some unconditionally fearful stimulus. A problem with this interpretation is that people with phobias are, on average, no more likely than people without phobias to recall having had any specific experiences with the feared object or situation that could have provided the basis for such learning (Craske, 1999; McNally & Steketee, 1985). It also does not explain such findings as this: In a survey in Burlington, Vermont, where there are no dangerous snakes, the single most common specific phobia was of snakes (Agras & others, 1969). If phobias are acquired by conditioning, why aren't phobias of such things as automobiles or (in Burlington) icy sidewalks more common?

Another theory, first proposed by Martin Seligman (1971), is that people are genetically prepared to be wary of—and to learn easily to fear—objects and situations that would have posed realistic dangers during most of our evolutionary history (discussed in Chapter 4). This idea is helpful in understanding why phobias of snakes, darkness, and heights are more common than those of automobiles and electric outlets. Research has shown that people can acquire strong fears of such evolutionarily significant objects and situations through means other than direct conditioning (Öhman & Mineka, 2001). Simply observing others respond fearfully to them, or reading or hearing fearful stories about them, can initiate or contribute to a phobia. The fact that some people acquire phobias and others don't in the face of similar experiences probably stems from a variety of predisposing factors, including genetic temperament and prior experiences (Craske, 1999).

Laboratory research shows that people with specific phobias are hypervigilant specifically for the category of object that they fear. For example, people with spider phobias can find spiders in photographs that contain many objects more quickly than can other people (Pflugshaupt & others, 2005). People with phobias also, however, turn their eyes away from the feared objects in such pictures more quickly than do other people (Pflugshaupt & others, 2005). The strong tendency

of phobia sufferers to avoid looking at or being anywhere near the objects they fear is believed to be the primary perpetuating cause of the disorder. As discussed in Chapter 17, the best way to overcome an irrational fear is to expose oneself to the feared object or situation for increasingly long periods. Maybe the high rate of snake phobia in Burlington, Vermont, occurs because there are so few snakes of any kind there. Without exposure to snakes, there is little opportunity to overcome a fear of them.

Obsessive-Compulsive Disorder

An *obsession* is a disturbing thought that intrudes repeatedly on a person's consciousness even though the person recognizes it as irrational. A *compulsion* is a repetitive action that is usually performed in response to an obsession. Most people experience moderate forms of these, especially in childhood (Steketee & Barlow, 2002). I remember a period in sixth grade when, while reading in school, the thought would repeatedly enter my mind that reading could make my eyes

Reprinted from Health Magazine



fall out. The only way I could banish this thought and go on reading was to close my eyelids down hard—a compulsive act that I fancied might push my eyes solidly back into their sockets. Of course, I knew that both the thought and the action were irrational, yet the thought kept intruding, and the only way I could abolish it for a while was to perform the action. Like most normal obsessions and compulsions, this one did not really disrupt my life, and it simply faded with time.

People who are diagnosed with **obsessive-compulsive disorder** are those for whom such thoughts and actions are severe, prolonged, and disruptive of normal life. To meet *DSM-IV* criteria for this disorder, the obsessions and compulsions must consume more than an hour per day of the person's time and seriously interfere with work or social relationships (American Psychiatric Association, 2000). By these criteria, the disorder occurs in roughly 1 to 2 percent of people at some time in their lives (Kessler & others, 2005). Obsessive-compulsive disorder is similar to a phobia in that it involves a specific irrational fear; it is different from a phobia primarily in that the fear is of something that exists only as a thought and can be reduced only by performing some ritual. People with obsessive-compulsive disorder, like those with phobias, suffer also from their knowledge of the irrationality of their actions and go to great lengths to hide them from other people.

The obsessions experienced by people diagnosed with this disorder are similar to, but stronger and more persistent than, the kinds of obsessions experienced by people in the general population (Rachman & DeSilva, 1978; Steketee & Barlow, 2002). The most common obsessions concern disease, disfigurement, or death, and the most common compulsions involve checking or cleaning. People with checking compulsions may spend hours each day repeatedly checking doors to be sure they are locked, the gas stove to be sure it is turned off, automobile wheels to be sure they are on tight, and so on. People with cleaning compulsions may wash their hands every few minutes, scrub everything they eat, and sterilize their dishes and clothes in response to their obsessions about disease-producing germs and dirt. Some compulsions, however, bear no apparent logical relationship to the obsession that triggers them. For example, a woman obsessed by the thought that her husband would die in an automobile accident could in fantasy protect him by dressing and undressing in a specific pattern 20 times every day (Marks, 1987).

More than is true for the other anxiety disorders, obsessive-compulsive disorder has been linked to observable abnormalities in the brain. In some cases, the disorder first appears after known brain damage, from such causes as a blow to the head, poisons, or diseases (Berthier & others, 2001; Steketee & Barlow, 2002). In many other cases, neural imaging reveals brain abnormalities stemming from unknown causes (Szeszko & others, 2004). The brain areas that seem to be particularly involved include portions of the frontal lobes of the cortex and parts of the underlying limbic system and basal ganglia, which (as discussed in Chapter 5) are believed to work together in a circuit to control voluntary actions—the kinds of actions that result from conscious thoughts. Damage in these areas may produce obsessive-compulsive behavior by interfering with the brain's ability to inhibit recurring, worrisome thoughts or to produce the psychological sense of closure or safety that normally occurs when a protective action is completed (Chamberlain & others, 2005; Szechtman & Woody, 2004). Consistent with this theory, people with obsessive-compulsive disorder often report that they do not experience the normal sense of task completion that should come after they have washed their hands or inspected the gas stove, so they feel an overwhelming need to perform the same action again (Szechtman & Woody, 2004).

>>> **15**
How is an obsessive-compulsive disorder similar to and different from a phobia?

>>> **16**
How might damage to certain areas of the brain result in obsessive-compulsive disorder?

Panic Disorder

Panic is a feeling of helpless terror, such as one might experience if cornered by a predator. In some people, this sense of terror comes at unpredictable times, unprovoked by any specific threat in the environment. Because the panic is unrelated to any specific situation or thought, the panic victim, unlike the victim of a phobia or

Panic This woman has a reason to feel panic. She has just been attacked by a thief. People with panic disorder feel panic such as this at unpredictable times, without provocation.

© Jutta Klee / Corbis



an obsessive-compulsion, cannot avoid it by avoiding certain situations or relieve it by engaging in certain rituals. *Panic attacks* usually last several minutes and are usually accompanied by high physiological arousal (including rapid heart rate and shortness of breath) and a fear of losing control and behaving in some frantic, desperate way (Barlow, 2002). People who have suffered such attacks often describe them as the worst experiences they have ever had—worse than the most severe physical pain they have felt.

To be diagnosed with **panic disorder**, by *DSM-IV* criteria, a person must have experienced at least two such attacks, each followed by at least 1 month of debilitating worry about having another attack or by life-constraining changes in behavior

(such as quitting a job or refusing to travel) motivated by fear of another attack (American Psychiatric Association, 2000). By these criteria, somewhere between 2 and 5 percent of North Americans suffer from panic disorder at some time in their lives (Kessler & others, 2005; White & Barlow, 2002). Most people diagnosed with panic disorder also suffer from a condition called *agoraphobia*, a fear of public places. The agoraphobia develops at least partly because panic victims fear the embarrassment and humiliation that might follow loss of control (panic) in front of others (Craske, 1999).

17

What learned pattern of thought might be a perpetuating cause of panic disorder?

<<< As is true with all other anxiety disorders, panic disorder often manifests itself shortly after some stressful event or life change (White & Barlow, 2002). Panic victims seem to be particularly attuned to, and afraid of, physiological changes that are similar to those involved in fearful arousal. In the laboratory or clinic, panic attacks can be brought on in people with the disorder by any of various procedures that increase heart rate and breathing rate—such as lactic acid injection, a high dose of caffeine, carbon dioxide inhalation, or intense physical exercise (Barlow, 2002). This has led to the view that a perpetuating cause, and possibly also predisposing cause, of the disorder is a learned tendency to interpret physiological arousal as catastrophic. One treatment, used by cognitive therapists, is to help the person learn to interpret each attack as a temporary physiological condition rather than as a sign of mental derangement or impending doom.

Posttraumatic Stress Disorder

18

How does posttraumatic stress disorder differ from other anxiety disorders?

<<< All anxiety disorders can be brought on or exacerbated, to some degree, by stressful experiences, but **posttraumatic stress disorder** is the only anxiety disorder that is *necessarily* brought on by such experiences. By definition, the symptoms of posttraumatic stress disorder must be linked to one or more emotionally traumatic incidents that the affected person has experienced. The disorder is most common in torture victims, concentration camp survivors, people who have been raped or in other ways violently assaulted, people who have survived dreadful accidents, and soldiers who have experienced the horrors of battle.

The most characteristic features of this disorder, according to *DSM-IV*, are frightening, uncontrollable re-experiences of the traumatic events, in nightmares and in daytime thoughts or “flashbacks.” Other common symptoms are sleeplessness, high arousal, irritability, guilt (perhaps for surviving when others didn’t), emotional numbing, and depression (American Psychiatric Association, 2000). In an effort to relieve such symptoms, posttraumatic stress victims may turn to alcohol or street drugs, which often compound the problem.

Not surprisingly, the prevalence of this disorder depends very much on the population sampled. Surveys in the United States suggest that up to 7 percent of people may develop the disorder, by *DSM-IV* criteria, at some point in their lives (Kessler & others, 2005). In contrast, in present-day Algeria, where nearly everyone has been a victim of terrorist attacks, 40 percent of a random sample of adults were found to suffer from this disorder (Khaled, 2004). Similarly, rates of posttraumatic distress disorder ranging from 15 to 30 percent have been found in war-torn regions of Cambodia, Ethiopia, and Palestine (de Jong & others, 2003).

People who are exposed repeatedly, or over long periods of time, to distressing conditions are apparently more likely to develop posttraumatic stress symptoms than are those exposed to a single, short-term, highly traumatic incident. One study revealed that the incidence of posttraumatic stress disorder among Vietnam War veterans correlated more strongly with long-term exposure to the daily stressors and dangers of the war—such as heat, insects, loss of sleep, sight of dead bodies, and risk of capture by the enemy—than with exposure to a single atrocity (King & others, 1995). Likewise, children exposed to repeated abuse in their homes are particularly prone to the disorder (Roesler & McKenzie, 1994).

Not everyone exposed to extraordinary stress develops posttraumatic stress symptoms. Social support, both before and after the stressful experiences, seems to play a role in reducing the likelihood of the disorder (Ozer & others, 2003). Genes also play a role (Paris, 2000). For instance, a study of twin pairs who fought in Vietnam revealed that identical twins were considerably more similar to each other in the incidence of the disorder and in the types of symptoms they developed than were fraternal twins (True & others, 1993).

>>> 19
What conditions are particularly conducive to development of posttraumatic stress disorder?

SECTION REVIEW

Anxiety disorders have fear or anxiety as their primary symptom.

Generalized Anxiety Disorder • This disorder, characterized by excessive worry about real or imagined threats, may be predisposed by genes or childhood trauma and brought on by disturbing events in adulthood. • Hypervigilance—automatic attention to possible threats—may stem from early trauma and may lead to generalized anxiety. • Levels of generalized anxiety have risen sharply in Western culture since the mid-twentieth century.	Phobias • Social phobias involve intense fear of evaluation by others. Specific phobias involve intense fear of specific nonsocial objects (e.g., spiders) or situations (e.g., heights). • Phobia sufferers usually know that the fear is irrational but cannot control it. • The difference between a normal fear and a phobia is one of degree. • Natural selection may have prepared us to fear some objects and situations more than others.	Obsessive-Compulsive Disorder • This disorder—involving repetitive, disturbing thoughts (obsessions) and repeated, ritualistic actions (compulsions)—is associated with abnormalities in an area of the brain that links conscious thought to action. • Obsessions and compulsions are often extreme versions of normal safety concerns and protective actions, which the sufferer cannot shut off despite being aware of their irrationality.	Panic Disorder • People with this disorder experience bouts of helpless terror (panic attacks) unrelated to specific events in their environment. • The disorder may be predisposed and perpetuated by a learned tendency to regard physiological arousal as catastrophic. • Caffeine, exercise, or other methods to induce high arousal can trigger panic attacks in susceptible people.	Posttraumatic Stress Disorder • This disorder is characterized by the re-experiencing—in nightmares, daytime thoughts, and flashbacks—of an emotionally traumatic event or set of events that occurred in the person's life. Other symptoms include sleeplessness, irritability, guilt, and depression. • Genetic predisposition, repeated exposures to traumatic events, and inadequate social support increase the risk for the disorder.
---	--	---	--	--

Edvard Munch (1863–1944), *Despair*, SuperStock

Depression This work is by Edvard Munch, a Norwegian artist who himself suffered from episodes of severe depression. Like many creative people who suffer from depression, Munch took much of his subject matter from his memories of his depressed mood, but created most of his works during periods of remission (Rothenberg, 2001).

20

What are some similarities and differences between depression and generalized anxiety?

<<< Much has been written about the similarities and differences between depression and generalized anxiety disorder. The two disorders seem to be predisposed by the same genes. The identical twins of people suffering from either generalized anxiety disorder or major depression show equally enhanced rates of either disorder (Kendler & others, 1992, 1995). Other studies reveal that the two disorders often occur in the same individuals. Approximately 60 percent of people diagnosed with generalized anxiety disorder also suffer from a depressive disorder at some point in their lives (Overbeek & Griez, 2005). Typically, generalized anxiety occurs before, rather than after, the onset of major depression (Hunt & others, 2004; Wittchen & others, 2000).

One way to conceive of the difference between generalized anxiety and depression is to think of the former as a frantic, relatively ineffective attempt to cope with life's real and imagined threats, through worry and hypervigilance, and to think of the latter as a kind of giving up, which occurs when one has despaired of coping and concludes that life is not worth living. Cognitively, anxious individuals worry about what might happen in the future, while depressed individuals feel that all is already lost (see Table 16.2). Depressed individuals are much more likely than anxious individuals to stop caring for themselves and to stay in bed all day. The sense

Mood Disorders

Mood refers to a prolonged emotional state that colors many, if not all, aspects of a person's thought and behavior. It is useful (though oversimplified) to think of a single dimension of mood, running from depression at one end to elation at the other. Because we all have tasted both, we have an idea of what they are like. Both are normal experiences, but at times either of them can become so intense or prolonged as to promote harmful, even life-threatening, actions. Extreme depression can keep a person from working, lead to withdrawal from friends, or even provoke suicide. Extreme elation, called *mania*, can lead to outrageous behaviors that turn other people away or to dangerous acts that stem from a false sense of security and bravado. *DSM-IV* identifies two main categories of **mood disorders: depressive disorders**, characterized by prolonged or extreme depression, and **bipolar disorders**, characterized usually by alternating episodes of mania and depression.

Depression

Depression is characterized primarily by prolonged sadness, self-blame, a sense of worthlessness, and absence of pleasure. Other common symptoms include decreased or increased sleep, decreased or increased appetite, and either retarded or agitated motor symptoms. Retarded motor symptoms include slowed speech and slowed body movements. Agitated symptoms, which are less common than retarded symptoms, include repetitive, aimless movements such as hand wringing and pacing. To warrant a diagnosis of a depressive disorder, the symptoms must be either very severe or very prolonged and not attributable just to a specific life experience, though they may be triggered or exacerbated by such an experience.

Two main classes of depressive disorders are distinguished in *DSM-IV*. **Major depression** is characterized by very severe symptoms that last essentially without remission for at least 2 weeks. **Dysthymia** [dis-thigh-me-uh] is characterized by less severe symptoms that last for at least 2 years. Quite often, bouts of major depression are superimposed over a more chronic state of dysthymia, in which case the person is said to have *double depression*. Surveys in North America suggest that as many as 15 percent of people suffer from major depression, and 2 to 3 percent suffer from dysthymia, at some time in their lives (Kessler & others, 2005).

TABLE 16.2 |
Comparison of anxious thoughts with depressive thoughts

Anxious thoughts	Depressed thoughts
What if I get sick and become an invalid?	I'm worthless.
Something will happen to someone I care about.	I'm a social failure.
Something might happen that will ruin my appearance.	I have become physically unattractive.
I am going to have a heart attack.	Life isn't worth living.

Source: Adapted from Beck & others, 1987.

of giving up is commonly accompanied by strong feelings of self-blame, of not deserving to live. The feelings of worthlessness are captured in the following quotation from Norman Endler (1982), a highly respected psychologist describing his own bout with depression:

I honestly felt subhuman, lower than the lowest vermin. . . . I could not understand why anyone would want to associate with me, let alone love me. . . . I was positive that I was a fraud and a phony. . . . I couldn't understand how I had written the books and journal articles that I had and how they had been accepted for publication. I must have conned a lot of people.

Heredity and Stressful Experiences as Causes of Depression

As is the case for nearly every other mental disorder, there is good evidence that heredity plays a role in predisposition for depression. The more closely related two people are genetically, the more similar they are, on average, in their history of depression or lack of it (Kendler & others, 1996; Serretti, 2005). Some research suggests that genetic predisposition influences depression largely by influencing the way a person responds to stressful life events.

In one study supporting such a conclusion, a team of researchers identified over 1000 women who had twin sisters (Kendler, 1998; Kendler & others, 1995). They studied each woman to identify (a) whether or not she had recently experienced a highly stressful life event (defined as assault, serious marital problems, divorce or breakup of a marriage or other romantic relationship, or death of a close relative), (b) whether or not a period of major depression began within a month after that stressful event (or within a comparable time period for those who had not experienced a serious stressor), and (c) her level of genetic predisposition for depression. The level of genetic predisposition was judged on the basis of whether or not the woman's twin sister (co-twin) did or did not have a history of major depression and whether she was an identical twin or a nonidentical twin. The results of the study are depicted in Figure 16.3 on page 598. Among women who had not recently experienced a highly stressful life event, the incidence of depression was very low regardless of level of genetic predisposition. However, among those who had recently experienced such an event, the incidence of depression was strongly related to the level of genetic predisposition.

21
How did a study of twins demonstrate that genes may influence depression primarily by influencing the way a person responds to stressful experiences?

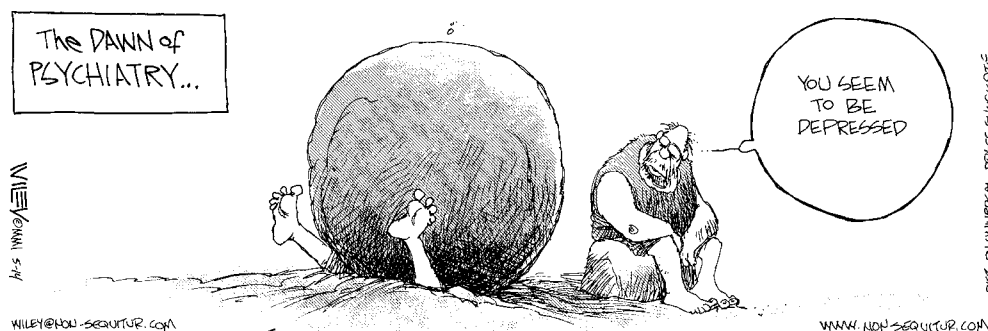
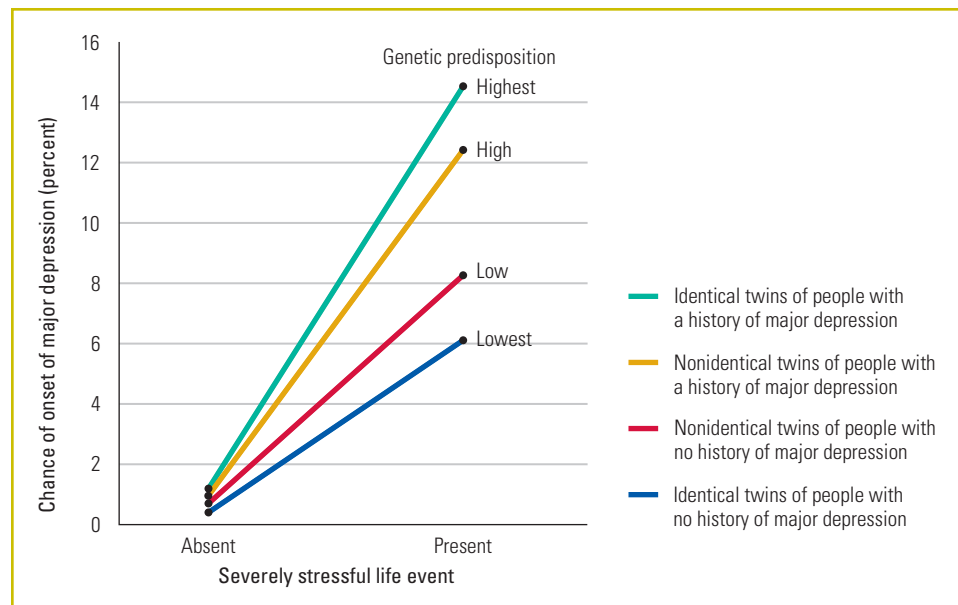


FIGURE 16.3 Roles of genetic predisposition and stressful experiences in the incidence of depression Women who had recently experienced a severely stressful event were far more likely to experience the onset of major depression than were those who had not. This effect of stress was greater for those who were more genetically predisposed for depression (based on the status of their twin sisters, as shown in the key) than for those who were less genetically predisposed. (Adapted from Kendler & others, 1995, p. 837.)
Reproduced by permission of *American Journal of Psychiatry*.



22

Why should we be somewhat skeptical about interpreting correlations between stressful life events and depression in cause-effect terms? What sorts of observations seem to provide the best evidence that stressful events can play a causal role in depression?

<<< Many other studies, too, have shown strong correlations between the onset of depression and stressful life events reported as occurring within weeks or a few months prior to the depression. Such studies are usually taken as evidence that stressful events help to bring on depression, but we should not automatically accept that interpretation. In most cases the measure of stressful events comes from reports made by depressed and non-depressed people who are asked to recall or check off events that happened to them over a specified period of time. Depression itself can lead people to recall more negative events or to interpret events more negatively than they otherwise would (Cohen & others, 1988). For example, depression might lead a person to remember a cold as a “serious illness” and to place a check next to that item on an inventory of stressful experiences. Moreover, people who are predisposed for depression may also be predisposed to behave in ways that bring on stressful life events (Hammen, 1991; Kendler, 1998). A stressful event such as being in a traffic accident, or being fired from a job, or being abandoned by a lover or spouse may not just “happen” but may be partly caused by one’s own actions. Thus, at least in theory, an observed relationship between reported stressful events and depression could be due to biased reporting or to some general behavioral difference between those who are prone to depression and those who are not.

Perhaps the best evidence that stressful events can bring on depression comes from studies that follow people, over time, after they have experienced some stressful event that was presumably out of their control, such as the death of a spouse (Umberson & others, 1992) or job loss due to general layoff rather than lack of ability (Dew & others, 1987). Such studies provide rather compelling evidence that stressful events can indeed bring on depression (Kessler, 1997). The kinds of experiences that appear to be most often associated with depression are *losses* that permanently change the nature of one’s life—such as loss of a spouse or other close daily companion, loss of a job that one has held for a long time, loss in social status, loss in income, or permanent loss in health. Such events interrupt one’s life routines and render ineffective well-established ways of satisfying one’s needs and desires.

Cognitive Style as a Cause of Depression

From a cognitive perspective, depression is characterized by negative patterns of thought and may be brought on, in part, by negative ways of interpreting life events. One of the first to emphasize the role of cognition in depression was Aaron Beck (1967), a psychiatrist who observed that his depressed clients distorted their perceptions of their experiences in negative ways, contributing to consistently pes-

simistic views of themselves, their world, and the future. They would mentally exaggerate bad experiences and minimize or overlook good ones. Beck developed a mode of therapy (discussed in Chapter 17) that centers on training depressed people to assess their experiences more optimistically.

Following Beck, Martin Seligman and Lyn Abramson developed a more specific cognitive theory of depression, referred to as the *hopelessness theory* (Abramson & others, 1989). According to this theory, depression results from a pattern of thinking about negative experiences that reduces or eliminates hope that life will get better. Part of the theory holds that people who are most prone to depression consistently attribute negative experiences in their lives to causes that are stable and global. A cause is *stable* to the degree that it is unlikely to change, and it is *global* to the degree that it applies to a wide sphere of endeavors.

As an illustration, consider the different attributions that two students might make concerning a failing grade on a test. One student might say, “I failed because I lack intelligence.” That attribution is stable and global; lack of intelligence is the kind of trait that stays with a person and affects all realms of life. Another student might say, “I failed that test because I didn’t study hard enough for it.” That attribution is unstable and not global; study habits can change and the lack of study may apply only to that one test. According to the hopelessness theory, the first student in this example is much more likely to become depressed than the second. The first student sees no hope, no route for passing the next test. The second student, in contrast, finds hope in the idea that he or she can pass the next test by studying harder.

Seligman and his colleagues developed an *attributional-style questionnaire* to assess people’s routine ways of explaining their negative and positive experiences. Research with this questionnaire shows that depressed people do clearly make attributions in a manner consistent with the theory (Riso & others, 2003; Seligman & others, 1988), but that observation by itself does not show that the negative attributional style is a cause of depression; it could simply be a symptom of depression. Evidence that the negative style of thinking does serve as a cause of depression is found in studies that have followed individuals over time (Abela & others, 2004).

In one such study, first-year, non-depressed college students with no prior history of clinical depression filled out the attributional-style questionnaire and then were assessed for depression at various times over a 2-year follow-up period (Alloy & others, 1999). Of those whose attributional style placed them in the “high risk for depression” category, 17 percent developed a major depressive disorder during the follow-up period. In contrast, only 1 percent of those in the low-risk category developed such a disorder. In other studies, people who learned, through cognitive therapy, to change their way of thinking to the more positive style were significantly less likely than otherwise comparable control subjects to become depressed during a follow-up period of many months (DeRubeis & Hollon, 1995; Gillham & others, 1995).

Possible Brain Mechanisms of Depression

Depression, like every other psychological state, is a product of the brain. Thus, one route to understanding depression is to try to identify changes in the brain that might cause bouts of depression to appear and disappear.

All the drugs that are used regularly to treat depression (discussed in Chapter 17) have the effect of increasing the amount or activity of one or both of two neurotransmitters in the brain: norepinephrine and serotonin. For that reason, an early theory of depression posited that the disorder results from a brain deficiency in either or both of these transmitters (Cowen, 2005; Schildkraut, 1965). Today, however, that theory is much doubted (Hirschfeld, 2000; Cowen, 2005). One problem is that it does not explain the delayed effectiveness of drug treatments. Antidepressant drugs begin to enhance the activity of norepinephrine or serotonin (or both) in the brain immediately, yet they do not begin to relieve depression until at least 2 weeks of continuous treatment have elapsed. Moreover, by various

23

According to the hopelessness theory, what patterns of thinking predispose a person for depression? What is some evidence for the theory?

24

What early evidence supported the theory that depression results from a deficit in the neurotransmitters norepinephrine and serotonin? Why is that theory now doubted?

25

According to a new theory, how might stressful experiences alter the brain in a way that brings on depression?



© David Sipress

Thank God they have pills for that now.

measures, most depressed people do not appear to have unusually low levels of these neurotransmitters (Cowen, 2005).

<<< At present, neuroscientists interested in depression are focusing on the ways that the brain changes during periods of psychological distress. Stress and worry are often associated with an increased release of certain hormones from the pituitary and adrenal glands (discussed in Chapter 5). Research with animals shows that these hormones can act on the brain to shut off certain growth-promoting processes there. Over periods of weeks or months, such effects can result in a small but measurable shrinkage in some portions of the brain, including portions of the prefrontal cortex and the hippocampus (Jacobs, 2004). These brain changes are reversible. During periods of reduced stress, the shrunken brain areas may regain their former size. Moreover, increases in norepinephrine and serotonin over periods of weeks can stimulate growth in these brain areas, and this observation may explain the delayed effects of drug treatments in relieving depression.

Thus, one current theory—which is far from proven—is that depression in humans results from a stress-induced loss of neurons or neural connections in certain parts of the brain and that recovery from depression results from regrowth in those brain areas (Cowen, 2005; Jacobs, 2004). This theory is quite consistent with the evidence that depression often follows a rather prolonged period of anxiety. Anxiety stimulates production of the hormones that interfere with brain growth. The theory is also consistent with the idea that altered ways of thinking can alter one's predisposition for depression. Hopefulness reduces psychological distress, which reduces the production of the growth-inhibiting hormones and thereby protects the brain from the changes that lead to depression.

Possible Evolutionary Bases for Depression

26

How might moderate depression, following a loss, be adaptive?

<<< Psychologists viewing depression from an evolutionary perspective have suggested that it may be an exaggerated form of a response to loss that in less extreme form is adaptive (Allen & Badcock, 2003; Nesse, 2000; Nesse & Williams, 1994). Only a minority of people develop clinically severe depression, but most of us, at some periods in our lives, experience a low mood or moderate depression. A depressed mood slows us down, makes us think realistically rather than optimistically, leads us to turn away from goals that we can no longer hope to achieve, and signals to others that we are no threat to them and need their help. The signals of helplessness by depressed persons resemble the appeasement displays used by other animals to signal submissiveness and need for care (Price & others, 2004). A depressed mood can also lead to a kind of soul-searching, the end result of which may be the establishment of new, more realistic goals and a new approach to life (Welling, 2003).

27

According to Keller and Nesse, how might depressed moods vary, adaptively, depending on the situation?

<<< Matthew Keller and Randolph Nesse (2005) have recently suggested that depressed moods may come in a variety of different forms, each adapted for different survival purposes. Most people who live in northern latitudes experience some degree of depressed mood during the winter (Dam & others, 1998). When it occurs in extreme form, such winter-limited depression is diagnosed, according to *DSM-IV*, as *seasonal affective disorder*, or *SAD*. This form of depression is accompanied by increased appetite, increased sleepiness, and lethargy—all responses that, in less extreme form, may have been useful to our evolutionary ancestors for building layers of fat and conserving energy to survive the harsh winter. That form of depression is generally not accompanied by the degrees of sadness, crying, and self-reproach that occur in other forms of depression (Keller & Nesse, 2005). In contrast, Keller and Nesse found evidence that depressed mood following death of a loved one or loss of a romantic partner is especially characterized by crying and other expressions of sadness, which may signal the need for help from others; and depressed



© Keren Su / Corbis

A possible precursor of depression
The appeasement display, with bowed head and turned-away eyes, shown by the langur monkey on the left, resembles the expression of a depressed person. One function of depression in humans may be to signal helplessness and lack of threat to others.

mood following repeated failure is especially characterized by self-blame and pessimism, which may motivate the person to withdraw from futile activities and begin a period of realistic reappraisal of life goals. It remains for future research to test further the idea that depressed moods come in different forms that serve different adaptive purposes.

Bipolar Disorders

Major depression and dysthymia are sometimes called *unipolar disorders*, because they are characterized by mood changes in only one direction—downward from normal. Bipolar disorders (formerly called *manic-depression*) are characterized by mood swings in both directions: downward in *depressive episodes* and upward in *manic episodes*. Such episodes may last anywhere from a few days to several months, often separated by years or months of relatively normal mood.

DSM-IV identifies two main varieties of bipolar disorders. **Bipolar I disorder** is the classic type, characterized by at least one manic episode and at least one depressive episode. **Bipolar II disorder** is similar to bipolar I disorder except that its high phase is less extreme and is referred to as *hypomania* rather than mania. Bipolar I disorder occurs, at some time in life, in about 1 percent of the population, and Bipolar II disorder occurs in about 2 to 3 percent of the population (Belmaker, 2004; Kessler & others, 2005). In some, relatively rare, cases, episodes of mania occur without intervening episodes of depression (Belmaker, 2004).

Research with twins and adoptees has shown that the predisposition for bipolar disorder is strongly influenced by genes (Kelsoe, 1997). Stressful life events may help bring on manic and depressive episodes in people who are predisposed (Ambelas, 1987; Hlastala & others, 2000), but the evidence for such effects is not as strong as it is for unipolar depression. Bipolar disorder, unlike unipolar depression, can usually be controlled with regular doses of the element *lithium*, used as a drug, but how lithium works is as yet unknown (Pilcher, 2003).

The Manic Condition

Manic episodes are typically characterized by expansive, euphoric feelings; elevated self-esteem; increased talkativeness; decreased need for sleep; and enhanced energy and enthusiasm, which may be focused on one or more grandiose projects



© Bizarro—Dan Piraro. King Features Syndicate

or schemes (American Psychiatric Association, 2000). The inordinate feelings of power, confidence, and energy are illustrated by the following quotation from a woman describing her own disorder (Fieve, 1975):

When I start going into a high, I no longer feel like an ordinary housewife. Instead, I feel organized and accomplished, and I begin to feel I am my most creative self. I can write poetry easily. I can compose melodies without effort. I can paint. . . . I have countless ideas about how the environmental problem could inspire a crusade for the health and betterment of everyone. . . . I don't seem to need much sleep. . . . I feel sexy and men stare at me. Maybe I'll have an affair, or perhaps several. I feel capable of speaking and doing good in politics.

During hypomania and the early stages of a manic episode, the high energy and confidence may lead to an increase in productive work, but, as a manic episode progresses, judgment becomes increasingly poor and behavior increasingly maladaptive. Full-blown mania may be accompanied by bizarre thoughts and dangerous behaviors, such as jumping off a building in the false belief that one can fly; and even hypomania may be accompanied by spending sprees, absence from work, or sexual escapades that the affected person later regrets (Akiskal, 2002). Moreover, not all people with bipolar disorders experience the manic state as euphoric. Some experience it as a time of extraordinary irritability, suspiciousness, or destructive rage (Carroll, 1991).

Possible Relation of Hypomania to Enhanced Creativity

28

How are manic states experienced?
What is some evidence linking mild manic (hypomanic) episodes to heightened creativity?

<<< Analyses of biographies and historical documents concerning eminently creative writers, artists, and musicians suggest that a disproportionate percentage of them suffered from a bipolar disorder (Andreasen, 1987; Jamison, 1995). The same studies suggest that those people were most productive during hypomanic phases of their illness—when their mood was elevated, but not to such an extreme as to prevent coherent thought and action. As one example, the composer Robert Schumann—who often suffered from severe depression—produced an extraordinary number of his valued musical works during two episodes of apparent hypomania, one in 1840 and the other in 1849 (Jamison, 1995). As another example, the poet Emily Dickinson wrote much of her best poetry during episodes of apparent hypomania that followed bouts of winter depression (Ramey & Weisberg, 2004).

Such analyses have been criticized on the grounds that the diagnoses of mood disorders and the judgments concerning hypomania were generally made after the fact, on the basis of written material that may not have been entirely accurate (Rothenberg, 2001). It is also possible that hypomania in highly creative people is more a *result* of high creativity than a cause: People who become strongly absorbed in and excited by their work may, as a result, exhibit manic-like behaviors (Ramey & Weisberg, 2004). Still another possibility is that people who suffer from extremes of mood may not, on average, be naturally more creative than others, but may be more drawn to creative activities. Such activities may provide a means for them to deal with or express the unusual feelings and thoughts that accompany their low and high moods.

Further evidence for a correlation between creativity and hypomania—regardless of causal direction—comes from studies of people who were selected on the basis of their clinical diagnoses, rather than because of their creative accomplishments, and then were assessed for creativity. In one study, people who manifested moderate mood swings—states of hypomania alternating with moderate levels of depression—were judged to be more creative in their regular work and home life than were people who exhibited greater stability in their moods (Richards & others, 1988). In another study, people who frequently experienced bouts of hypomania were found, on personality tests, to score higher than others on the trait of openness to experience (discussed in Chapter 15), which, itself, correlates rather strongly with measures of creativity (Nowakowska & others, 2005). All in all, the weight of the evidence seems to favor the idea of a link between hypomanic mood states and heightened creativity, but the nature of that link has yet to be determined.



Corbis-Bettman

Ups and downs of a composer Like many highly creative people, composer Robert Schumann suffered from a mood disorder. He attempted suicide twice during bouts of severe depression and eventually died, in 1856, from self-starvation in an asylum. His depression waxed and waned over long cycles, and during two prolonged periods, in 1840 and 1849, he was hypomanic (exhibited moderate mania). Those were his most productive years by far. He composed 24 musical works in 1840 and 27 in 1849. (Jamison, 1995.)

SECTION REVIEW

Mood disorders involve intense, prolonged moods that can disrupt or threaten life.**Depression**

- Prolonged sadness, self-blame, a sense of worthlessness, and an absence of pleasure are the hallmarks of depression; sleep, appetite, and speed of movement may be affected as well.
- Depression and generalized anxiety are predisposed by the same genes and often occur in the same individuals.
- Genes may predispose individuals to depression by affecting the way they respond to stressful experiences. There is strong evidence that stressful life events, especially those involving losses, can bring on depression.
- A cognitive style involving negative ways of interpreting life events may predispose people to depression. According to the hopelessness theory, people prone to depression attribute negative experiences to stable, global causes.
- Depression may arise from reversible brain changes that occur during periods of psychological distress, possibly induced by pituitary and adrenal hormones.
- Moderate depression may be adaptive. Different forms of moderate, normal depression may be adaptive reactions to different situations.

Bipolar Disorders

- Bipolar disorders—marked by mood changes in an upward (manic) as well as downward (depressive) direction—are highly heritable.
- Episodes of bipolar disorder may be brought on by stressful life events, but the evidence for this is not as clear as it is for unipolar depression.
- Manic episodes generally involve expansive, euphoric feelings and elevated self-esteem, talkativeness, and energy. Some may experience extreme irritability, suspicion, or destructive rage instead of euphoria.
- Full-blown mania involves poor judgment, bizarre thoughts, and self-harmful or dangerous actions. Milder mania (hypomania), however, may be a time of heightened creativity.

Psychological Influences on Physical Symptoms and Diseases

The brain is responsible not just for thoughts and emotion but also for sensation, movement, pain, and the regulation of many biological systems within the body. It is not surprising, therefore, that experiences that affect our thoughts and emotions can also affect our bodily functions and our manner of experiencing those functions.

Somatoform Disorders

Somatoform literally means “bodily form,” and **somatoform disorders** are those in which the person experiences bodily ailments in the absence of any physical disease that could cause them. Such disorders are believed to result from psychological distress that manifests itself in the form of physical pain, discomfort, or interferences with normal bodily processes. *DSM-IV* distinguishes among a number of different categories of somatoform disorders. The two categories I will discuss here are somatization disorder, which is one of the most prevalent types, and conversion disorder, which is the most dramatic type.

Somatization Disorder, and the Role of Culture in Somatoform Disorders

Somatization disorder is characterized by a long history of dramatic complaints about many different medical conditions, most of which are vague and unverifiable, such as dizziness, general aching of the body, heart palpitations, and nausea. For a firm diagnosis of somatization disorder, there must be strong reason to believe that the symptoms stem from psychological distress and not from any physical disease or injury. The true prevalence of this disorder is unknown, as it is often difficult to know the true source of a patient's symptoms. From what data are available, the American Psychiatric Association (2000) estimates that the prevalence in the United States is

29

Why are somatoform disorders apparently more common in non-Western cultures than in Western cultures?

somewhere between 0.2 and 1 percent of the population. The prevalence is reported to be much higher in some other cultures, particularly in China, other parts of Asia, and Latin America (Gureje, 2004; López & Guarnaccia, 2005).

<<< Somatization appears, to many clinical researchers, to be closely related to depression. Indeed, somatization and depression may be alternative ways of feeling and expressing the same underlying problems. A sense of hopelessness is usually experienced as psychological depression in North America and Western Europe today but in the past was more often experienced as physical aches and pains, as it still is today in China and many other parts of the world (Kleinman, 1988, 2004; López & Guarnaccia, 2005). As disorders purely of mood or emotion become more legitimized by a culture, the incidence of somatoform disorders of all types tends to go down and that of anxiety and mood disorders tends to go up.

Regardless of culture, it appears that a high percentage of people diagnosable with somatoform disorders of any type are people who find a psychological diagnosis unacceptable. Most people with these disorders seek help from general practice doctors, not from psychiatrists or psychologists, and a diagnosis implying that the discomfort is “all in their head” is often unacceptable or even insulting to them (Sharpe & Mayou, 2004). For this reason, great tact is needed in helping people with somatoform disorders. The helpful clinician must keep in mind that, regardless of the source of suffering, the person really is suffering. Often such patients find explanations in terms of presumed brain pathways, which connect parts of the brain involved with emotion to other parts of the brain that produce their symptoms, to be more satisfying than purely psychological explanations (Sharpe & Mayou, 2004). The ability to see that ultimately there is no real distinction between “physical” and “psychological”—all suffering involves the physical brain—helps patients accept the idea that psychological treatments might help.

Conversion Disorder

The most extreme and disabling category of somatoform disorder is **conversion disorder**, in which the person temporarily loses some bodily function, perhaps (in the most dramatic cases) becoming blind, deaf, or partially paralyzed, in a manner that cannot be explained in terms of physical damage to the affected organs or their neural connections (for example, see Figure 16.4). The term *conversion*, in this context, refers to the conversion of psychological distress into a physical disability.

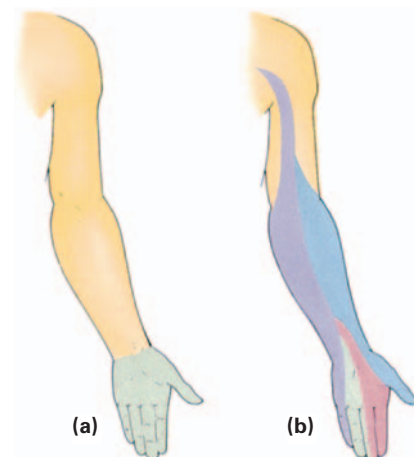
30

How did Freud interpret conversion disorders? What conditions appeared to produce conversion reactions in Cambodian survivors of the Khmer Rouge and in certain children raised in dysfunctional families?

<<< Conversion disorders are rare in North America and Western Europe today—with prevalence estimated at roughly 4 cases per 100,000 people (Krem, 2004)—but they were relatively common in those places 100 years ago and earlier. Such disorders were prominent in Sigmund Freud’s practice in the late nineteenth and early twentieth centuries, and they played a major role in his development of psychoanalytic theory. As a neurologist, Freud saw many patients with physical complaints and disabilities that appeared to have no medical explanation. He concluded that such symptoms were products of the patients’ unconscious minds that served to

FIGURE 16.4 A conversion disorder

Glove anesthesia—the experience of no sensation in the hand but continued sensation in all other parts of the arm, as shown in (a)—cannot result from nerve damage, because no nerves innervate the hand without innervating part of the arm. The actual areas of sensory loss that would occur if specific nerves were damaged are shown in (b). Thus, whenever glove anesthesia occurs, it is most likely a conversion disorder.



protect them from anxiety-producing activities and experiences (discussed further in Chapter 17).

Today, the most dramatic examples of conversion disorder are generally found in people raised in non-Western cultures who have been exposed to highly traumatic events. An example is the relatively high rate of psychologically based blindness discovered among Cambodian women who managed to emigrate to the United States after the Khmer Rouge reign of terror in the 1970s (Rozée & Van Boemel, 1989). All these psychologically blind women manifested posttraumatic stress syndrome brought on by the tortures they had suffered and witnessed. Many of them had seen their own children being slowly tortured and murdered. One woman described the onset of her blindness in these words: “My family was killed in 1975 and I cried for 4 years. When I stopped crying I was blind.” Another described becoming blind “from the smoke” as she was looking at the cooking pot in a forced-labor camp where she was held—a pot that, she knew, often contained human flesh, rats, and worms, which the women were forced to cook for themselves and other inmates to eat.

In Western countries, conversion disorder may be more common in children than is generally recognized. In Sydney, Australia, Kasia Kozłowska (2001, 2003) identified a number of children who suffered from partial paralysis, brought on, apparently unconsciously, as a way of protecting themselves from their overly demanding, abusive parents. Kozłowska describes these patients as “good children,” who tried desperately to please their parents, but who could not meet their parents’ impossible demands and who, for relief, became partially paralyzed. The paralysis brought them to the hospital, where they felt safe. Treatment in these cases involved a combination of family counseling, aimed at improving the family environments, and psychotherapy for the children, aimed at helping them understand that their symptoms resulted from their stressful experiences at home and would subside as their homes became safer.



Bart Bartholemew / NYT Pictures

Traumatic blindness This woman is one of many Cambodian women whose acute vision problems have been tied to seeing atrocities committed by, and being tortured by, the Khmer Rouge.

Brain Responses in Conversion Disorder

To learn more about the neural pathways that may “convert” psychological distress into sensory or motor deficits, several neuroimaging studies have been conducted in recent years, using the fMRI and PET methods described in Chapter 5 (Ron, 2001; Vuilleumier & others, 2001; Werring & others, 2004). One clear finding is that the patterns of brain activity underlying conversion disorder are not mimicked in control subjects who are merely instructed to “pretend” that they are paralyzed, blind, or in some other way incapacitated. The brain activity patterns may, however, be at least partly mimicked in hypnotized subjects who are given the temporary suggestion that they have that disorder (Halligan & others, 2000; Ward & others, 2003).

Typically, both in hypnosis and in actual conversion disorder, relevant motor or sensory areas of the brain become less active than normal, and areas of the limbic system and prefrontal cortex that are known to inhibit those motor or sensory areas become unusually active. For example, one study of patients with medically unexplained partial blindness showed reduced neural activity in the visual cortex in response to light, but increased neural activity in portions of the limbic system, prefrontal cortex, and thalamus, all of which have inhibitory connections to the visual cortex (Werring & others, 2004). Such work promises to help demystify conversion disorder and may someday help clinicians distinguish true cases of conversion from other sources of motor and sensory impairment and from cases of malingering, in which the person pretends to be incapacitated in some way.

>>> 31

How might neuroimaging methods be used, in the future, to help identify true cases of conversion disorder?

Psychological Factors Affecting Medical Condition

You have just read of disorders in which bodily pain or disability is experienced in the absence of any medical explanation. We turn now to a different category of psychological effects on the body—cases in which one's behavior or emotions precipitate or influence the course of a disease that clearly does have a physical, medical basis. The harm-producing behaviors or emotions do not technically constitute mental disorders, but a category—called *psychological factors affecting medical condition*—is reserved for them in a section of *DSM-IV* devoted to clinically significant conditions other than mental disorders.

32

What evidence did a study of widows and widowers provide that psychological state affects medical condition? What are possible mechanisms for such effects?

<<< Typical of the evidence that psychological state can affect physical health are studies showing that recent widowers and widows are more vulnerable than others to essentially all types of physical diseases (Kaprio & others, 1987; Ray, 2004). In one study (Prigerson & others, 1997), 150 men and women whose spouses were terminally ill were assessed for psychological state and physical health from the time just before they were widowed until 25 months afterward. On the basis of interviews conducted 6 months after their spouses died, a subgroup was identified whose members manifested what the researchers called *traumatic grief*—a syndrome similar to posttraumatic stress syndrome combined with extraordinary yearning for the deceased spouse. Although the members of this subgroup were not physically less healthy than the other subjects before the spouse's death or at the 6-month interview, by 25 months of widowhood they had developed significantly more incidences of flu, heart disease, and cancer than had the less-traumatized group.

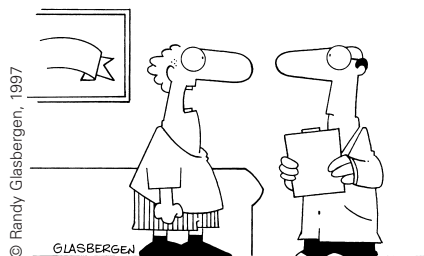
It is not hard to imagine possible mechanisms through which traumatic grief or other disruptive psychological states can affect a person's health. The most obvious possibility is that the affected person may fail to take good physical care of himself or herself. Bereavement is often associated with poor diet, irregular sleep patterns, increased smoking, increased alcohol consumption, and failure to take prescribed medicines (Cohen & Herbert, 1996). Such decline in self-care typically occurs more strongly in men than in women, which may help explain why men, on average, exhibit more physical decline and higher rates of mortality after the death of their spouse than do women (Ray, 2004). There is also evidence, however, that psychological states can affect health directly, through the brain's effects on internal organs and the immune system.

Negative Emotions and Heart Disease

In the 1970s and 1980s a popular concept, supported by some research (M. Friedman & Rosenman, 1974), held that people with a *Type A personality* are at heightened risk for heart attacks and other cardiovascular disorders (disorders of the heart and its blood vessels). Type A people were identified as competitive, aggressive, easily irritated, impatient workaholics, constantly concerned with deadlines and getting ahead. In some corners, "I'm Type A" became a kind of boast, meaning "I'm so hard-working and competitive that I'm likely to have a heart attack."

Subsequent research has led to a modification of the original Type A theory. There is now strong evidence that people who frequently experience negative emotions—specifically, anger, anxiety, or sadness—have an enhanced risk for cardiovascular diseases, regardless of the degree to which they are competitive, work-oriented, or have a hurried style of life (H. Friedman & others, 1994; Suls & Bunde, 2005). The hardworking, constantly rushed person who enjoys what he or she is doing is apparently not at heightened risk for heart disease, but the emotionally upset person is. This is confirmed in studies that control for other known risk factors, such as smoking and high cholesterol.

Nobody knows just how negative emotions promote heart disease, but most researchers assume that the effect is mediated through the autonomic nervous system and endocrine system, both of which have direct influences on the circulatory system and are altered during emotional states. Among other things, negative emotions increase blood pressure (Ewart & Kolodner, 1994), which, in turn, can contribute to heart disease.



"I'm learning to relax, doctor—but I want to relax better and faster! I want to be on the cutting edge of relaxation!"

Emotional Distress and Suppression of the Body's Defenses Against Disease

Emotional distress also tends to shut down some of the body's defenses against viruses and other pathogens, increasing our vulnerability to infectious diseases such as colds and flu (Robles & others, 2005; Segerstrom & Miller, 2004).

In one study demonstrating such increased vulnerability, 394 healthy, rather heroic men and women agreed, for the sake of science, to have a fluid containing known respiratory viruses dribbled into their nostrils and then to remain quarantined for 6 days while researchers assessed their medical conditions (Cohen & others, 1991). At the outset of the study, each volunteer filled out a set of questionnaires to assess the degree of psychological distress he or she had experienced recently. The result was that the more distress people reported at the outset of the study, the more likely they were to develop a cold within the 6-day period (see Figure 16.5). The colds were real, not imagined; they were assessed by direct viral counts as well as by external symptoms. Moreover, the relation between distress and the incidence of colds could not be accounted for by any other risk factors the researchers measured, including smoking, alcohol consumption, diet, and quality of sleep.

Other research has shown, more directly, that emotional distress can suppress the body's immune response (Cohen & others, 2001; Segerstrom & Miller, 2004). In one study, people given frustrating cognitive tasks, including difficult mental arithmetic, manifested a temporary decline in the production of T-cells, a class of white blood cells known to be involved in fighting disease organisms (Manuck & others, 1991). This effect occurred only in subjects who showed other physiological signs of distress in response to the tasks. In other studies, laboratory animals subjected to various long-term stressful conditions showed long-term declines in immune responses to infectious agents (Ader & Cohen, 1993; Cohen & others, 1992).

If you think about this from an evolutionary perspective, you might wonder why we and other mammals are constructed in such a manner that our disease-fighting capacity is reduced in times of emotional distress. A plausible answer, proposed by Steven Maier and his colleagues (1994), begins with a reminder that the body responds to stress by shifting as many resources as possible toward muscular energy to enable fight or flight. To do so, the body partly shuts down many physiological functions that are not essential for dealing with the immediate emergency. The physiological mechanisms for resisting disease use a great deal of bodily energy, so it makes sense that these would be suppressed when dealing with something like a charging tiger. The sympathetic nervous system and adrenal hormones mediate much of the body's response to stress, and, as part of that response, they inhibit energy-consuming aspects of the body's defense against disease, including the production of white blood cells. Unfortunately, that ancient stress response continues to be triggered by the anger, fear, or dread that we experience in response to situations or events that can't be dealt with by fight or flight, and that affect us for longer periods of time than a charging tiger would. The immune response is more severely damaged by long-term or chronic stress than by acute stress (Segerstrom & Miller, 2004).

FIGURE 16.5 | Relationship between psychological distress and subsequent development of a cold The psychological distress score was based on a combination of recent stressful life events and negative emotions reported by volunteer subjects on questionnaires. As the degree of distress increased, so did the percentage of subjects who developed a cold within 6 days of deliberate exposure to respiratory viruses. (From Cohen & others, 1991.)

Reprinted with permission from *The New England Journal of Medicine*.



What is some evidence that one's emotional state can affect the immune system and alter one's chance of catching a cold?

From an evolutionary perspective, why might psychological distress inhibit immune function?

SECTION REVIEW

For psychological reasons, the body may be experienced as unhealthy or even become so.

Somatoform Disorders

- In somatoform disorders, individuals experience bodily ailments that originate from psychological distress rather than physical disease. Examples are somatization disorder and conversion disorder.
- Somatization disorder—characterized by a history of dramatic complaints about vague aches and pains unrelated to known medical conditions—may be an alternative way of experiencing what others experience as depression. A resistance to psychological diagnoses may underlie cultural and individual tendencies toward somatization.
- In conversion disorder, the person temporarily loses some bodily function (experiencing blindness or paralysis, for example) for psychological reasons. Though rare in Western countries, it may be more common in children than previously recognized. Dramatic cases have been found in people from non-Western cultures who experienced highly traumatic events, such as Cambodian women who survived the Khmer Rouge in the 1970s.
- Neuroimaging shows that the brain activity patterns of people with conversion disorder are unlike those of people pretending to have the physical disorder, but are similar to those of people given a hypnotic suggestion.

Psychological Factors Affecting Medical Condition

- This diagnostic category refers not to mental disorders, but to the influence of psychological distress on actual medical conditions.
- Widows and (especially) widowers are unusually vulnerable to physical diseases in the months after the death of their spouse. One possible mechanism is decreased self-care, but more direct effects on health are possible.
- Frequent negative emotions (anger, anxiety, sadness) may increase the risk for cardiovascular disease, most likely through changes in autonomic neural activity and hormone secretions.
- Under controlled conditions, emotionally distressed people were more likely than others to develop a cold. In other research, emotional distress was shown to suppress the normal immune response to disease organisms.
- Stress-induced immune suppression may have an evolutionary explanation in the fight-or-flight mechanism that shifts bodily resources to systems needed to handle emergencies.

Schizophrenia

John sits alone in his dismal room, surrounded by plastic bags of garbage that he has hoarded during late-night excursions onto the street. He is unkempt and scrawny. He has no appetite and hasn't eaten a proper meal in weeks. He lost his job months ago. He is afraid to go out during the day because he sees in every passing face a spy who is trying to learn, by reading his mind, the secret that will destroy him. He hears voices telling him that he must stay away from the spies, and he hears his own thoughts as if they were broadcast aloud for all to hear. He sits still, trying to keep calm, trying to reduce the volume of his thoughts. John's disorder is diagnosed as schizophrenia.

Schizophrenia is a serious disorder that is found in roughly 1 percent of people at some time in their lives (Walker & others, 2004). It accounts for a higher percentage of the inpatient population of mental hospitals than do disorders in any other diagnostic category. The disorder is equally prevalent in males and females, but, for unknown reasons, it is typically more severe and strikes earlier in males (Salyers & Mueser, 2001). It usually first manifests itself in late adolescence or early adulthood, and the average age for first diagnosis is about 4 years later in women than in men (Riecher-Rossler & Hafner, 2000). Sometimes people make a full recovery from schizophrenia, sometimes they make a partial recovery, and sometimes the disorder takes a deteriorating course throughout the person's life.

Adolf Wolfi, *Untitled* (1915). The Prinzhorn Institute, University of Heidelberg.



Art by a person with schizophrenia This piece by Adolf Wolfi exemplifies the unusual personal symbolism and rhythmical forms that characterize the artwork of many people with schizophrenia.

The label *schizophrenia* was first used by the Swiss psychiatrist Eugen Bleuler (1911/1950), whose writings are still a valuable source of information and insight about the disorder. The term comes from the Greek words *schizo*, which means “split,” and *phrenum*, which means “mind,” so it literally means “split mind.” Bleuler believed, as do many theorists today, that schizophrenia entails a split among such mental processes as attention, perception, emotion, motivation, and thought, such that these processes operate in relative isolation from one another, leading to bizarre and disorganized thoughts and actions.

Diagnostic Characteristics of Schizophrenia

No two sufferers of schizophrenia have quite the same symptoms. But to receive the *DSM-IV* diagnosis of **schizophrenia**, the person must manifest a serious decline in ability to work, care for himself or herself, and connect socially with others. The person must also manifest, for at least 1 month, two or more of the following five categories of symptoms: disorganized thought and speech, delusions, hallucinations, grossly disorganized or catatonic behavior, and negative symptoms—all described in this section. These symptoms are usually not continuously present; the person typically goes through episodes of active phases of the disorder, lasting for weeks or months, separated by periods of comparative normalcy.

Disorganized Thought and Speech

Many people with schizophrenia show speech patterns that reflect an underlying deficit in the ability to think in a logical, coherent manner. In some cases, thought and speech are guided by loose word associations. A classic example is this greeting to Bleuler (1911/1950) from one of his patients: “I wish you a happy, joyful, healthy, blessed and fruitful year, and many good wine-years to come as well as a healthy and good apple-year, and sauerkraut and cabbage and squash and seed year.” Once the patient’s mind hooked onto fruit (in “fruitful year”), it entered into a chain of associations involving fruits and vegetables that had little to do with the original intent of the statement.

In all sorts of formal tests of logic, people with schizophrenia do poorly when in an active phase of their disorder (Docherty & others, 2003; Simpson & Done, 2004). They often encode the problem information incorrectly, fail to see meaningful connections, or base their reasoning on superficial connections having more to do with the sounds of words than with their meanings.

Delusions

A *delusion* is a false belief held in the face of compelling evidence to the contrary. Common types of delusions in schizophrenia are *delusions of persecution*, which are beliefs that others are plotting against one; *delusions of being controlled*, such as believing that one’s thoughts or movements are being controlled by radio waves or by invisible wires in puppet-like fashion; and *delusions of grandeur*, which are beliefs in one’s own extraordinary importance—for example, that one is the queen of England or is the love object of a famous movie star.

Delusions may result, in part, from a fundamental difficulty in identifying and remembering the original source of ideas or actions (Moritz & others, 2005). For example, delusions of being controlled may derive from a failure to mentally separate voluntary actions from involuntary actions; people with schizophrenia may often find themselves performing actions without remembering that they willfully initiated those actions (Broome & others, 2005). Delusions of persecution may stem, in part, from attempts to make sense of their horrible feelings and confusion. Delusions of grandeur may derive from an inability to separate fantasies from real-world experiences. Moreover, all sorts of delusions may be buttressed by deficits in logical reasoning, as in the case of a woman who supported her claim to be the Virgin Mary this way: “The Virgin Mary is a virgin. I am a virgin. Therefore, I am the Virgin Mary” (Arieti, 1966).

>>> **35**
What are the five main classes of symptoms of schizophrenia?

Hallucinations

Hallucinations are false sensory perceptions—hearing or seeing things that aren't there. By far the most common hallucinations in schizophrenia are auditory, usually the hearing of voices. Hallucinations and delusions typically work together to support one another. For example, a man who has a delusion of persecution may repeatedly hear the voice of his persecutor insulting or threatening him.

A good deal of evidence supports the view that auditory hallucinations derive from the person's own intrusive verbal thoughts—thoughts that others of us would experience as disruptive but self-generated. People with schizophrenia apparently “hear” such thoughts as if they were broadcast aloud and controlled by someone else. When asked to describe the source of the voices, they typically say that they come from inside their own heads, and some even say that the voices are produced, against their will, by their own speech mechanisms (Smith, 1992). Consistent with these reports, people with schizophrenia can usually stop the voices by such procedures as humming, counting, or silently repeating a word—procedures that, in everyone, interfere with the ability to imagine vividly any sounds other than those that are being hummed or silently recited (Bick & Kinsbourne, 1987; Reisberg & others, 1989). In addition, recent brain-imaging research, using fMRI, has revealed that verbal hallucinations in people with schizophrenia are accompanied by neural activity in the same brain regions that are normally involved in subvocally generating and “hearing” one's own verbal statements (Shergill & others, 2004).



Grunnitus / Photo Researchers, Inc.

A person in a catatonic stupor
People with schizophrenia withdraw from their environment in various ways. One of the most extreme forms of withdrawal is the catatonic stupor, in which the person may remain motionless for hours in an uncomfortable position.

Grossly Disorganized Behavior and Catatonic Behavior

Not surprisingly, given their difficulties with thoughts and perceptions, people in the active phase of schizophrenia often behave in very disorganized ways. Many of their actions are strikingly inappropriate for the context—such as wearing many overcoats on a hot day or behaving in a silly manner at a solemn occasion. The inability to keep context in mind and to coordinate actions with it seems to be among the basic deficits in schizophrenia (Broome & others, 2005). Even when engaging in appropriate behaviors, such as preparing a simple meal, people with schizophrenia may fail because they are unable to generate or follow a coherent plan of action.

In some cases, schizophrenia is marked by **catatonic behavior**, defined as behavior that is unresponsive to the environment (American Psychiatric Association, 2000). Catatonic behavior may involve excited, restless motor activity that is not directed meaningfully toward the environment; or, at the other extreme, it may involve a complete lack of movement for long periods—a form referred to as a *catatonic*

stupor. Such behaviors may be means of withdrawing from a world that seems frighteningly difficult to understand or control.

Negative Symptoms

The so-called *negative symptoms* of schizophrenia are those (other than catatonic stupor) that involve a lack of, or reduction in, expected behaviors, thoughts, feelings, and drives (American Psychiatric Association, 2000). They include a general slowing down of bodily movements, poverty of speech (slow, labored, spontaneous speech), flattened affect (reduction in or absence of emotional expression), loss of basic drives such as hunger, and loss of the pleasure that normally comes from fulfilling drives.

The majority of people with schizophrenia manifest negative symptoms to some degree, and for many these are the most prominent symptoms. I once asked a dear friend, who was suffering from schizophrenia and was starving himself, why he didn't eat. His answer, in labored but thoughtful speech, was essentially this: “I have no appetite. I feel no pleasure from eating or anything else. I keep thinking

that if I go long enough without eating, food will taste good again and life might be worth living.” For him, the most painful symptom of all was an inability to experience normal drives or the pleasure of fulfilling them.

Variations of the Disorder

Two people diagnosed with schizophrenia may differ greatly from each other, with little overlap in the kinds of symptoms they show. Studies using factor analysis to determine which symptoms correlate most strongly with each other suggest three different clusters of symptoms: *positive symptoms* (primarily delusions and hallucinations), *disorganized symptoms* (disorganized thought, speech, and behavior), and *negative symptoms* (as defined earlier) (Toomey & others, 1997). Positive symptoms and, to a lesser degree, disorganized symptoms tend to come and go; patients often experience remission from them for many months between episodes (Docherty & others, 2003). These symptoms are also most responsive to antipsychotic drugs. The negative symptoms, in contrast, are more constant and less responsive to drugs.

Various subcategories of schizophrenia have been proposed, based on the prevalence of specific symptoms or symptom clusters. The subcategories described in *DSM-IV* include *paranoid type*, characterized mainly by delusions of persecution and grandeur and often by hallucinations; *catatonic type*, characterized mainly by non-reaction to the environment; and *disorganized type*, characterized by disorganized speech, disorganized behavior, and either inappropriate or flattened affect. However, no system of subcategories has proved very satisfactory. The symptoms appear in a huge number of different combinations and may change over time in a given person, thereby defying attempts at classification.

Underlying Cognitive and Neural Deficits in Schizophrenia

Schizophrenia is typically characterized as primarily a cognitive disorder, brought on by deleterious changes in the brain (Broome & others, 2005; Heinrichs, 2005). To say that it is a cognitive disorder is to say that the fundamental psychological deficits have to do with the ability to process information.

Deficits in Information Processing

People with schizophrenia appear to suffer from deficits in essentially all the basic processes of attention and memory that were discussed in Chapter 9. They perform particularly poorly at tasks that require sustained attention over time or that require them to respond only to relevant information and to ignore irrelevant information (Bozikas & others, 2005; Nuechterlein & others, 1992; K. Wang & others, 2005). They are abnormally slow at bringing perceived information into their working-memory stores, and poor at holding onto the information once it is there (Barch, 2003; Fuller & others, 2005). They also perform more poorly than other subjects at encoding and recalling new long-term explicit memories, and they are especially deficient in remembering the source of new information (Kerns & Berebaum, 2002; Moritz & others, 2005). Because of their failure to recall the source, they have difficulty knowing whether or not something they learned was presented to them as truth or fiction or was something they merely imagined—a problem all of us occasionally experience to a much lesser degree. All these kinds of mental deficits appear to contribute to the whole set of diagnostic symptoms of schizophrenia—the disordered thought, delusions, hallucinations, disorganized behavior, and negative symptoms (Broome & others, 2005).

Such information-processing deficits are usually greatest during the active phase of the disorder, but they are also found in people who are in remission—that is, who are no longer manifesting the diagnostic symptoms (Green, 1993; Heinrichs, 2005). To a lesser degree, such deficits also occur in the biological relatives of people with schizophrenia (Park & others, 1995). In a long-term study of

>>> **36**
What general cognitive deficits have been observed in laboratory studies of people with schizophrenia?

people known to be genetically at risk for schizophrenia, deficits in attention and memory in childhood correlated positively with the subsequent development of schizophrenia in adulthood (Erlenmeyer-Kimling & others, 2000). Such findings suggest that deficits in attention and memory are not just symptoms of schizophrenia, but are, to some degree, preexisting conditions that contribute to a person's vulnerability to the disorder.

Disruptions in Brain Chemistry

37

What early evidence supported the dopamine theory of schizophrenia? Why is the simple form of that theory doubted today?

<<< The mental problems observed in schizophrenia derive, presumably, from deficits in the brain. By the early 1970s evidence had accrued for the *dopamine theory of schizophrenia*—the theory that schizophrenia arises from too much activity at brain synapses where dopamine is the neurotransmitter. The most compelling evidence was the observation that the clinical effectiveness of drugs in reducing the positive symptoms of schizophrenia was directly proportional to the drug's effectiveness in blocking dopamine release at synaptic terminals (Seeman & Lee, 1975). Other support came from the finding that drugs such as cocaine and amphetamine, which increase the action of dopamine in the brain, can greatly exacerbate the symptoms of schizophrenia in people with the disorder (Davis, 1974) and, at higher doses, can even induce such symptoms in people who do not have the disorder (Griffith & others, 1972).

Today few, if any, researchers accept the original, simple form of the dopamine theory. One major flaw in the theory is that it does not explain the negative symptoms of schizophrenia, which are not well treated by drugs that act solely on dopamine and are not typically exacerbated by drugs that increase that action of dopamine. Newer drugs used to treat the negative as well as positive symptoms of schizophrenia have effects beyond blocking dopamine release (discussed in Chapter 17). Modern theories suggest that schizophrenia may involve unusual patterns of dopamine activity, perhaps including overactivity in some areas of the brain and underactivity in others, and may also involve disruptions in the activity of other neurotransmitters (Goldsmith & others, 1997; Javitt & Coyle, 2004).

38

What evidence supports the theory that a defect in glutamate neurotransmission may play a role in schizophrenia?

<<< Currently, a good deal of attention is being devoted to the role of *glutamate* in schizophrenia. As was noted in Chapter 5, glutamate is the major excitatory neurotransmitter at fast synapses throughout the brain. Some research suggests that one of the major receptor molecules for glutamate is defective in people who have schizophrenia, resulting in a decline in the effectiveness of glutamate neurotransmission (Javitt & Coyle, 2004; Phillips & Silverstein, 2003). Such a decline could account for the general cognitive debilitation that characterizes the disorder. Consistent with this theory, the dangerous, often-abused drug phencyclidine (PCP)—known on the street as “angel dust”—interferes with glutamate neurotransmission and is capable of inducing, for varying periods of time, in otherwise normal people, the full range of schizophrenic symptoms, including the negative and disorganized symptoms as well as hallucinations and delusions (Javitt & Coyle, 2004).

Alterations in Brain Structure

Studies using brain-imaging techniques have shown structural differences between the brains of people with schizophrenia and those of other people (Heinrichs, 2005; Walker & others, 2004). The most common finding is enlargement of the cerebral ventricles (fluid-filled spaces in the brain) accompanied by a reduction in neural tissue surrounding the ventricles. Decreased neural mass has been found in various parts of the brain, including portions of the limbic system, basal ganglia, thalamus, cerebellum, and the temporal and frontal lobes of the cerebral cortex (Marsh & others, 1997; Walker & others, 2004). These differences are usually small and vary from person to person; they are not sufficiently reliable to be useful in diagnosing the disorder (Heinrichs, 2005).

© The New Yorker Collection 2005 Charles Barsotti from cartoonbank.com. All rights reserved.





© Joe McNally

Identical twins with nonidentical brains
Steven, on the right, has schizophrenia. His genetically identical twin, on the left, does not. Brain scans revealed that Steven's brain has larger cerebral ventricles (seen here as Y-shaped black areas within the brain) than does his brother's. Such discordance for schizophrenia in identical twins sometimes occurs when one of the two twins (usually the one born second) suffers from lack of oxygen during the birth process and the other does not.

Some researchers have attempted to explain the timing of the onset of schizophrenia in terms of maturational changes in the brain. During adolescence, the brain normally undergoes certain structural changes: Many neural cell bodies are lost, a process referred to as *pruning*, and many new neural connections grow. The result is a decrease in gray matter (masses of cell bodies) and an increase in white matter (bundles of axons running from one brain area to another). Some researchers have suggested that an abnormality in pruning, which leads to the loss of too many cell bodies, may underlie at least some cases of schizophrenia (Lewis & Levitt, 2002). Consistent with this view, neuroimaging studies of people at risk for schizophrenia revealed a larger decline in gray matter during adolescence or early adulthood in those who subsequently developed the disorder than in those who did not (Lewis & Levitt, 2002; Pantelis & others, 2003).

>>> **39**
How might an exaggeration of a normal developmental change at adolescence help bring on schizophrenia?

Genetic and Environmental Causes of Schizophrenia

Why do the neural and cognitive alterations that bring on and constitute schizophrenia occur in some people and not in others? Genetic differences certainly play a role, and environmental differences do, too.

Predisposing Effects of Genes

Schizophrenia was one of the first mental disorders to be studied extensively by behavior geneticists (Gottesman, 1991). In such studies, the first step is to identify a group of people, referred to as *index cases*, who have the disorder. Then the relatives of the index cases are studied to see what percentage of them have the disorder. This percentage is referred to as the **concordance** for the disorder, for the class of relatives studied. The average concordances found for schizophrenia in many such studies, for various classes of relatives, are shown in Table 16.3 on page 614.

All in all, the results indicate that genetic differences among individuals play a substantial role in the predisposition for schizophrenia. The more closely related a person is to an index case, the greater is the chance that he or she will develop schizophrenia. Other research, conducted with people who were adopted at an early age, shows high concordance for schizophrenia between biological relatives but not between adoptive relatives (Owen & O'Donovan, 2003). The results of one such study are shown in Figure 16.6 on page 614. Such results indicate that it is the genetic similarity, not the environmental similarity, between relatives that produces high concordance for schizophrenia.

>>> **40**
How do the varying rates of concordance for schizophrenia among different classes of relatives support the idea that heredity influences one's susceptibility for the disorder?

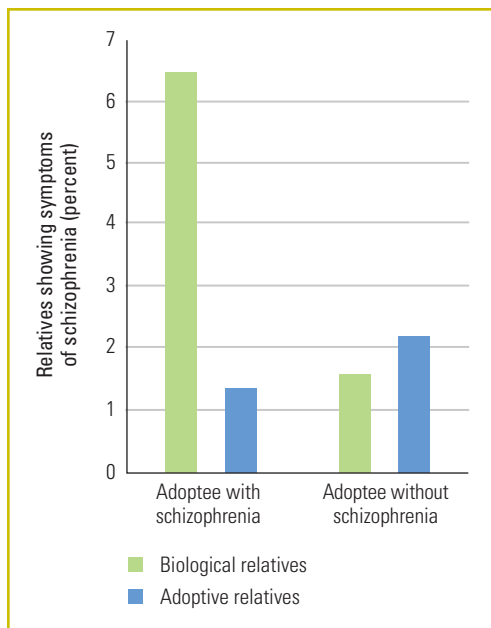


FIGURE 16.6 Results of a classic study of the heritability of schizophrenia The researchers looked for signs of schizophrenia in the biological and adoptive relatives of people who had been adopted at an early age and either did or did not subsequently develop schizophrenia. The results here are the percentage of relatives who showed either schizophrenia or a milder disorder now called schizotypal personality disorder. (Data from Kety & others, 1976.)

TABLE 16.3

Concordance rates for schizophrenia

Relationship to a person who has schizophrenia	Average percentage found to have schizophrenia (concordance)
Relatives in same generation	
Identical twin	48%
Fraternal twin	17
Non-twin brother or sister	9
Half-sibling	6
First cousin	2
Relatives in later generation	
Child of two parents with schizophrenia	46
Child of one parent with schizophrenia	13
Grandchild of one person with schizophrenia	5
Niece or nephew of one person with schizophrenia	4

Sources: I. I. Gottesman, 1991. *Schizophrenia genesis: The origins of madness*, p. 96. New York: Freeman.

Recent research has been aimed at identifying individual genes that contribute to the development of schizophrenia. Many different genes appear to be involved, and progress has been made in identifying the chromosomal locations of some of them (Owen & O'Donovan, 2003; Walker & others, 2004). It is still too early to say just how these genes contribute to the brain alterations that predispose a person for schizophrenia, but, consistent with current chemical theories of schizophrenia, at least one of the identified genes is believed to influence dopamine neurotransmission in some parts of the brain and several of them are believed to influence glutamate neurotransmission (Broome & others, 2005).

Effects of the Prenatal Environment and Early Brain Traumas

The data in Table 16.3 show that genes are heavily involved in the predisposition for schizophrenia, but they also show that genes are not the only determinants of the disorder. Of particular interest is the fact that the average concordance for schizophrenia for identical twins, 48 percent, is much less than the 100 percent that would be predicted if genes alone were involved. Another noteworthy observation, in the table, is that the concordance for schizophrenia in fraternal twins is considerably higher than that for non-twin pairs of full siblings—17 percent compared to 9 percent. This difference cannot be explained in terms of genes, as fraternal twins are no more similar to each other genetically than are other full siblings, but it is consistent with the possibility of a prenatal influence. Twins share the same womb at the same time, so they are exposed to the same prenatal stressors and toxins. That fact, of course, applies to identical twins as well as to fraternal twins, so some portion of the 48 percent concordance in identical twins may result not from shared genes but from shared prenatal environments.

41

What sorts of early disruptions to brain development have been implicated as predisposing causes of schizophrenia?

<<< A number of studies have pointed to specific prenatal variables that can contribute to the liability for schizophrenia. One such variable is malnutrition. People born in the western Netherlands between October 15 and December 31, 1945, immediately following a severe famine brought on by a Nazi blockade of food supplies, are twice as likely as others born in the Netherlands to have developed schizophrenia (Susser & others, 1996). A more recent study, illustrated in Figure 16.7, revealed a similar effect for people who were born in China during or shortly after the Chinese famine of 1960–1961, which resulted from massive crop failures (St. Clair & others, 2005).

Prenatal viral infections and birth complications may also contribute to a predisposition for schizophrenia. Heightened rates of schizophrenia have been found in

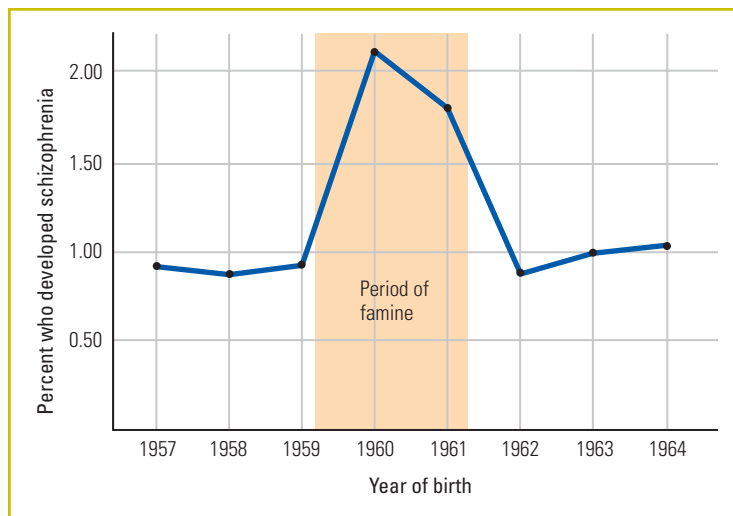


FIGURE 16.7 | Effect of malnutrition during pregnancy on rate of schizophrenia in offspring Children who were in their mothers' wombs during the period of great famine in China later developed schizophrenia at roughly twice the usual rate. (Data from St. Clair & others, 2005.)

people whose mothers had rubella (also known as German measles) or certain other viral diseases during pregnancy and in people who had difficult births, involving oxygen deprivation or other trauma to the brain (Lewis & Levitt, 2002; Tsuang & others, 2001). There is also evidence that head injury later on in childhood, before age 10, can increase the liability for schizophrenia (AbdelMalik & others, 2003).

Effects of Life Experiences

The effects discussed so far are predisposing effects. Given a certain level of predisposition, the actual manifestation and course of schizophrenia may be influenced by one's daily experiences. There is evidence that stressful life events of all sorts can exacerbate the symptoms of schizophrenia or precipitate its active phases (Fowles, 1992; Pallanti & others, 1997). Most research on such effects has focused on the family environment.

The most thorough investigation, to date, of the effect of the family environment on schizophrenia is a 21-year-long study, in Finland, of two groups of adopted children (Tienari & others, 2004). One group was genetically at high genetic risk for schizophrenia because their biological mothers were diagnosed with either schizophrenia or a milder schizophrenia-like disorder, and the other group was at low risk because neither of their biological parents showed any evidence of schizophrenia or a related disorder. The main finding was that those high-risk children whose adoptive parents communicated in a relatively disorganized, hard-to-follow, or highly emotional manner were more likely to develop schizophrenia or a milder disorder akin to schizophrenia than were high-risk children whose adoptive parents communicated in a calmer, more organized fashion. This relationship was not found among the low-risk children. The results suggest that a degree of disordered communication at home that does not harm most children may have damaging effects on those who are genetically predisposed for schizophrenia.

Other research has centered on a concept referred to as *expressed emotion*, defined as criticisms and negative attitudes or feelings expressed about and toward a person with a mental disorder by family members with whom that person lives. Families high in expressed emotion tend to blame the diagnosed person for his or her disruptive or ineffective behavior rather than seeing it as something that he or she cannot control. Many studies indicate that, other things being equal, the greater the expressed emotion, the greater the likelihood that the active symptoms will return or worsen and the person will require hospitalization (Hooley, 2004). Moreover, a number of experiments have shown that support and training for family members, aimed at creating a more stable and accepting home environment and reducing expressed emotion, significantly reduce the rate of relapse (Hooley & Hiller, 2001).

>>> 42
What evidence suggests that the family environment may promote schizophrenia, but only in those who are genetically predisposed for the disorder?

>>> 43
What evidence suggests that a hostile family environment can be a perpetuating cause of schizophrenia?

Living and playing with schizophrenia

Despite having schizophrenia since he was a young man, Tom Harrell has developed into one of the world's top jazz composers and trumpeters. According to his wife, composing and playing music help to focus Harrell's mind and to reduce the symptoms of schizophrenia.



Gabe Palacio / Image Direct

44

What cross-cultural difference has been observed in rate of recovery from schizophrenia? What are some possible explanations of that difference?

Influences of Culture on the Course of Schizophrenia

Perhaps the most striking evidence for environmental influences on the course of schizophrenia comes from comparisons made across cultures. In the late 1970s, the World Health Organization (WHO) initiated an ambitious cross-cultural study of schizophrenia involving various locations in 10 countries (Jablensky & others, 1992). Three of the countries—Colombia, Nigeria, and India—were identified as relatively nonindustrialized, *developing* countries. The others were identified as *developed* countries—England, Ireland, Denmark, the United States, Japan, Czechoslovakia (before it was divided), and Russia (then part of the Soviet Union). Using agreed-upon criteria and cross-cultural reliability checks, the researchers diagnosed new cases of schizophrenia in each location, classed them according to symptom types and apparent severity, and reassessed each case through interviews conducted at various times in subsequent years.

The results showed considerable cross-cultural consistency. The relative prevalence of the various symptoms, the severity of the initial symptoms, the average age of onset of the disorder, and the sex difference in age of onset (later for women than for men) were similar from location to location despite wide variations in the ways that people lived. However, the study also revealed a remarkable cross-cultural difference, which had been strongly hinted at by earlier, less systematic studies: The patients in developing countries were much more likely to recover from their symptoms during the follow-up period than were those in developed countries. This was true for every category of schizophrenia that was identified and for nearly every index of recovery that was used. At the 2-year follow-up, for example, 63 percent of patients in the developing countries, compared with only 37 percent in the developed countries, showed full remission from schizophrenic symptoms (Jablensky & others, 1992). Subsequent follow-ups and smaller-scale studies of other developing countries have revealed similar results (Hopper, 2004; Kulhara & Chakrabarti, 2001).

Why should people in less industrialized countries fare better after developing schizophrenia than people in more industrialized countries? Nobody knows for sure, but some interesting hypotheses have been put forth (Jenkins & Karno, 1992; Lin & Kleinman, 1988; López & Guarnaccia, 2005). Family members in less industrialized countries generally place less value on personal independence and more on interdependence and family ties than do those in industrialized countries, which may lead them to feel less resentful and more nurturing toward a family member who needs extra care. They are more likely to live in large, extended families, which means that more people share in providing the extra care. Perhaps

for these reasons, family members in less industrialized countries are more accepting and less critical of individuals diagnosed with schizophrenia than are those in industrialized countries. They show less expressed emotion, as defined earlier, which might account for the better recovery rate. They are also less likely to call the disorder “schizophrenia” or to think of it as permanent, and are more likely to refer to it as “a case of nerves,” which sounds more benign and ties it to experiences that everyone has had. Finally, in less industrialized countries those with schizophrenia are more able to play an economically useful role. The same person who could not hold a nine-to-five job can perform useful chores on the family farm or at the family trade. Being less stigmatized, less cut off from the normal course of human activity, and better cared for by close family members may increase the chance of recovery.

Another possible explanation has to do with drugs. Over the course of the WHO study, patients in the developed countries were generally treated with antipsychotic drugs for prolonged periods, as they are today. In contrast, patients in the developing countries were more often not treated with drugs or were treated only for short periods to bring the initial symptoms under control (Jablensky & others, 1992). Some controversial studies in Europe and the United States, some years ago, suggested that prolonged use of antipsychotic drugs, while dampening the positive symptoms of schizophrenia, may impede full recovery (Warner, 1985). This ironic possibility—that the drugs we use to treat schizophrenia might prolong the disorder—has not been followed up with systematic research.

A Developmental Model of Schizophrenia

The more scientists learn about schizophrenia, the more apparent it becomes that the disorder has no simple, unitary cause. Figure 16.8 provides a useful framework (similar to one developed by Tsuang & others, 2001) for thinking about the multiple causes. The disorder is brought on by some combination of genetic predisposition, early physical disruptions to normal brain development, and stressful life experiences. Once the disorder begins, the disorder itself can have effects that help to prolong it, depending at least partly on how family members and others in one’s community respond. Distressful effects of the disorder itself, or of others’ reactions to it, can cause further deterioration of the brain and a more chronic course for the disorder.

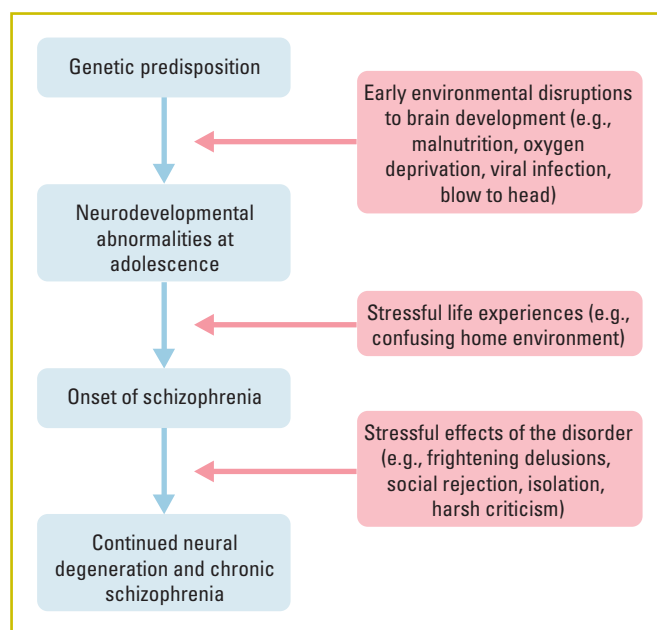


FIGURE 16.8 | A developmental model of schizophrenia This model depicts researchers’ understanding of the contributions of genes, environmental experiences, and the person’s present neural and behavioral condition to the development of schizophrenia and, in some cases, its progression into a chronic condition. (Based on a model depicted by Tsuang & others, 2001, p. 22.)

SECTION REVIEW

Schizophrenia is a cognitive disorder with wide-ranging symptoms and multiple causes.**Diagnostic Characteristics**

- Schizophrenia is characterized by various symptoms classified as positive or negative.
- Positive symptoms include (a) disorganized, illogical thought and speech; (b) delusions (false beliefs held despite compelling contrary evidence); (c) auditory hallucinations (which may come from the person's own intrusive verbal thoughts); and (d) disorganized behavior or catatonic behavior (behavior that is unresponsive to the environment).
- Negative symptoms include slowed movement, poverty of speech, flattened affect, and the loss of basic drives and the pleasure that comes from fulfilling them.
- Symptoms occur in a great variety of combinations and may change over time in the same person.

Underlying Cognitive and Neural Deficits

- The fundamental deficits in schizophrenia are cognitive, including problems with attention, working memory, and long-term memory.
- Abnormalities in brain chemistry, such as a decline in the effectiveness of glutamate (the major excitatory neurotransmitter at fast synapses) or unusual patterns of dopamine activity, may help to explain cognitive deficits.
- Structural differences in the brains of people with schizophrenia include enlarged cerebral ventricles and reduced neural mass in some areas. One theory suggests that the normal pruning of neural cell bodies in adolescence may overshoot the mark.

Genetic and Environmental Causes

- High concordance for schizophrenia indicates high heritability. Some of the specific genes that appear to be involved have effects that are consistent with brain chemistry theories of the disorder.
- Predisposition may also arise from early environmental injury to the brain—including prenatal viruses or malnutrition, birth complications, or early childhood head injury.
- Stressful life experiences and aspects of the family environment can bring on the active phase of schizophrenia or worsen symptoms in predisposed people.
- Recovery from schizophrenia has been shown to occur at higher rates in developing countries than in developed countries. Differences in living conditions, cultural attitudes, and the lower use of antipsychotic drugs in developing countries may play a role.

Concluding Thoughts

Here are three final ideas that may help you organize your thoughts as you reflect on what you have read in this chapter:

1. The multiple causes of mental disorders By this time in your study of psychology, you are no doubt used to the idea that human feelings, thoughts, and actions emerge from the interplay of many causes. This idea also applies to the feelings, thoughts, and actions that lead to the diagnosis of a mental disorder. The differences among us that are considered disorders—no less than the differences that are considered normal—are caused by differences in our genes and in our past and present environments. Any claim to have found *the* cause of generalized anxiety, depression, schizophrenia, or another major class of disorder is of doubtful validity.

One way to review this chapter would be to think about each disorder in relation to the three classes of causes—predisposing, precipitating, and perpetuating—that were introduced near the beginning of the chapter. How might genes and early physical assaults on the brain operate through the person's physiology to predispose him or her for the disorder? How might learned ways of thinking or acting predispose a person for the disorder? How might specific stressful events in life interact with the predisposition to precipitate the disorder? How might the disordered behavior itself, or people's reactions to it, help perpetuate the disorder once it begins? For each disorder, think about the relationship between the research or ideas described

and the possible answers to these questions. Note, too, that similar stressful events can result in different kinds of symptoms—and different diagnosed disorders—for different individuals or at different times for any given individual.

2. The continuum between normality and abnormality

Although a diagnosis of a mental disorder is categorical (all or none), the symptoms on which diagnoses are made are not; they vary in degree throughout the population. Thus, any decision as to whether a particular person does or does not have a mental disorder is based on arbitrary criteria describing how severe or prolonged each symptom must be in order to call the syndrome a disorder.

As you review each disorder described in the chapter, think about its symptoms in relation to the moods, emotions, thoughts, and behaviors that all of us manifest to some degree. Doing so helps remove some of the mystique from the concept of mental disorder and helps us identify with people whose troubles are like ours, only stronger. Thinking of disorders in terms of extremes of normal processes has also helped scientists understand them better. For example, in this chapter you read about evidence (a) relating the symptoms of ADHD to normal childhood impulsiveness and exuberance; (b) relating the symptoms of general anxiety disorder, phobias, and obsessive-compulsive disorders to normal worries, fears, thoughts, and actions; (c) relating the symptoms of depression to normal, possibly adaptive mood

changes following loss or failure; and (d) relating the auditory hallucinations experienced in schizophrenia to the intrusive thoughts that all of us experience at times.

3. An evolutionary view of mental disorder The evolutionary theme running through this book maintains that behavior is generally functional: It promotes survival and reproduction. But mental disorder is dysfunctional: It reduces one's ability to work and interact with others effectively in survival-promoting ways. Why, then, do mental disorders exist?

One partial answer may be that mental disorders are a cost the species pays for the general advantages that come with di-

versity. The previous chapter, on personality, dealt with individual differences within the range generally considered healthy, or not disordered, and the case was made that natural selection may have favored diversity because those who are different from average can exploit unfilled niches and reap their rewards. Such diversity stems from variations in genes and from behavioral mechanisms that are capable of being modified through experience. Variation in capacity for anxiety, or compulsiveness, or sadness may be beneficial within a certain range; but the coin tosses that distribute genes and experiences will sometimes produce those characteristics at pathological levels.

Further Reading

FLOYD SKLOOT (2005). *A world of light*. Lincoln: University of Nebraska Press.

In 1988, at age 41, Floyd Skloot lost much of his memory, linguistic ability, and ability for coherent thought as a result of a viral infection that destroyed parts of his brain. In 2003 he published a book, *In the Shadow of Memory*, describing his slow but active struggle to reconstruct his sense of self as he regained some of his mental capacities. Now, in *A World of Light*, he reconstructs some of his own past and describes his ongoing encounters with his mother, who is in her nineties and has Alzheimer's disease. The book offers a fascinating glimpse into the world of dementia, from a poetic soul who himself has experienced some of the effects of dementia.

TEMPLE GRANDIN (1995). *Thinking in pictures and other reports from my life with autism*. New York: Doubleday Dell.

In this remarkable book, Temple Grandin, a highly intelligent person with autism, tells us what it is like to think in pictures and translate them, with much effort, into words. She also describes her struggles growing up in a world of people whom she did not understand and who did not understand her, and her fascination with animal thought, which led her

to develop new, more humane ways of handling cattle.

KAY JAMISON (1995). *An unquiet mind*. New York: Knopf.

In this autobiography, Kay Jamison—a leading researcher of mood disorders—describes her own experiences as a sufferer of bipolar disorder. The book includes an especially vivid description of her first full-blown manic episode, which happened in her first year as an assistant professor, and a moving discussion of how she came to terms with the disorder and has managed to live with it.

NANCY C. ANDREASEN (2001). *Brave new brain: Conquering mental illness in the era of the genome*. Oxford: Oxford University Press.

Andreasen is a leading researcher into the biological foundations of mental disorders and an excellent writer. Here, for nonspecialists, she conveys what it is like to have a serious mental disorder or to have a loved one with such a disorder, and she also describes how biological research is leading to new understandings of such disorders and new methods of treatment. She focuses particularly on schizophrenia, mood disorders, dementias (including Alzheimer's disease), and anxiety disorders.

STEPHEN SCHWARTZBERG (2000). *Casebook of psychological disorders: The human face of emotional distress*. Boston: Allyn & Bacon.

This book introduces readers to mental disorders by describing case histories, including one each involving posttraumatic stress disorder, panic disorder, bipolar disorder, conversion disorder, dissociative identity disorder, and schizophrenia. Each case is discussed respectfully and compassionately, with a description of the person's psychological difficulties and history and the therapist's diagnosis and treatment plan.

DALE PETERSON (Ed.) (1982). *A mad people's history of madness*. Pittsburgh: University of Pittsburgh Press.

In this fascinating collection of excerpts from autobiographies written over the past 500 years, people who either were, or were regarded as, seriously mentally disordered describe their suffering, the reactions of others, and their own efforts toward recovery. Some of the excerpts are from well-known books that have helped shape reforms in the understanding and treatment of people with mental disorders.



Treatment

CARE AS A SOCIAL ISSUE

What to Do with the Severely Disturbed? A Brief History
Structure of the Mental Health System

BIOLOGICAL TREATMENTS

Drugs
Other Biologically Based Treatments

PSYCHOTHERAPY I: PSYCHODYNAMIC AND HUMANISTIC THERAPIES

Principles of Psychodynamic Therapies
A Case Example: Freud's Analysis of "the Rat Man"
Principles of Humanistic Therapy
A Case Example: Rogers's Meetings with a Silent and Bitter Young Man

PSYCHOTHERAPY II: COGNITIVE AND BEHAVIORAL THERAPIES

Principles of Cognitive Therapy
A Case Example: Beck's Cognitive Treatment of a Depressed Young Woman
Principles of Behavior Therapy
A Case Example: Miss Muffet Overcomes Her Spider Phobia

EVALUATING PSYCHOTHERAPIES

Is Psychotherapy Helpful, and Are Some Types of It More Helpful Than Others?
The Role of Common Factors in Therapy Outcome

1

How has Western society's response to people with serious mental disorders changed since the Middle Ages? What were the goals of the moral-treatment and deinstitutionalization movements?

Throughout this book you have read of psychology as a *science*. Perhaps you have discovered that psychology is a vast, complex, and fascinating science in which every finding generates far more questions than it answers. If your experience reading this book has been anything like my experience writing it, your attitude right now may be one of respect for what psychologists have discovered through their research, coupled with awe concerning the amount they have yet to learn about the human mind and behavior.

But psychology is not just a science; it is also a *practice*. Practitioners in psychology attempt to apply psychological ideas and findings in ways designed to make life more satisfying for individuals or society as a whole. This last chapter is about *clinical psychology*, the field of practice and research that is directed toward helping people who suffer from psychological problems and disorders. In this chapter you will see how some of the basic knowledge of the brain, mind, and behavior that you have read about in previous chapters has been applied in efforts to help people in psychological need. The chapter begins with a section on the social problem of providing care for the seriously disturbed, and then it progresses through sections that deal with various biological and psychological treatments for mental disorders. The final section is concerned with questions about the effectiveness of psychotherapy.

Care as a Social Issue

The existence of mental suffering, like that of physical suffering, raises social and moral questions. Who, if anyone, is responsible for caring for those who cannot care for themselves?

What to Do with the Severely Disturbed? A Brief History

Through most of history, Western cultures felt little obligation toward people with mental disorders. During the Middle Ages, and even into the seventeenth century, people with serious mental disorders—the kind called “madness” or “lunacy” (and today most often diagnosed as schizophrenia)—were often considered to be in league with the devil, and “treatment” commonly consisted of torture, hanging, burning at the stake, or being sent to sea in “ships of fools” to drown or be saved, depending upon divine Providence. By the eighteenth century, such “religious” views had waned somewhat and a more secular attitude began to prevail, which attributed mental disorders



Mary Evans Picture Library / Photo Researchers

Life in a nineteenth-century mental hospital One source of pressure for mental hospital reform in early-nineteenth-century England were portraits, such as this one, drawn by George Cruikshank. The man shown here had been bound to the wall by foot-long chains for 12 years at the time of the portrait.



Culver Pictures

Dorothea Dix This Boston schoolteacher's crusade resulted in new asylums for the mentally ill and improved conditions in existing institutions.

not to supernatural powers but to the basic degeneracy and unworthiness of the disordered people themselves. The principal treatment for those who couldn't care for themselves was to put them out of the way of decent society in places that were called "hospitals" but in reality were dark, damp, miserable dungeons, where inmates were frequently kept chained to the walls, alive but in a state that was perhaps worse than death.

Early Attempts at Reform, and the Failure of Asylums

Not until the beginning of the nineteenth century did significant humanitarian reform begin to occur. The best-known leader of reform in Europe was Philippe Pinel (1745–1826), who, as director of a large mental hospital in Paris, unchained the inmates, transferred them to sunny and airy rooms, and gave them access to the hospital grounds for exercise. Under these conditions some inmates who had been deemed permanently and hopelessly deranged recovered sufficiently to be released from the hospital.

In the United States, the leading reformer was Dorothea Dix (1802–1887), a Boston schoolteacher who visited dozens of jails and almshouses where people with mental disorders were housed and who then publicized the appalling conditions she found. As this *moral-treatment movement* grew, it spurred the building of large, state-supported asylums for the mentally disordered. The idea behind such institutions was high-minded: to provide kindly care and protection for those unable to care for themselves. Unfortunately, public sympathy was rarely sustained at this high level, at least not in the tangible form of financial support. Almost invariably, the asylums became overcrowded and understaffed, and reverted to conditions

not unlike those that had appalled Pinel and Dix. As recently as the 1940s, the following report could be written about a state mental institution in Philadelphia (Deutsch, 1948):

The male "incontinent ward" was like a scene out of Dante's *Inferno*. Three hundred nude men stood, squatted and sprawled in this bare room. . . . Winter or summer, these creatures never were given any clothing at all. . . . Many patients [in another ward] had to eat their meals with their hands. . . . Four hundred patients were herded into a barn-like day room intended for only 80. There were only a few benches; most of the men had to stand all day or sit on the splintery floor. . . . The hogs in a nearby pigpen were far better fed, in far greater comfort than these human beings.

Deinstitutionalization, and Continuing Deficits in Care

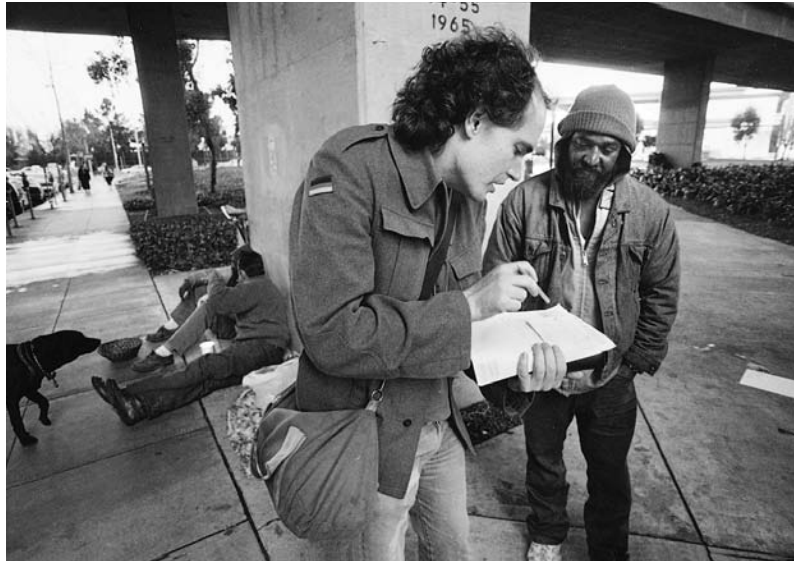
By the mid-1950s, disenchantment with large state institutions led to a new kind of reform movement in the United States—a movement to *deinstitutionalize* people with mental disorders, to get them back into the community. This new movement was inspired partly by the development of antipsychotic drugs and partly by a general mood of optimism in the nation, a feeling that everyone could "make it" if given the chance. President John F. Kennedy gave the movement a boost in 1961 by encouraging Congress to pass legislation to establish community-based mental health centers (Bassuk & Gerson, 1978). By the early 1970s, hundreds of such centers were in operation, offering transitional homes and outpatient care to patients who were capable of living in the community.

Unfortunately, the dream of the community mental health movement, like the earlier dream of asylums, has remained mostly unrealized. The number of chronic patients in state mental institutions has been greatly reduced—from about 600,000 in 1955 to about 60,000 now (Lamberg, 2004)—but it is debatable whether the quality of life of those who would formerly have been in asylums has been improved. They have generally not been integrated into the community but are living on its fringes. Roughly 200,000 of them are homeless, and a similar number are in prisons, usually for minor crimes such as trespassing or theft (Fleishman, 2004; Lamberg, 2004). Many thousands of others are living in rundown rooming houses and understaffed nursing homes (Lamb, 2000). The more fortunate minority are

living in long-term residential-care facilities, or *group homes*, which provide room and board, supervise medication, and offer assistance with problems of daily living (Fleishman, 2004).

A Positive Development: Assertive Community Treatment

Over the past 30 years or so, an increasing number of communities have developed outreach programs, often referred to as *assertive community treatment* programs, aimed at helping the severely mentally ill wherever they are in the community. Each mentally ill person in need is assigned to a multidisciplinary treatment team, which typically includes a case manager, psychiatrist, general physician, nurse, and social workers. Someone on the team is available at any time of day, seven days a week, to respond to crises. Each patient—whether living on the street or in a boarding house or with family—is visited at least twice a week by a team member, who checks on his or her health, sees if any services are needed, and offers counseling when that seems appropriate. In addition, the team meets frequently with family members who are involved with the patient, to support them in their care for the patient. Several well-designed research studies have shown that such programs can be highly effective in reducing the need for hospitalization and increasing patients' satisfaction with life (Bustillo & others, 2001; Rothbard & others, 2004). Such programs are expensive to operate, but generally, in the long run, save money by keeping the mentally ill out of hospitals, where their care is much more expensive (Lehman & others, 1998). Despite such evidence, the majority of people with schizophrenia and their families in the United States do not receive such services (Drake & others, 2001).



How do assertive community treatment programs attempt to help the severely mentally disordered and their families?

Reaching out In some communities social workers seek out people who are in need, get to know them as individuals, and help them gain access to useful services and means of support.

Structure of the Mental Health System

Public concern about mental health generally centers on the most severely disordered individuals, those who cannot care for themselves. But most people who seek mental health services have milder problems. The kinds of services they seek depend on such factors as the severity of their disorders, what they can afford, and the services available in their community. They may seek help from self-help groups organized by others who suffer from similar problems or disorders (such as Alcoholics Anonymous), from religious organizations, from general-practice physicians, or from mental health professionals.

Mental Health Professionals

Mental health professionals are those who have received special training and certification to work with people who have psychological problems or mental disorders. The primary categories of such professionals are the following:

- *Psychiatrists* have medical degrees, obtained through standard medical school training, followed by special training and residency in psychiatry. They are the only mental health specialists who can regularly prescribe drugs. They typically work in private offices, mental health clinics, and hospitals.
- *Clinical psychologists* usually have doctoral degrees in psychology, with training in research and clinical practice. Some are employed by universities as teachers and researchers in addition to having their own clinical practices. Like psychiatrists, they most often work in private offices, clinics, or hospitals.

What are the major categories of mental health providers?

Stone / Getty Images



Group support In group therapy, the therapeutic elements of support, hope, and motivation may derive not just from the therapist but also from all of the other participants.

- *Counseling psychologists* usually have doctoral degrees in counseling. Their training is similar to that of clinical psychologists but generally entails less emphasis on research and more on practice. In general, counseling psychologists are more likely than are psychiatrists or clinical psychologists to work with people who have problems of living that do not warrant a diagnosis of mental disorder.
- *Counselors* usually have master's degrees in counseling. They receive less training in research and psychological diagnostic procedures than do doctoral-level clinical or counseling psychologists. They often work in schools or other institutions, with people who are dealing with school- or job-related problems. They may also conduct psychotherapy in private practice.
- *Psychiatric social workers* usually have master's degrees in social work, followed by advanced training and experience working with people who have psychological problems. They are most often employed by public social-work agencies. They may conduct psychotherapy sessions or visit people in their homes to offer support and guidance.
- *Psychiatric nurses* usually have degrees in nursing, followed by advanced training in the care of mental patients. They usually work in hospitals and may conduct psychotherapy sessions as well as provide more typical nursing services.

Where People with Common Mental Disorders Go for Treatment

4

According to a survey conducted in the United States, where do people with mental disorders typically find treatment, and what types of treatment do they find?

◀◀ In the United States, several large-scale household surveys have been conducted to identify people who have mental disorders and to find out where they have sought treatment. The most recent such survey was conducted from early 2001 to early 2003 (P. S. Wang & others, 2005). That survey identified a representative sample of thousands of people who were suffering from clinically significant anxiety disorders, mood disorders (including major depression and bipolar disorders), substance use disorders (alcoholism and other drug-abuse or dependence disorders), or intermittent explosive disorder (a disorder involving uncontrolled anger). Of these people, 22 percent had received some form of treatment from a mental health professional within the past year, 59 percent had received no treatment at all, and most of the remainder had received treatment from a medical doctor or nurse who did not have a mental health specialty.

Support for friends and family of alcohol and drug abusers

The family and friends of people with a debilitating mental or behavioral disorder often need psychological support themselves, in order to be helpful to their loved one while at the same time preserving their own well-being. Al-Anon is a self-help group run by and for people who have a friend or family member who is addicted to alcohol or another drug. Self-help groups are valuable components of the mental health system that are not directed by mental health professionals.



Mary Kate Denny / PhotoEdit, Inc.

The survey also revealed that the typical person with a mental disorder who saw a general-practice physician saw that person just once or twice over the course of the year, usually to receive a prescription for drug treatment and/or a few minutes of counseling. In contrast, those who saw a mental health professional met with that person for an average of seven sessions of at least a half hour's length, mostly for counseling or psychotherapy (P. S. Wang & others, 2005). Not surprisingly, the wealthier and more educated a person with a mental disorder was, the more likely he or she was to have met for a series of sessions with a mental health professional for psychotherapy (P. S. Wang & others, 2005).

SECTION REVIEW

Caring for mentally suffering people raises moral, social, and practical issues.

How Society Has Responded to the Severely Disturbed

- Severely disturbed people were once considered to be allies of the devil or, later, degenerate and unworthy. They were often tortured, killed, or imprisoned in horrific so-called hospitals.
- In the early nineteenth century, reformers like Pinel in France and Dix in the United States successfully worked toward kindly treatment of the severely disturbed in large mental hospitals, but these soon returned to deplorable conditions.
- Deinstitutionalization—began in the 1950s in the United States—was a response to both the gross failure of large mental institutions and the apparent success of antipsychotic drugs. The alternative envisioned, community-based care, was never fully realized.
- Some assertive community treatment programs, however, have offered extensive, well-integrated, multidisciplinary care on an ongoing basis. These programs are quite effective, but expensive.

Structure of the Mental Health System

- Mental health professionals include psychiatrists (the only category that regularly prescribes drugs), clinical psychologists, counseling psychologists, counselors, psychiatric social workers, and psychiatric nurses. They differ in level and type of training, in the severity of problems they deal with, and in work setting.
- A recent survey of people with mental disorders in the United States found that, in the previous year, 22 percent had received treatment from a mental health professional, 59 percent had received no treatment, and most others briefly saw a medical doctor or nurse.

Biological Treatments

A person diagnosed with a mental disorder might be treated by biological means (most often drugs), psychological means (psychotherapy of one form or another), or both. Biological treatments attempt to relieve the disorder by directly altering bodily processes. In the distant past, such treatments included drilling holes in the skull to let out bad spirits and bloodletting to drain diseased humors. Today, in decreasing order of extent of use, the three main types of biological treatments are drugs, electroconvulsive shock therapy, and psychosurgery.

Drugs

A new era in the treatment of mental disorders began in the early 1950s when two French psychiatrists, Jean Delay and Pierre Deniker (1952), reported that they had reduced or abolished the psychotic symptoms of schizophrenia with a drug called *chlorpromazine*. Today, a plethora of drugs is available for treating essentially all major varieties of mental disorders.

Drugs for mental disorders have been far from unmixed blessings, however. They are not magic bullets that zero in on and correct a disordered part of the mental machinery while leaving the rest of the machinery untouched. Like drugs used in general medicine, drugs used to treat mental disorders nearly always produce undesirable side effects. Some of the drugs are also addictive, and the attempt to withdraw from them sometimes produces symptoms worse than those for which the drug was prescribed. As you read of the three categories of drugs described below, notice their drawbacks as well as their benefits.



Dan McCoy / Rainbow

Madness as possession This twelfth-century painting probably depicts trephination (piercing the skull to permit evil spirits to escape). With the revival of the medical model in the late Middle Ages, people gave up the use of trephination and other brutal treatments that were based on the belief that abnormal behavior indicates possession.

5

What is known about the mechanisms, effectiveness, and limitations of drugs used to treat schizophrenia, generalized anxiety, and depression?

<<< **Antipsychotic Drugs**

Antipsychotic drugs are used to treat schizophrenia and other disorders in which psychotic symptoms (such as hallucinations and delusions) predominate. Chlorpromazine (sold as Thorazine) was the first such drug, but now many others exist as well. Many well-designed experiments have shown that such drugs really do reduce or abolish the hallucinations, delusions, and bizarre actions that characterize the active phase of schizophrenia and that they prevent or forestall the recurrence of such symptoms when taken continually (Ashton, 1987; Freedman, 2003). The antipsychotic drugs most often prescribed today come in two broad categories, usually referred to, respectively, as the “typical” and “atypical” antipsychotics.

The *typical* antipsychotic drugs—including chlorpromazine and haloperidol—were the first to be developed and are still at least as effective as newer drugs in relieving the positive symptoms of schizophrenia (hallucinations, delusions, and bizarre behavior) (Hefti, 2005). As noted in Chapter 16, all of these drugs decrease the activity of the neurotransmitter dopamine at certain synapses in the brain, and that effect is believed to be responsible, directly or indirectly, for the reduction in positive symptoms. These drugs do not, however, reduce the negative symptoms (blunted emotions, poverty of speech and thought, and lack of motivation) of schizophrenia; in fact, in some cases they worsen the negative symptoms (Fowles, 1992). Another major problem is that they interfere with motor-control processes in the brain and thereby produce symptoms akin to Parkinson's disease, including shaking and difficulty in controlling voluntary movements (Klein & others, 1980). Roughly 40 percent of patients who take typical antipsychotic drugs for many years eventually develop a serious and often irreversible motor disturbance called *tardive dyskinesia*, manifested as involuntary jerking of the tongue, face, and sometimes other muscles (Karon, 1989).

The newer, *atypical* antipsychotic drugs—including risperidone, olanzapine, and several others—have the advantage that they partially relieve the negative symptoms of schizophrenia as well as the positive ones, and they are less likely than the typical drugs to produce Parkinson's-like symptoms or tardive dyskinesia. For these reasons, despite still being called “atypical,” these drugs have gradually replaced the “typical” drugs as the treatment of first choice (Freedman, 2003; Walker & others, 2004). The atypical drugs, like the typical drugs, reduce the activity of dopamine at certain synapses, and this is presumed to account for their effectiveness against the positive symptoms of schizophrenia. Unlike the typical antipsychotics, however, these drugs also alter the activity of a number of other neurotransmitters, which may help explain their ability to reduce the negative symptoms of schizophrenia (Hefti, 2005). In addition, the atypical antipsychotics generally interfere less strongly with dopamine neurotransmission than do the typical drugs, and this may explain why they are less likely to produce motor-control problems (Hefti, 2005).

All antipsychotic drugs, however, have serious and unpleasant side effects. They can produce dizziness, confusion, nausea, dry mouth, blurred vision, constipation, sexual impotence in men, and disrupted menstrual cycles in women (McKim, 2000). They also commonly cause weight gain, and the atypical antipsychotics in particular produce metabolic changes that heighten the risk for diabetes (Freedman, 2003). Although antipsychotic drugs reduce many of the symptoms of schizophrenia, they do not cure the person of the disorder (Hefti, 2005). Most people with schizophrenia who take such drugs remain unable to keep a job or to function fully normally in their daily life. Moreover, as noted in Chapter 16, there is as yet no convincing evidence to refute the contention of some critics that antipsychotic drugs, given for prolonged periods, may have effects on the brain that reduce the chance of full recovery from schizophrenia (Breggin, 1997).

Antianxiety Drugs

Drugs used to treat anxiety are commonly referred to as *tranquilizers*. At one time, barbiturates such as phenobarbital were often prescribed as tranquilizers, and

many people became seriously addicted to them. During the 1960s, barbiturates were replaced by a new, safer group of antianxiety drugs belonging to a chemical class called *benzodiazepines* [**ben**-zoh-di-az-uh-peens], including chlordiazepoxide (sold as Librium) and diazepam (sold as Valium). According to some estimates, by 1975 more than 10 percent of adults in the United States and Western Europe were taking these drugs on a regular basis (Lickey & Gordon, 1991; Lipman, 1989). Subsequently their use declined somewhat (Pincus & others, 1998), partly because of growing recognition that they are not as safe as they were once thought to be. Today, benzodiazepines are still the most frequently prescribed drugs for generalized anxiety disorder, and they are also often prescribed for panic disorder (Pies, 2005). They are generally not effective, however, for other anxiety disorders. Obsessive-compulsive disorder and posttraumatic stress disorder are more likely now to be treated with an antidepressant drug (in the SSRI class, described below) than with an antianxiety drug (Pies, 2005).

Biochemically, benzodiazepines appear to produce their tranquilizing effects by augmenting the action of the neurotransmitter GABA (gamma-aminobutyric acid) in the brain (Hefti, 2005). GABA is the brain's main inhibitory neurotransmitter, so its increased action decreases the excitability of neurons almost everywhere in the brain. Side effects of benzodiazepines at high doses include drowsiness and a decline in motor coordination. More important, the drugs potentiate the action of alcohol; an amount of alcohol that would otherwise be safe can produce a coma or death in people taking a benzodiazepine. In addition, benzodiazepines are now known to be at least moderately addictive, and very unpleasant withdrawal symptoms—sleeplessness, shakiness, anxiety, headaches, and nausea—occur in those who stop taking them after having taken high doses for a long time (Bond & Lader, 1996). These symptoms subside within about 2 weeks, but they drive many people back to the pills before the end of that period.

Antidepressant Drugs

From the 1960s into the early 1980s, the drugs most commonly used to treat depression belonged to a chemical class referred to as *tricyclics*, of which imipramine (sold as Tofranil) and amitriptyline (sold as Elavil) are examples. Tricyclics block the reuptake of the neurotransmitters serotonin and norepinephrine into pre-synaptic neurons after their release into the synapse, thereby prolonging the action of the transmitter molecules on post-synaptic neurons (Hefti, 2005). Beginning in the mid-1980s, a newer class of antidepressants, referred to as *selective serotonin reuptake inhibitors* (SSRIs), which block the reuptake of serotonin but not that of other monoamine transmitters, overtook the tricyclic drugs as the first line of treatment for depression (Parott & others, 2004). The first and still most often prescribed SSRI is fluoxetine (Prozac). To visualize how such drugs act at the synapse, see Figure 17.1 (on p. 628).

Many research studies have demonstrated that the tricyclics and SSRIs are about equally effective in treating depression. With either type of drug, approximately 50 percent of people suffering from major depression show a clinically significant improvement in mood, compared with about 30 percent who improve over the same time period if given a placebo (inactive substance) instead (Hollon & others, 2002). The SSRIs are preferred because of their milder side effects. Tricyclics are much more likely to cause death if taken in overdose than are SSRIs, and they are also more likely to produce such disruptive and discomforting side effects as fatigue, dry mouth, and blurred vision. Still, in some people, the SSRIs do produce unwanted side effects, including reduced sexual drive, headache, nausea, and diarrhea (Hollon & others, 2002).

As noted in Chapter 16, the effects of antidepressant drugs on neurotransmitters occur immediately, but the antidepressant effects take several weeks to develop. This fact suggests that some gradual process—maybe including the growth of new neurons in



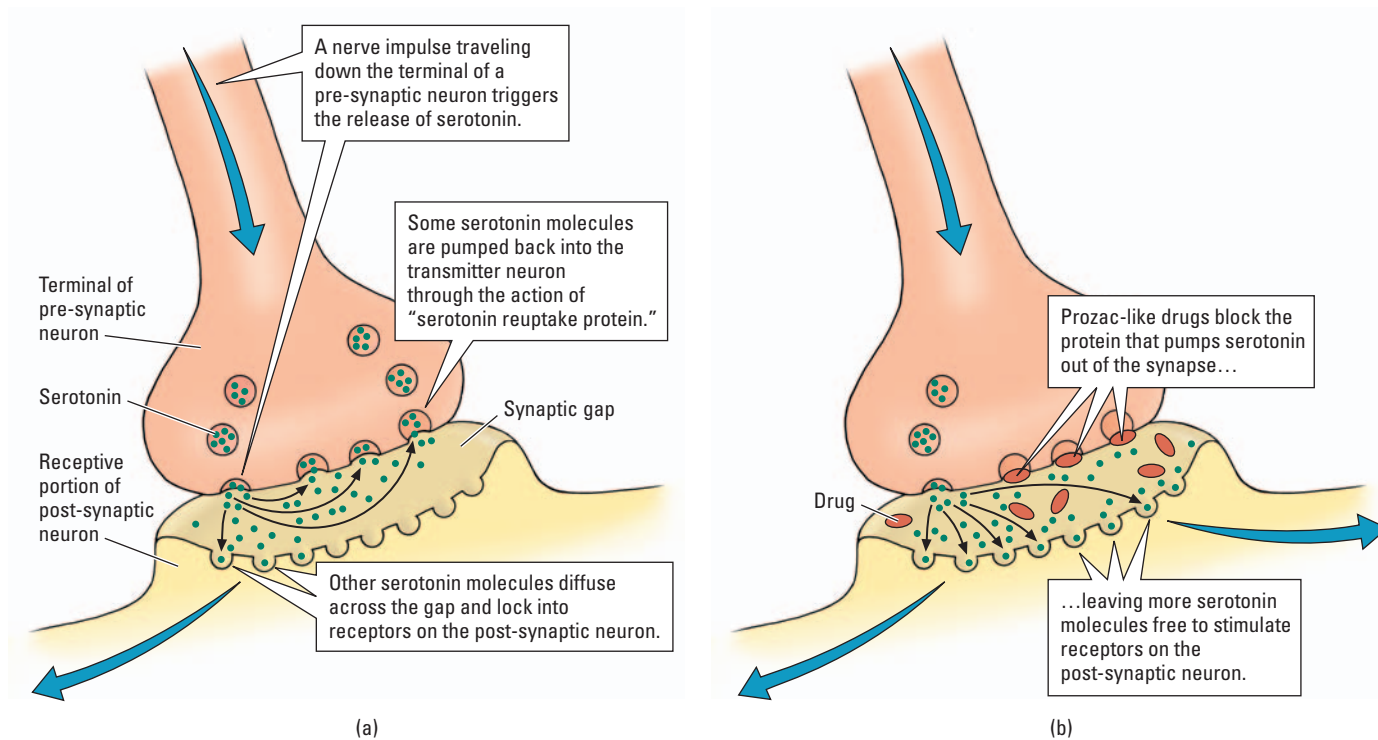


FIGURE 17.1 | How Prozac and other SSRI antidepressant drugs work (a) When serotonin is released into a synapse, its action is normally cut short by reuptake into the pre-synaptic neuron. (b) Selective serotonin reuptake inhibitors block this reuptake and thereby increase the action of serotonin on the post-synaptic neuron. Reproduced by permission of *New Scientist*.

the brain—underlies the therapeutic effect (Jacobs, 2004). However, despite many theories and much study, researchers still do not know just how antidepressants reduce depression. If they knew, they could probably develop new drugs that would be more effective in reducing depression and would produce fewer side effects.

Placebo Effects

To be approved for clinical use, a drug must be shown to be more effective than a placebo in treating the condition for which it is intended. A *placebo* (as discussed in Chapter 2) is an inactive substance that is indistinguishable in appearance from the drug. To test a drug's efficacy in treating a mental disorder, a large number of people diagnosed with the disorder must agree to take part in an experiment, and then each person is assigned randomly to either the drug or the placebo condition. The experiment is double-blind (as discussed in Chapter 2)—that is, neither the subjects nor the researchers who evaluate them are told who is receiving the drug or the placebo. When such experiments are conducted to test drugs for anxiety disorders or depression, the typical result is that both groups improve considerably over time, but the drug group improves somewhat more than does the placebo group (Kirsch, 2000; Oosterbaan & others, 2001; Wampold & others, 2005).

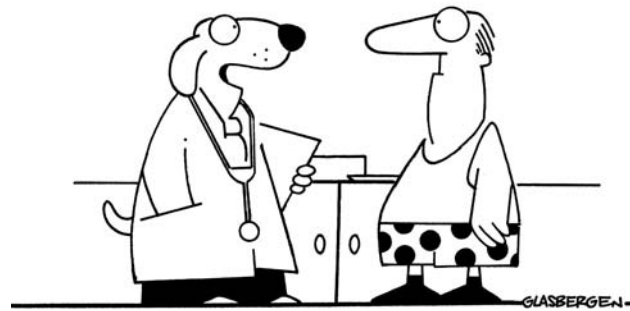
6 How might strong placebo effects in the treatment of anxiety and depression be explained? Why has it been difficult to prove that the effects of drug treatments are not simply placebo effects?

Why do anxious and depressed people who receive a placebo improve as much as they do? Anxiety and depression are characterized by feelings of helplessness and hopelessness. In fact, for some people those feelings *are* the disorder. Simply participating in a treatment experiment—meeting regularly with someone who seems to care and taking what *might* be a useful drug—may restore feelings of control and hope and produce expectations of improvement. These feelings and expectations may promote life changes that lead to partial or full recovery from the disorder. All such effects, which derive from the belief that one's disorder is being treated, are referred to as *placebo effects*.

In addition to placebo effects, some improvement typically occurs over time even with no treatment at all. Depression and anxiety are disorders that tend to come and go. Improvement that would occur without any treatment, not even

placebo treatment, is referred to as *spontaneous remission*. According to one analysis, roughly 25 percent of the improvement shown by groups of people receiving antidepressant drugs may be attributable to spontaneous remission, 50 percent may be attributable to the placebo effect, and the remaining 25 percent may be attributable to the specific effects that antidepressant drugs have on the brain (Kirsch & Sapirstein, 1999).

A serious limitation of placebo-controlled studies of psychoactive drugs is that patients are often able to figure out—according to the presence or absence of side effects—whether they are receiving the active drug or the placebo (Wampold & others, 2005). A few researchers (e.g., Kirsch & Sapirstein, 1999) have gone so far as to suggest that antidepressant drugs may work better than placebos only because they are more effective than typical placebos in inducing placebo effects! People experience the side effects and feel all the more certain that they are receiving an active drug, a belief that gives them hope for recovery. The best way to determine whether a drug effect is more than a placebo effect is to conduct experiments with so-called active placebos—that is, with substances that mimic the side effects of the drug without producing the specific effect in the brain that is presumed to be the basis for the drug's effectiveness. The few such studies that have been conducted have generally shown that active placebos are as effective as standard antidepressant drugs in relieving depression, but the experiments have been criticized on the grounds that the substances used as active placebos might indeed have antidepressant effects through their action on the brain (Hamburg, 2000; Kirsch, 2000).



"Play some Frisbee, chew on an old sock, bark at a squirrel. If that doesn't make you feel better, eat some cheese with a pill in it."

© 2001 by Randy Glasbergen.
www.glasbergen.com

Other Biologically Based Treatments

The increased use of drugs, coupled with the increased understanding and acceptance of psychotherapy, has led to the abandonment of most non-drug biological therapies for mental disorders. Two such treatments still used, however, are electroconvulsive therapy and, in very rare cases, psychosurgery.

Electroconvulsive Therapy

Electroconvulsive therapy, or **ECT**, is used primarily in cases of severe depression that do not respond to psychotherapy or antidepressant drugs. To the general public this treatment often seems barbaric, a remnant of the days when victims of mental disorders were tortured to exorcise the demons, and indeed it once was a brutal treatment. The brain seizure induced by the shock would cause muscular contractions so violent that they sometimes broke bones. Today, however, ECT is administered in a way that is painless and quite safe (Fink, 2000; Mathew & others, 2005). Before receiving the shock, the patient is given drugs that block nerve and muscle activity so that no pain will be felt and no damaging muscle contractions will occur. Then an electric current is passed through the patient's skull, triggering a seizure in the brain that lasts a minute or so. Usually such treatments are given in a series, one every 2 or 3 days for about 2 weeks.

Overall, about 60 percent of people who are suffering from major depression and have not been helped by psychotherapy and drug treatments experience remission with ECT (Hollon & others, 2002). In some cases, the remission is permanent; in others, depression recurs after several months or more, and then another series of treatments may be given. Nobody knows how ECT produces its antidepressant effect. In non-human animals, such shocks cause immediate release of all varieties of neurotransmitters, followed by longer-lasting changes in transmitter production and in the sensitivity of post-synaptic receptors (Nutt & Glue, 1993). The shocks also stimulate the growth of new neurons in the brain, which some believe may contribute to their antidepressant effect (Jacobs, 2004).

Under what conditions and how is ECT used to treat depression? Why is it often administered just to the right half of the brain?

7

Electroconvulsive therapy ECT is the treatment of choice for very severe depression that does not respond to drug therapy. This man has been anesthetized (so he will be unconscious), given a muscle relaxant (so he will not show muscle spasms when the shock is given), and provided with a tube so that his breathing will not be impeded. The shock will be applied to the right side of his brain through the two black leads shown in the photo. Unilateral right-hemisphere ECT disrupts memory much less than does ECT applied across the whole brain.

W. & D. McIntyre / Photo Researchers



Clinicians and researchers have long been concerned about the possibility that ECT can produce permanent brain damage. This concern is fueled by evidence of brain damage in laboratory animals subjected to intense electroconvulsive shocks. To date there is no evidence that such damage occurs with the shock levels used in clinical treatment, but temporary disruptions in cognition, especially in memory, do occur (Mathew & others, 2005). When ECT is applied in the traditional manner—*bilaterally*, with the current running across both of the brain's hemispheres—the patient typically loses memory for events that occurred a day or two before the treatment. With a series of treatments, some memories for earlier events may also be lost, but these usually return within a month after the last treatment (Calev & others, 1993).

Many clinicians today apply ECT *unilaterally*, just to the right half of the brain. With a sufficiently high shock level, that procedure appears to be as effective, or nearly as effective, as the bilateral procedure in reducing depression, and it causes little or no loss of conscious, verbal memories (Sackeim & others, 2000; UK ECT Review Group, 2003). Unilateral right-brain treatment may, however, cause some loss of pictorial memories, as indexed by reduced recognition of geometric designs seen before the shocks (Sachs & Gelenberg, 1988). Such findings are consistent with other evidence (discussed in Chapter 5) that the left half of the brain is more involved than the right in verbal processes and the right half is more involved in visuospatial processes and emotion.

Psychosurgery and Deep Brain Stimulation

A treatment of last resort today is *psychosurgery*, the surgical cutting or production of lesions in portions of the brain to relieve a mental disorder. From the late 1930s into the early 1950s, tens of thousands of men and women were subjected to an operation called *prefrontal lobotomy*, in which the anterior (front) portions of the frontal lobes were surgically separated from the rest of the brain. Individuals with severe cases of schizophrenia, bipolar disorder, depression, obsessive-compulsive disorder, and pathological violence were subjected to the operation. Prefrontal lobotomy was so highly regarded that in 1949 the Portuguese neurologist who developed the technique, Egas Moniz, was awarded the Nobel prize.

8 Why are prefrontal lobotomies no longer performed, and why are less drastic forms of psychosurgery occasionally performed?

<<< By the mid-1950s, however, prefrontal lobotomies had gone out of style, partly because newly developed drug treatments offered an alternative and partly because of mounting evidence that, although lobotomy relieved people of their incapacitating emotions, it left them severely incapacitated in new ways (Valenstein, 1986). The prefrontal lobes are a critical part of the brain's circuitry for integrating plans with action (discussed in Chapter 5), and lobotomized patients showed life-long deficits in the ability to make plans and behave according to them.

Refined versions of psychosurgery were developed in the 1960s and continue to be used in rare cases today. The new procedures involve destruction of very small areas of the brain by applying radio-frequency current through fine wire electrodes implanted temporarily into the brain. Today this procedure is used primarily for treatment of highly incapacitating cases of obsessive-compulsive disorder that have proven, over many years, to be untreatable by any other means. As was noted in Chapter 16, obsessive-compulsive disorder is often associated with abnormal amounts of activity in a neural circuit that is involved in converting conscious

thoughts into actions. This circuit includes a portion of the prefrontal cortex, a portion of the limbic system called the cingulum, and parts of the basal ganglia. Surgical destruction either of a portion of the cingulum or of a specific neural pathway that enters the basal ganglia reduces or abolishes obsessive-compulsive symptoms in about 50 percent of people who could not be successfully treated in any other way (Mashour & others, 2005). These procedures produce quite serious side effects in some patients, however, including confusion, weight gain, depression, and, very occasionally, epilepsy (Mashour & others, 2005; Trimble, 1996).

In recent years a number of brain surgeons have been experimenting with a new, safer procedure, called *deep brain stimulation*, for treating intractable cases of obsessive-compulsive disorder (Abbott, 2005). In this procedure a hair-thin wire electrode is implanted permanently into the brain—usually in the cingulum or in a portion of the basal ganglia. The electrode can be activated in order to electrically stimulate, rather than destroy, the neurons lying near it. High-frequency but low-intensity stimulation through the electrode is believed to desynchronize and disrupt ongoing neural activity and in that way have an effect comparable to producing a lesion (Abelson & others, 2005). This effect, unlike that of a lesion, can be reversed just by turning off the electrical current. Trials with deep brain stimulation suggest that it may be as effective as psychosurgery, without the negative side effects (Abelson & others, 2005; Husted & Shapira, 2004).

What is the advantage of deep brain stimulation over current forms of psychosurgery?



Preparation for deep brain stimulation

As a thin wire electrode is inserted into the patient's brain, the surgeon keeps track of its position through brain imaging.

Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Vol. 436/7, July 2005; Abbott, Allison: Neuroscience: Deep in Thought.

SECTION REVIEW

Biological treatments target the brain physically in order to alleviate mental disorders.

Drugs

- Antipsychotic drugs treat psychotic symptoms, but do not cure people. Typical antipsychotic drugs (the original class) effectively treat the positive symptoms of schizophrenia but not the negative ones; atypical antipsychotic drugs treat both. All are believed to work by affecting neurotransmission. All have unpleasant side effects, some quite serious.
- Antianxiety drugs, in a chemical class called benzodiazepines, are effective mainly for generalized anxiety disorder and panic disorder. They increase inhibitory activity in the brain, thus reducing excitability; but they are addictive, have side effects (including alcohol potentiation), and cause unpleasant withdrawal symptoms.
- Antidepressant drugs include tricyclics and selective serotonin reuptake inhibitors (SSRIs); the former prolong the action of serotonin and norepinephrine in the brain, while the latter affect only serotonin. They are equally effective in treating depression, but SSRIs have milder side effects and are therefore more often prescribed.
- Placebos are quite effective in reducing anxiety and depression; antianxiety and antidepressant drugs appear to be only moderately more effective than placebos.

Other Biological Treatments

- In electroconvulsive therapy (ECT), used to treat depression not helped by other means, electrical current is applied to the skull to induce brain seizures. It is quite safe and effective, but causes some loss of recent memories when given across the whole brain rather than just across the right hemisphere.
- Prefrontal lobotomies, once common, are no longer performed. Today psychosurgery involving small, localized lesions is used occasionally for incapacitating obsessive-compulsive disorder. It is often effective, but potential side effects are serious.
- Deep brain stimulation, a possible (but experimental) alternative to psychosurgery, uses electrical current to disrupt activity rather than destroy tissue at specific brain locations.

Psychotherapy I: Psychodynamic and Humanistic Therapies

While biological treatment for mental disorder is aimed at improving moods, thinking, and behavior through altering the chemistry and physiology of the brain, *psychological* treatment, referred to as *psychotherapy*, is aimed at improving the same through talk, reflection, learning, and practice.

The two approaches—biological and psychological—are not incompatible. Indeed, most clinicians agree that the best treatment for many people who suffer from serious mental disorders involves a combination of drug therapy and psychotherapy. In theory as well as in practice, the biological and the psychological are tightly entwined. Changes of any sort in the brain can alter the way a person feels, thinks, and behaves; and changes in feeling, thought, and behavior can alter the brain. The brain is not a “hard-wired” machine. It is a dynamic biological organ that is constantly growing new neural connections and losing old ones as it adapts to new experiences and thoughts.

Psychotherapy can be defined as any theory-based, systematic procedure, conducted by a trained therapist, for helping people to overcome or cope with mental problems through psychological rather than directly physiological means. Psychotherapy usually involves dialogue between the person in need and the therapist, and its aim is usually to restructure some aspect of the person's way of feeling, thinking, or behaving. If you have ever helped a child overcome a fear, encouraged a friend to give up a bad habit, or cheered up a despondent roommate, you have engaged in a process akin to psychotherapy, though less formal.

By one count, more than 400 nominally different forms of psychotherapy have appeared over the years (Karasu, 1986). Psychotherapists may work with groups of people, or with couples or families, as well as with individuals. In this chapter, however, our discussion is limited to four classic varieties of individual therapy. We examine psychodynamic and humanistic therapies in this section, and cognitive and behavioral therapies in the next.

There was a time when many psychotherapists believed that their own approach was “right” and other approaches “wrong.” Today, however, most therapists recognize strengths and weaknesses in each of the classic schools of thought in psychotherapy. As shown in Figure 17.2, nearly 30 percent of psychotherapists today consider themselves to be “eclectic” or “integrative” in orientation—that is, they do not identify with any one school of thought but use methods that are gleaned from various schools. Even among those who do identify with a particular school of thought, most borrow techniques and ideas from other schools.

As you will discover in this section and the next, each major approach in psychotherapy draws on a set of psychological principles and ideas that apply to adaptive as well as to maladaptive behavior. The psychodynamic approach focuses on the idea that unconscious memories and emotions influence our conscious thoughts and actions. The humanistic approach focuses on the value of self-esteem and self-direction, and on the idea that people often need psychological support from others in order to pursue freely their own chosen goals. The cognitive approach focuses on the idea that people's ingrained, habitual ways of thinking affect their moods and behavior. The behavioral approach focuses on the roles of basic learning processes in the development and maintenance of adaptive and maladaptive ways of responding to the environment.

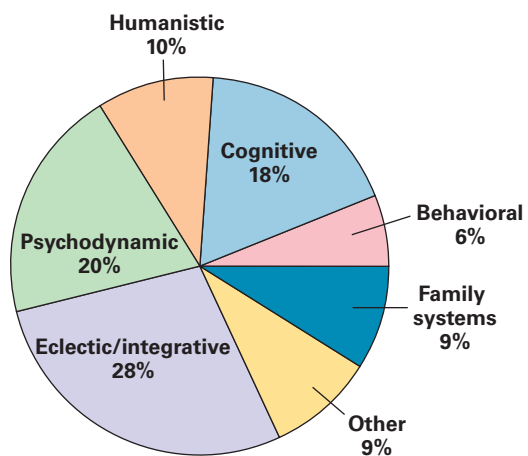


FIGURE 17.2 Theoretical orientations of a sample of psychotherapists Shown here are the results of a survey in which North American clinical and counseling psychologists were asked to name the theoretical orientation with which they most closely identified (Bechtoldt & others, 2001). (Note: The “humanistic” category includes those who labeled their practice as humanistic, Rogerian, or existential; and the “psychodynamic” category includes those who labeled their practice as psychodynamic or psychoanalytic.)

Principles of Psychodynamic Therapies

Sigmund Freud, whose theory of personality was discussed in Chapter 15, was the primary founder of psychodynamic psychotherapy; indeed, he is widely regarded as the founder of all of psychotherapy. As noted in Chapter 15, Freud used the

term *psychoanalysis* to refer to both his theory of personality and his methods of therapy. Today the term ***psychoanalysis*** is generally used to refer to those forms of therapy that adhere most closely to the ideas set forth by Freud, and the broader term ***psychodynamic therapy*** is used to include psychoanalysis and therapies that are more loosely based on Freud's ideas. In the survey depicted in Figure 17.2, only 10 percent of those who categorized their therapy as psychodynamic identified their specific method as psychoanalysis (Bechtoldt & others, 2001). In what follows, I set forth some of the main principles and methods that are common to most if not all psychodynamic therapies.

The Idea That Unconscious Conflicts, Often Deriving from Early Childhood Experiences, Underlie Mental Disorders

The most central, unifying idea of all psychodynamic theories is the idea that mental problems arise from unresolved mental conflicts, which themselves arise from the holding of contradictory motives and beliefs. The motives, beliefs, and conflicts may be unconscious, or partly so, but they nevertheless influence conscious thoughts and actions. The word *dynamic* refers to force, and *psychodynamic* refers to the forceful, generally unconscious influences that conflicting motives and beliefs can have on a person's consciously experienced emotions, thoughts, and behavior.

Freud himself argued, in some of his writing (e.g., Freud, 1933/1964), that the unconscious conflicts that cause trouble always originate in the first five or six years of life and have to do with infantile sexual and aggressive wishes, but few psychodynamic therapists today have retained that view. Most psychodynamic therapists today are concerned with conflicts that can originate at any time in life and have to do with any drives or needs that are important to the person. Still, however, most such therapists tend to see sexual and aggressive drives as particularly important, as these drives often conflict with learned beliefs and societal constraints. They also tend to see childhood as a particularly vulnerable period during which frightening or confusing experiences can produce lasting marks on a person's ways of feeling, thinking, and behaving. Such frightening or confusing experiences might derive from such sources as sexual or physical abuse, lack of security, or lack of consistent love from parents.

In general, psychodynamic theorists see their approach as tightly linked, theoretically, to the field of developmental psychology. In their view, growing up necessarily entails the facing and resolving of conflicts; conflicts that are not resolved can produce problems later in life (Thompson & Cotlove, 2005).

The Idea That Patients' Observable Speech and Behavior Provide Clues to Their Unconscious Conflicts

To a psychodynamic therapist, the symptoms that bring a person in for therapy and that are used to label a disorder, using *DSM-IV* or any other diagnostic guide, are just that—*symptoms*. They are surface manifestations of the disorder; the disorder itself is buried in the person's unconscious mind and must be unearthed before it can be treated. For example, consider a patient who is diagnosed as having anorexia nervosa (an eating disorder discussed in Chapter 16). To a psychodynamic therapist, the fundamental problem with the person is not the failure to eat but is something deeper and more hidden—some conflict that makes her want to starve herself. Two people who have similar symptoms and identical diagnoses of anorexia nervosa may suffer from quite different underlying conflicts. One might be starving herself because she fears sex, and starvation is a way of forestalling her sexual development; another might be starving herself because she feels unaccepted for who she is. To a psychodynamic therapist, the first step in treatment of either woman is that of finding out *why* she is starving herself, and, quite likely, she herself is not conscious of the real reason (Binder, 2004).

To learn about the content of a patient's unconscious mind, the psychodynamic therapist must, in detective-like fashion, analyze clues found in the patients' speech

10
According to psychodynamic therapists, what are the underlying sources of mental disorders?

11
According to psychodynamic therapists, what is the relationship between *symptoms* and *disorders*?

and other forms of observable behavior. This is where Freud's term *psychoanalysis* comes from. The symptoms that brought the patient in for help, and the unique ways those symptoms are manifested, are one source of clues that the therapist considers. What other clues might be useful? Since the conscious mind usually attempts to act in ways that are consistent with conventional logic, Freud reasoned that the elements of thought and behavior that are *least* logical would provide the most useful clues. They would represent elements of the unconscious mind that leaked out relatively unmodified by consciousness. This insight led Freud to suggest that the most useful clues to a patient's unconscious motives and beliefs are found in the patients' free associations, dreams, and slips of the tongue or behavioral errors. These sources of clues are still widely used by psychodynamic therapists.

12

How do psychodynamic therapists use patients' free associations, dreams, and "mistakes" as routes to learn about their unconscious minds?



Free Associations as Clues to the Unconscious The technique of *free association* is one in which the patient is encouraged to sit back (or, in traditional psychoanalysis, to lie down on a couch), relax, free his or her mind, refrain from trying to be logical or "correct," and report every image or idea that enters his or her awareness, usually in response to some word or picture that the therapist provides as an initial stimulus. You might try this exercise yourself: Relax, free your mind from

what you have just been reading or thinking about, and write down the words or ideas that come immediately to your mind in response to each of the following: *liquid, horse, soft, potato*. Now, when you examine your set of responses to these words, do they make any sense that was not apparent when you produced them? Do you believe that they give you any clues to your unconscious mind?

Dreams as Clues to the Unconscious Freud believed that dreams are the purest exercises of free association, so he asked patients to try to remember their dreams, or to write them down upon awakening, and to describe them to him. During sleep, conventional logic is largely absent, and the forces that normally hold down unconscious ideas are weakened. Still, even in dreams, according to Freud

and other psychodynamic therapists, the unconscious is partially disguised. Freud distinguished the underlying, unconscious meaning of the dream (the *latent content*) from the dream as it is consciously experienced and remembered by the dreamer (the *manifest content*). The analyst's task in interpreting a dream is the same as that in interpreting any other form of free association: to see through the disguises and uncover the latent content from the manifest content.

Psychodynamic therapists hold that the disguises in dreams come in various forms. Some are unique to a particular person, but some are common to many people. Freud himself looked especially for sexual themes in dreams, and the common disguises that he wrote about have become known as *Freudian symbols*, some of which he described as follows (from Freud, 1900/1953):

The Emperor and Empress (or King and Queen) as a rule really represent the dreamer's parents; and a Prince or Princess represents the dreamer himself or herself. . . . All elongated objects, such as sticks, tree-trunks, and umbrellas (the opening of these last being comparable to an erection) may stand for the male organ—as well as all long, sharp weapons, such as knives, daggers, and pikes. . . . Boxes, cases, chests, cupboards, and ovens represent the uterus, and also hollow objects, ships, and vessels of all kinds. Rooms in dreams are usually women; if the various ways in and out of them are represented, this interpretation is scarcely open to doubt. In this connection interest in whether the room is open or locked is easily intelligible. There is no need to name explicitly the key that unlocks the room.

Mistakes and Slips of the Tongue as Clues to the Unconscious Still other clues to the unconscious mind, used by psychodynamic therapists, are found in the mistakes that patients make in their speech and actions. In Freud's rather extreme view, mistakes are never simply random accidents but are always expressions of un-

Edmund Engelman



The original therapeutic couch This photograph of Freud's consulting room shows the couch on which patients reclined while he sat, out of their line of sight, listening to their free associations. Contemporary psychodynamic therapists have generally abandoned the couch in favor of a more egalitarian face-to-face encounter.

Freudian symbols in a work of art painted four centuries before Freud This is a detail from *The Garden of Earthly Delights*, painted in the late fifteenth century by the Dutch artist Hieronymus Bosch. It is believed to represent Bosch's conception of decadence or Hell. Notice the dreamlike (nightmarish) quality and the numerous Freudian symbols—the protruding, phallic-like objects and the womb-like enclosures.



Hieronymus Bosch (1450–1516), *The Garden of Earthly Delights* (detail of Hell). Prado / Art Resource

conscious wishes or conflicts. In one of his most popular books, *The Psychopathology of Everyday Life*, Freud (1901/1990) supported this claim with numerous, sometimes amusing examples of such errors, along with his interpretations. For example, he reported an incident in which a young woman, complaining about the disadvantages of being a woman, stated, “A woman must be pretty if she is to please the men. A man is much better off. As long as he has his *five straight limbs* he needs no more.” According to Freud, this slip involved a fusion of two separate clichés, *four straight limbs* and *five senses*, which would not have occurred had it not expressed an idea that was on the woman's mind (either unconscious or conscious mind) that she consciously would have preferred to conceal. In another context the same statement could have been a deliberate, slightly off-color joke; but Freud claims that in this case it was an honest slip of the tongue, as evidenced by the woman's great embarrassment upon realizing what she had said.

The Roles of Resistance and Transference in the Therapeutic Process

Psychodynamic therapists, from Freud's time through today, have found that patients often resist the therapist's attempt to bring their unconscious memories or wishes into consciousness (Thompson & Cotlove, 2005). The **resistance** may manifest itself in such forms as refusing to talk about certain topics, “forgetting” to come to therapy sessions, or arguing incessantly in a way that subverts the therapeutic process. Freud assumed that resistance stems from the general defensive processes by which people protect themselves from becoming conscious of anxiety-provoking thoughts. Resistance provides clues that therapy is going in the right direction, toward critical unconscious material; but it can also slow down the course of therapy or even bring it to a halt. To avoid triggering too much resistance, the therapist must present interpretations gradually, when the patient is ready to accept them.

Freud also observed that patients often express strong emotional feelings—sometimes love, sometimes anger—toward the therapist. Freud believed that the true object of such feelings is usually not the therapist but some other significant person in the patient's life whom the therapist symbolizes. Thus, **transference** is the phenomenon by which the patient's unconscious feelings about a significant person in his or her life are experienced consciously as feelings about the therapist. Psychodynamic therapists to this day consider transference to be especially useful in psychotherapy because it provides an opportunity for the patient to become aware of his or her strong emotions (Thompson & Cotlove, 2005). With help from the analyst, the patient can gradually become aware of the origin of those feelings and their true target.

The Relationship Between Insight and Cure

Psychodynamic therapy is essentially a process in which the analyst makes inferences about the patient's unconscious conflicts and helps the patient become

>>> 13
In psychodynamic therapy, how do resistance and transference contribute to the therapeutic process?

>>> 14
According to psychodynamic theories, how do insights into the patient's unconscious conflicts bring about a cure?

Play therapy In this extension of psychodynamic therapy, children use dolls and other toys to express their feelings. The therapist uses the child's actions as clues for understanding the child's feelings and motives. If the therapist joins in the play, as in this example, he must be careful to play only in a way that is responsive to and does not direct the child's actions.



Michael Heron / Woodfin Camp

conscious of those conflicts. How does such consciousness help? According to psychodynamic therapists, from Freud on, it helps because, once conscious, the conflicting beliefs and wishes can be experienced directly and acted upon or, if they are unrealistic, can be modified by the conscious mind into healthier, more appropriate beliefs and pursuits (Thompson & Cotlove, 2005). At the same time, the patient is freed of the defenses that had kept that material repressed and has more psychic energy for other activities. But for all this to happen, the patient must truly accept the insights, viscerally as well as intellectually. The therapist cannot just tell the patient about the unconscious conflicts but must lead the patient gradually to actually experience the emotions and to arrive at the insights himself or herself.

A Case Example: Freud's Analysis of "the Rat Man"

All the just-described ideas about psychoanalysis can be made more vivid with a case example. Here, in much abbreviated form, is one of Freud's most famous and illustrative cases, that of a 29-year-old man referred to in the case history as "the Rat Man." This sketch necessarily leaves out many of the details and much of the subtlety of Freud's analysis, but it is useful nevertheless. You can decide for yourself from the sketch or, better yet, from Freud's published report of the case (Freud, 1909/1963) whether or not Freud's interpretation seems plausible.

The patient—I'll refer to him as RM—came to Freud complaining of various fears, obsessions, and compulsions that had begun 6 years earlier and had prevented him from completing his university studies and going on to a career. One of his most revealing symptoms—which gave the case its name—was an obsessive, extremely disturbing fantasy of a horrible torture applied both to RM's father and to the woman RM had been courting for the past 9 years, in which a pot of hungry rats strapped against the victims' buttocks would chew their way out through the only available opening. The fact that RM's father and beloved were both victims in this fantasy suggested to Freud that RM's problems lay in some conflict that had to do with both of these people and that a great deal of anger was associated with that conflict.

15

How did Freud's analysis explain the Rat Man's symptoms? How did free association and transference contribute to Freud's analysis?

Freud's Analysis of the Immediate Cause of RM's Symptoms

Through dialogue in therapy, Freud learned that RM had, sometime before his symptoms had begun, expressed a desire to marry the woman he was courting, but his father had opposed this wish. This discovery led Freud to suppose that the man's conflict between his desire to marry his beloved and his desire to please his father was the immediate, precipitating cause of his symptoms. Unable to resolve this conflict consciously, he had allowed his unconscious mind to resolve it by making him too psychologically ill to complete his studies and start a career, which

he viewed as a prerequisite for marriage. RM's psychological symptoms, according to Freud, protected him from having to choose between his own wish to marry the woman he loved and his father's wish that he not.

The relation between the rat fantasy and RM's father's opposition to the marriage was confirmed, in Freud's mind, by some of RM's free associations. When Freud asked him to free-associate to the concept *rats* (*ratten* in German), RM immediately came up with *rates* (*raten* in German), meaning "installments" or "money." RM had previously mentioned that the woman had little money and that his father had wanted him to marry a certain wealthy relative.

Freud's Analysis of the Deeper, Childhood Cause of RM's Symptoms

Why was RM so reluctant to defy his father and proceed with a plan for marriage? By the time RM had started therapy, his father had already been dead for several years, so there was no realistic barrier to marriage. Freud guessed that the answer to this problem must lie in RM's early-childhood relationship with his father.

In one therapy session, RM described his mother's account of an event that took place when he was about 4 years old. RM's father had begun to beat him for having bitten his nurse, and RM had responded with such a torrent of angry words that his father, shaken, stopped the beating and never struck him again. In Freud's analysis, this incident had great significance. Freud assumed that biting the nurse was a sexual act for the boy and that the beating consolidated in RM a strong fear of his father's reactions to his sexual wishes—a fear that lasted to this day and was preventing him from marrying the woman he loved.

Moreover, according to Freud's analysis, RM's apparent power over his father at the age of 4—the fact that RM's anger had made his father stop the beating—helped stamp into RM the unconscious fear that he could kill his father through anger. From that time until his psychoanalysis, RM had never consciously experienced anger toward his father, and even now, with his father dead, he feared unconsciously that he could kill his father again if he expressed his anger. In addition, Freud saw direct symbolic links between this early childhood incident and the rat-torture obsession: Rats, biting into and destroying his woman friend and his father, symbolized another small, biting beast—RM himself, who had once bitten first his beloved nurse and then, with angry words, his father.

Transference also entered into Freud's analysis. At one point in therapy, RM reported a fantasy in which it was Freud, not RM's father, who was trying to control whom he married. At another point—which Freud regarded as the major breakthrough—RM became irrationally angry at Freud, jumping up from the couch and shouting abusive words at him. Reflecting on this incident later, RM recalled that his sudden anger was accompanied by a fear that Freud would beat him and that he had jumped up to defend himself. By this time in the analysis, it took little convincing for RM to share Freud's view that this was a reenactment of the incident in which his father had beaten him and he had responded so angrily.

According to Freud, this transference experience helped RM overcome his resistance to the idea that his problems arose from deep-seated fear and anger that he held toward his father. RM's conscious acceptance of those feelings led, according to Freud, to recovery from his symptoms. In a sad footnote to the case, Freud (1923/1963) added that after a short period as a healthy man, RM was killed as an officer in World War I.

Principles of Humanistic Therapy

Two fundamental psychological ideas that have been emphasized repeatedly in previous chapters of this book are these: (1) People have the capacity to make adaptive choices regarding their own behavior—choices that promote their survival and well-being. (2) In order to feel good about themselves and to feel motivated to move

16

What is the primary goal of humanistic therapy? How does that goal relate to existentialist philosophy?



Michael Rougier / LIFE Magazine, © Time Warner, Inc

Carl Rogers The inventor of client-centered therapy (second from right) was a charismatic individual who personally embodied the empathy and genuineness that are the essence of his method of therapy.

forward in life, people need to feel accepted and approved of by others. No approach to psychotherapy takes these two ideas more seriously than does the humanistic approach.

<<< As described in Chapter 15, the humanistic view of the person emphasizes the inner potential for positive growth—the so-called *actualizing potential*. For this potential to exert its effects, people must be conscious of their feelings and desires, rather than denying or distorting them. Such denial and distortion occur when

people perceive that others who are important to them consistently disapprove of their feelings and desires. The goal of **humanistic therapy** is to help people regain awareness of their own desires and control of their own lives.

The humanistic movement in psychology originated in the mid-twentieth century. It grew partly out of *existentialist philosophy*, which focuses on the idea that human beings create their own life meanings (Hazler & Barwick, 2001). From an existentialist perspective, each person must decide for himself or herself what is true and worthwhile in order to live a full, meaningful life; meaning and purpose cannot be thrust upon a person from the outside. The primary founder of humanistic psychotherapy was Carl Rogers, whose theory of personality was introduced in Chapter 15. Rogers (1951) called his therapeutic approach *client-centered therapy*, a term deliberately chosen

to distinguish it from psychodynamic therapies. Client-centered therapy focuses on the abilities and insights of the *client* rather than those of the therapist. Practitioners of this approach today, however, more often refer to it as **person-centered therapy**, because they see the therapeutic process as involving a relationship between two unique persons—the client and the therapist (Mearns & Thorne, 2000; Wilkins, 2003). The therapist must attend to his or her own thoughts and feelings, as well as to those of the client, in order to respond in a supportive yet honest way to the client.

The core principles of person-centered therapy, elaborated upon below, all have to do with the relationship between the therapist and the client. The therapist lets the client take the lead in therapy, strives to understand and empathize with the client, and strives to think positively and genuinely of the client as a competent, valuable person. Through these means, the therapist tries to help the client regain the self-understanding and confidence necessary to control his or her own life.

17

How do the therapist's non-directive approach, empathic listening, and genuine positive regard contribute to the client's recovery, according to humanistic therapists?

<<<

Allowing the Client to Take the Lead

Humanistic psychotherapists usually refer to those who come to them for help as *clients*, not as patients, because the latter term implies passivity and lack of ability. In humanistic therapy, at least in theory, the client is in charge; it is the therapist who must be “patient” enough to follow the client's lead. The therapist may take steps to encourage the client to start talking, to tell his or her story, but once the client is started it is the therapist's task to understand the story, not to direct it or interpret it in ways that the client does not. While psychodynamic theorists may strain their patients' credulity with far-reaching interpretations, humanistic therapists more often just repeat, in their own words, what the client said as a way of checking to be sure that they understood correctly.

Listening Carefully and Empathetically

The humanistic therapist tries to provide a context within which clients can become aware of and accept their own feelings and learn to trust their own decision-making abilities. To do so, the therapist must first and foremost truly *listen* to the client in an empathetic way (Hazler & Barwick, 2001; Rogers, 1951). *Empathy* here refers to the therapist's attempt to comprehend what the client is saying or feeling at any given moment from the client's point of view rather than as an outsider. As

part of the attempt to achieve and manifest such empathy, the therapist frequently reflects back the ideas and feelings that the client expresses. A typical exchange might go like this:

Client: *My mother is a mean, horrible witch!*

Therapist: *I guess you're feeling a lot of anger toward your mother right now.*

To an outsider, the therapist's response here might seem silly, a statement of the obvious. But to the client and therapist fully engaged in the process, the response serves several purposes. First, it shows the client that the therapist is listening and trying to understand. Second, it distills and reflects back to the client the feeling that seems to lie behind the client's words—a feeling of which the client may or may not have been fully aware at the moment. Third, it offers the client a chance to correct the therapist's understanding. By clarifying things to the therapist, the client clarifies them to himself or herself.

Providing Unconditional but Genuine Positive Regard

Unconditional positive regard implies a belief on the therapist's part that the client is worthy and capable even when the client may not feel or act that way. By expressing positive feelings about the client regardless of what the client says or does, the therapist creates a safe, nonjudgmental environment for the client to explore and express all of his or her thoughts and feelings. Through experiencing the therapist's positive regard, clients begin to feel more positive about themselves, an essential step if they are going to take charge of their lives. Unconditional positive regard does not imply agreement with everything the client says or approval of everything the client does, but it does imply faith in the client's underlying capacity to make appropriate decisions. Consider the following hypothetical exchange:

Client: *Last semester in college I cheated on every test.*

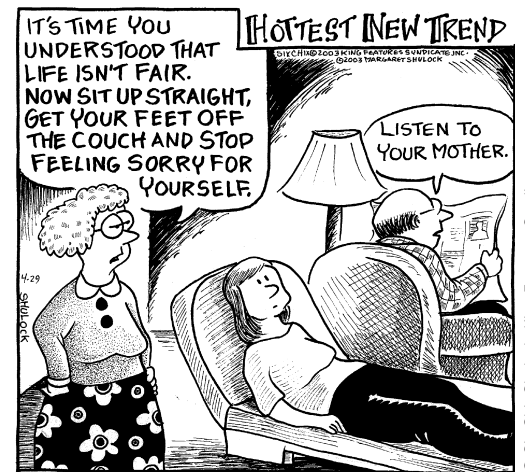
Therapist: *I guess what you're saying is that last semester you did something against your values.*

Notice that the therapist has said something positive about the client in relation to the misdeed without condoning the misdeed. The shift in focus from the negative act to the client's positive values affirms the client's inner worth and potential ability to make constructive decisions.

Humanistic therapists must work hard to be both positive and *genuine* in the feedback they supply to clients. They contend that it is impossible to fake empathy and positive regard, that the therapist must really experience them. If the therapist's words and true feelings don't match, the words will not be believable to the client. The capacity for genuine empathy and positive regard toward all clients might seem to be a rare quality, but Rogers suggests that it can be cultivated by deliberately trying to see things as the client sees them.

A Case Example: Rogers's Meetings with a Silent and Bitter Young Man

Rogers's greatest challenge as a therapist came when, as part of a research study, he conducted client-centered therapy with severely withdrawn inpatients in a state mental hospital (Rogers, 1967). One of these individuals was Jim, a 28-year-old man who had been hospitalized for 2 years with a diagnosis of schizophrenia. Prior to the two sessions excerpted below, Rogers had met with Jim once a week for months, and most of that time had been spent in silence. According to Rogers, these two sessions marked a turning point, because for the first time Jim began to express how he felt. As you read the exchange, notice the qualities of patience, empathy, and genuine positive regard reflected in Rogers's words.



"Mom and Pop Psychology."
It'll make you feel like a kid again.

© Six Chix-Shulock. King Features Syndicate

>>> **18**
How did Rogers use empathy, unconditional positive regard, and genuineness to help a man who had been diagnosed with schizophrenia?

[At about 40 minutes into the session, marked mostly by silence.]

Jim: *I just ain't no good to nobody, never was and never will be.*

Rogers: *Feeling that now, hm? . . . Those really are lousy feelings. Just feel that you're no good at all, hm?*

Jim: *Yeah.* [Muttering in a low, discouraged voice.] *That's what this guy I went to town with just the other day told me.*

Rogers: *This guy that you went to town with really told you that you were no good? Is that what you're saying? Did I get that right?*

Jim: *M-hm.*

Rogers: *I guess the meaning of that, if I get it right, is that here's somebody that meant something to you and what does he think of you? Why, he's told you that he thinks you're no good at all. And that just really knocks the props out from under you.* [Jim weeps quietly.] *It just brings the tears.*

[Silence of 20 seconds.]

Jim: [Defiantly.] *I don't care though.*

Rogers: *You tell yourself you don't care at all, but somehow I guess some part of you cares because some part of you weeps over it.*

[This is followed by more silence, more weeping, and more reflection by Rogers on how it must hurt to feel that someone thinks you're no good. The next session also begins with lots of silence, punctuated by bitter statements in which Jim says that he wants to die. We pick up the conversation as Rogers breaks a long silence by asking whether Jim's wish to die is related to the comment discussed in the previous session.]

Rogers: *Can't help but wonder whether it's still true that some things this friend said to you—are those still part of the thing that makes you feel so awful?*

Jim: *In general, yes.*

Rogers: *M-hm.*

[Silence of 47 seconds, interrupted by another comment from Rogers. Then:]

Jim: *I ain't no good to nobody, or I ain't no good for nothin', so what's the use of living?*

Rogers: *M-hm. I guess a part of that is—here I'm kind of guessing and you can set me straight, I guess a part of that is that you felt, "I tried to be good for something as far as he was concerned. I really tried. And now—If I'm no good to him, if he feels I'm no good, then that proves I'm just no good to anybody." Is that, uh—anywhere near it?*

Jim: *Oh, well, other people have told me that, too.*

Rogers: *Yeah, m-hm. I see. So you feel if, if you go by what others—what several others have said, then, you are no good. No good to anybody.*

[This is followed by more silence, interrupted by a few comments along the same track. Then:]

Jim: [Muttering in discouraged tone.] *That's why I want to go, 'cause I don't care what happens.*

Rogers: *M-hm, m-hm . . . You don't care what happens. And I guess I'd just like to say—I care about you. And, I care what happens.*

[Silence of 30 seconds, and then Jim bursts into tears.]

Rogers: [Tenderly.] *Somehow that just—makes the feelings pour out.* [Silence of 35 seconds.] *And you just weep and weep. And feel so badly.*

Commenting on this transaction, Rogers (1967) wrote:

Jim Brown, who sees himself as stubborn, bitter, mistreated, worthless, useless, hopeless, unloved, unlovable, experiences my caring. In that moment his defensive shell cracks wide open, and can never be quite the same. When someone cares for him, and when he feels and experiences this caring, he becomes a softer person whose years of stored up hurt come pouring out in anguished sobs. He is not the shell of hardness and bitterness, the stranger to tenderness. He is a person hurt beyond words, and aching for the love and caring which alone can make him human. This is evident in his sobs. It is evident too in his returning to my office [shortly after the session], partly for a cigarette, partly to say spontaneously that he will return.

After this session, according to Rogers, Jim gradually became more open, spontaneous, and optimistic at their meetings. After several months, he was able to leave the hospital and support himself with a job. Eight years later, on his own initiative, Jim called Rogers to tell him that he was still employed, had friends, was content with life, and that his feelings toward Rogers were still important, even though they had not been in touch during all that time (Meador & Rogers, 1973).

SECTION REVIEW

Talk, reflection, learning, and practice are the tools psychotherapists use to help people.

Principles of Psychodynamic Therapies

- An assumption of this approach is that mental disorders arise from unresolved mental conflicts that, though unconscious, strongly influence conscious thought and action.
- The psychodynamic therapist's job is to identify the unconscious conflict from observable clues, such as dreams, free association, mistakes, and slips of the tongue. The goal is to help the patient become aware of the conflicting beliefs and wishes and thus be able to deal with them.
- The patient's resistance to the uncovering of unconscious material suggests that the therapist is on the right track. And transference of feelings about significant persons in the patient's life to the therapist helps the patient become aware of these strong emotions.
- The case of "the Rat Man" illustrates all the preceding principles, as well as Freud's belief that problematic conflicts generally originate in early childhood and involve sexual and aggressive wishes.

Principles of Humanistic Therapies

- Humanistic therapists strive to help clients accept their own feelings and desires, a prerequisite for self-actualization (positive, self-directed psychological growth).
- Clients may fail to accept (instead, denying and distorting) their feelings and desires because they perceive that these are disapproved of by valued others.
- The humanistic therapist lets clients take the lead, listens carefully and empathetically, and provides unconditional but genuine positive regard.
- A case example shows Carl Rogers, the founder of this approach, helping a client called Jim emerge from silence, pain, and bitterness to live a satisfying life.

Psychotherapy II:
Cognitive and Behavioral Therapies

Cognitive and behavioral therapies are in many ways similar. Whereas psychodynamic and humanistic therapies tend to be holistic in their approach, focusing on the whole person, cognitive and behavioral therapies typically focus more directly and narrowly on the specific symptoms and problems that the client presents. Cognitive and behavioral therapists both are also very much concerned with data; they use objective measures to assess whether or not the treatment given is helping the client overcome the problem that is being treated. Many therapists today combine cognitive and behavioral methods, in what they call **cognitive-behavioral therapy**. Here, however, we will discuss the two separately, as they involve distinct assumptions and methods.

>>> 19
In what ways are cognitive and behavioral therapies similar to each other?

Principles of Cognitive Therapy

Cognitive therapy arose in the 1960s as part of a shift in the overall field of psychology toward greater focus on the roles of thoughts, beliefs, and attitudes in controlling

>>> 20
According to cognitive therapists, what is the source of clients' behavioral and emotional problems?



Albert Ellis

Founder of rational-emotive therapy Albert Ellis is a colorful individual who, in his practice, took a highly assertive role with clients. He would refer to their maladaptive, irrational thoughts with such terms as “musturbation” and “awfulizing.”

21

How does Ellis explain people's negative emotions in terms of their irrational beliefs?

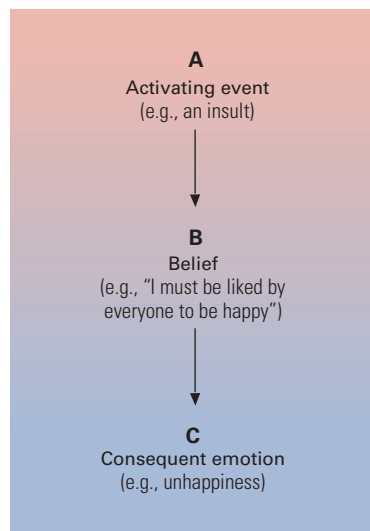


FIGURE 17.3 The ABCs of Ellis's theory Ellis and other cognitive therapists contend that our emotional feelings stem not directly from events that happen to us, but from our interpretation of those events. By changing our beliefs—the cognitions we use for interpreting what happens—we can alter our emotional reactions.

behavior. In previous chapters (especially Chapter 13) you have read of evidence that people's beliefs and ingrained, habitual ways of thinking affect their behavior and their emotions. Cognitive therapy begins with the assumption that people disturb themselves through their own, often illogical beliefs and thoughts. Maladaptive beliefs and thoughts make reality seem worse than it is and in that way produce anxiety or depression. The goal of cognitive therapy is to identify maladaptive ways of thinking and replace them with adaptive ways that provide a base for more effective coping with the real world. Unlike psychoanalysis, cognitive therapy generally centers on conscious thoughts, though such thoughts may be so ingrained and automatic that they occur with little conscious effort.

The two best-known pioneers of cognitive therapy are Albert Ellis and Aaron Beck. Both started their careers in the 1950s as psychodynamic therapists but became disenchanted with that approach (Bowles, 2004; Ellis, 1986). Each was impressed by the observation that clients seemed to distort, mentally, their experiences in ways that quite directly produced their problems or symptoms. Ellis refers to his specific brand of cognitive therapy as *rational-emotive therapy*, highlighting his belief that rational thought will improve clients' emotions.

Here we will look briefly at three general principles of cognitive therapy: the identification and correction of maladaptive beliefs and thoughts; the establishment of clear-cut goals and steps for achieving them; and the changing role of the therapist, from teacher early in the process to consultant later on.

Identifying and Correcting Maladaptive Beliefs and Habits of Thought

<<< Different cognitive therapists take different approaches to pointing out their patients' irrational ways of thinking. Beck, as you will see later (in the case example), tends to take a Socratic approach, in which, through questioning, he gets the patient himself or herself to discover and correct the irrational thought. Ellis, in contrast, takes a more direct and blunt approach. He gives humorous names to certain styles of irrational thinking. Thus, *musturbation* is the irrational belief that one *must* have some particular thing or *must* act in some particular way in order to be happy or worthwhile. If a client says, “I must get all A's this semester in college,” Ellis might respond, “You're musturbating again.” *Awfulizing*, in Ellis's vocabulary, is the mental exaggeration of setbacks or inconveniences. A client who feels bad for a whole week because of a dent in her new car might be told, “Stop awfulizing.”

The following dialogue between Ellis (1962) and a client not only illustrates Ellis's blunt style but also makes explicit his theory of the relationship between thoughts and emotions—a theory shared by most cognitive therapists. The client began the session by complaining that he was unhappy because some men with whom he played golf didn't like him. Ellis, in response, claimed that the client's reasoning was illogical—the men's not liking him couldn't make him unhappy.

Client: Well, why was I unhappy then?

Ellis: It's very simple—as simple as A, B, C, I might say. A in this case is the fact that these men didn't like you. Let's assume that you observed their attitude correctly and were not merely imagining they didn't like you.

Client: I assure you that they didn't. I could see that very clearly.

Ellis: Very well, let's assume they didn't like you and call that A. Now, C is your unhappiness—which we'll definitely have to assume is a fact, since you felt it.

Client: Damn right I did!

Ellis: All right, then: A is the fact that the men didn't like you, and C is your unhappiness. You see A and C and you assume that A, their not liking you, caused your unhappiness. But it didn't.

Client: It didn't? What did, then?

Ellis: *B did.*

Client: *What's B?*

Ellis: *B is what you said to yourself while you were playing golf with those men.*

Client: *What I said to myself? But I didn't say anything.*

Ellis: *You did. You couldn't possibly be unhappy if you didn't. The only thing that could possibly make you unhappy that occurs from without is a brick falling on your head, or some such equivalent. But no brick fell. Obviously, therefore, you must have told yourself something to make you unhappy.*

In this dialogue, Ellis invokes his famous *ABC theory of emotions*: *A* is the *activating event* in the environment, *B* is the *belief* that is triggered in the client's mind when the event occurs, and *C* is the emotional *consequence* of that belief (illustrated in Figure 17.3). Therapy proceeds by changing *B*, the belief. In this particular example, the man suffers because he believes irrationally that he must be liked by everyone (an example of *musturbation*), so if someone doesn't like him, he is unhappy. The first step will be to convince the man that it is irrational to expect everyone to like him and that there is little or no harm in not being liked by everyone.

Establishing Clear-Cut Goals and Steps for Achieving Them

Once a client admits to the irrational and self-injurious nature of some belief or habit of thought, the next step is to help the client get rid of that belief or habit and replace it with a more rational and adaptive way of thinking. That takes hard work. Long-held beliefs and thoughts do not simply disappear once they are recognized as irrational. They occur automatically unless they are actively resisted.

To help clients overcome their self-injurious ways of thinking, cognitive therapists often assign homework. For example, clients might be asked to keep a diary, or to fill out a form every day such as that depicted in Figure 17.4 (on page 645), in which they record the negative emotions they felt that day, describe the situations and automatic thoughts that accompanied those emotions, and describe a rational alternative thought that might make them feel less upset. Such exercises help train clients to become more aware of their automatic thoughts and, through awareness, change them. The diaries or charts also become a record of progress, by which the therapist and client can see if positive ways of thinking are increasing over time and negative emotions are decreasing.

Moving from a Teaching Role to a Consulting Role with the Client

Cognitive therapists, unlike humanistic therapists, are quite directive in their approach. At least at the beginning, the relationship between cognitive therapist and client is fundamentally like that of a teacher and student. The therapist helps the client identify a set of goals, develops a curriculum for achieving those goals, assigns homework, and assesses the client's progress using the most objective measures available. With time, however, as the client becomes increasingly expert in spotting and correcting his or her own maladaptive thoughts, the client becomes increasingly self-directive in the therapy, and the therapist begins to act more like a consultant and less like a teacher (Hazler & Barwick, 2001). Eventually, the client may meet with the therapist just occasionally to describe continued progress and to ask for advice when needed. When such advice is no longer needed, the therapy has achieved its goals and is over.

Although cognitive therapists are directive with clients, most of them acknowledge the value of some of the other tenets of humanistic therapy, particularly the value of maintaining a warm, genuine, and empathic relationship with their clients (Beck & others, 1993). In fact, increasingly, psychotherapists of all theoretical persuasions are acknowledging that empathy and understanding of the patient are an essential part of the therapeutic process (Hazler & Barwick, 2001).

>>> **22**
What is the purpose of homework in cognitive therapy?

>>> **23**
In what sense is a cognitive therapist first a teacher and then a consultant?

A Case Example: Beck's Cognitive Treatment of a Depressed Young Woman

Aaron Beck, who at this writing has been applying and researching cognitive therapy for more than 40 years, is by far the most influential cognitive therapist today. His earliest and best-known work was with depressed patients, but he has also developed cognitive therapy methods for treating anxiety, uncontrollable anger, and, most recently, schizophrenia (Bowles, 2004). In the dialogue that follows (from Beck & Young, 1985) we see him at work with a client referred to as Irene—a 29-year-old woman with two young children who was diagnosed with major depression.

24

How does Beck's treatment of a depressed woman illustrate his approach to identifying and correcting maladaptive, automatic thoughts?

<<< Irene had not been employed outside her home since marriage, and her husband, who had been in and out of drug-treatment centers, was also unemployed. She was socially isolated and felt that people looked down on her because of her poor control over her children and her husband's drug record. She was treated for three sessions by Beck and then was treated for a longer period by another cognitive therapist.

During the first session, Beck helped Irene to identify a number of her automatic negative beliefs, including: *things won't get better*; *nobody cares for me*; and *I am stupid*. By the end of the session, she accepted Beck's suggestion to try to invalidate the first of those thoughts by doing certain things for herself, before the next session, that might make life more fun. She agreed to take the children on an outing, visit her mother, go shopping, read a book, and find out about joining a tennis group—all things that she claimed she would like to do. Having completed that homework, she came to the second session feeling more hopeful. However, she began to feel depressed again when, during the session, she misunderstood a question that Beck asked her, which, she said, made her "look dumb." Beck responded with a questioning strategy that helped her to distinguish between the *fact* of what happened (not understanding a question) and her *belief* about it (looking dumb) and then to correct that belief:

Beck: OK, what is a rational answer [to why you didn't answer the question]? A realistic answer?

Irene: I didn't hear the question right; that is why I didn't answer it right.

Beck: OK, so that is the fact situation. And so, is the fact situation that you look dumb or you just didn't hear the question right?

Irene: I didn't hear the question right.

Beck: Or is it possible that I didn't say the question in such a way that it was clear?

Irene: Possible.

Beck: Very possible. I'm not perfect so it's very possible that I didn't express the question properly.

Irene: But instead of saying you made a mistake, I would still say I made a mistake.

Beck: We'll have to watch the video to see. Whichever. Does it mean if I didn't express the question, if I made the mistake, does it make me dumb?

Irene: No.

Beck: And if you made the mistake, does it make you dumb?

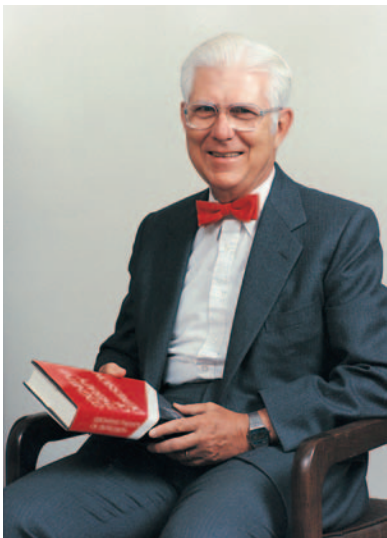
Irene: No, not really.

Beck: But you felt dumb?

Irene: But I did, yeah.

Beck: Do you still feel dumb?

Irene: No. Right now I feel glad. I'm feeling a little better that at least somebody is pointing all these things out to me because I have never seen this before. I never knew that I thought that I was that dumb.



Will McIntyre / Photo Researchers

A pioneer of cognitive therapy Aaron Beck's approach to therapy is gentler than Ellis's, but not less teacher-like. Beck typically leads clients, through a Socratic style of questioning, to discover the irrationality of their thoughts.

	SITUATION	EMOTION(S)	AUTOMATIC THOUGHT(S)	RATIONAL RESPONSE	OUTCOME
	Describe: 1. Actual event leading to unpleasant emotion, or 2. Stream of thoughts, daydream, or recollection, leading to unpleasant emotion.	1. Specify sad/anxious/angry, etc. 2. Rate degree of emotion, 1–100.	1. Write automatic thought(s) that preceded emotion(s). 2. Rate belief in automatic thought(s), 0–100%.	1. Write rational response to automatic thought(s). 2. Rate belief in rational response, 0–100%.	1. Rate belief in automatic thought(s), 0–100%. 2. Specify and rate subsequent emotions, 0–100.
DATE	7/15 Store clerk didn't smile at me when I paid for purchase.	Sad - 60 Anxious - 40	Nobody likes me - 70% I look awful - 80%	Maybe the clerk was having a bad day or maybe she never smiles at customers - 70%	1. 20% 30% 2. Pleasure - 25

Explanation: When you experience an unpleasant emotion, note the situation that seemed to stimulate the emotion. (If the emotion occurred while you were thinking, daydreaming, etc., please note this.) Then note the automatic thought associated with the emotion. Record the degree to which you believe this thought: 0% = not at all, 100% = completely. In rating degree of emotion: 1 = a trace, 100 = the most intense possible.

As homework between the second and third sessions, Beck gave Irene the assignment of catching, writing down, and correcting her own dysfunctional thoughts, using the form shown in Figure 17.4. Subsequent sessions were aimed at eradicating each of her depressive thoughts, one by one, and reinforcing the steps she was taking to improve her life. Progress was rapid. Irene felt increasingly better about herself. During the next several months, she joined a tennis league, got a job, took a college course in sociology, and left her husband after trying and failing to get him to develop a better attitude toward her or to join her in couple therapy. By this time, according to Beck, she was cured of her depression, had created for herself a healthy environment, and no longer needed therapy.

FIGURE 17.4 | Homework sheet for cognitive therapy The purpose of this homework is to enable clients to become aware of and correct the automatic thoughts that contribute to their emotional difficulties. (Adapted from Beck & Young, 1985.)

Principles of Behavior Therapy

If a cognitive therapist is a teacher, a behavior therapist is a trainer. While cognitive therapy deals with maladaptive habits of thought, behavior therapy deals directly with maladaptive behaviors. This approach is rooted in the research on basic learning processes initiated by such pioneers as Ivan Pavlov, John B. Watson, and B. F. Skinner (discussed in Chapter 4). Unlike all the other psychotherapy approaches we have discussed, behavior therapy is *not* fundamentally talk therapy. Rather, in behavior therapy clients are exposed by the therapist to new environmental conditions that are designed to retrain them, so that maladaptive habitual or reflexive ways of responding become extinguished and new, healthier habits and reflexes are conditioned.

In other regards, behavior therapy is much like cognitive therapy, and, as I noted earlier, the approaches are often combined, in cognitive-behavioral therapy. Behavior therapy, like cognitive therapy, is very much symptom-oriented and concerned with immediate, measurable results. Two of the most common types of

>>> 25
How is behavior therapy distinguished from cognitive therapy?

treatment in behavior therapy, discussed below, are *contingency management therapy* to modify habits and *exposure therapy* to overcome unwanted fears.

Contingency Management: Altering the Relationship Between Actions and Rewards

The basic principle of operant conditioning, discussed in Chapter 4, is that behavioral actions are reinforced by their consequences. People, like other animals, learn to behave in ways that bring desired consequences and to avoid behaving in ways that do not bring desired consequences. If a behavior therapist learns that an individual is behaving in ways that are harmful to himself or herself or to others, the first question the therapist might ask is this: What reward is this person getting for this behavior, which leads him or her to continue it? The next step, where possible, is to modify the behavior–reward contingency, so that rewards are provided for desired behaviors and not for undesired behaviors. The broad term for all therapy programs that alter the contingency between actions and rewards is **contingency management**.

26

How has contingency management been used to improve children's behavior, promote self-care in residents of mental institutions, and motivate abstinence in drug abusers?

<<< For example, if parents complain to a behavior therapist that their child is acting in aggressive and disruptive ways at home, the therapist might ask the parents to keep a record, for a week or more, of each instance of such misbehavior and of how they or others in the family responded. From that record, the therapist might learn that the child is gaining desired attention through misbehavior. The therapist might then work out a training program in which the parents agree to attend more to the child when he is behaving in desired ways and to ignore the child, or provide some clearly negative consequence (such as withdrawing some privilege), when he is behaving in undesired ways. This sort of behavioral work with families, aimed at altering the contingencies between actions and rewards at home, is referred to as *parent management training* (Kazdin, 2003).

Contingency management is also sometimes used, in programs referred to as *token economies*, in mental hospitals and group homes for people with schizophrenia or other serious mental disorders (Ayllon & Azrin, 1968; Dickerson & others, 2005). Patients may receive tokens for such activities as making their beds, helping out in the kitchen, or helping other patients in specific ways, and they can cash in the tokens for desired privileges, such as movies or treats at the hospital commissary. This technique not only helps build desired habits but also helps combat the lethargy, boredom, and dependence that are so common in mental institutions. In a number of controlled studies, patients in token economy programs improved more quickly, and were able to leave the institutions sooner, than otherwise similar patients who were not in such programs (Dickerson & others, 2005).

In recent years, contingency management has been instituted in many community drug rehabilitation programs. In these programs, patients who remain drug-free for specified periods of time, as measured by regular urine tests, receive vouchers that they can turn in for valued prizes. Such programs have proven quite successful in encouraging cocaine and heroin abusers to go many weeks without taking drugs, and in some cases these programs are less expensive than other, more standard treatments (Higgins & others, 2004; Petry & Martin, 2002).

Exposure Treatments for Unwanted Fears

27

What is the theoretical rationale for treating specific phobias by exposing clients to the feared objects or situations?

<<< Behavior therapy has proven especially successful in treating specific phobias, in which the person fears something well defined, such as high places or a particular type of animal (Emmelkamp, 2004). From a behavioral perspective, fear is a reflexive response, which through classical conditioning (discussed in Chapter 4) can come to be triggered by various non-dangerous as well as dangerous stimuli. An unconditioned stimulus for fear is one that elicits the response even if the individual has had no previous experience with the stimulus; a conditioned stimulus for fear is one that elicits the response only because the stimulus was previously paired with some fearful event in the person's experience. Opinions may differ as to whether a particular fear, such as a fear of snakes, is unconditioned or condi-

tioned (unlearned or learned), but in practice this does not matter because the treatment is the same in either case.

A characteristic of the fear reflex, whether conditioned or unconditioned, is that it declines and gradually disappears if the eliciting stimulus is presented many times or over a prolonged period in a context where no harm comes to the person. In the case of an unconditioned fear reflex—such as the startle response to a sudden noise—the decline is called *habituation*. In the case of a conditioned fear reflex, the decline that occurs when the conditioned stimulus is presented repeatedly without the unconditioned stimulus is called *extinction* (discussed in Chapter 4). For example, if a person fears all dogs because of once having been bitten, then prolonged exposure to various dogs (the conditioned stimuli) in the absence of being bitten (the unconditioned stimulus) will result in loss of the fear. Any treatment for an unwanted fear or phobia that involves exposure to the feared stimulus to habituate or extinguish the fear response is referred to as an **exposure treatment**. Behavior therapists have developed three different means to present feared stimuli to clients in such treatments (Krijn & others, 2004).

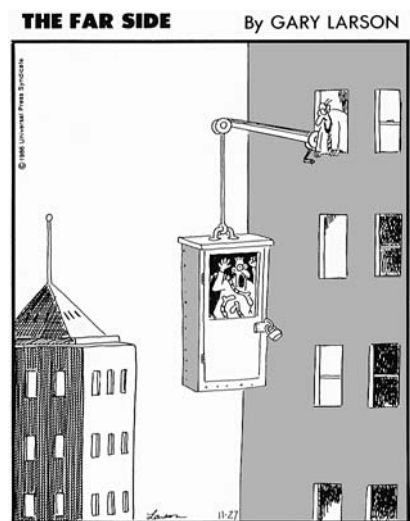
One means is *imaginative exposure*. A client in this form of treatment is instructed to imagine a particular, moderately fearful scene as vividly as possible, until it no longer seems frightening. Then the client is instructed to imagine a somewhat more fearful scene, until that no longer seems frightening. In this way the client gradually works up to the most feared scene. For example, a woman afraid of heights might be asked first to imagine that she is looking out a second-floor window, then a third-floor window, and so on, until she can imagine, without strong fear, that she is looking down from the top of a skyscraper. An assumption here is that the ability to remain calm while imagining the previously feared situation will generalize, so that the person will be able to remain calm in the actual situation. Research involving long-term follow-up has shown that imaginative exposure techniques can be quite effective in treating specific phobias (Zinbarg & others, 1992).

The most direct exposure technique is *in vivo exposure*, that is, real-life exposure. With this technique the client, usually accompanied by the therapist or by some other comforting and encouraging helper, must force himself or herself to confront the feared situation in reality. For example, to overcome a fear of flying, the client might, as a first assignment, go to the airport and watch planes take off and land until that can be done without fear. Then, through a special arrangement with an airline, the client might enter a stationary plane and sit in it for a while, until that can be done without fear. Finally, the client goes on an actual flight—perhaps just a short one the first time. Research suggests that *in vivo* exposure is generally more effective than imaginative exposure when both are possible (Emmelkamp, 2004). However, *in vivo* exposure is usually more time-consuming and expensive, and is often not practical because the feared situation is difficult to arrange for the purpose of therapy.

Today, with advanced computer technology, a third means of exposure is possible and is rapidly growing in use—*virtual reality exposure*. With this technique, the client wears a helmet that contains two goggle-sized TV screens, one in front of each eye. A slightly different image of the virtual environment is sent to each eye, such that the client sees that environment in three dimensions and experiences it as if he or she were truly immersed in it. The helmet contains sensors that track head movements, and these sensors control the display in the TV screens so that the view of the virtual environment changes, in appropriate ways, as the client's head moves. Thus the client can scan the virtual environment, seeing different parts of it, by rotating his or her head left and right or up and down. The client can also experience himself or herself as moving through the virtual environment, either through actually walking or through use of a joystick. At the same time, sound effects that add to the reality of the client's experience can be played through speakers in the helmet.

Virtual worlds have been developed for exposure treatments for many different phobias, including fear of heights, fear of flying, claustrophobia (fear of being in

28 What are three ways of exposing clients to feared objects or situations, and what are the advantages of each?



Professor Gallagher and his controversial technique of simultaneously confronting the fear of heights, snakes, and the dark.

"The Far Side" © 1986 Universal Press Syndicate. Reprinted with permission. All rights reserved.

© Mary Levin, U. W., with permission from Hunter Hoffman, U. W.



Virtual reality exposure treatment As treatment for her spider phobia, this woman wears a headset that allows her to walk through a virtual environment in which she confronts spiders, some of which are monstrosly large. The screen in the background depicts the scene that she is currently seeing, but her view is three-dimensional. The fuzzy object held in front of her will add a tactile component to her spider experience; she will be asked to touch it with her real hand as she touches the virtual spider with her virtual hand.

gradually through virtual experiences of going to an airport, entering a plane, flying, looking out the plane window, and feeling and hearing turbulence while flying (Botella & others, 2004). After the therapy, all of them were able to take an actual plane flight with relatively little fear.

A Case Example: Miss Muffet Overcomes Her Spider Phobia

Our case example of behavior therapy is that of a woman nicknamed Miss Muffet by her therapists (Hoffman, 2004). For 20 years prior to her behavioral treatment, Miss Muffet had suffered from an extraordinary fear of spiders, a fear that had led to a diagnosis not just of spider phobia but also obsessive-compulsive disorder. Her obsessive fear of spiders prompted her to engage in such compulsive rituals as regularly fumigating her car with smoke and pesticides, sealing her bedroom windows with duct tape every night, and sealing her clothes in plastic bags immediately after washing them. She scanned constantly for spiders wherever she went, and she avoided places where she had ever encountered spiders. By the time she came for behavior therapy her spider fear was so intense that she found it difficult to leave home at all.

29

How did virtual exposure treatment help Miss Muffet overcome her fear of spiders?

Miss Muffet's therapy, conducted by Hunter Hoffman (2004) and his colleagues, consisted of ten 1-hour sessions of virtual reality exposure, using the helmet system described earlier. In the first session she navigated through a virtual kitchen, where she encountered a huge virtual tarantula. She was asked to approach the huge spider as closely as she could, using a handheld joystick to move forward in the three-dimensional scene. In later sessions, as she became less fearful of approaching the virtual tarantula, she was instructed to reach out and "touch" it. To create the tactile effect of actually touching the spider, the therapist held out a fuzzy toy spider, which she felt with her real hand as she saw her virtual hand touch the virtual spider.

In other sessions, she encountered other spiders of various shapes and sizes. On one occasion she was instructed to pick up a virtual vase, out of which wriggled a virtual spider. On another, a virtual spider dropped slowly to the floor of the virtual kitchen, directly in front of her, as sound effects from the horror movie *Psycho* were played through her headphones. The goal of the therapy was to create spider experiences that were more frightening than any she would encounter in the real world, so as to convince her, at a gut emotional level, that encountering real spiders would not cause her to panic or go crazy. As a test, after the ten virtual reality sessions, Miss Muffet was asked to hold a live tarantula (a large, hairy, fearsome looking, but harmless spider), which she did with little fear, even while the creature crawled partway up her arm.

SECTION REVIEW

Cognitive and behavioral therapy—both problem-centered—are often combined.**Principles of Cognitive Therapy**

- Cognitive therapy is based on the idea that psychological distress results from illogical, maladaptive beliefs and thoughts.
- The cognitive therapist works to identify the maladaptive thoughts and beliefs, convince the client of their irrationality, and help the client eliminate them along with the unpleasant emotions they provoke.
- To help the client replace the old, habitual ways of thinking with more adaptive ways, cognitive therapists assign homework (e.g., writing down and correcting irrational thoughts). Progress is objectively measured. As the client can assume more of the responsibility, the therapist becomes less directive.
- A case example shows Aaron Beck's cognitive treatment of a depressed young woman who came to see how *her* automatic negative thoughts (e.g., "I am stupid") were affecting her.

Principles of Behavior Therapy

- The goals of behavior therapy are to extinguish maladaptive responses and to condition healthier responses by exposing clients to new environmental conditions.
- Contingency management programs (such as the token economies used in some mental institutions) are based on operant-conditioning principles. They modify behavior by modifying behavior–reward contingencies.
- Exposure treatments, based on classical conditioning, are used to habituate or extinguish reflexive fear responses in phobias. Three techniques are imaginative exposure, *in vivo* (real-life) exposure, and virtual reality exposure.
- In a case example, virtual reality exposure therapy helped "Miss Muffet" overcome a debilitating fear of spiders.

Evaluating Psychotherapies

You have just read about four major varieties of psychotherapy. Do they work? >>> That might seem like an odd question at this point. After all, didn't Freud cure the Rat Man; didn't Rogers help the silent young man; didn't Beck and his associates cure the depressed young woman; and didn't behavior therapy cure Miss Muffet? But case studies—even thousands of them—showing that people are better off at the end of therapy than at the beginning cannot tell us for sure that therapy works. Maybe those people would have improved anyway, without therapy.

An adage about the common cold goes like this: "Treat a cold with the latest remedy and you'll get rid of it in 7 days; leave it untreated and it'll hang on for a week." Maybe psychological problems or disorders are often like colds in this respect. Everyone has peaks and valleys in life, and people are most likely to start therapy while in one of the valleys (see Figure 17.5). Thus, even if therapy has no effect, most people will feel better at some time after entering it than they did when they began (a point that was made earlier in this chapter in the discussion of drug treatments). The natural tendency for both therapist and client is to attribute the improvement to the therapy, but that attribution may often be wrong.

The only way to know if psychotherapy really works is to perform controlled experiments, in which groups of people undergoing therapy are compared with otherwise-similar control groups who are not undergoing therapy.

30
Why must we rely on experiments rather than case studies to assess the effectiveness of psychotherapy?

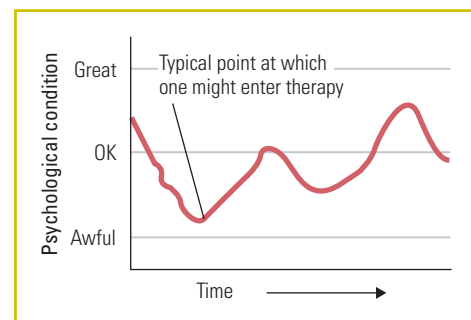


FIGURE 17.5 | The peaks and valleys of life If a person enters psychotherapy while in a valley, he or she is likely to feel better after a time even if the therapy is ineffective.

Is Psychotherapy Helpful, and Are Some Types of It More Helpful Than Others?**A Classic Example of a Therapy-Outcome Experiment**

One of the earliest well-controlled experiments on therapy outcome was conducted >>> at a psychiatric outpatient clinic in Philadelphia (Sloane & others, 1975). The subjects were 94 men and women, ages 18 to 45, who sought psychotherapy at the clinic. Most of these individuals suffered from diffuse anxiety, the type that today would probably be diagnosed as generalized anxiety disorder. Each was assigned by a random procedure to one of three groups. Members of one group received once-a-week

31
How did an experiment in Philadelphia demonstrate the effectiveness of behavior therapy and psychodynamic therapy?

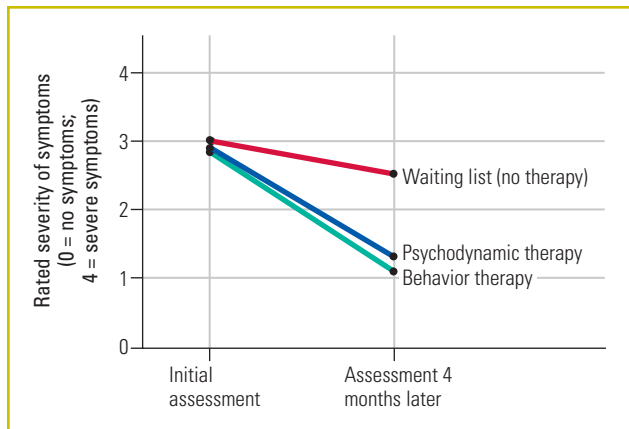


FIGURE 17.6 Results of the Philadelphia experiment Psychiatrists rated the severity of each subject's symptoms before and after a 4-month treatment period. As shown here, those in the two therapy groups improved more than did those who were placed on the waiting list. (Adapted from Sloane & others, 1975.)

Redrawn, with permission of the publishers, from *Psychotherapy versus behavior therapy*.

sessions of *behavior therapy* for 4 months (including such procedures as imaginative exposure to various feared situations and training in assertiveness) from one of three highly experienced behavior therapists. Members of the second group received the same amount of a type of *psychodynamic therapy* (including such procedures as probing into childhood memories, dream analysis, and interpretation of resistance) from one of three highly experienced psychodynamic therapists. The members of the third group, the *no-therapy* group, were placed on a waiting list and given no treatment during the 4-month period but were called periodically to let them know that they would eventually be accepted for therapy.

To measure treatment effectiveness, all subjects, including those in the no-therapy group, were assessed both before and after the 4-month period by psychiatrists who were not informed of the groups to which the subjects had been assigned. As illustrated in Figure 17.6, all three groups improved during the 4-month period, but the treatment groups improved significantly more than did the control group. Moreover, the two treatment groups did not differ significantly from each other in degree of improvement.

This early study anticipated the result of hundreds of other psychotherapy-outcome experiments that have followed it. Overall, such research provides compelling evidence that psychotherapy works, but provides little if any evidence that one variety of therapy is regularly better than any other standard variety (Lambert & Ogles, 2004).

Evidence That Psychotherapy Helps

32

What is the evidence that psychotherapy works and that for some disorders it works as well as, or better than, standard drug treatments?

<<< When hundreds of separate therapy-outcome experiments are combined, the results indicate, very clearly, that therapy helps. In fact, about 75 to 80 percent of people in the psychotherapy condition in such experiments improved more than did the average person in a non-therapy condition (Lambert & Ogles, 2004; Lipsey & Wilson, 1993; Smith & others, 1980). Experiments comparing the effectiveness of psychotherapy with standard drug therapies indicate, overall, that psychotherapy is at least as effective as drug therapy in treating depression and generalized anxiety disorder and is more effective than drug therapy in treating panic disorder (Gloaguen & others, 1998; Gould & others, 1995; Mitte, 2005).

A note of caution should be added, however, for interpreting such findings. Psychotherapy-outcome experiments are usually carried out at clinics associated with major research centers. The therapists are normally highly experienced and, since they know that their work is being evaluated, they are presumably functioning at maximal capacity. It is quite possible that the outcomes in these experiments are more positive than are those in the larger clinical environment, where therapists are not always so experienced or so motivated to succeed (Lambert & Ogles, 2004).

Evidence That No Type of Therapy Is Clearly Better, Overall, Than Other Standard Types

33

What general conclusions have been drawn from research comparing different forms of psychotherapy?

<<< There was a time when psychotherapists who adhered to one variety of therapy argued strongly that their variety was greatly superior to others. Such arguments have largely been quelled by the results of psychotherapy-outcome experiments. Taken as a whole, such experiments provide no convincing evidence that any of the major types of psychotherapy is superior to any other for treating the kinds of problems—such as depression and generalized anxiety—for which people most often seek treatment (Lambert & Ogles, 2004; Wampold, 2001). Most such experiments have pitted some form of cognitive or cognitive-behavioral therapy against some form of psychodynamic therapy. When the two forms of therapy are given in equal numbers of sessions, by therapists who are equally well trained and equally passionate about the value of the therapy they administer, the usual result is no

difference in outcome; people treated by either therapy get better at about the same rate. Relatively few studies have compared humanistic forms of therapy with others, but the results of those few suggest that humanistic therapy works as well as the others (Lambert & Ogles, 2004; Smith & others, 1980).

There are some exceptions to the generalization that all types of therapy are equally effective. Almost all experts agree that behavioral exposure treatment is the most effective treatment, by far, for specific phobias (Antony & Barlow, 2002). There is also reason to believe that cognitive and behavior therapies generally work best for clients who have quite specific problems—problems that can be zeroed in on by the specific techniques of those therapies—and that psychodynamic and humanistic therapies may be most effective for clients who have multiple or diffuse problems or problems related to their personalities (Lambert & Ogles, 2004; Weston & others, 2004).

One reason for the general equivalence in effectiveness of different therapies may be that they are not as different in practice as they are in theory. Analyses of recorded therapy sessions have revealed that good therapists, regardless of the type of therapy they claim to practice, often use methods that overlap with other types of therapy. For example, effective cognitive and behavior therapists, who are normally quite directive, tend to become non-directive—more like humanistic therapists—when working with clients who seem unmotivated to follow their lead (Barlow, 2004; Beutler & Harwood, 2000). Similarly, psychodynamic therapists often use cognitive methods (such as correcting maladaptive automatic thoughts), and cognitive therapists often use psychodynamic methods (such as asking clients to talk about childhood experiences), to degrees that seem to depend on the client's personality and the type of problem to be solved (Weston & others, 2004).

The Role of Common Factors in Therapy Outcome

Ostensibly different types of psychotherapy are similar to one another not just because the therapists borrow one another's methods but also because all or at least most therapies, regardless of theoretical orientation, share certain *common factors*. These factors may be more critical to the effectiveness of therapy than are the specific, theory-derived factors that differentiate therapies from one another (Lambert & Ogles, 2004; Wampold, 2001). Many common factors have been proposed, but I find it useful to distill them into three fundamental categories: support, hope, and motivation.

Support

Support includes acceptance, empathy, and encouragement. By devoting time to the client, listening warmly and respectfully, and not being shocked at the client's statements or actions, any good psychotherapist communicates the attitude that the client is a worthwhile human being. This support may directly enhance the client's self-esteem and indirectly lead to other improvements as well.

Many studies have demonstrated the value of such support. One long-term study, for example, revealed that psychodynamic therapists provide more support and less insight to their clients than would be expected from psychoanalytic theory (Wallerstein, 1989). The same study also showed that support, even without insight, seemed to produce stable therapeutic gains. More recently, a study of cognitive-behavioral therapy for depression revealed that the bond of rapport that clients felt with their therapists early in treatment was a good predictor of their subsequent improvement (Klein & others, 2003). In yet another study, college professors with no training in psychology or methods of therapy, but with reputations for good rapport with students, proved able to help depressed college students in twice-a-week therapy sessions as effectively as highly trained and experienced clinical psychologists (Strupp & Hadley, 1979).

Results such as these led one prominent psychodynamic therapist, Hans Strupp (1989), to the following conclusion: "The first and foremost task for the therapist is

>>> 34

How might support, hope, and motivation be provided within the framework of every standard type of psychotherapy? What evidence suggests that these common factors may be crucial to the therapy's effectiveness?

© The New Yorker Collection 2005 Tom Cheney
from cartoonbank.com. All rights reserved.



"So, would anyone in the group care to respond to what Clifford has just shared with us?"

to create an accepting and empathic context, which in itself has great therapeutic value because for many people it is a novel and deeply gratifying experience to be accepted and listened to respectfully."

Hope

Hope, for a suffering person, is the expectation that things will get better, that life will become more enjoyable. In psychotherapy, hope may come partly from the sense of support, but it may also come from faith in the therapy process.

Most psychotherapists believe in what they do. They speak with authority and confidence about their methods, and they offer scientific-sounding theories or even data to back up their confidence. Thus, clients also come to believe that their therapy will work. Many studies have shown that people who believe they will get better have an improved chance of getting better, even if the specific reason for the belief is false. As was discussed earlier in

the chapter, patients who take an inactive placebo pill are likely to improve if they believe that they are taking a drug that will help them. Belief in the process is at least as important for improvement in psychotherapy as it is in drug therapy (Wampold & others, 2005). Hope derived from faith in the process is an element that all sought-after healing procedures have in common, whether provided by psychotherapists, medical doctors, faith healers, or folk healers.

Many psychotherapy-outcome experiments have compared accepted forms of psychotherapy with made-up, "false" forms of psychotherapy—forms that lack the specific ingredients that mark any well-accepted type of therapy. For example, in one such study, cognitive-behavioral therapy for problems of anxiety was compared with a made-up treatment called "systematic ventilation," in which clients talked about their fears in a systematic manner (Kirsch & others, 1983). A general problem with such studies is that it is hard for therapists to present the "false" therapy to clients in a convincing manner, especially if the therapists themselves don't believe it will work. Yet, analyses of many such studies indicate that when the "false" therapy is equivalent to the "real" therapy in credibility to the client, it is also generally equivalent in effectiveness (Ahn & Wampold, 2001; Baskin & others, 2003).

Motivation

Psychotherapy is not a passive process; it is not something "done to" the client by the therapist. The active agent of change in psychotherapy is the client himself or herself. This is an explicit tenet of humanistic therapy and an implicit tenet of other therapies (Hazler & Barwick, 2001). To change, the client must work at changing, and such work requires strong motivation. Motivation for change comes partly from the support and hope engendered by the therapist, which make change seem worthwhile and feasible, but it also comes from other common aspects of the psychotherapy process.

Every form of psychotherapy operates in part by motivating the client to confront his or her problems directly, to think about them in new ways, and, in so doing, to take charge of them and work at resolving them. Almost any caring therapist, regardless of theoretical orientation, will start sessions by asking the client how things have gone since the last meeting. The anticipation of such reporting may cause clients to go through life, between therapy sessions, with more conscious attention to their actions and feelings than they otherwise would, so that they can give a full report. Such attention leads them to think more consciously, fully, and constructively about their problems, which enhances their motivation to improve and may also lead to new insights about how to improve. Such reporting may also make the client aware of progress, and progress engenders a desire for more progress.

A Final Comment

Through most of this book our main concern has been with normal, healthy, adaptive human behavior. It is perhaps fitting, though, and certainly humbling, to end on this note concerning psychotherapy. We are social animals who need positive regard from other people in order to function well. We are thinking animals, but sometimes our emotions and disappointments get in the way of our thinking. We are self-motivated and self-directed creatures, but sometimes we lose our motivation and direction. In these cases, a supportive, hope-inspiring, and motivating psychotherapist can help. The therapist helps not by solving our problems for us, but by providing a context in which we can solve them ourselves.

SECTION REVIEW

Psychotherapy's effectiveness, and reasons for its effectiveness, have been widely researched.	
Determining the Effectiveness of Psychotherapy - Many controlled experiments, comparing patients in psychotherapy with comparable others not in psychotherapy, have shown that psychotherapy helps. - Research also indicates that no form of therapy is more effective overall than any other, perhaps because they differ more in theory than in practice. - Some therapies are better than others for particular situations, however. For example, behavioral exposure therapy is more effective than other treatments for specific phobias.	The Role of Common Factors in Therapy Outcome - Common factors, shared by all standard psychotherapies, may account for much of the effectiveness of psychotherapy. These include support, hope, and motivation. - Support (acceptance, empathy, and encouragement) helps to build the client's confidence. Research suggests that support, in and of itself, has considerable therapeutic value. - Hope comes in part from faith in the process, bolstered by the therapist's own faith in it. People who believe they will get better have a greater chance of getting better. - Clients must have strong motivation to work toward change; they are the only ones who can bring about change. Therapy motivates clients to confront and resolve problems, to think and act in new ways.

Concluding Thoughts

The best way to review this chapter is to think about the principles underlying each approach to treatment, about the potential problems of each, and about the evidence concerning the effectiveness of each. As a start in that review, the following ideas may be useful.

1. Self-knowledge and self-acceptance as goals of psychodynamic and humanistic therapies The psychodynamic and humanistic approaches to treatment focus less on the person's specific symptoms or problems, and more on the person as a whole, than do the other approaches. A psychoanalyst or other psychodynamic therapist who is asked to describe the purpose of therapy might well respond with the Socratic dictum: Know thyself. The goal of such therapies is to enable clients to learn about aspects of themselves that were previously unconscious, so that they can think and behave in more rational and integrated ways.

Most humanistic therapists would agree with the Socratic dictum, but they would add, and place greater emphasis on, a second dictum: Accept thyself. Humanistic therapists argue that people often learn to dislike or deny important aspects of themselves because of real or imagined criticism from other people.

The task for the humanistic therapist is to help clients regain their self-esteem, so that they can regain control of their lives.

2. Biological, behavioral, and cognitive therapies as derivatives of basic approaches to psychological research Biological, behavioral, and cognitive therapies focus more closely on clients' specific symptoms than do psychodynamic and humanistic therapies. These three approaches differ from one another, however, in that they emphasize different levels of causation of behavior. In that respect, they mirror the approaches taken by research psychologists who focus on (a) physiological mechanisms, (b) the role of environmental stimuli and learned habits, and (c) the role of cognitive mediators of behavior.

Biological treatments are founded on the knowledge that everything psychological is a product of the nervous system. Drugs, electroconvulsive therapy, psychosurgery, and deep brain stimulation all involve attempts to help a person overcome psychological problems or disorders by altering the nervous system in some way. Behavioral treatments are founded on the knowledge that people acquire, through conditioning, automatic and sometimes maladaptive ways of responding to stimuli in the world around them. The goal of behavior therapy is to

eliminate the maladaptive responses and replace them with useful responses. Cognitive treatments are founded on the knowledge that people interpret and think about stimuli in their environments and that those interpretations and thoughts affect the way they feel and behave. The goal of cognitive therapy is to replace maladaptive ways of thinking with useful ways of thinking.

3. Psychotherapy and science Two questions can be asked about the relationship between psychotherapy and science: (a) Is psychotherapy a science? (b) Has science shown that psychotherapy works? These are fundamentally different questions.

The first question concerns the degree to which the techniques used in psychotherapy are based on scientific principles and can be described objectively. Most psychotherapists would respond that their practice is a blend of science and art—that it is based on theories that stem from scientific research but that it also involves a great deal of intuition, not unlike the sort of intuition that is critical to any prolonged interaction between

two human beings. Each client is a distinct individual with distinct problems and needs, who does not necessarily fit snugly with the statistically derived principles that have emerged from scientific research. One way to compare the various psychotherapy approaches is on the degree to which they explicitly, as part of the theories behind them, emphasize the roles of empathy and intuition compared with the rigorous application of laboratory-derived principles. Rogers's humanistic therapy lies at one end of this spectrum and behavior therapy lies at the other.

The second question concerns the use of scientific methods to evaluate psychotherapy. Carefully controlled therapy-outcome experiments suggest that all the well-established psychotherapies work about equally well, on a statistical basis, though some may be more effective than others in treating certain kinds of problems. Such findings have led to increased recognition of the nonspecific therapeutic factors shared by the various approaches and have also inspired a movement toward eclecticism, in which therapists draw from each tradition those methods that seem most appropriate to the client's specific needs.

Further Reading

RICHARD J. HAZLER & NICK BARWICK (2001). *The therapeutic environment: Core conditions for facilitating therapy*. Philadelphia: Open University Press.

This brief, well-written, thought-provoking book begins with the assumption that psychotherapy works by providing a therapeutic environment—a setting in which clients can find workable solutions to their problems. The authors describe both the unique and the shared ways by which good practitioners of psychodynamic, humanistic, and cognitive-behavioral therapies create such an environment.

FRANK DUMONT & RAYMOND J. CORSINE (Eds.) (2000). *Six therapists and one client* (2nd ed.). New York: Springer.

Six different well-known therapists—one of whom is Albert Ellis—describe here how they would treat the same client, a socially withdrawn man who suffers from a variety of fears and other psychological problems. Each account of treatment is followed by a critique by another expert therapist. Reading this book will give

you a good idea of the various ways that psychotherapists think about their work.

BRUCE E. WAMPOLD (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.

Wampold is a leading advocate of the view that psychotherapy works largely because of the common factors that cut across the various types of therapy, not because of the specific factors that distinguish each therapy from others. In this book he carefully documents the evidence for this view. It is a scholarly book, but clearly enough written to be understood by first-year students.

ANDREW PAROTT, ALUN MORINAN, MARK MOSS, & ANDREW SCHOLEY (2004). *Understanding drugs and behaviour*. Chichester, UK: Wiley.

This is a well-written, relatively brief introduction to psychopharmacology. The book begins by describing the various ways by which drugs can act on the brain to influence behavior and emotions. In that context it goes on to describe

what is known about the mechanisms of action of the major drugs used to treat mental disorders as well as such other drugs as caffeine, cocaine, nicotine, cannabis, heroin, and alcohol.

ROBERT WHITAKER (2002). *Mad in America: Bad science, bad medicine, and the enduring mistreatment of the mentally ill*. Cambridge, MA: Perseus.

Whitaker, a science reporter, begins this book by reviewing the neglect and brutal treatment of people with schizophrenia from the mid-eighteenth to the mid-twentieth century. Through the rest of the book he argues that the neglect and brutality continue. His most controversial claim is that an unholy alliance between the pharmaceutical industry and the medical research community has suppressed evidence concerning the harmful effects of antipsychotic drugs, including evidence that the drugs may prevent full recovery from schizophrenia. Whitaker may overstate the case, but his argument deserves serious consideration.