



大夏心理·心空间



CENGAGE
Learning

Critical
thinking
in psychology

心理学带你 走出思维误区

◀ (美) John Ruscio◎著 ▶

王 湘 李欢欢◎主译



华东师范大学出版社
EAST CHINA NORMAL UNIVERSITY PRESS

- 营养品的骗局是如何形成的？
- 神秘的心灵感应真的存在吗？
- 磁疗对人体健康确实有益吗？
- 测谎仪能准确识别说谎行为吗？

.....

本书告诉我们，我们的思维有着怎样的缺陷，各种思维捷径会导致怎样的误差；如何进行理性思考，享受思维的力量。

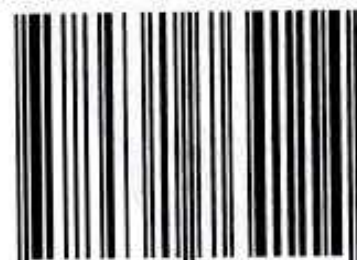
Critical thinking in psychology

www.cengageasia.com



上架建议：大众心理 热点话题

ISBN 978-7-5617-6411-4



9 787561 764114 >

定价：32.00元

www.dxjy.com

大夏心理·心空间

心理学带你 走出思维误区

(美) John Ruscio◎著
王 湘 李欢欢◎主译



华东师范大学出版社
EAST CHINA NORMAL UNIVERSITY PRESS

图书在版编目 (CIP) 数据

心理学带你走出思维误区/(美)罗斯科 (Ruscio, J.) 著;
王湘, 李欢欢主译. —上海: 华东师范大学出版社, 2008
ISBN 978 - 7 - 5617 - 6411 - 4
I. 心... II. ①罗... ②王... ③李... III. 思维科学—普及
读物 IV. B80-49
中国版本图书馆 CIP 数据核字 (2008) 第 146643 号

大夏心理·心空间

心理学带你走出思维误区

作 者 (美) John Ruscio
主 译 王 湘 李欢欢
项目编辑 任红瑚 汪明帅
封面设计 回归线视觉传达
责任印制 殷艳红

出版发行 华东师范大学出版社
社 址 上海市中山北路 3663 号 邮编 200062
电话总机 021 - 62450163 行政传真 021 - 62572105
网 址 www.ecnupress.com.cn

印 刷 者 北京振兴华印刷有限公司
开 本 700 × 1000 16 开
印 张 16
字 数 265 千字
版 次 2009 年 1 月第一版
印 次 2009 年 1 月第一次
印 数 6 000
书 号 ISBN 978 - 7 - 5617 - 6411 - 4/B · 439
定 价 32.00 元

出 版 人 朱杰人

(如发现本版图书有印订质量问题, 请寄回本社市场部调换或电话 021 - 62865537 联系)

译者序

作为一门学科，批判性思维大约形成于 20 世纪 70 年代。在国外尤其是在北美，批判性思维是一门相对较大的学科，许多高校都专门开设了这一课程。批判性思维与其说是一种理论，还不如说是一种技能，一种对各种观念和信息进行解释、分析、评价、推断、质疑和论证的能力。无论是对知识的获取，还是创造性思维活动，或是对认识成果的辩证分析和综合，都离不开批判性思维。

另一方面，批判性思维的分析、研究对象是日常推理和论证中的大量事例，其本质是考虑真实的问题，将科学推理与逻辑论证的理论和方法落实到生活中，学习者也因此会体验到思维训练的乐趣。

本书作者 John Ruscio 在判断与决策心理学、心理统计学以及心理评估等方面颇有建树，他的社会心理学专长以及对于心理病理学诊断分类、区分科学与伪科学的强烈兴趣，又使得他在运用多种学科门类的著名事例及经典研究实例来讲解批判性思维的理论与应用时，显得游刃有余。

本书分为四个不同的部分，分别强调了在评估各种观点时批判性思维运用的不同方面：科学推理的基本元素以及他人误导我们的几种可能的方式；有关自我欺骗的几种情况；心理陷阱与心理捷径的使用；判断和决策过程中的批判性思考以及相关伦理学问题。在每一部分，都着重运用了大量实例，强调了科学与伪科学的区分，通过谬误分析和批判性问题的解答，对批判性思维的一系列原则与技巧进行了详细说明与建议。

研究人类思维与行为特征是心理学的主要目的之一。因此，无论是心理

学的学习者还是专业的研究人员，阅读此书都将大有裨益，可以帮助我们形成更积极理智的怀疑与反思精神，严谨审慎的思考态度，追求合理性、准确性、可信性的思维习惯，以及独立自主和自我校正的思维技能。同时，这本书写作流畅，通俗易懂，大量实例十分贴近生活，对于一般大众而言，若想了解批判性思维与心理学、提高合理判断和明智决策的能力，本书也是不可多得的极好读物。

本书各部分的译者依序为：王湘、蒋莉、李欢欢、蚁金瑶和钟明天。其中王湘负责翻译前言、附录以及第一篇；蒋莉负责翻译第二篇；李欢欢负责翻译第三篇；蚁金瑶和钟明天负责翻译第四篇。由王湘负责统审全部译稿并做修改。

本书涉及的学科门类及专业术语较多，且有许多具体的心理学研究实例及司法案例，在翻译过程中已逐一进行详细了解及查证，并尽量统一了在不同章节中有关内容的译名。然而，翻译之中疏漏及不甚完善之处在所难免，恳请各位读者批评指正。

王湘

2008年6月

前言

1973 年，David Rosenhan 发表了一项充满争议的名为“精神病院中的正常人”（On Being Sane in Insane Places）的研究结果。在这个实验中，有 8 名精神健全的正常人（包括 Rosenhan 本人在内）假扮成精神病患者，到几家不同的精神病院就诊。他们抱怨说自己有烦人的幻听，要求入院治疗。除了虚构的名字和职业之外，其他相关背景和健康状况的问题，他们都据实回答。结果，所有的假病人都被诊断为精神病患，收进了精神病院住院治疗。其中一位被诊断为“双相情感障碍”，其他人都被诊断为“精神分裂症”。入院以后，所有的假病人都停止了表演。也就是说，那些假病人除了密切观察和留意收集数据以外，实际上表现得相当正常。他们想看看医护人员是否会发现他们“精神正常”，并且让他们出院。但实际的结果是：平均住院 19 天之后，假病人们才得以出院，其中绝大多数人的出院诊断被修改为“精神分裂症缓解期”。

这个实验的结果似乎表明，从事精神卫生工作的专业人员也不能准确地区分“精神健全”与“精神错乱”的状态，这听上去真是让人难以接受。事实上，Rosenhan 的研究在精神卫生领域引起了轰动：“我们的医院必须进行改进！”但是，暂时先别急着做什么改变，让我们来考虑一下另一种解释，并对这项研究进行一些批判性的思考。

挑战已广为接受的观念

1975 年，在《变态心理学杂志》（*Journal of Abnormal Psychology*）的一期特刊上，许多精神卫生专家就 Rosenhan 的研究进行了探讨，认为该研究存在严重

的方法论上的缺陷，忽视了一些相关的实验数据，得出的结论也不合理。《变态心理学杂志》的这一期特刊，以及 Spitzer（1976）针对该研究所进行的详尽批判，都指出了该项研究中的许多问题：

- 这项研究有潜在的验证性偏见（confirmation bias）。因为 Rosenhan 以及那些扮演假病人的助手事先已经知道了这一研究的假设，因此，他们可能会选择性地观察和记录那些与其假设相符而不是相反的信息，或者通过某种表现方式来得到他们所想要的结果，也就是自我验证预言。

- 有关 Rosenhan 的研究，争论的重要议题之一就是忽视了相关的实验性数据。他对诊断结果进行了过度概括，而且没有提到任何证据或资料，来说明诊断可以通过可信、有效的方式及过程进行，诊断可以帮助病人得到适当的治疗，可以促进精神卫生领域的医护人员、管理者及研究者之间的相互交流。

- 同样，在讨论诊断是否无效的问题上，Rosenhan 的逻辑也存在错误，因为那些假病人事实上就是精神健全的正常人。Spitzer 指出：（1）病人抱怨有幻听的症状；（2）一般来说，只有在人们感到十分苦恼不安的时候，才会主动提出要住精神病院；（3）假如没有足够好的理由，当然不会去怀疑别人是装病。因此，在这种情况下，最可能给出的诊断就是精神分裂症。在该研究中，假病人都被诊断为精神分裂症，只有一个人例外，这其实恰恰说明在不同的精神病院之间，存在着令人惊叹的诊断一致性。

- Rosenhan 的其他结论也存在证据不足的情况。他指出那些医护人员采取了有偏见的态度看待假病人，大胆地做出了医护人员有贴标签式行为的结论，并进一步阐述了其负面效应，但与此同时，没有提供任何必需的支持性证据。事实上，他所说的有关医护人员的偏差态度的唯一证据就是，有几个护士在“病人观察”那栏记录了“病人总是表现出书写行为”，这一点事实上完全是真实的记录，并没有任何病理意义的暗示。因此，要说在认知和解释病人行为方面存在所谓的贴标签式的偏差效应，实际上是缺乏证据的。

- 当指责医护人员所贴的标签时，Rosenhan 再次犯了错误，即将一种行为状态的名称与其行为所可能表明的含义完全等同。也就是说，他没有考虑到那些医护人员的做法可能并不是贴什么标签，其本质只是人们对待精神疾病的态度度的反映，更有可能仅仅是对与疾病本身相关的、不可预知的、甚至有时是危险性的行为的描述。

• 另外，Rosenhan 还使用了迷惑性的语言。“精神健全”（sane）与“精神错乱”（insane）这两个词，并不是在心理学晤谈或精神病学诊断中所用的术语，它们纯粹是一种法律解释用语，通常用于对一个犯罪嫌疑人进行判断，看他有没有真实反映现实状况的能力。Rosenhan 既是心理学教授又是法律学教授，因此他应当很清楚这样的分别。他使用这两个词，只能说是为了达到某种目的而有意进行的一种修辞手段的选择。

• 最后，Rosenhan 并没有提供比较性的数据来支持他的主要结论——医护人员不能区分精神健全和不健全的人。要做出这一结论，他必须提供一些资料，说明医护人员对待那些假病人和住院的精神分裂症病人是完全一样的。但实际上，与那些精神分裂症的病人相比，假病人在精神病房的住院时间明显短一些（几乎没有哪个精神分裂症发作期的病人住院时间是在一周左右），而且出院时候的诊断——精神分裂症缓解期——也说明了这一点（Spitzer 指出精神分裂症病人出院时很少被诊断为“精神分裂症缓解期”）。因此，Rosenhan 的数据实际上就是对他的结论的最大反驳。医护人员其实对于那些假病人所表现出的不同寻常的状况十分敏感，很明显假病人并没有被错看为普通精神病人。

拥有自行评估各种观点的能力

在许多（也许不是绝大多数）心理学教材中提到 Rosenhan 的研究时，对其总的研究方法以及结论并没有持批判性的态度。前文已经讨论了这一研究中颇有争议的几个方面，以及 Rosenhan 所做结论的一些不实之处，目的就是为了举例说明写这本书的作用：运用逻辑性分析方法仔细地评估所有的观点——即使是那些已广为人知的，或者与自身观念非常吻合的观点，也应如此对待。为了使大家掌握这种批判性思维的方法，本书将从以下四个相关的方面来安排内容：

1. 区分科学与伪科学

在本书所讲的每一部分内容中，我都会着重强调科学与伪科学的区分。要这么做，需要在一直备受 Carl Sagan 推崇的两种相互补充的思考模式——惊叹和怀疑（wonder and skepticism）中进行仔细的权衡。本书鼓励对于一切事物都持一种开放的态度，这样所有的观点都能被认真地接受。与此同时，也应当对所有观点

都秉持公正、不带任何偏见的态度：任何观点都必须有充分的证据才能被接受。这两种标准同时使用，最大的好处就是可以保证所有的观点都有公平的机会得到关注，最终那些有足够支持性证据的观点得以保留。

2. 教授批判性思维的技巧

批判性思维包含了一系列的技巧，任何希望自己能进行更周详、更审慎的思考和推理的人，都能掌握这些技巧。我认为所有的读者都有能力做出个人认为最好的决策。我在这本书里所提出的建议，并不是让读者去相信什么观点，而是怎样才能做出与自己的价值观最为一致的、真正明智的选择。通过认识别人的推理谬误，避免自己犯同样的错误，读者们就更可能免受与他们的真正利益并不相符的游说或者劝诱的左右，可以获得并坚持正确的信念，还可以在阐述自己的观点时构建更有说服力的、更合理的论证。

3. 运用生动的事例进行说明

在这本书里，我会运用大量生动的事例来帮助读者理解、掌握与应用所学到的原则。比如在绪论一章中，就是以 Patricia Burgus 的著名案例作为开头的。在那个经典案例中，Patricia 的抑郁和焦虑症状被当作多重人格障碍的症状表现，而且人们还确信她是一个崇拜恶魔的国际邪教组织的高级女祭司。同样，在第6章，是以一名妇女撰书推广神奇的尿液疗法的故事作为开场白。应该说，一系列著名案例与经典研究的介绍，使得本书十分生动活泼，给人印象深刻。

4. 开发多学科的相关材料

在选择案例的时候，我特别注意从心理学以及其他学科的各个独立领域中广泛收集材料，以表明科学的推理过程并不仅仅限于我们生活中的某一个领域。选修了心理学导论、心理学研究方法、实验心理学等课程的学生，都可以将本书作为一本介绍科学推理技巧的增补读物，并从中获益良多。对于学习健康心理学、变态心理学、临床心理学，以及与伪科学或批判性思维有关的课程，本书所提供的相当丰富的关于伪科学性质的心理及躯体治疗方法的案例讨论，也将会有很大的帮助。另外，我本人也曾用这本书中的某些章节作为统计学与心理评估课程的教材之一。

目 录

CONTENTS

译者序	1
-----	---

前 言	3
-----	---

第一篇 错觉与欺骗	1
-----------	---

1. 绪论：伪科学的存在及批判性思维的必要性 / 3	
区分有意义与无意义的资料、信息 / 6	
运用批判性思维这一工具 / 9	
伪科学的十大特征 / 10	
本书的计划 / 15	
2. 科学：从知识性的角度来评价各种观念 / 19	
科学推理 / 20	
反科学的信念 / 26	
巫毒科学与法律标准 / 32	
3. 语言：误导和回避战术的工具 / 36	
模棱两可的语言和架构效应 / 37	
合作以及对话准则 / 41	
最后的思考：错误的两分法和滑坡谬误 / 50	
4. 魔法：来自异国宗教、幻想以及神秘主义的诱惑 / 52	
相信魔法 / 53	

一厢情愿的想法 / 53
相信生命力 / 55
将“人类能量场”付诸检验 / 57
迷信与奇特仪式的产生 / 58
作为典型思维方式的迷信 / 59
迷信行为的维持 / 60
假科学之名支持神秘主义 / 62
现代科技的魔力 / 63

5. 权威：要求盲从 / 65
 Milgram 的研究 / 65
 对 Milgram 研究结果的分析 / 68
 我们应当在什么情况下服从 / 69
 靠不住的权威 / 70
 将政治游说作为警示旗 / 74

第二篇 自我欺骗

79

6. 经验：作为证据的局限性 / 81
 自我欺骗 / 82
 竞争性解释 / 83
 概率的基本规律 / 91
7. 貌似合理：并不是所有信念都具有同样的价值 / 94
 磁疗法：假说 VS 现实 / 95
 同种疗法：空洞的承诺 / 99
 电磁场与癌症 / 102
 外星人劫持的故事：现代社会的误解 / 104
8. 关联：建立和解释相关 / 109
 小心媒体的误导 / 117
 占星术真能预示什么吗 / 119

9. 风险：知觉偏差和媒体悖论 / 123

逼真性 / 124

可获得性 / 125

媒体悖论 / 127

重新分析坠机残片 / 128

激进的推测 / 129

对媒体报道的谨慎关注 / 130

10. 信念：验证性意见、事后解释和过度自信 / 133

发现 VS 证实：预测的重要性 / 134

验证性意见 / 139

缺乏反馈或误导性的反馈 / 145

后见之明偏差和过分自信 / 146

阻碍成功的事后解释 / 147

信心问题的答案 / 148

11. 图式：大阴谋理论的诱惑力 / 149

在一个阴谋理论中的三种角色 / 149

大阴谋理论：“第 51 区” / 151

大阴谋理论存在的问题 / 154

“隐匿疗法”的阴谋 / 155

质疑阴谋理论 / 158

12. 错觉：控制感 / 160

控制错觉 / 161

确定效应 / 162

控制错觉的危险 / 163

心理意象的负面作用 / 165

“选择健康”的谬误 / 166

祈祷和宗教信仰下的医疗忽视 / 168

- 13. 评估：经典决策理论 / 173
 - 多通道记录仪测试 / 175
 - 经典决策理论和频率树 / 176
 - 基本比率：试图探测的结果有多罕见 / 178
 - 效度：证据的强度如何 / 180
 - 阈值：用于决策的分界线要定在哪里 / 182
 - 实验室效度与领域效用的比较 / 184
 - 将所有因素都加在一起 / 185
 - 有关营养品的骗局 / 188
 - 防止专业误导 / 192
- 14. 决策：临床方法与统计学方法的比较 / 194
 - 进行决策的两种方法 / 195
 - 为什么统计学方法更加优越 / 197
 - 坚持使用临床方法的原因 / 202
 - 接受统计学决策方法 / 206
- 15. 伦理学问题：使用并推广未经证实的治疗方法 / 209
 - 忽略显而易见的事实，赞成不太可能出现的情况 / 209
 - 压抑记忆与多重人格 / 210
 - 投入—收益比 / 217
 - 真正的知情同意权以及患者自主权 / 219
 - “健康自由”的谬论 / 221
- 16. 工具：关于批判性思维的建议 / 223
 - 在医疗保健行业中伪科学方法的危险性 / 225
 - 进行批判性思维 / 227
 - 一厢情愿的结束性思考 / 234

第一篇

错觉与欺骗

1. 绪论：伪科学的存在及批判性思维的必要性

Patricia Burgus 并非一直都是一个国际邪教组织的高级女祭司。1982 年 4 月，她在生第二个孩子时差点死去，此后，她因为出现产后抑郁的症状开始接受心理治疗。经过一位精神病学社会工作者长达数月的治疗，Patricia 的病情几乎没有什么好转。这时，附近有位心理治疗家接手治疗一位患有多重人格障碍（multiple personality disorder, MPD）的病人，Patricia 的治疗师开始怀疑她是否也患有这种疾病。为此，Patricia 与她的治疗师又一起探讨了几个月。Patricia 知道她自己暗示性很高（具有高度的催眠易感性），清醒时也容易处于恍惚状态并迷失自我，而且她的治疗师也经常对她施用催眠术。Patricia 深入地检查了她自己的情绪波动，并努力寻找那些可以提示另一人格存在的记忆空白。在她接受治疗的 3 年后，Patricia 开始假设自己还有叫 “Kathy Love” 的另一人格的存在，那是一个 11 岁的女孩。

“当我以另一种不同的人格出现时，并不是我想要欺骗任何人，它更像是一种内心的释放。” Patricia 这样解释她自己的这种特殊体验。“刚开始时它就像是一个压力阀，当我表现为另一种人格时，我不必为自己所说的任何话负责。我不再是那个经历过可怕的分娩过程而且差点死去的妈妈。我也不再是那个必须要对两个孩子负责、必须打扫和清洁这座巨大的维多利亚式老房子的可怜人。如果我想成为一个甜蜜的 11 岁女孩，我就可以变成她。我简直有些沉迷于这种逃避的游戏了。” (Ofshe & Watters, 1996)

直到这时为止，尽管在心理治疗上没有取得多少进展，Patricia 还是坚持从事她的工作，并且照顾两个孩子。可是自从她被诊断出患有多重人格障碍（MPD）后，她的情况就很快恶化了，她开始寻找精神科住院治疗的机会。1986年3月，她住进了芝加哥 Rush 长老会医院的分离性障碍治疗中心，开始接受 Bennett Braun 医生的治疗。开始治疗以后，Braun 医生问她是否在儿童时期遭受过性虐待，Patricia 十分确定地告诉医生她并没有受过性虐待，她认为除了产后抑郁之外，一位邻居的自杀以及姐姐的过早去世都是造成她心理问题的原因。不过，Braun 并没有兴趣听她说这些，而且还告诉她，她以前一定受过性虐待，尽管现在她自己都已经不记得了。随后，Braun 与他的同事们逐渐运用催眠术、抗精神病药物、具有高度暗示性和诱导性的问题，以及冗长的“情感宣泄”期（在这期间，Patricia 的躯体运动受限，被迫忍受身体与情感均感到痛苦的“记忆工作”），使得 Patricia 形成了有关性虐待的“记忆”，却全然不顾在此过程中 Patricia 本人的一再否认。

由于住院期间远离家人与朋友，唯一能接受的只是医院工作人员的强制性治疗技术，Patricia 变得更加抑郁了。她急于重新控制她的生活，因此很快就意识到在这个环境里要幸存下来，关键就在于对医生们说那些他们想要听到的东西。Braun 相信，不仅所有的 MPD 病人在童年时都曾遭受过性虐待，而且还认为所有的人类社会都受到一种崇拜恶魔的、跨越世代的、国际性的邪教组织的影响。而对 Patricia 及其他许多病人的治疗，使得他更坚信他那种古怪的想象。因此，随后的几个月里，在医生们压迫性的治疗技术的影响下，Patricia 逐渐“恢复”了所谓的被压抑的记忆。这些记忆的内容包括：被多次强奸；参与了崇拜恶魔的宗教仪式，在这些仪式中有多名成人与婴儿被杀害，并且有骇人听闻的食人行为；还有其他数不清的暴行。尽管 Patricia 的家人一直认为这些都是无稽之谈，但 Braun 与他的同事仍然坚信 Patricia 就是一个国际邪教组织的高级女祭司，而且在九个州策划了邪教的宗教仪式活动。

即使是在自己描述参与邪教组织的故事时，Patricia 也意识到了这些故事前后不一致甚至相互矛盾，而且经常是一些几乎不可能发生的事情。她回忆的一些特殊场景，其人物与背景总是在一种梦幻般的形式中变换。她说点燃的火炬被作为性工具；某次她被埋葬了数天；她一年必须要吃 2000 个人的部分身体……她的工作是高级女祭司，因此她必须经常去其他州出差，而

这些都是瞒着她的丈夫进行的。

对于 Patricia 所说的，Braun 总是不断地帮她将所有新想法的不可能之处加以解释……而对于那些实在太荒谬的细节，Braun 仅仅是简单地跟她说，只要有充足的时间，事情的真相都会弄明白的，以此来打消她的疑虑（Ofshe & Watters, 1996）。

Braun 想象有一个国际性的阴谋，而恶魔崇拜者、美国电话电报公司、Hallmark 贺卡公司、中央情报局、FTD 花商协会等都参与其中，而当美国联邦调查局发现 Braun 所宣称的阴谋并没有任何证据支持时，他又认为美国联邦调查局也涉及这个阴谋。Braun 的另一个同事则运用催眠术，使 Patricia 回忆起她的前世是一个在一场火灾中去世的年轻的英国女孩，以及伟大的凯瑟琳女皇。出人意料的是，最终将 Patricia 从这种极端的治疗中拯救出来的，正是由于她精神状态的极度恶化。有一次，Patricia 企图放火，之后她就被转到了一个急性疾病治疗中心，然后又转到了另一个精神科病房。在那里，她坚持认为自己是多种人格中的某个年轻女孩，而拒绝恢复她的成人人格。那里的工作人员为了惩罚她，将她分到了另一个治疗小组，被当作普通病人一样治疗，并没有受到特别的关注，也没有被当作一个高级女祭司，更没有要求她进行任何记忆工作。慢慢地，Patricia 正常的感觉开始恢复了，她开始对原来所说的那些新的体验产生了怀疑。到 1990 年 12 月，Patricia 完全停止服用任何抗精神病药物（“冷火鸡”现象*）。1992 年夏天，Patricia 开始反抗 Braun 医生，并且告诉他，她不再相信治疗过程中所描述的任何故事了。“我告诉他那些全是胡说，我的问题就是医院的治疗造成的。”Patricia 还强调说：“我还告诉他，他做了一件可怕的事，对我和我的家人造成了极大的伤害”（Ofshe & Watters, 1996）。

由于这段错误的治疗，Patricia Burgus 损失了她生命中宝贵的十年。她的确患有抑郁，但这是一种常见的情感障碍，好几种治疗方法均可有效治疗。而 Patricia 被下了一个备受争议的诊断——多重人格障碍，然后被强迫通过一些有问题的治疗技术来恢复记忆，并接受了多种未经试验的治疗方法。虽然 Patricia 的案例是一个极端的例子，但这样的事情并不少见。这说明，当某些理论的提倡者

* 冷火鸡（cold turkey）是一美国俚语，表示突然完全停止某种成瘾习惯，例如使用毒品、抽烟、酗酒等。——译者注

坚持自己的假设，而这些假设又缺乏甚至根本没有足够的科学研究的支持，存在着根本的缺陷和疑点时，类似 Patricia 的遭遇就有可能发生。简单地说，Patricia 就是这种伪科学的牺牲品。

区分有意义与无意义的资料、信息

我写这本书，一方面是为了介绍批判性思维这一技能，另一方面也是为了演示批判性思维的具体应用。所谓批判性思维，并不是意味着什么都不相信（犬儒主义*），也不是持过度的批评态度甚至拒绝承认合理的信念。它指的是努力寻求在完全不持偏见与怀疑论之间的适当协调的一种态度。也就是说，对所有的观点都持公正的态度去听取，但只有那些符合高可信度标准的观点才会被接受或保留。由于这一平衡的过程比简单地全盘接受（过分思想开放，不持偏见）或全盘否定（过分怀疑）要难得多，因此要掌握批判性思维这一技巧，需要不断地努力尝试与练习。

聚焦批判性思维，首先需要辨认或避免（这一点也许更重要）错误推理中的主观陷阱。在这一方面，首先要强调的就是区分有意义与无意义的信息的能力，其次是如何运用这一技能对现存的一些伪科学进行评估。要掌握一项新的技能，练习是必不可少的，而伪科学的存在为我们学习批判性思维提供了绝佳的练习机会。事实上，伪科学背后的推理过程往往存在缺陷，通过仔细检查其逻辑推理过程，我们就可以学到许多东西。

研究人类的推理过程

要弄清楚人类的推理过程，首先得知道如何才能科学地研究它。我们的目的是探索人类判断与决策的心理学问题，包括我们每天都会运用的心理捷径。但是我们怎么才能了解这些心理捷径呢？我们可以像那些研究感知觉的科学家一样，首先找出常见的错误，然后倒推出在正常状况下导致我们犯错的一系列逻辑

* 犬儒主义是古希腊的一个哲学流派，其主要教条是，人要摆脱世俗的利益而追求唯一值得拥有的善。早期的犬儒主义者是坚持内在的美德和价值，鄙视外在的、世俗的功利；后期的犬儒主义者依旧蔑视世俗的观念，但是却丧失了赖以为准绳的道德原则，看什么都是虚伪的，认为世上没有道义可言，是一种极端消极的世界观。——译者注

推理过程。例如，在人类如何将光能转换成对颜色、形状、距离以及运动的知觉方面，视觉研究者们已取得了令人叹为观止的知识成果，而他们取得如此丰硕成果的一个主要方法就是研究视错觉。也就是说，通过研究我们的视觉系统是如何被蒙蔽而形成错觉的，再来推出它原本是如何正常运作的。

请看图 1，你看到些什么呢？绝大多数人会报告看到了一个连接在一起的螺旋状的图案。但是再仔细看一看，你就会发现这个螺旋的各个部分实际上并没有连在一起，是每一个同心圆中精确组合的黑白条纹造成了这种螺旋图案的错觉。那么，从这个现象又可以做出什么样的推论呢？那就是：我们的视觉系统一定对于明显的边界或分界线十分敏感。再看看图 1，每一条黑色或白色的条纹似乎都沿着一个自图形中央发出的螺旋形状的边缘排列，因此在我们的大脑中就形成了螺旋的印象。就这样，一个简单的错觉就引导我们对视觉加工的过程进行了一次非常有力的探索与推理。

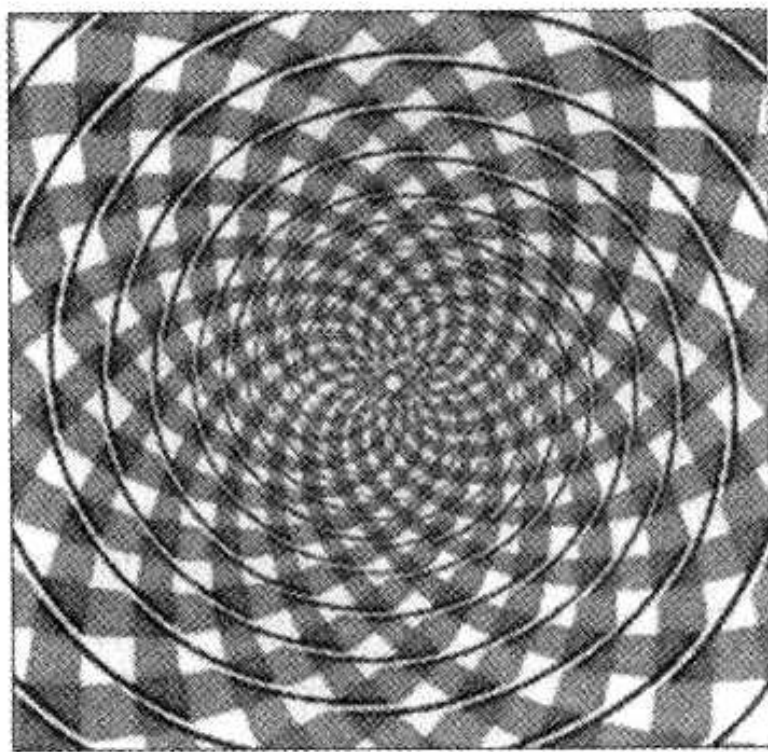


图 1 在 Fraser 螺旋中，每一个同心圆的黑色和白色条纹排列的模式会给人造成一种错觉，以为这是一个连接在一起的螺旋图案

心理捷径导致的系统误差

同样的研究策略在研究推理过程中也很有用。通过研究人们判断或推论中常出现问题的过程及原因，我们就可以找到那些系统的错误，从而对心理捷径有所了解。举个例子来说，请在五秒钟内猜出你认为下面这个简单算式的最佳答案，然后再阐述得出这一答案的心理过程。

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 = ?$$

你得到的估计值是多少呢？如果你猜出一个接近于 512 的值，那么你和那些参与 Tversky 及 Kahneman 这一实验（1974）的学生差不多，这是他们所得答案的平均值。这个答案当然是错的，比正确答案要低得多。但是，得出这个值并不仅仅意味着算错了一道题。在得到答案的过程中反映了可预见的错误。当以相反的排列方式给出同样的算式（ $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ ）时，学生们的估计值就高得多，平均值为 2250。

那么，这一偏差的出现可以提供给我们什么样的信息呢？一方面，人类在解决这类问题时只是进行了一种部分的运算。当有严格的时间限制，即要求很快得出答案时，人们往往会将序列中的头几个数相乘，然后锚定这个初始值，以其为参照点再向上调整，计算出序列中剩下的那些数值。这就是人们在第二个算式所得的计算结果要比第一个算式高得多的原因，因为在这个算式中，用作基数的初步结果（ $8 \times 7 \times 6 \times \dots$ ）要比第一个算式（ $1 \times 2 \times 3 \times \dots$ ）高很多。另一方面，人们在锚定初步结果以后所作的调整也远远不够。这一问题的答案原本是 40320，很明显几乎所有人都大大地低估了计算结果。因此，通过这样一个常见的简单判断失误，我们就可以得出一些关于推理过程的十分重要的研究结论。心理学家们将这一心理捷径称为采用锚定与调整法则的启发式推理方法*：当要求在很短时间内解决复杂的问题时，人们通常依赖于锚定的初始值，再通过一定的调整来得出最后的结论，但后期的调整通常都是不足的。

依赖心理捷径的利与弊

在日常生活中，我们每天接受的信息数不胜数，因此只能依靠心理捷径来将我们有限的注意力及心理能量集中在我们认为最重要的那些事情上。在这种情况下，大多数时候我们是可以进行相当准确的判断与合理的决策的。但是追求这种效率必须付出一定的代价，那就是必须在效率与准确性之间做一权衡与取舍。使用心理捷径的缺点就是，这种简化思维过程的方法有时会导致一些可预见性的错误出现。充分理解这些错误的性质会帮助我们尽量避免在推理过程中出现上述问

* 启发式推理方法是决策理论的重要概念之一，即利用非常简单的方法简化复杂的问题，形成一种单一的决策过程，其中主要有代表性法则、可利用性法则及锚定与调整法则等。——译者注

题，同时也能及时注意到别人所犯的该类错误。要做到这一点，我们必须超越这些心理捷径，以一种更审慎和更全面的方式对所接收的信息进行思考与加工。

进行批判性的思考

在这本书的每一章里，我们都会探讨一些常见的心理捷径，并运用经典的或现代的科学研究实例，来揭示这些心理捷径会给我们的判断与决策过程带来怎样深远的影响。由此你可以学会辨认那些因过度依赖心理捷径所造成的错误推理，并知道什么时候应当超越这些捷径，进行更仔细的思考。通过找出其他人的推理谬误、预防自己出现推理谬误，你就能在自己并不擅长的领域，更好地避免那些表面上特别令人信服的观点的影响。另外，你也可以获得并坚持正确的信念，进行更有说服力的、更合理的论证。清醒的思考会给人以无穷的力量。

运用批判性思维这一工具

要掌握一项新的技能，最好的方法就是尽量多地去运用它。本书将会在检视伪科学观念与实践的过程中，不断以关于人类推理过程的研究作为材料，来帮助你练习批判性思维。几乎所有的科学学科都会面临一些伪科学的问题。例如，天文学与占星术；化学与炼金术；进化生物学与创世论科学（creation science）以及智慧设计论（intelligent design theory）；史学与毁灭否定论（Holocaust denial）；医学与替代治疗（alternative medicine）；心理学与压抑记忆（repressed memories）、MPD 诊断以及思维场治疗（thought field therapy）等问题。由于存在于物理与心理健康领域的这些伪科学很有可能会给个人造成直接的伤害，因此本书会花相当的篇幅来探讨这方面的问题。

纵观历史，许多人都曾被那些自称是专家而实际上并不是专家的人、那些为了推销其产品或服务而做出惊人的但根本实现不了的许诺的人欺骗过。绝大多数伪科学观点都具有一种表面的可信性，然而一旦揭开这层表面可信性的面纱，我们往往可以发现其推理过程的缺陷以及缺乏合理证据。由于那些宣传伪科学观念的人都具有一个明显的动机，就是要利用你并从中获益，因此抱着一种健康的怀疑论态度去进行某些评估是十分有必要的。这就是为什么我们将在本书中运用所有相关的心理学原理，来对那些伪科学支持者的论点进行仔细的检视和验证。

伪科学的十大特征

要给伪科学下一定义，最好的办法可能就是从它的对立面——科学的定义来进行描述。所谓科学，是使人们了解事物真相的一种方法，是从谎言中提取事实的过程，是帮助人们以一种明智的、全面了解信息的方式进行艰难抉择的思维方法。它的对立面——伪科学，则缺乏对研究证据的基本尊重，而那正是任何一种真正科学的标志性特征。伪科学总是伪装成科学来获得尊重，但却不像真正的科学那样提供实际的研究结果。尽管在科学与伪科学之间的界线并不总是那么明晰，但还是有很多方法可以用来对它们进行区分。下面我要列举伪科学的十大特征，这些特征彼此相互关联，它们就是伪科学的一系列危险信号。

1. 披着科学的外衣

伪科学必须运用听上去十分具有科学性的语言来表达，但是在这些语言的背后却缺乏实质性的研究证据的支持。科学家们使用的专业术语，实际上是一种可以使专家们交流起来更便利、简洁的语言。而伪科学，则使用一种含糊的、不确定的语言来蒙蔽人或逃避详细审查。例如，一位物理学家可以对“能量”一词做出明确的定义，并且用无数精确的方法来证明能量的守恒性。而一位替代医学的治疗师则会以一种含糊不清的、不准确的方式来使用诸如“人体能量场”这类术语，但却根本不能证明人体能量场的存在。就这样，通过披上科学的外衣，伪科学就登堂入室，获得了原本只属于科学的可信性与合法性，但却没有经过科学要获得公众承认所必需的严格论证过程。

2. 缺乏同行的评议

无论是科学还是伪科学，都有可能会授予学位、拥有自己的学术会议以及出版物。在上述活动中，科学家们通常会经受某种程度的持怀疑态度的考验，这一点往往会让非科学工作者感到惊奇。比如，当一篇论文被投到一个科学期刊时，必须经过几位匿名专家的严格审查，而他们全部的工作就是尽可能多地挑毛病：文中所提的假设是否具有理论依据？所有的数据收集过程是不是毫无偏差，并且在方法学上合理可信？分析过程是否恰当？结果是否还可以作其他解释？对于论

文作者来说，同行评议往往是一个极其痛苦的过程。那些评阅意见可能直率得近于残酷，而当某篇文章存在一些严重的缺陷时，评阅人还会直接建议编辑拒绝发表。但是，作者们不仅都会忍受这一过程，而且还积极鼓励这种行为。并不是科学家比其他人更具有受虐狂的倾向，而是他们的确意识到，只有通过这样严格的方式审查每一项研究，那些真正对知识的积累有所贡献的文章才有可能跻身于科学文献之列。健康的怀疑态度在各类科学学术会议上、在取得科学学位的答辩过程中，同样表现得十分明显。然而，与此正好相反，伪科学虽然也有自己的期刊、会议甚至学位，但那其中仅仅只有象征性的怀疑与论证，甚至根本就没有这一过程。任何观点，无论是多么愚蠢、未经证明，或者没有说服力，都可以被伪科学接受，因为它根本就缺乏科学所具有的检查、纠正错误的良好机制。

3. 依赖个人经验

科学是以系统的经验论为基础的，也就是说，科学家会进行一些控制了影响因素的研究，来检验他们依据假设所得的观测结果是否真实可靠。根据奇闻轶事、个人经验，以及推荐与证明的方式都只能得出假设，却不足以证明假设，因此通常不会被认为是可靠的科学证据。当然，设计合理的实验需要相当有创造力，而完成实验又需要耗费大量的时间与金钱，但只有通过那些有良好控制与对照的实验来排除与确定假设，才能真正做到去伪存真。如果没有相关的实验证据，科学家会认为某种假设仅仅只是一种推测而已，根本谈不上什么事实与真相。而伪科学家则与此相反，有时他们甚至不遵循最基本的研究设计原则，而且对于那些与他们的观点对抗或相左的其他种种观点视而不见，也不愿意为有效的实验投入必要的时间与精力。换句话说，他们非常相信奇闻轶事，以及推荐与证明之类的东西，并且鼓励人们依据个人经验去评估那些观点是否真正具有知识性与科学性。在后面的章节中，你会发现仅仅依靠个人经验去检验某种观点，实在是太危险了。

4. 逃避有风险的检验

一个好的科学实验应当遵循这样一种思路：除非结果的确验证了原有的假设，否则绝对不能得出正性的结论。因此，每一个实验都应当提供一种有风险的检验，使得错误的假设不能通过。例如，对于某种理论来说，考察它是否能解释

过去已发生的事实，是一种相对无力的检测方式，而能否用其预测未来将发生的事情，就是一种更有风险的检验方法。科学家们往往根据检验的风险性所占的直接比例来评价所得证据的有效性。而当那些唯利是图的伪科学观点出现时，其倡导者往往迫切希望回避有风险的检验：通过这样的检验固然好，但那无助于产品或观念的营销；而一旦检验没有通过，又被普通民众了解到这一点的话，就可能对生意造成致命的打击。因此，当一名伪科学家不出示相关的有说服力的证据，往往就说明他所持的观点可能没什么价值，因为那是基于一个十分无力的风险性检验的结果所做出的结论。

5. 用超自然力量来自我保护

科学是人们用来揭示自然界真相的一种方法，即使某种机制还尚未为人所知，客观观测到的结果也可以作为科学有效性的证据。而许多伪科学正与此相反，当它们所鼓吹的观点看起来有可能出错时，它们往往求助于超自然力量的解释。比如，接触疗法是在护理行业中颇为风行的一种治疗方法，它的主要理论依据就是“人类能量场”的概念，而这一能量场的存在从未被证实过。当研究证明那些自称专家的人根本检测不到这一能量场，更不用说用其进行诊断或治疗的时候，接触疗法的从业者并不会关张大吉，而是开始致力于超自然力量的宣传（例如整体效应或精神效应这类根本无法观察或测量到的东西）。与可观测的现实状况完全无关的超自然的解释，并不会使其解释的内容变得更具可体验性或可证实性，恰恰相反，它们改变了科学证据的标准，使得那些伪科学者有了更多逃避合理检验的借口。

6. 为整体论大唱赞歌

随着有用的理论越来越多，科学的重要任务之一就是对导致这些理论产生的事物作越来越具体或专门的区分，例如科学家们要辨别亚原子粒子、化学元素，躯体疾病以及心理障碍等等。而伪科学则往往打着整体论的旗号，拒绝和反感进行这样的区分与鉴别。正如整体论这一术语其意义本身的模糊不清，这一理论的倡导者也总是在含糊其辞。他们运用这一术语来坚持要人们考虑更多超出心理允许范围的信息。当一个整体论取向的伪科学家声称是以一种完全非特异性的方式来考虑一切事物时，他在做出重要判断与决策时究竟运用了何种信息、以何种方

式思考，就彻底成了一个谜。整体论因此也就成了拒绝认真进行有意义的区分与鉴别的一把保护伞，最终流于华而不实的、空洞的言辞。

7. 允许矛盾现象存在

科学家们往往赞同并且运用许多有关形式逻辑的法则，其中最重要的一条法则就是“矛盾对立的两方不可能都是正确的”。而伪科学家们则以允许逻辑上相互对立的观念存在而著称，而且这种矛盾的观点不仅可以是同一学科的，还可以是跨学科的。例如，黄道十二宫（signs of the zodiac）这一概念是基于一年中的季节来定的，一些占星家就将南半球的黄道十二宫的顺序颠倒过来，但另一些占星家却并没有这样做。此外，分子矫正医学（orthomolecular medicine）主张进行特大剂量的饮食补充来增强效能，而同种疗法（homeopathy）则主张将有效成分稀释至极微弱的水平来增强效能。然而，我们很难听到或看到伪科学家们对这样矛盾的理论进行讨论，即使是随意进行的、表面的探讨也没有，这就很能说明问题了。伪科学家们从不相互质疑，而是将矛盾弃之一旁，视而不见。从这一点上看，他们就与那些为追求真理、消除矛盾而总是积极进行讨论的科学家们有着极大的不同。

8. 求助于权威

科学是建立在实验数据的基础之上的。任何人都可以读取这些实验记录，并且根据数据从中得出结论，而伪科学则几乎没有什么数据可以提供。伪科学的从业者总是极力劝说人们依据他们个人的意见、他们所声称的内容来信任他们的所作所为。而事实上，仅仅根据权威的说法就盲从于某种观念是十分危险的，历史早已证明了这一点。

9. 做出不可能实现的许诺

科学意味着尊重现有知识与当前技术的局限性。伪科学则与此相反，它不理睬现实的限制，往往做出根本不可能实现的、极其夸张的许诺以获取利益。一些伪科学家对于那些物理常数（如光速）、物理法则（如能量守恒法则）、人们易犯的错误（如心理捷径）等等，一概予以忽略。例如，他们会宣称不明飞行物（UFO）穿越星际距离的速度要大于光速；永动机是存在的；多通道生理记录仪

(测谎仪) 对于欺骗行为的测试百分之百准确等等。

10. 停滞不前

健康的科学学科所共有的最显著的特征之一就是：内容总是更新得很快。也就是说，随着知识的不断增长，它也不断更新。新的研究结果总是对旧的知识进行改进——不断提炼、去伪存真，使其更接近于事物的真实面目，最终体现为可见的实际利益。

举例来说，千年以来人类的平均寿命数都保持稳定的水平，而在 20 世纪却翻了将近一番。是什么原因导致了这一空前的、令人震惊的、人道主义方面的变化呢？应该说关于疾病的微生物学说、公共卫生测量水平、药物以及医疗技术的发展等等，均对这一改观有所贡献。寿命的长短可能是衡量物质生活质量的单一性指标中最好的一种了（如果你死了，当然你也就不可能做那些让你感到开心的事了）。可以说，这是科学赋予人类最宝贵的礼物，一点也不亚于生命的馈赠（Sagan, 1995）。

与科学恰恰相反，伪科学的显著特点就是停滞不前。荒谬的是，伪科学理论的支持者往往还把这一特点作为优点加以大肆鼓吹。正因为这样，某种伪科学可能将“远古智慧”作为具有宝贵价值的证据，丝毫不认为这是缺乏进步与改善、与现实的科学进展脱节的表现。例如关于东方医药的远古治疗技术的神话就是其一，实际上在延长寿命方面，它从未取得过像西方科学那样惊人的效果。还有一种由来已久的错误观念：人类只利用了自己大脑的十分之一。人们将这一观念代代相传，却对堆积如山的、与之相反的神经科学证据置之不理。实际上，最初的预感或猜测很少是完全正确的，如果你不对那些原始的观念与实践进行论证和提炼，往往得到的就是伪科学的垃圾。

通过上述种种方式，伪科学制造出了某种假象，而其从业者也总是装扮成专家，但实际上伪科学缺乏必要的实验证据，也不愿通过艰苦的工作来获取合理的研究结果。伪科学的支持者们总是抱着对他们有利的观点死死不放，尽可能地避免真实性检验，也否认证据之间的关联性。而当这一切都不能遮掩其真实面目时，他们就捏造一些所谓超自然力量的解释，以使其观点避开与现实相互矛盾的雷区。虽然科学家们也可能存在类似的缺点，也一定会有某种程度的错误，但他

们总是趋于将自己的观点建立在可观测的、可重复的实验证据之上，而一旦获得新的数据以后，他们也总是明智地修正自己的观点。

本书的计划

本章前面的内容对目前伪科学存在的一些普遍性问题进行了概述。在随后的章节里，我们还将进行更深入的讨论，结合各种伪科学的例子来说明合理推论的基本原则，并且给出一些应用实例。这些章节，以一种不存任何偏见的、持健康怀疑论的态度，将理论说明与案例相结合，最终的目的就是为了帮助读者掌握批判性思维的工具。下面首先将各章节的内容作一简介。

第一篇 错觉与欺骗

第1章首先介绍了科学推理的基础，然后讨论了伪科学家们用于影响大众的几种方法。第2章围绕任何一种知识都需面对的六种挑战，对科学推理的过程进行了简要的归纳说明。这六种挑战分别是：可证伪性、逻辑性、综合性、诚信性、可重复性和充分性。如果某种观点能够通过上述六种挑战的检验，那么至少在一段时间内可以保证它具有合理性和可信性。除了架构科学推理的基本框架以外，第2章还回顾了科学哲学所面临的一些挑战——科学信念的前提和结果。虽然科学是由人类创造并具体实现的，总难免会存在错误，大家也公认它是一种并不完善的工具，但相对于其他方式来说，科学在解答关于自然界的重要问题方面还是更具有优越性的。

建立了科学推理的基础以后，随后几章深入探讨了伪科学用于影响和愚弄大众的诸多手段。第3章概要说明了在构架问题和形成观点的过程中“更佳词语选择”的作用。比如，在堕胎的争论中，使用“生命优先”和“选择优先”的术语，就使得问题超出了准确捕捉观点或看法之实质的原有目的。换句话说，这些术语的运用反映的是一种有意通过语言的精心选择和组织来影响人们的信念、而不是寻求逻辑论证或研究证据的企图。

在第4章中，作者描述了历史悠久的魔力思维对人们的吸引力。从物种的角度来说，人类总是孜孜不倦地寻求自己在某些方面更出类拔萃、人类生命中存在某些奥秘的证据。纵观历史，人们也总是倾向于将自己解释不了的现象归结为超

自然的力量。本章运用了广为人知的关于学习的心理学原理，来阐明迷信观念是如何形成并且维持下去的。此外，还探讨了关于无法证实的生命力的魔力信念，并给出了一个案例，说明如何通过一个简单的试验，从根本上质疑某种所谓的人类能量场是否真的存在。

第5章则回顾了权威在引导普通大众的观念及行为方面的力量。虽然很多时候服从具有合法资格的权威是正确的，而且可以大大简化我们的日常生活，不需费时费力去做艰难的抉择，但我们偶尔还是会犯两种极端的错误：一是过分轻视权威的意见，即使是那些具备专家资格的权威；二是过分服从权威的意见，即使是那些可疑的权威。在第5章里，我们说明了可信的权威和可疑的权威通常都有些什么特征，并提醒人们，伪科学家们总是企图通过政治手段，而不是科学研究的途径来使其权威取得合法性，而这种方式事实上恰恰有损于其观点的权威性。

第二篇 自我欺骗

第二篇基于人们都具有不自知的自欺行为，对进行评估与判断时不同类型论据的优劣进行了讨论。在第6章中列举了很多种情境，说明即使是诚实的、出自好意的人，有时也会欺骗自己或者他人。例如，当我们过分相信那些推荐与证明、奇闻轶事，或者个案研究时，就有可能被严重误导。虽然上述信息可以为假设服务，但是它们均不足以检验或证明假设。造成这种情况的部分原因是因为我们所听到的案例是经过选择的，而那些经过精心挑选的案例往往只是我们需要学习的原理的一些特例而已。

第7章介绍了从知识性的角度初步评估某种观点的时候，一些貌似合理的理论所起的作用。当有人在宣扬某种与公认的科学原理不一致的观点时，这些观点就特别需要仔细的检查 and 批判性的分析。第七章以磁疗法、同种疗法以及外星人劫持的故事为例，深入探讨这些伪科学的观点是如何与公认的科学原理相违背的。

第8章探讨了在两件事物之间建立联系并阐明其意义的各种方法，比如治疗方法与治疗结果之间的关系如何确定。在每天的日常生活中，我们都会接收大量不完全的信息，不能最终确定事物之间是否存在关联。而且，即使证实了AB两件事之间存在相关关系，也并不意味着那是一种因果关系——A可以引起B，或者B可以引起A。要建立因果关系，必须符合几项标准，在第八章中我们就详细

讨论了这一问题。

第三篇 心理陷阱

在第三篇里我们对一系列心理陷阱进行了分析，这些心理陷阱可以导致一些无根据的伪科学观念的产生，并使其继续维持下去。第9章讨论的是选择性的媒体报道对于人们的信念所产生的影响。大众传媒总是倾向于报道一些罕见的或具有轰动效应的事件，而且总是以一种生动逼真的方式进行报道，使得这类信息更容易被记住。然后，由于提取这些生动鲜明的信息相对比较容易，就会造成我们在对某种潜在的危险或威胁、某种信念的普遍程度或者某种结论的有效性进行判断时，出现一定程度的偏差或歪曲。

第10章所讨论的问题是，无论我们选择哪种信念，对该信念的信任度都很容易越来越高。我们总是倾向于寻求和辨认那些与我们自身观点一致的信息，而对那些不一致的信息则采取更为严格的检测方法。类似的，我们也总是过于相信那些对已经发生的事件所进行的理论解释，却较少要求那些理论能正确预测未来的事件。此外，具有讽刺意味的是，往往是那些最缺少科学研究支持的观念的倡导者，对其理论会表现出最强的自信。

第11章讨论了伪科学家们编造阴谋理论的倾向。相信“大阴谋理论”会使得人们将自己的日常活动都赋予某种特殊的意义或目的，而且也拒绝对其进行理性的思考或探究。但对于绝大部分大阴谋理论而言，一旦其表面可信性被识破，逻辑上的自相矛盾及研究证据的极度匮乏就会使人质疑其真实程度，进而对其准确性产生怀疑。

第12章详细探讨了关于控制错觉形成的一些方式。控制错觉是指我们有时会高估自己对于生活事件的控制能力，相信个人具有某种控制力可以增强个体的能力感。然而，控制错觉也会使那些不可预知、不能避免的灾祸的受害者遭受无情的伤害，例如认为遭受强奸的妇女对其遭遇的不幸应负有某种责任。在第12章里通过详细的论证，说明了许多伪科学所提供的控制错觉，其实从各个方面来说都是很危险的。

第四篇 决策和伦理学问题

在第四篇里，我们探讨了决策的过程，以及伪科学实践所导致的一些伦理学

问题。其中，第13章介绍了经典决策理论，认为它是一种评价各种评估或决策程序的有效工具。例如，当人们需要检测的医学问题十分罕见时，即使是有效的诊断性试验也可能很难准确地加以检测。在这一章中还介绍了一种简单的方法，可以让你将某种情况出现的频率，以及支持或反对的证据的强度进行恰当的结合，从而提高决策的正确性。伪科学的从业者为了将其实践描述得特别实用或百分百有效，通常不会以一种负责任的态度去设计或使用有效的检测手段。

第14章回顾了以往的研究，指出在决策过程中运用统计学方法（精算法）要比临床方法（直觉法）更为准确，至少不会更差。随后我们对统计学方法更为优越的各种原因进行了详细解释，又列举和驳斥了几种常见的坚持使用临床方法的理由。在这一章里，我们还论证了在伦理学上最站得住脚的决策方法，就是那些有最佳正确决策记录的方法。

第15章整合了多种推理方式，就使用和推广某项未经证实的治疗方法所必需的伦理学方面的考虑进行了探讨。各种治疗方法或理论的支持者经常声称他们的疗法是完全无害的，因而试用该疗法不必冒任何风险。但事实上，他们总是倾向于低估该疗法造成危害的可能性与风险，高估可能的获益，同时完全不考虑当这一方法应用于临床领域时，需要放弃以往已经证实有效的治疗方法所需的机会成本。在这一章里，我们通过论证指出：使用或者推广某种未经证实其疗效的治疗方法的行为，都不符合最基本的道德伦理原则。

最后一章即第16章，对于如何科学推理给出了一些普遍性的建议，以分辨错误，避免犯错。我们回顾了人们日常生活中常用于简化思考过程的各种心理捷径，并列举一些方法来避免推理谬误或偏差的出现。我在进行这一总结性章节的写作时主要是基于这样一种观点：每位读者都有能力进行明智的判断，并且最终做出个人最满意的决策。本书并不是建议人们应该相信什么，不应该相信什么，而是如何能得到与个人价值观一致的真正合情合理的选择。我真诚地希望本书关于人类推理过程的心理学方面的探讨，可以帮助读者在面临人生重要抉择时作出更为明智的选择。

2. 科学：从知识性的角度来评价各种观念

从前，有一匹马会解数学题。那是在 20 世纪初，一位德国教师向世人展示了她的这匹马——聪明的汉斯。当汉斯的训练者（或者观众）问汉斯 5 加 3 等于多少，汉斯就会用蹄子轻叩地面 8 次；如果问 4 乘以 7 等于多少，汉斯就用蹄子轻叩地面 28 次。惊叹不已的目击者将汉斯的这种特殊能力迅速地传开，聪明的汉斯也因此一夜成名。请想象一下，一匹拥有神秘的数学解题能力的马会引起怎样的轰动吧！

后来的调查证实了汉斯的确是一匹拥有特殊才能的马，不过情况并不像大多数人想的那样。目击者们说得没错，对任何向它提出的数学问题，汉斯会用蹄子轻叩出答案。但要是认为它是通过思维运算过程来得到这一答案的，那就大错特错了。心理学家 Oskar Pfungst 想弄清楚事情的真相，就在实验控制的条件下，对汉斯的能力进行了测试。他发现汉斯只有在两种关键条件都符合的情况下，才会做出正确回答。第一，提出数学问题的人自己必须知道正确的答案；第二，汉斯必须能看到这个提问的人。当其中任何一个条件没有满足时，汉斯都解答不了题目。虽然并不了解是什么原因，但很显然，汉斯拥有敏锐捕捉、判断非言语线索与信息的能力。

当有人提问时，汉斯可以分辨提问者的声调，然后适时地开始叩蹄子。它还会仔细观察问它问题的人，一旦达到正确答案的数目时，这人就会不知不觉地轻轻点头，这就是汉斯停止叩蹄的信号，而提问者自己通常根本察觉不到。就是这个过程，使人们得出了“汉斯会正确解答数学计算题”的结论。而 Pfungst 运用

了科学的推理，排除了这个错误的结论——汉斯会算数，同时还证实了一个同样令人惊叹不已的假设——汉斯是一位敏锐的人类非言语行为的“观察家”。

科学推理

也许是一直以来学校教授科学的方法问题，使得许多人都有这样一种误解：科学就是收集事实资料。但实际上，科学远不止收集事实资料这么简单。它是一种方法，一个过程，一种思维的方式，是使人们可以更接近和了解事实真相的有力工具。科学是在严格控制的实验条件下获得观察结果，因而避免了人们一厢情愿的主观推断，最终确定究竟引起某种效应的原因是什么。在获取知识的各种途径和手段中，科学也有独特的优势，它包含内在的检测错误的机制和程序，可以发现和排除错误的观点。前面所说的汉斯拥有超凡能力的案例，清楚地说明了科学推理是如何帮助我们排除和修正不成熟的结论，从而对现象加以正确解释。通过系统地操控汉斯解题环境中的影响变量，Pfungst 排除了汉斯会进行加法和乘法的假设，到最后唯一站得住脚的假设就是：汉斯是通过观察提问者的非言语行为来做出解答的。

科学总是通过一系列成功的逼近真实状况的估计，力求不断提炼和完善我们对于现实的理解。当某种有趣的或惊人的理论被证实是不对的时候，通常就会被同样令人惊叹或更令人着迷的理论所替代。当然，无论科学理论是否令人着迷，其真实有效性都是首先要得到保证的，因为这完全是依据该理论与所有可获得的证据之间的一致性来进行判断的。科学的力量就在于它具有增加人类知识的能力，同时也使来自不可知的神秘世界的威胁减少，并为技术革新作好准备。不过，只有那些愿意以开放性的怀疑态度进行思考，给所有观点以同等的机会，同时又对那些伪科学理论坚持严格评估和取舍标准的人，才能获得科学的回报。

然而，科学推理的工具到底指的是什么呢？James Lett（1990）提出了一种较好的标准，将他认为可信的观点所必备的条件概述为六种检验方式。每种检验方式都反映了科学思维应包含的一种基本成分，借助这些检验，我们可以免受错误观点的愚弄。这六种检验方式是：可证伪性、逻辑性、综合性、诚信性、可重复性和充分性。

可证伪性

对于任何具有科学意义或可能具有科学意义的观点来说，原则上都必须提出一种可能被证明是不对的假设。也就是说，如果该观点所指的确是错的，那么一定有某种方法来证明其错误，这就被称为观点的可证伪性。之所以说可证伪性是科学观点的必备条件，是因为如果没有这一特点，那么该观点就完全与现实脱节了。换言之，如果完全不存在对某种观点不利的任何证据，那么该观点也就不可能被证伪，因此完全没有科学意义。

为了说明不可证伪的观点是多么的空虚无意义，Sagan（1995）曾举过一个例子：如果他宣称在他的车库里有一条会喷火的龙，会是什么样的结果。当然，你会试图检验他的这一说法。当你要求看这条龙时，他就告诉你龙是看不见的。能否用着色的方法再现这条龙呢？不行，因为龙是无实体的，所有的颜料都无从附着。用面粉洒在地板上，以再现龙的脚印呢？这也不行，因为它悬浮在空中，根本不着地。那么测量一下它呼吸过程中喷火燃烧的温度？没有这么方便的事，它根本就是无热的。你不断地提出合理的检验方法，但得到的推诿和托辞却越来越多。最后你只能开始怀疑，讨论一条看不见的、无实体的、盘旋在空中、呼吸中能喷出无热的火焰的龙，究竟有什么意义？这与根本没有龙存在，或者说根本没有任何东西存在有什么区别？

Sagan 的例子说明了不可证伪的观点所存在的问题：由于根本没有办法对它进行检验，那么绝大部分的重要证据就依赖于某人的说辞，而这一点在科学推理的领域里是万万不可接受的。科学家要证明某种观点的方法就是“拿出证据来给我看”，如果做不到这一点，那么该观点就会被纳入伪科学的范畴。伪科学家们可能通过构建一些不可证伪的假设来逃避现实的检验，而科学家们取得进展的方式则是通过精心的实验，一个个地排除错误观点。可证伪性并不是说所提出的假设本身就是错误的，没有人会把时间浪费在已知是错误的假设或观点上，它只是指一定存在某种方法可以用于检验这些假设，也就是它们具有可被证伪的特征。

逻辑性

任何知识性的观点当然都必须在逻辑上是合理的。这种逻辑论证的合理性必

须同时符合两个标准：首先，逻辑论证的基本前提必须全部是正确的；第二，所得的结论必须是在那些逻辑前提下，经过恰当有效的推理过程产生的。如果这两个标准中有任何一个不符，那么该逻辑论证就有缺陷。

举例来说，我们可以用逻辑论证来检验“麦田怪圈”的科学性。所谓麦田怪圈，是指出现在农场麦地中，由于农作物被压倒而产生的巨大的几何图形。长期以来，这一现象都被当作是外星生物造访地球的证据：

1. 麦田怪圈的图案都是非常复杂而且为数众多的；
2. 人类没有制造出如此大规模、如此复杂的图案的能力；
3. 因此，麦田怪圈是外星生物制造的。

在上述推理过程中，第一个前提的确是对的。几十年来，这种精心制作的“麦田怪圈”在全世界许多地方都出现过。第二个前提则不那么正确了：有两个英国人 Doug Bower 和 Dave Chorley，就曾公开承认自己制造了麦田怪圈愚弄人们足足 15 年（Schnabel, 1994）。他们不仅承认自己干过这些事情，而且还很热心地向记者们示范了制造麦田怪圈是一件多么简单的事情——仅凭一些 5 厘米厚、10 厘米宽的窄条木板和绳索就可以完成。请想象一下，当他们听说自己的创造被认为是人类不可能完成的太过复杂的任务时，他们心里该有多么骄傲和自豪啊。

让我们再来看看，如果假设第二个前提是正确的，是否可以合乎逻辑地得出“麦田怪圈由外星生物制造”这一结论？那也不一定。自然界的力量经常会戏弄人类，给我们造成一些错觉，例如形成某种错综复杂的、类似于火星上的金字塔一样的图形。那么在一些特殊的情况下，风或者其他自然现象也有可能造成小麦田里出现一些奇怪的图形。综上所述，对“麦田怪圈是外星生物造访地球的证据”的逻辑论证，在好几个地方都站不住脚，当然它也就不能为这一观点提供逻辑上合理的支持了。

综合性

综合性指的是作为一种具有知识性的观点，它必须要能说明所有的相关数据，而不仅仅是选择性地说明其中一些零散的部分。比如，一个至今还广为流传的、与 1965 年大停电事件相关的观点，就不符合综合性的检验。1965 年，美国东海岸大部分地区突然断电，这一断电事件持续了数小时。当时经历这一事件的

绝大多数人相信，这就是后来婴儿出生率显著增高的原因。人们津津乐道且广为流传的一个故事是：纽约市的婴儿出生率正好在大断电事件九个月后的某个周一和周二达到高峰，可想而知人们在没有电的时候都在干些什么。这一观点所涉及的数据都是正确的，但它只考虑了相关证据中的一小部分。调查表明，事实上每个周一和周二婴儿出生率都会出现一个小的高峰，直至今日还是如此。这种现象的原因其实很简单：医生们不愿意在周末工作。因此，药物诱导分娩及剖宫产这类择期手术就会被排在每个星期的头两天。虽然关于医生安排工作的习惯这一解释听上去平淡无奇，一点也不像“在一次大面积长时间断电后九个月出现了生育高峰”那么有趣，但它可能才是真正的原因。因此说，如果某种观点不能解释所有相关的证据，那么它就通不过综合性的检验。

诚信性

毫无疑问，任何知识性观点必须经过诚实可靠的评估，不能有半点自我欺骗，但这一点往往说起来容易做起来难。需要判断的情况有时候的确是模糊不清，难以取舍。尤其是在评估者已持有某种信念的时候，更容易出现诚信性方面的偏差。例如，有两名心理学家的研究发现，幼时在日托中心待的时间较长的儿童，比那些待的时间较短的儿童成绩更差，家长及老师对其进行的行为评定结果也更差，这些儿童在三年级进行的标准化测试中得分也更低（Vandell & Corasan-tini, 1991）。这一研究其实在方法学上与其他同类研究一样严谨，但仍然在其领域的顶尖学术期刊——《发展心理学》上遭到了同行们的强烈反对。让我们来分析一下这个问题：为这一研究做同行评议的那些学者，每天可能都要工作很长时间，而且还很可能是夫妻双方都外出工作的家庭，那么他们自己的孩子就有可能在日托中心待很长时间。有可能正是因为这一点，使他们本能地对上述研究结果感到厌恶吗？这种影响其实是不太可能被证实或弄个水落石出的，但是我们仍然可以想象得到，这一研究结果的性质本身就决定了它很难获得公正、诚实的评价。想想看，我们自己平日也总是自觉不自觉地在使用很多方式“欺骗”自己，所以说，任何观点要真正通过诚信性这一标准的检验，难度其实比表面看上去的要大很多。

可重复性

一种知识性观点还应当有多个实验得出的一致性结果作为基础，并且最好这

些实验结果来自于不同的实验室。侥幸成功的现象可能在某个单一的实验中出现，但在重复性实验中以同样的方式出现同一问题的可能性就小之又小了。因此，通过重复性实验得到的一致性结果，要比单一的或者存在不一致的实验结果可信度高很多。

在得不到重复实验的验证方面，超感官知觉（extrasensory perception）可能是个最好的例子了。超感官知觉的研究者们通常总是会进行大量的实验，并用各种各样的方法来分析他们的数据。而且，即便显著性检验显示其成功率与机会率无异，他们也总是以一种令人惊奇的方式得出结论，使其表面上看来效果非常显著。问题的关键在于，要能用重复性实验的方法证明这不是一种统计学上的特例，并不只是侥幸成功或偶然事件。但事实上，尽管超感官知觉的研究者们共同致力于这一领域的时间已经超过了—个世纪，他们还是不能对哪怕是一个单一现象进行重复和演示。下面一些在超感官知觉研究领域的常用术语就可以表明，到目前为止，他们在可重复性方面的努力仍然是多么失败和无效（Gilovich, 1991）：

- **超心理学消失效应（psi missing）**。通常在某人表演的成功率甚至比一般正常的机会率还低时，就会用到这一术语。超感官知觉的拥护者，或超心理学家往往辩解说这种差劲的表现其实也是其能力超常的证据，因为成功率比机会率还低，可见是存在某种能力的作用，以避免此事的发生。但是那些非常拙劣的预测，真的可以作为超自然心理能力的支持证据吗？

- **实验者效应（experimenter effect）**。这一效应指的是当个体表演超感官知觉能力的时候，不能有一个持怀疑态度的观察者在场，否则就不能成功。超感官知觉的拥护者将这一现象命名为“负性能量”，认为其会对超心理现象产生某种抑制。无论这一要求是出于有意还是无意，可能它真正的意思就是：必须有一个便于进行欺骗活动的实验环境，因为这时就不会出现一个怀疑者来让大家保持诚实清醒的头脑了。

- **衰退效应（decline effect）**。这一效应是指某人的超感官知觉能力原本表现得很好，后来开始变差，最终只能达到机会率水平的现象。这一效应其实就是不能有效地重复，但超感官知觉的拥护者们拒绝承认这一点。实际上，像“新手好运”之类的侥幸成功，在重复测验中就会出现类似衰退效应的迅速消失现象。

充分性

James Lett 所提出的检验标准最后一条是：一种具有知识性的观点必须有足够多的证据来证明其科学性。在评价有效证据的充分性时，需要考虑以下三点：

1. **申请者有义务提出证据：**这一原则其实与我们的司法审判系统的操作程序是类似的。提出某种观点的人就好比是原告，他有责任和义务去搜集足够多的证据来证明其观点是正确的。我们根本用不着忙于反驳或推翻这一观点，因为如果不能提出充分的证据，那就说明这种观点是不足为信的。

2. **特别的观点需要特别的证据：**这一原则与我们的常识是相当符合的，它是指所提出的观点越是稀奇古怪，就越是需要提供更有说服力的证据来获得支持。

3. **仅仅有来自权威的证据是不够的：**这一点是指，除了得到某些权威的首肯之外，还有没有其他类型的证据？历史告诉我们，任何人都可能犯错误，因此无论支持该观点的人是多么有声望，他还是会有可能在某件事上出错。例如，像爱因斯坦这样鼎鼎大名的科学家，也在他那个时代的一个重要科学理论上犯了错误：他不支持量子力学的测不准原理（uncertainty principle），理由是“上帝不会掷骰子”。但是目前物理实验已经证实，根据测不准原理所做的预测其实是相当精确的。

一种观点要通过充分性的检验，其支持性的证据必须同时满足上述三个条件。请记住下面这句话：“缺乏证据不能构成否定的证据。”例如在关于外星访客的争论中，那些相信不明飞行物（UFO）的人认为目击者就是最好的证据，而且如果你不能证明他们是错的，那么他们所说的就是对的。但是，你需要谨记的一点是：不能对每一个宣称看见 UFO 的事件提出合理的解释，不代表就不存在合理的解释。事实上，我们可以列出许多曾经被人们错误地认为是天外来客的东西，其中一些是已知的自然现象，另外也包括一些人类制造的机器等等（Sagan, 1995）：

- 有意的欺骗或恶作剧，比如用绳子吊起某物，使其悬浮在漆黑的夜空，或者伪造照片或视频资料等。
- 常规的或非常规的飞行器，有可能在军事实验飞行过程中被涂成有斑点的图案，而且因为安全原因不会被当局证实。

- 高空热气球，比如 1947 年在新墨西哥州被错当成坠毁飞碟的热气球。
- 在特殊的大气条件下所看到的行星。
- 发光的昆虫
- 光学幻影
- 凸透镜云
- 球状闪电
- 流星和绿火球
- 重返大气层的卫星、火箭的头锥或助推火箭等等。每天都会有两到三个这样的物体在重返地球大气层的过程中被毁坏，而其中很多仅用肉眼就可以看到。

你用不着去判断究竟是上面所说的这些物质，还是其他的什么造成了那些 UFO 的特殊现象。正如前面所说，提出这一观点的人应当责无旁贷地提出足够多的证据，来证明他所看到的现象的确不能被上述任何一种情况所解释。特别是在天空中本来就有可能看到许多东西，而且也容易因此产生误会，那么宣称有 UFO 存在并推断有天外访客的人就更应该负起举证的责任，并且还要提出特别的证据来证实这一不同寻常的观点。

如果有哪种观点通过了上述六种测试，那么你可以尝试相信这种观点。如果其中有一种或几种未能做出明确的结论，或者根本没通过检验，那么你就应当明智一点，保持谨慎。只有那些总是对各种观点的相关证据坚持严格标准的人，才有可能形成最准确可信的信念，并尽量避免做出愚蠢、错误的决策。

反科学的信念

看到这里你可能会想，科学推理的好处是不言而喻的，而且那些经由科学推理过程形成的已经证实的知识以及令人惊叹的技术革新，也足以说明它的用处。但是，仍然有那么一些人，有意或无意地直接抨击这些科学的方法，而且其信念或观点总是显得对科学有敌意。

后现代主义

在学术研究领域的一些角落，还有那么一些人赞成某些密切相关的思维方

式——如后现代主义与文化相对论——的极端形式，认为科学仅仅是“获取认识的方法之一”而已，并不见得比其他方法更有效可信。虽然与这种观念进行彻底的辩驳已经超出了本书的范围，但在这里讨论一下后现代主义的几点最致命的缺陷，我认为还是很有必要的（更广泛的驳斥与批判见 Norris, 1993；误用科学及数学的综述见 Sokal & Bricmont, 1998；有关后现代主义学者的学术标准降低的简要讨论见 Englebrechtsen, 1997）。

后现代主义的思潮发源于文学批判领域，该领域的学者们认为文本（一个术语，用来特指任何书面的材料或用来传递信息的图像）在没有结合不同的上下文背景时，是没有确切的含义的。也就是说，作者或者创造文本的人想要传递的意思是不确定的；所有的内容都是个体对于文本的个人反应。极端后现代主义的两个核心信条是：（1）没有客观存在的外部现实，我们每个人都在建构属于自己的个人现实世界。（2）各种解释之间的逻辑性矛盾不是问题，因为知识的有效性原本就取决于我们愿意相信它有效的程度。

1996年，Alan Sokal 写了一篇题为“超越边界：迈向量子重力论的一个转换诠释学（Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity）”的论文，题目晦涩难懂，内容也十分令人费解。在这篇文章中，他运用了后现代主义文学批判的工具来解释物理学上的复杂问题，在对科学的批判性攻击中着重强调了世界的主观性。文章顺利地通过了审阅和同行专家评议，最终发表在一本顶尖的后现代主义研究期刊——《社会文本》（*Social Text*）上。文章发表一段时间后，Sokal 公布了一个令众人大为震惊的消息：这篇文章只是一个恶作剧。并且，他还详细地说明了自己是怎样有意误用科学和数学概念，构成不合逻辑的、毫无意义的荒谬论点，并得出错误的结论。

Sokal 问道：作为一本著名的后现代主义研究期刊，《社会文本》的审稿人和编辑怎么会如此轻易地受此愚弄？是不是由于他们个人赞同他的结论，而且在情感上支持他无理由地引用后现代主义学者的学术著作？故事并没有结束，Sokal 和 Bricmont 随后又写了一本书，来详尽阐述这一恶作剧事件，书名是“时髦的废话——后现代主义知识分子对科学的误用”（*Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*）。在那本书里，他们总结并讨论了一种不在乎、怎样都行的态度在学术领域造成的令人不安的后果：科学、数学以及任何偏重客观性或可证实的外部现实的东西，都一律被藐视和误用。

通过思考“是否存在没有外部客观现实的可能性”的问题，我们就可以对外部现实的存在与否进行一次最令人信服的论证。事实上，是否真的有人相信后现代主义如此空洞的“没有外部客观现实，只有主观现实”的理论，的确很值得怀疑。因为该理论既不能解释我们所体验到的世界的一致性或连贯性，也不能为任何抉择过程提供依据，更不会对我们的行为得出任何逻辑上的结论。在现实生活中，我们是依照某种确定的顺序来获得信息的，而且我们的行为也会产生一些可预测的结果，就是这些帮助我们进行决策并计划将来，而缺乏外部客观现实就会导致完全没有任何方向的情况出现。

如果真正相信外部客观现实不存在的话，还会导致教育变得毫无意义。如果没有现实，也就是说没有任何外在的东西在那儿等着我们去了解和探索，那我们当然就不能教授或者学习任何超出我们自己直接的个人经验的东西。因此，假如那些相对较少的支持极端后现代主义观念的人受雇于学术机构，情况就会变得既有趣又麻烦。有趣的是他们选择了一个以他们的观点看来完全无用或者说徒劳的方向来作为毕生的事业，麻烦的则是他们使学生也受到了这种毫无根据的观点的影响。

后现代主义对于逻辑性矛盾的容忍，是它的另一让人感到迷惑之处。如果事实的确是“真理只是相对的”，那么对某一人来说是对的事情对另一人来说就可能是错的，问题是在这种情况下如何才能建立和执行社会功能呢？想象一下，我们的刑事司法系统中就可能出现这样的情况：被告坚持无罪申诉，理由是这才是与他自己的个人现实最协调一致的方式；而陪审团则根据他们的个人现实来评估所有证据，最终得出该被告有罪的结论。显然，司法审判过程根本容不得这样的逻辑矛盾。

再让我们来看看在医疗领域应用后现代主义的思考方式又会如何吧。下面是可能出现的一种情况：一位医生认为你体内有一个恶性肿瘤，必须切除；而另一位医生则认为肿瘤是良性的，对你不会构成任何伤害。你可以同时接受他们两方所认为的真相吗？当然不能，他们本来就是相互矛盾的两方。你的肿瘤要么会导致躯体受损，要么不会导致躯体受损；要么会危及生命，要么不会危及生命。接受“那两位医生的观点对他们自身而言可能都是真相”这一可能性，根本就没有任何意义。

事实上，包含矛盾双方的模糊思考在任何真实世界的情境下都会分崩离析。

想要通过纯粹意志的力量，将所有观点置于同一可接受水平，从而减少或消除误差，实际上是一种可能导致危险后果的草率举动。没有人能在两种观点本身就相互矛盾的情况下，既证实了其中一种观点的理论或实际利益，也同时获得另一种观点的利益。那么，为什么我们还要忍受这种荒谬可笑的操作呢？任何现实的逻辑矛盾都有一个清楚的寓意：其中至少有一种观点是错的。

与“最容易的方法”的态度，一厢情愿的“只要感觉好”的信念，以及极端后现代主义的识时务、合时宜的态度进行对比，是此处探讨人类推理过程的一个核心原则：作为一名批判性思考者，一个根本的目的就是要试图弄清楚现实究竟是什么。而这需要我们基于科学数据，建立逻辑论证，不断寻找更接近真理的结论。

虽然如前所述，驳斥后现代主义的所有不可能的前提与错误的逻辑方式已超出了本书的范围，我还是请读者自己判断一下，什么样的思考方式最可能使我们更好地了解这个世界，从而得到理论上或实际水平的提高。科学有错误自检的机制，而后现代主义则从未提供一种检测或排除无效观点的方法。例如，你遇到一位后现代主义思想家，要求他建立一个逻辑论证，比如“用环钻在颅骨上钻孔做手术，以释放邪恶灵魂的行为，从医疗保健的角度来说是不被允许的”；或者“鹤是不能生出幼鸟的”，或者就是“地球不是平的”这样的论证。那么，在一个极端后现代主义的思想家看来，根本就不可能建立这样的论证，因为“怎样都行”。所有人的观点都被认为有“个人有效性”而予以承认，因此就不可能做出批判性的评价或者得出一般性的结论。

伪科学伪装成科学但不提供任何保证，会造成一定的危害，而后现代主义则可能引起更深远的威胁。现在人口越来越多，我们也在生活的各个领域越来越依赖于科学技术（如食物的生产和分配，通讯和信息技术，身心健康保健，监测全球环境等），因此，拒绝科学的推理就必然会阻止人类进步，将我们带回黑暗时代（欧洲中世纪的早期）。

关于“什么才是有用的，对每个人来说都不同”的谬误

这里还要提到另一种反科学的观念，是来自许多伪科学家中非常流行的一句话，“什么才是有用的，对每个人来说都不同”。也就是说，对于不同的人，要用不同的治疗方法才会有效。要注意的是，这句话所隐藏的含意是科学家们没有

认识到这一点，而伪科学家们认识到了。事实上，科学研究已经建立了调查个体差异的一套机制。比如说，你怀疑某种特殊的治疗方法对女性的疗效要优于男性，那么就可以把患者分组，将不同性别患者的疗效分开来记录，列出表格并加以比较。研究者们可以利用这样的科学研究方法检验任何潜在的影响因素，并且提供更详细的信息来说明“各种治疗方法究竟对何种类型的患者以及何种疾病最有效果”。

伪科学的从业者们提出了这个关于个体差异的特殊性问题，但完全没有任何确切的根据。说明某一特殊个体应当选择某种特殊治疗方法的证据在哪儿？就如我们在第3章里要深入讨论的，伪科学家总是试图躲进整体论的迷雾中。问题在于，当某人对所谓“整体的人”进行评估时，究竟可以做些什么？没有可靠的个体差异信息作指导，所作出的系统决策实际上是任何人都可以作的不负责任的猜测。而个体化的信息在伪科学家那里总是极度缺乏的，因为那只能通过精心设计的科学研究来获得。

虽然事实上，伪科学家们开出的所谓整体治疗的处方的确缺乏任何证据的支持，但是请想一想，“什么才是有用的，对每个人来说都不同”这种精心设计的说法是掩饰无知的多么美妙的外衣呀。它使得伪科学的从业者们可以以任何他们喜欢或愿意的方式来诊治疾病，而且任何不一致的结果或者治疗的失败，都可以用谜一般神秘莫测的、从不详细说明的“个体差异”来遮掩过去。如果考虑的是你自己的健康问题，你一定会觉得还是不要找那种言语间充斥着奇闻逸事和一厢情愿的想法的医生，而是坚持跟着一个可以用研究数据来证明其治疗方案的有效性、也可以用数据说明你的预后问题的医生。

关于“几率与单一个体无关”的谬误

在伪科学家们强调个体化治疗的必要性时，另一个常见的谬误就是，由于科学家们研究的总是成组的病人，那么科学研究的结果也就与个案无关。虽然这一说法的前提通常是正确的——科学家们总是倾向于研究组与组之间的差异，但是其结论却是完全错误的，科学研究的结果对于个案的理解也至关重要。事实上，唯一能揭示出普遍规律，再将其用于对新个体进行合理预测的办法就是：对包含各种样本、各种趋势或其他规律性的大量数据进行研究。而提出合理建议的唯一途径，则是根据以往的知识进行归纳和概括。如果没有那些普适性的法则和规

律，那么就没有人能做出命中率高于机会率的预测。

认为“在研究中发现的模式与单一个体无关”的观点实在是很荒谬可笑。如果真是这样，我们又可以从中学到些什么新东西呢？每一个新的个案都是违反规律的特例吗（如果真是这样，那么最初是如何得到这一规律的）？有必要考虑更多的变量吗？如果有必要，那么需要考虑些什么？更重要的是，你怎么知道另一些变量很重要，如果它们实际上从未被研究过的话？除非是仅凭猜测，否则我们就必须运用研究的方法，来判定是否其他变量的确很重要。

不幸的是，许多人并不去考虑逻辑性的问题，也不管否认这一点只会带来更多麻烦的事实，就声称“几率与单一个体无关”的说法是对的。他们的说法是，“将几率应用于个案是毫无意义的。现在在我面前的病人就是一个单一的个体”。但事实上，一个简单的思维实验就可以让你相信，几率与每一个个体都有关联（Meehl, 1973）。

现在请你设想你即将玩一次俄罗斯轮盘赌，并且你将要选择两把左轮手枪中的一把来玩。第一把只有一颗子弹在旋转弹膛里，有五个弹膛是空的；而第二把有五颗子弹在旋转弹膛里，只有一个弹膛是空的。你要选择其中的一把，再用枪口指着你自己的头，扣动扳机一次。这是一个单一的事件，一个不可能再重复的个案——要么你死了，要么还活着。如果你真的相信几率是与单一个案无关的话，你就不会在乎应当选哪一把手枪了。不过，我还从来没有见过哪个人真的可以这样无所顾忌。唯一合理的选枪方法就是：从被打死的几率考虑，选择六分之一的机会肯定会比六分之五的机会危险性小得多。事实上，所有人都会选那把只有一粒子弹的手枪，这也表明，我们每个人实际上都意识到了：要对所面临的实际情况做一个明智的决策，考虑几率的问题是最重要的，而且对每一个单独的个案也都是如此。

当可以用科学研究的证据来指导决策时，对它视而不见，不管不顾，其实就违反了最基本的医学伦理学原则，因为那些科研结果实际上预示着某种可能性的大小，就像我们上面提到的那两把左轮手枪的差别一样。而突然跳出来声称“几率与单一个体无关”，唯一的作用就是制造混乱。很明显，在处理自己的日常事务时，我们每个人都知道如何做才是最好的，因此很难相信真的有人发自内心地、一贯地支持和宣扬“几率与单一个体无关”，除非他别有用心。

巫毒科学与法律标准

我们在前面已经介绍了，后现代主义、推理的谬误以及其他很多这类问题的变种，都在冲击着科学的推理过程。当我们在评价某种知识的准确性时，思考一下究竟谁会在这种反对科学的过程中获益，可能是有助于我们保持清醒的头脑，并最终做出正确判断的最好办法。科学包含了一系列完善的程序，让我们可以从那些令人兴奋但却模糊不清的故事中分辨出冷静的事实。而很多人在科学研究的结论威胁到他们自己的利益时，就企图忽略或者质疑科学的方法，这一点也不让人奇怪。

2000年，Robert Park 在他的著作《巫毒科学——从愚蠢到诈欺的道路》（*Voodoo Science: The Road from Foolishness to Fraud*）中，提出了四种他称之为“巫毒科学”的理论，这四种理论都背离了真正的科学的重要原则。

首先是对伪科学（pseudoscience）的声讨，这是该书的重点内容。他列举了许多披着科学的外衣，却缺乏严格的方法学和怀疑性推理过程的伪科学例子。其次是病态科学（pathological science），一种自我欺骗性的科学。病态科学的研究者们可能没有意识到，他们是在以一种有偏差的方式来进行研究设计或数据评估。他们只看到了对其所持的理论观点有利的一面，也就是只看到自己希望看到的，因而导致了错误的结论。再次是欺诈科学（fraudulent science），指研究者故意捏造或选择性地报告那些有利于自己的信息，有意地欺骗或愚弄他人。最后是垃圾科学（junk science），其理论都建立在诸如“可能比那些已经过验证并得到支持的理论更好”的假设上。Park 指出垃圾科学经常被告上法庭。思考法庭是如何进行判决，并制止向公众传播这些垃圾科学及其相关的犯罪行为，对我们会有很大的启发。这一过程会让我们对科学的本质有更清楚的了解，也会帮助我们正确地认识到科学在现代社会中所处的地位。

1923年，美国一个巡回法庭处理了一个案件，后来被称为“弗赖依诉合众国案”（*Frye v. United States*），其中包括裁决一名证人是否有资格解释审判中所涉及的多通道心理生理记录仪（俗称“测谎仪”）的测试数据。根据法庭的判决，该证人不能提交心理生理测试的结果，因为这项技术还没有“充分地发展与完善，并在其所属的特殊领域中得到普遍的认可与接受”（更多有关多通道心理

生理记录技术的知识及功能参见第10章与第13章)。正是这种要求“普遍接受”的标准,帮助法官判断了什么才是可以接受的科学证据。70年后,美国最高法院又就此制定了一套更为明确的规则与标准。

1993年,在道伯特诉梅里尔·道药品公司案(Daubert v. Merrell Dow Pharmaceuticals, Inc.)中,起诉人控告一种由美国食品与药品管理局批准用于治疗妊娠期间晨起呕吐症状的药物——苯丹克汀(Bendectin),导致了两名儿童的先天性肢体残疾。根据Foster和Huber的综述(1997),从许多大样本的流行病学研究数据来看,并不支持“母亲怀孕期间服用苯丹克汀的儿童,发生先天残疾的风险高于母亲没有服药的儿童”的结论。在整个诉讼的过程中,问题的焦点正是:作为证人的专家是否会提出关于“怀孕期间服用苯丹克汀会造成胎儿先天残疾”的证据。在案件的一份书面陈述中,一名可能作证的专家Shanna Swan这样说道:

以我个人的观点来看,不能下结论说“事实上在苯丹克汀和肢体缺失性畸形之间没有统计性的联系”。更确切地说,是从那些研究数据中不能得出“苯丹克汀不会造成肢体缺失性畸形”的结论。(Swan, 1992)

在检验法庭的审判标准之前,请注意这一声明,它显示了该专家并没有理解在科学领域中观点的可证伪性的重要性。科学研究可以用于评估一项干预的潜在风险程度,但它永远也不能证明一种药物(或其他治疗方法)完全没有风险,理由就是任何研究都有误差的成分,误差的大小通常很大程度上受到研究样本数目的影响。

一项研究可能提出:服用某种药物会产生特别的副作用,与未服药的对照组相比,服药组出现副作用的频率可能从低于三分之一到高出三倍。这里所说的低于三分之一到高出三倍就是误差的范围。要降低这种不确定性,并对药物是否改变了副作用出现的风险进行评估,就必须进行一项样本量更大的研究,或者综合多个研究的数据来做分析。按照惯例,流行病学家对一项干预的风险进行判断并认为其临床意义显著的标准是:至少是原始风险值的两倍以上。而在类似先天性缺陷的小概率事件中,要检测到即使只是两倍的风险,也要多做几万例新生儿的研究。

就苯丹克汀这一药物来说,当时已有的资料并未说明它会导致先天性缺陷的数量增加。专家们在误差的实际范围大小方面的意见并不一致,但研究结果还是

表明，可以排除有临床显著意义的风险存在。而 Swan 的证词是在要求科学家们做不可能的事情，要他们证明没有任何风险。问题在于，在科学研究中没有人能完全消除误差，即使研究了几百万的新生儿，误差范围说明服药后胎儿出现先天性缺陷的风险不会比对照组高 10%，那么还是有出现缺陷的可能性；再增大样本，可能也只是得到风险比对照组高 5% 或 1% 的结论。无论如何也没有办法确定地说，服药以后风险增加为 0，即完全没有风险，就像 Swan 所要求的结果那样。而且，我们通常不会对每一种治疗方法都进行几百万人的研究，并检验其所有正性和负性的效应，我们只能做到能力所至之极，以有限的时间、金钱以及其他资源，来尽量缩小误差的范围，并根据可获得的最好的数据做出最明智可靠的判断。假如美国食品与药品管理局坚持说，药品公司必须提供零风险增加率的证据作为药品获准销售的依据，那么无论是过去、现在，还是将来，都不会有任何药物获得批准。

也许正是由于认识到了提供具有可证伪性的证据的重要性，美国最高法院最终裁决，Swan 的证词，以及其他准备出于原告的利益立场来解释流行病学证据的专家证词，均不予接受。而且最高法院提出，法官应当充当专家证据的“看门人”，依据一系列灵活的标准来决定是否允许证人作证。这些标准后来被称为道伯特标准，用于辨别合法的科学证据以及来自垃圾科学的专家意见。法院所提出的道伯特标准，是基于其对科学哲学的仔细研究，现将这一系列标准概述如下 (Grove & Barden, 1999)：

1. 所提出的理论或技术是否可以检验？那些没有可证伪性、不能被检验的理论，不应当作为科学依据在法庭上被采纳。
2. 所提出的理论或技术是否经过有效、可信的程序进行检验，并且得到正性的结果了吗？一项理论可以被检验并不等于它能够通过那些检验，因此需要提供相关的实验证据。
3. 该理论或技术已经经过同行评议了吗？法庭认为，同行评议的过程是将真正的科学与垃圾科学区分开来的一项必需的质量监督机制。
4. 对于该科学理论或技术来说，已知的和潜在的错误率是多少？除了要在准确性方面证明，应用该理论或技术的成功率较机会率更高之外，法庭还特别希望能对它出现错误的频率做出可信的评价。
5. 该技术或方法是否具有可控制的程序和操作标准？此标准是否最大限度

地保证了该技术的有效性？法庭认定的标准是：在某特殊个案中所涉及的理论事实，应当存在一种逻辑上的相关。如果是作为诉讼的证据，还必须能够提交关于这一科学理论与技术对于该案例的适用性的证据。

6. 该理论或技术在相关的科学领域中是否被普遍认同和接受？这与前面所说的弗赖依案中的要求是类似的，但是现在它并不是唯一的标准，只是需要满足的几个条件之一。

最高法院强调，在律师想要提交科学专家的证词时，法官必须按照上述标准询问，并作出相应的判断。提出这些标准并不是要追根究底，实际上是在绝大多数案例中要做到全面考虑所必须注重的几个最主要的问题。法庭的推理或论证过程是这样的：考虑到实际上有那四种巫毒科学的存在，要求提供关于科学的方法与科学的推理过程的证据，会对那些企图提出没有足够证据支持的观点的人起到阻止和威慑作用。设置这些标准所针对的是，那些随便按照自己的喜好来进行选择和行动、还认为可以在诉讼过程中钻空子的人，最终的目的是限制他们没有诚信的、投机取巧的行为。

Robyn Dawes (1994) 运用了一个术语来描述那些没有足够证据支持的信念——“自真空中产生的论证” (argument from a vacuum)。伪科学就总是依赖于这种“自真空中产生的论证”。而与伪科学随心所欲地提出或放弃论证不同，科学是利用我们已知现实的局限性来限制我们的思维。不过，总有些人觉得受此限制十分麻烦，他们想要得到超出现实边界的、完全意义上的自由，并且否认科学所带来的这种限制的合法性。

学习科学推理的基本原理，会帮助你很好地抵御那些毫无根据的观点所带来的危害。你不可能指望别人会时刻顾及你的利益，因此只能采取自我保护的措施。按可证伪性、逻辑性、综合性、诚信性、可重复性和充分性的标准来进行检验是科学推理能力的重要组成部分，你可以在日常生活中经常进行这种有益的思考训练。

同样，对于那些声称具有科学证据支持的各种观点来说，美国联邦最高法院所提出的道伯特标准，也是评估其价值的十分有用的方法。不过，在法律的界限之外，我们必须自己给自己充当“看门人”，区分真正的科学和冒名顶替、滥竽充数的东西。只有当你运用了收集证据的最严格的标准，才能形成和保持最正确的信念。如果所用的标准比较宽松，你就可能犯错，而且经常被人利用。

3. 语言：误导和回避战术的工具

有一次，我参加了一个销售神奇锅具的展示会。那些炒锅、平底锅之类的产品真是令人惊叹。销售者推销这套锅具时有一整套行话。在谈到与钱有关的部分时，他强调精明的消费者会购买他们的产品，因为用这种神奇的锅具，食物不会因烹饪过程而有任何缩水或损耗，“如果你买了一块五磅的肉，那么当你端上桌时，它还是五磅。食物的天然汁液和香味都会被完美地保留下来”。当时在场的很多人看起来都十分乐于接受这样的论断。然后，销售者又谈到了健康饮食，并且配合现场烹饪演示，来说明其产品是如何促进健康饮食的。他当着我们的面烹饪了一块去骨鸡胸肉，还从锅里倒出了满满一茶杯鸡油，并且说如果做的是牛肉的话，会出来两茶杯的油。

销售者接下来解释了一下关于脂肪、胆固醇以及心脏疾病之间的关系，得出了一个看上去显而易见的结论：与其他产品相比，使用这种锅具进行烹饪，会大大降低一些食物导致的致死性疾病发生的可能性。这一结论被反复强调，直到在开车回家的路上还一直在我耳边回响。需要补充说明的一点是，当看到那些难看的鸡油被盛在塑料杯里的时候，许多观众看上去已经彻底地接受了销售者的结论。

不过，现在可能你已经发现了，在销售者的上述论据中，存在着一个明显的逻辑上的自相矛盾：不可能同时做到既完整地保留食物的天然汁液和香味，又榨出那么多油。

模棱两可的语言和架构效应

虽然正如前文所说，锅具销售者所声称的这两件事是绝对不可能同时做到的，但它们实际上是在销售的不同阶段分别被提及，而且各成一个独立的环节，因此它们之间的矛盾就被巧妙地隐藏起来了。除了在整个约 30 分钟的销售过程中，将它们进行时间上的分隔，更重要的可能是销售者的语言滥用。

如果“脂肪”这一术语在两个部分都被提到的话，就可能非常容易被发现：同一物质是不可能同时被保留又被去除的，因此销售者在不同时间点使用了不同的术语。其目的除了要有意地模糊这一自相矛盾之处外，还要促进人们作出他们所期望的反应。当要论证肉汁是否应该被保留时，用一种正面的方式提到它会比较有利——当然每个人都希望能够使食物更加汁鲜肉美。同样还是那种东西——肉汁，当焦点转向如何消除它时，就必须以一种负面的形式来进行引导了——时至今日，还有谁不厌恶脂肪和胆固醇呢？这种企图引导人们进行粗浅的思考和判断而故意滥用的语言，有一个术语就是模棱两可的语言（weasel words）。

上面这个例子可能会让你联想起广告中也有类似的情况，也是通过有意的语言操作，最终达到说服人们购买商品的目的。事实上，模棱两可的语言的作用不止于此，它对个人所坚持的信念也会产生影响。Rugg 曾经做过一个实验询问人们对于自由言论的看法（1941）。参与实验者被随机分为两组，两组都被问到同一个问题，但提问方式略有不同。对其中一组问的是“您是否认为美国应该允许有关反对民主的公开言论”，有多数被试（62%）的回答是“不”；而对另一组问的是“您是否认为美国应该禁止有关反对民主的公开言论”，仍有多数被试（54%）回答是“不”。被试既认为不应该允许发表反对民主的公开言论，又认为不应禁止这些言论，看上去似乎十分矛盾，那么应该如何来解释这一实验结果呢？

仅仅是替换了一个词，就影响了个体对这个重要问题的概念形成，这不得不说是十分令人惊讶的。询问“是否应该允许发表某种言论”，是从言论内容的可接受性来构架整个问题的；而询问“是否应该禁止发表某种言论”，则是从限制个人自由的角度来构架整个问题的。心理学上将这种针对可察觉的个人自由的丧

失所产生的强烈反应称为“心理抗拒”*（psychological reactance）。我们每个人都特别在意个人的自主性，因而对于这方面的威胁显得特别不能容忍，哪怕是比较温和的危险信息也会引起强烈的反应。

在关于堕胎或人工流产的争论中，也存在同样的问题。当争论的一方强调“选择优先”（pro-choice）时，就是企图激起我们本能的关于维护个人自由的倾向。而当另一方宣称“生命优先”（pro-life）时，则是企图激发我们尊重人类生命的神圣感情。通过使用这样的标签，争论的双方以一种高度暗示性的方式来有意地构架问题，使得这一问题看起来根本没有反对的可能。而实际上，从一般的哲学角度来考虑，你认为会有人反对选择或者反对生命吗？

在这里，堕胎问题的争论双方狡黠地运用了心理架构**（framing），使人们的注意力集中于某一相关语义的特定方面，其前提就是：即使是会产生深远影响的信念，也会因为言语的简单变化而发生转变。但这还不是心理架构带来的所有问题。Tversky 和 Kahneman（1981）进行了一项经典的实验，结果表明心理架构在特别重要的决策上也会产生影响。实验的参与者都被要求进行一个艰难的选择：

假设美国正在准备防御一种罕见疾病的爆发，预计会死亡 600 人。现有两种不同的处理方案，对于这两种方案的结局进行了如下准确、科学的估计：

- 采纳 A 方案，200 人将获救。
- 采纳 B 方案，三分之一的可能性是 600 人获救，三分之二的可能性是没人获救。

在你继续阅读之前，请稍做考虑你可能会选择哪一种方案？如果你已经有了决定，那么请按下面这两种新方案再重新选择：

- 采纳 C 方案，400 人将死去。
- 采纳 D 方案，三分之一的可能性是没人会死，三分之二的可能性是 600 人死去。

* 心理抗拒是一种对抗外部控制措施、企图恢复自由感的动机状态。——译者注

** 心理架构指人们做出选择的行为，会受到对问题所作的不同陈述或框架方式的诱导，产生有系统的偏好改变的现象。即一个问题的两种在逻辑意义上相似的说法会导致不同的决策判断。——译者注

现在，在继续阅读之前，请再次考虑一下你会选择哪一种方案？

Tversky 和 Kahneman 发现，结果十分令人困惑，受试者对这两个版本的问题反应差别极大：当在方案 A 和 B 之间选择时，有 72% 的人选择方案 A；但在方案 C 和 D 之间选择时，有 78% 的人选择方案 D。要弄清楚为什么这种结果会令人惊讶，你需要注意到，其实方案 A 和 C 是一样的（600 人里有 200 人将获救就意味着 400 人将死去），只是它们的措辞略有不同。同样，方案 B 和 D 也是一样的（三分之一的可能性是 600 人全部获救，三分之二的可能性是 600 人全部死去）。因此，上述实验的选择结果说明，被试在面临相同的问题时，做出了逻辑上完全不一致的两种选择。

对于这一研究结果，Tversky 和 Kahneman 从心理学的角度进行了解释：当从获得（拯救生命）的角度对问题进行措辞时，人们通常会规避风险，选择确定能救 200 人；而当问题是从丧失（生命死亡）的角度来进行描述时，人们则往往追求风险，选择碰碰运气，或者尝试冒一下险来挽救所有人。但是，这一现象不得不让人感到忧虑：仅仅是这种表面的问题架构方式的差异，就会在可能导致极其严重后果的问题上影响决策。因此，了解语言的架构效应对于我们来说是非常重要的。

Kahneman 和 Tversky（1979）提出了预期理论（prospect theory），用于解释架构效应如何导致我们做出表面上自相矛盾的选择。要理解预期理论所涉及的两个原则，请思考下面的两个问题：

1. 哪种情况让你感觉差异更大，是零和一百万元的差别，还是一百元和一百零一万元的差别？
2. 假设你几乎没有钱，下面哪种情况可以激起你更强烈的情感反应，是得到 1000 元，还是损失 1000 元？

绝大多数人报告零和一百万元之间的差别在心理上感觉比一百元和一百零一万元之间的差别更大，而损失 1000 元所造成的伤害其情感反应远大于得到 1000 元所带来的愉悦感受。这两种反应都是基于预期理论的原则。图 2 所示的 S 形曲线，就是用于描述实际的获取或丧失（客观现实）及其对于人们的意义（我们对价值的心理感受）之间的关系。

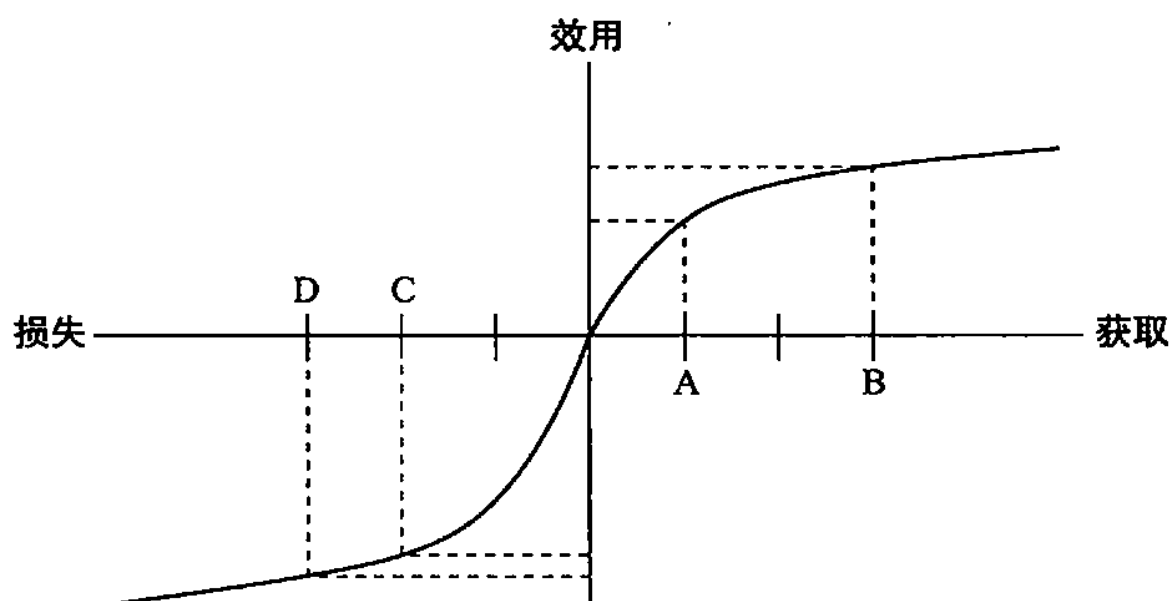


图2 预期理论的S形曲线说明，人们将初始的获取或损失看得比增加的获取或损失更重要，而且相对于获取而言，更看重相应的损失。

首先请看图2中的实线，曲线中段是最陡的，两端则趋于平缓。这一趋势表明，人们在进行决策或判断时，会认为初始的获取或损失（例如零到一百万元之间的差别）要比从远离其参照点的地方开始的实际获取或损失（例如一百元到一百零一万元之间的差别）重要得多。另外，请注意曲线的斜度在损失范围内比在获取范围内要陡，也就是说，在我们的思考和评价过程中，丧失（例如损失1000元）被认为比同等量的获取（如获得1000元）更受重视。这两个原则可以帮助我们解释，为什么我们在处理获益时总试图规避风险，而在处理丧失的问题时则倾向于追求风险。了解了这两个原则，让我们再来看看上面那个挽救生命还是丧失生命的悖论吧。

从A到D四种方案，在图中以虚线标识。我们首先来看看方案A与B之间的选择：方案A所说的是确定能救200人的命，这是一种相当高的效用。方案B则提供了可以救600人的可能性，但是请注意这种结果的效用并不是200人获救的三倍，事实上，整体来看，还根本没有高过方案A。如果方案B失败的话（而且失败的几率还是三分之二），就没有一个人能获救，这一结果的效用远远低于200人获救。人们会意识到，如果冒险的话，可能的收益很少，可能的损失则很大，因此选择方案A，这是为了获益而采取的规避风险。

但是，当我们比较方案C与D时，理解就完全相反了。根据方案C，确定有400人会丧命，这是一个很强的负性效用。而根据方案D，我们没有觉得600人全部丧命的风险与400人丧命的风险有很大不同——它们的效用是接近相等的，但还存在一种无一丧生的可能性，这种结局的效用则是非常大的。选择冒险只会

带来很少的损失，而可能的获益则很大，因此人们会选择方案 D，这是面临损失而采取的追求风险。

合作以及对话准则

除了模棱两可的语言和架构效应带来的影响以外，我们还容易受到其他语言陷阱的不良影响。绝大多数普通的对话都会受到一种隐藏的合作目的的影响，而伪科学家们就是利用了这种假设他们会遵守基本会话原则的心理，实际上他们并未遵守。Grice（1975）描述了人们交流时通常为了显得合作而遵循的四条法则，也被称为对话准则：

1. 质的准则：我们是诚实的，所说的话有足够的证据做基础。
2. 量的准则：我们提供交流所必需的信息，但不会事无巨细，超出所需要的信息量。
3. 关联准则：我们所提供的信息与当时的谈话主题有关。
4. 方式准则：避免含糊不清和歧义，保持简练、清晰、明白。

上述这些对话准则不仅会促进合作性的会谈，而且也是良好的科学交流的重要基础。但伪科学家们无一例外，总会违反其中的一条或几条准则。下面让我们来看看如果有意违背这些准则会产生什么后果。

质的准则：“健康食品”的神话

我们首先要从彻底的不诚实说起。伪科学家们由于缺乏证据，往往忽略对话中的质的准则，利用语言方面的伎俩向公众提供虚假的信息。在现代生活中，构架得最精巧、极富欺骗性和含糊性的短语可能就是“健康食品店”了。要注意的是，所谓的健康食品，就是含有营养物质，如蛋白质、碳水化合物、脂肪、维生素及矿物质等的食品。就这一含义来说，一个普通的杂货店里所卖的绝大多数食品，其实都是“健康食品”，而合理、均衡地摄入这些食品，就可以满足人体的所有需要。说到这里，你可能会想，那么营养品就应该是前面所列的那些物质了。不是这样的，在健康食品店里售卖的许多营养品的化学成分都超出了单纯的营养物质的概念，厂商声称它们可以取得单一营养物无法达到的显著疗效。例如：

1989年，美国联邦贸易委员会通过了一项决议，禁止印第安纳州埃尔克城的Miles公司为其生产的One-A-day牌多种维生素所做的未经证实的宣传。根据联邦贸易委员会的陈述，Miles公司在广播和电视上宣传其产品时说道：“要保护你的肺部不受空气污染的影响，需要维生素A、E和C。而维生素B则可缓解你的日常压力；跑步等高强度的身体锻炼还会使人体系统中必需的矿物质流失。One-A-day牌多种维生素具有独特的配方，可以补充上述流失或不足的物质，帮助你进行身体的调节。”（Barrett & Herbert, 1994）

事实上，根本没有证据支持上述陈述中的任何一条。即使是试着从“这条广告也许更多的是出自比喻，而不是学术”的角度来理解，关于人体系统会流失矿物质之说也很难用目前所知的人体生物学的机制来解释。令人悲哀的是，现在诸如此类的观念已经借健康食品之名被广为接受，而实际上那只不过是厂家和消费者的一个借口，用以回避自己是在卖药或者买药的事实。

我们在第5章中还会详细讨论，许多人都不清楚，1994年出台的一项联邦法律已经规定，任何冠以“食品或营养品”的药物，如果没有经过美国食品与药品管理局的许可，都不允许出售。政府机构所能做的仅仅是在出现了不良效应之后，从消费者的利益角度提出诉讼，而不是在产品投放市场之前要求其出示安全性方面的证据。Robert Park曾在美国物理学会2000年6月23日的时事通讯中提出了一个重要的问题：像这种马后炮式的策略，是一种对公众健康问题行之有效的管理方法吗？显然，消费者在购买那些贴着营养品标签的未经检测的化学药品时，必须非常小心谨慎。

当那些健康食品的厂家在向不知情的消费者强行推销其产品时，其自由程度令人瞠目结舌。例如，由于当局对美国通用营养中心（General Nutrition Center, GNC）所采取的法律强制行动完全是例行公事，以至于GNC公司已经将这一项列入了其日常开支。在过去的30年里，美国联邦贸易委员会、美国食品与药品管理局都只是每年一次对GNC公司进行大体上的、粗略的谴责。而在1994年美国联邦贸易委员会的法案当中，GNC公司同意付240万美元作为解决其41个产品受到关于虚假广告控诉的代价。类似“健康食品”和“膳食补充剂”之类的欺诈性术语给了那些厂家相当大的空间，使他们可以打擦边球，炮制出实质上与虚假广告无异的一套市场营销计划，却又不至于受到法律的制裁。

量的准则：暗示性问题与诱导性问题

在我们讨论这个问题之前，你可能会认为通过略带暗示性的方式交流过多的细节，并不见得会产生什么害处。但实际上，违反量的准则提供暗示，可能会以破坏性的方式干扰我们的记忆。Loftus 和 Palmer (1974) 进行了一个关于诱导性问题的经典实验：实验的被试先看一段关于一辆轿车与一辆卡车相撞的车祸录像，然后回答一些问题。当要求他们估计当时的车速时，听到“当轿车碰到卡车时，你估计车速有多快”的问题时，被试回答结果的平均值为 31.8 公里/小时；而听到的问题是“当轿车撞到卡车，撞个粉碎时，你估计车速有多快”，被试回答结果的平均值为 40.8 公里/小时。很显然，与相对更中性的“碰到”一词相比，“撞个粉碎”的说法在被试的头脑中诱导出了一种不同的画面。也就是说，这种说法提供了一些额外的细节，而这些细节与实际情况可能相符，也可能不相符。除了这个估计车速的问题以外，Loftus 和 Palmer 还做了另一个实验：一星期后，被试者被问“你是否看到车上有任何玻璃碎片”，结果是“撞个粉碎”组所报告的肯定结果要高于“碰到”组，而实际上在这场事故的录像中并没有任何玻璃碎片出现。上述实验过程中，错误记忆就是其中的影响因素。

由于我们的记忆有时是混杂在一起的，因此错误的记忆如果和已有的记忆看上去一致的话，就有可能被当成真实的情况。为了检验这种可能性，Bransford 和 Frank (1971) 做了一个实验，他们在被试者的面前呈现一组句子：

- 蚂蚁吃了放在桌上的甜果冻。
- 蚂蚁在厨房里。
- 蚂蚁吃了甜果冻。
- 厨房里的蚂蚁吃了放在桌上的甜果冻。
- 桌上有甜果冻。
- 厨房里的蚂蚁吃了果冻。

五分钟后，被试拿到一组新的句子，然后被问到其中的哪一句曾在前面的那组句子中见过。有趣的现象出现了：无论一个句子是否在前一组句子中呈现过，被试都十分自信地选择了那些包含更多细节的句子。例如，实际上是“蚂蚁吃了甜果冻”这句话曾出现在第一组句子里，而不是“厨房里的蚂蚁吃了放在桌上的甜果冻”，但被试都会肯定地认为是后者。这就是运用细节造成了关于这一句

子曾被看到的虚假记忆。

如果看到这里，你还认为暗示性或诱导性问题所导致的仅仅是一种微不足道的影响的话，那么就请再看一看细节性的暗示带来的更为严重的记忆扭曲后果吧。在 Elizabeth Loftus 及其同事所做的实验中，被试“错误地认为看见了实际上并不存在的玻璃碎片和磁带录音机，将一名无须男子当成有胡须，将直发记成了卷发，甚至田园中像谷仓那样大且显著的标志都会被当作没有”（1997）。下面这些实验中出现的虚假记忆就更匪夷所思，具有戏剧色彩了：看见过一次毒品搜查（Nucci, Hoffman, et al. 由 Loftus 引用，1997）；五岁时在购物中心、百货商店或其他公共场合被弄丢（Loftus, & Pickrell, 1995）；由于可能存在的耳朵感染导致了高热，头天晚上接受了住院治疗；或者参加了一个提供比萨并有小丑表演的生日聚会（Hyman, Husband & Billings, 1995）；参加了一个婚宴，并失手将一杯潘趣酒洒到新娘父母的身上；或者在出现错误的消防警报时，疏散了整个杂货店的人（Hyman & Billings, 1995）；将某人的手夹到捕鼠器中，不得不上医院去把它弄掉；或者与同学一起乘热气球旅行（Ceci, Huffman, Smith & Loftus, 1994）；亲手打碎了一扇窗户（Garry, Manning, Loftus & Sherman, 1996）等等。在这些实验中，研究者都是事先设计了一种记忆，并通过细节性的暗示，成功地将其移植到了被试的头脑中。而且，许多被试对所获得的虚假记忆都描述得栩栩如生，详细具体，并且描述者本人自信满满。看了这些通过暗示性的细节造成记忆扭曲的例子，那些对目击者证词或“重现的”记忆信以为真的人应当会感到有些踌躇不安。

关联准则：补充和替代治疗中不相干的语言

如今，由于科学在治疗身心疾病方面已经取得了显著的成就，要推行一种有问题的、未经验证的治疗方法应该说是很复杂、很困难的。但是，有许多伪科学的从业者与病人恰恰相反，他们轻视有实验依据的健康保健措施，赞同那些补充和替代治疗的措施。这些人总是运用一些伪医学的术语来逃避法律的责任与义务，也因此违反了关联准则。那些宣扬替代治疗方法的伪科学家们，惯于用一些不相干的事物来骗取信任，并据此在健康保健领域开拓出十分可观的市场。例如，他们会将那些有实验依据的，或者说科学的健康保健措施说成是“传统的、常规的、固定模式的，或者西方的”，而称其未经确认的治疗方法是“选择性

的、补充的，或者综合性的”措施，是重新发现的、所谓东西方结合的新方法。请注意，上面所说的词语中，没有一个与该治疗方法是否有实验依据支持的事实有关。伪科学家们就是这样利用从根本上歪曲的语言，制造出他们想要人们形成的印象——当前整个科学事业都是落后的、因循守旧的、死气沉沉的，拒绝任何改变，只接受一种文化观念，而且对整个人类来说可能都是弊大于利。而现实情况呢，当然根本不是这样。

补充和替代治疗措施的存在，仅仅意味着它是多种具有同等价值的选择中的一个。个体可以选择某种治疗方法，也正好满足了民众在深深的民主意识的影响下，对于选择自由度的需要。但是，这种选择的真正本质是什么呢？著名的《新英格兰医学期刊》的一位编辑对于这种未经检验和控制的治疗方法所带来的危害进行了探讨，并且得出了以下结论：“根本不能将药物分为所谓‘常规药物’和‘替代性药物’，只能说是已通过适当检验的药物和没通过检验的药物，或者能治病的药物和可能有用也可能没用的药物。”（Angell & Kassirer, 1998）那些连简单的“给我看研究证据”的原则也满足不了的治疗方法，就不是什么替代治疗，而是健康保健的替代品。

“传统治疗”的术语指的是那些最初源自口传身授，然后忠实地流传下来的治疗方法。具有讽刺意味的是，这一描述其实精确地说明了替代治疗措施上百年、甚至上千年来从未改变的特点。例如，传统中医当中的中草药学与针灸学就不存在所谓“重新发现”的问题，只是一些很少接受并通过直接检验的古老观念。而与之相反，科学的健康保健是一门相对较新的人类成就，它包含了大量20世纪以后出现的研究内容，而且还在根据更新和更好的研究数据不断进行修正。这绝对是“非传统性的”：科学提供的是不断更新、不断发展的新的医学。

“标准治疗”或“常规治疗”的术语，或者“医学既成模式”，往往意味着一种不灵活的、严苛的、过于保守的系统，但其实这些术语的法律解释是更准确明晰的。这些语言表明有支持性的研究证据存在，而据此建立的措施是一种安全有效的治疗措施。

认为科学的健康保健措施就是西方的，这种想法存在许多误解。所谓“西方的”这一术语，在目前这样一个越来越意识到并重视文化多样性的年代里运用时，看起来就是指科学反对其他观念的存在。其实真实的情况完全与之相反：科学推理的标准是没有国界的，无论是美国还是中国的研究者提交的数据都会受到

同等的对待。恰恰是替代治疗者经常因信奉“东方的”治疗方法，反对“西方的”治疗方法，根本无视已有的科研证据，因而显露出其文化偏见。请设想如果一个人真的将所有文化的观点都置于同等的标准之下，那又怎么能断定其中的一种是完全没有价值的呢？换句话说，为什么西方的保健措施就应该遭受忽视呢？这种明显的矛盾，在目前科学研究已取得了显著成就的事实背景之下，尤其让人难以理解。

“东西方结合”这一术语，不仅仅是在表述上有问题，而且它与“补充的”或者“整合的”治疗措施这些术语一样，都是内在缺陷的某种借口或伪装。这些相对较新的术语，其实是针对替代治疗的不足所做的让步或妥协，暗示补充和替代治疗措施虽然单用可能无效，但将它与科学治疗措施相结合的时候就可能会有某种价值。这简直就是废话。他们所谓的补充治疗或者综合性治疗，仅仅是在某种已知有效的治疗措施（科学治疗措施）上，加了一些尚未经过验证的成分（替代治疗措施）而已。

最后我们还要提到一种语言滥用，它使得替代治疗者逃避了他们的行为所应负的法律 responsibility。简单来说，就是无照行医是非法的，而宣扬无确实依据的健康观念就构成了犯罪。虽然许多伪科学家经常由于主张错误的观点而违反了法律，但也有一些人小心翼翼地避开了诊疗科学用语，以某种合法的借口和遁词取而代之。

一些江湖医生从不说治愈疾病，而是许诺可以“解毒”、“净化”、或者“使机体恢复元气”；“平衡”机体内的化学物质，增进机体的和谐或协调性；“刺激”或“增强”机体的免疫系统；对体内各种器官提供“支持”或“使其恢复活力”；或者激发机体自愈的机能等等。当然，他们从来不会对上述过程进行甄别，或者进行前后对照的有效数据测量。这些庸医永不承诺只是出于两个目的：第一，既然他们所宣称的治疗过程根本不可能测量，那么你也就很难证明它是错的；第二，由于他们并不是内科医生，因此运用非医学术语也许可以帮助其规避无照行医的起诉——尽管这样做是不对的（Barrett & Herbert, 1994）。

这类回避性的语言除了可以用来与法律打擦边球之外，没有其他任何意义。那么，究竟是什么原因导致人们会推销未经证实的产品，或者发表尚未获肯定的

主张呢？健康保健市场里总是充斥着不同程度的欺诈行为，而保护自我的措施之一就是：当评价某种治疗方法的安全性与有效性时，要求对方提供证据，并且保持谨慎，避免被那些无关的言语所误导。

方式准则：“整体论”的空洞无力

方式准则要求我们在进行交流时保持言语的简洁和明晰。而那些鼓吹“整体论”的人，如从事整体语言介绍、整体医疗保健，或者整体占星术等的人，则故意违反这一准则，言语表述相当晦涩和含糊。整体论的观点实际上是摒弃了无数具有不可估量的价值的差异性。Gilovich（1991）曾引用 Roger Lambert——美国当代著名文学家 John Updike 的《罗杰教授的版本》（*Roger's version*）一书当中的主角——所说的话来说明这一问题：“除了不确定原则之外，最近这几年我还十分厌恶‘整体的’这个术语。这是一个毫无意义的表述，只会混淆两千年来人类思想所形成的一切有用的差异性和特征性，制造出一片混乱。”

在这里，我们可以用一个清楚明白的类比来说明整体论的主要谬误：如果你的车不幸损坏，你是希望机械师针对某一特殊部件的问题进行评估和修理，还是希望他整体地修理全部汽车部件？我想选择是不言而喻的，而这一选择就很能说明我们要讨论的问题。

美国兽医协会曾在其网站上刊登了一篇关于全面促进对宠物进行整体治疗的报道，其中做了一个常见的，但十分可笑的、歪曲的类比。虽然那篇文章的主题是赞成唇部护理是有实验依据的治疗措施，但通篇都表现出了赤裸裸的对科学的攻击。在提到“常规治疗仅仅是用来缓解症状的”之后，文章又做了一个明显针对科学的类比：“想象一下，如果汽车油量低的警示灯亮了的话，关掉这个灯的确是可以让警示信号消失不见，但实际上问题解决了么？”这等于是在清楚地暗示科学家们是那么的无能，只会“关掉警示灯”而不会“把油加满”。但问题是，这辆车并没有整体崩溃才是事实所在，那么持整体论的汽车修理师究竟想干什么呢？那些从事整体治疗的兽医对刚刚所提到的这一点保持了沉默。因此，即使用来诋毁科学的例子看上去毋庸置疑，包含了最聪明的逻辑和行动的理由——有效地找出了特异性的问题所在（油量低）和处理方法（加油），但最终由于其措施并没有实际的落脚之处，整体论还是陷入了一纸空谈。

整体论在许多涉及众所周知的健康原则的情境中，也显得毫无意义。例如，

你摔断了腿以后，是想将腿打上石膏，然后借助拐杖走路直到治愈呢，还是想让一位并不直接针对你的腿，而是将你作为完整的个体来考虑的整体治疗师来治疗呢？整体治疗师采取的方法可能是针灸、草药、同种疗法、磁疗、营养补充、反射学、治疗性接触，以及其他治疗措施的某种综合。换句话说，对于一条断腿是应该直接进行针对性的治疗呢，还是将其作为更广泛的问题加以考虑，进而采取更全面的治疗措施呢？即使是最坚决的整体医学拥护者，当考虑的是他自己的腿时，可能也会选择打石膏、拄拐杖。

不过大多数情况下，人们考虑的并不是那么直接就可以看透的事情，因此这种保持有意义的区分的逻辑要求就变得模糊不清了。那么，为什么必须进行有意义的区分呢？例如，一位现代占星家宣称其必须借助整体天宫图进行占卜，这恰好说明了整体论的无用。占星术的基本理念是这样的：所有自然现象都会受到各种天体的影响。按照 Kelly 所说，这其中不无道理：“每次当太阳升起，我们就醒来；或者我们会在月光普照的晚上准备烧烤，在涨潮的时候外出捕鱼，这些都是天体切实影响我们生活的最好例证。”（1997）但问题是，这些影响都是一些日常琐事，与占星术所做出的整体性的结论根本不是一回事。许多世纪以来，占星家们总是被当作重要人物，人们向他们咨询，在他们的帮助下做出重要的现实决策，例如：我是否应该与这个人结婚？是否应该与这个国家作战？诸如此类。大约从 20 世纪 50 年代起，科学界开始注意占星术的问题，并对占星家们的预言进行了检验，结果表明占星家预言的准确率根本不比胡乱猜测的机会率高（Dean, Mather, & Kelly 的综述，1996；Kelly, 1997, 1998）。

面对压倒性的负面证据，占星家们面临着一个艰难的抉择：是放弃他们广受怀疑的信念，还是创造一种事后处理性的解释？占星家们选择了一系列回避战术，其中就包括了整体论，而它们才是真正的伪科学的形式。现代占星术根本不理会那些不利证据，其原因就如批评家 Kelly 所说：“天宫图是一个整体的系统，其中任何一部分的变化都会影响到其他各个部分。”（1997）也就是说，占星家们反对科学研究每次只研究一个有关占星术的因素，认为这样做不是从“整体天宫图”的角度考虑，而“整体天宫图”据说恰恰是专业的占星家赖以作出预言的根本。

占星术与其他所有的经验性问题一样，其致命的缺陷就是采取了整体论的防御措施。人们可以直接询问关键的一点：这个“整体天宫图”到底包括什么？

根据其整体性来考虑问题的要求，是否仅仅只是一种心理学上的可能，实际上却根本办不到？

“整体天宫图”的边界究竟在哪儿？无论是十大行星（占星家们认为太阳与我们的月亮是行星，而我们的太阳系里其他的许多月亮并不算行星，地球也未被他们算入十大“行星”之列），还是十二星座、十二宫、中点、阿拉伯点，交点、相位，还是其他任何占星术所用的概念及术语，都决不可能解释清楚什么是“整体天宫图”。当占星家们声称他们运用了整体天宫图来预测时，仅仅是指他们综合运用了多种因素，而不是仅仅考虑一种因素。然而，无论他们同时考虑了多少因素，这些因素的数量还是有限的，因此只能说这只是天宫图的一部分。可以这么说，从来就没有人用过整体天宫图。那么，新的问题又出现了：如果不可能考虑到整体天宫图或者说全部的因素，那么究竟考虑多少因素合适呢？又应当考虑哪些因素呢？……假设我宣称自己同时考虑了20种因素，毫无疑问，马上就会有一位占星家站出来说我应当考虑21种（Van Rooij, 1994）。

根据一个与占星有关的因素的列表，Kelly（1997）计算出来大约存在 10^{28} 种天宫图。那么，即使不考虑计算的细节，一位占星家计算的次数也会达到 5.39×10^{68} （Doane, 1956）。这两个数字十分巨大，已经大大超过了曾经存在过的人类的数量了（大约为 10^{10} ）。因此，可能几乎每一个人（也许双生子例外）都会有一个独一无二的天宫图。这也就否定了关于占星术可用于预测的任何可能性：如果我们当中的任何一人，在出生前都没有人有过与他相同的天宫图，那么占星家怎么知道可以预测些什么？唯一的线索就是在天宫图中与个人有关的因素（这是与其他人有共同之处的地方），但是占星家们又不能用这些来解释其推理过程及预言，因为这样就不是根据整体天宫图来做预测了。事实上，整体论中“每一样东西都会对其他任何东西产生影响”的说法，总是会导致这种存在致命缺陷的悖论。

在第14章里，我们还会更为详尽地讨论到，实际上所有的整体论措施与进行决策的临床方法一样，存在着严重的缺陷。人类所作的任何独立的决策，都不可能有效地处理大量复杂的信息（Ruscio, 2003b）。当你要求整体论治疗家们提供证据，说明他们是否运用过任何一致性的决策策略时，他们总是坚决拒绝做出

如何综合考虑各种信息或事件的说明。整体论实际上是无意义地回避现实，它只是伪科学家们用于扰乱合理的思考，回避简洁明晰的交流，以达到只按自己的怪异信念随心所欲行事的目的是的一种方法。

最后的思考：错误的两分法和滑坡谬误

当你怀疑有人想用欺骗性的语言蒙蔽你的时候，你应当好好弄清楚他真正的意思是什么。我们已经讨论过那种策略性的问题架构，以及精心设计的违反会话准则的行为，它们可能都是只对语言进行微小的调整，但却会产生十分严重的后果。现在让我们再来看看语言陷阱影响思维过程的其他方式。再用堕胎问题来举例：有一天我在开车的时候，看到了前面汽车保险杠的贴纸上的标语，最上面的一行写着“微笑”。我加速上前想看个清楚，最后看到了底下一行“你的妈妈是生命优先派的”。这实在是一句空话，即使作为保险杠贴纸上的标语来看也是如此。所谓“生命优先”，就如“选择优先”一样，是一个空洞无物的词，整个表述的逻辑本身就存在缺陷。事实上，去猜测每个人的母亲是不是持“生命优先”的观点毫无意义，许多崇尚自由，认为应当“选择优先”的女性也都有孩子。难道你认为所有认同“选择优先”的妇女每次怀孕都必定堕胎吗？所以老实说，那条标语实在是愚蠢之极。

就堕胎问题来说，实际上凸显了最后的两个语言陷阱，这在人们试图说服别人的时候经常也会用到。首先，“生命优先”还是“选择优先”的议题可以说是一个“错误的两分法”的问题。无论你的父母、老师、朋友，以及大众媒体，在判断应当允许堕胎还是禁止堕胎的问题上，是如何推崇这种简单的两分法原则，你还是应当看到在所有的情况中，都存在第三种或其他的选择。例如，某人可能在某种情境下赞同“选择优先”（允许堕胎），比如在讨论强奸后的怀孕，或者怀孕过程将危害母亲生命的情况时；在另一些情况下则可能赞成“生命优先”（反对堕胎），如当人们并不是出于人口控制的需要，而是由于胎儿性别的原因要堕胎时。Sagan（1997）在与 Anne Druyan 合写的一篇文章中提出，我们每个人都反对的，实际上是杀死一个生命。这又迫使我们弄清一个复杂的问题：堕胎对于人类来说究竟意味着什么？Sagan 和 Druyan 持允许堕胎的观点，他们提出我们之所以被称为人类，是由于存在某些特殊类型的思维加工过程，然后列举

了一些证据，说明这些独特的思维加工至少要在胎儿发育到七个月后才会出现。这样，他们将堕胎问题的争论一直进行到了“并不会产生一个人类生命”这个点上，然而这已经超出了在这个点上所能讨论的范围。

在这里，我并没有对一些特定的看法采取赞成、批评或其他态度的意思。我的目的只是要指出，你根本用不着去赞同一个错误的二分法。在表达自己的观点和信念之前，你可以对许多可能性进行更为深入的思考。为什么还有如此多的人非要将自己划分为“生命优先”派或者“选择优先”派呢？一种可能的原因是他们害怕出现滑坡谬误*。在那种情况下，一方任何微小的让步都会被另一方（与他们持相反立场的人）所利用，用一句老话来说，就是害怕对方会得寸进尺。例如，“生命优先”派害怕只要他们在“受精卵还谈不上任何人类特征”的问题上稍做让步，允许流产就成了必然，而这一逻辑推理又经由滑坡谬误的途径，变成了“孩子应当在什么时候出生”的问题。对于“选择优先”派来说，他们也有类似的恐惧，只不过他们害怕的是一旦在“胎儿也是人类生命，是出生前阶段的生命”的问题上做出让步，就会使得这场关于堕胎的争辩被用来攻击他们自己，而问题则容易倒退到“什么时候怀孕合适”的问题。

就这样，对于滑坡谬误的恐惧使得人们继续维持错误的两分法。我们也不难看到，为防止在如枪支管理、环境保护、削减核武器等一系列具有重大意义的问题上发生严重的争论，人们都采取了同样的处理措施。当合理的争论被扼杀的时候，就一定会出现误导性和回避性语言的运用。

* 滑坡谬误（the slippery slope）是由于不合理地使用一连串的因果关系而导致的一种逻辑错误，其中假定了一连串的“可能性”为“必然性”，最终导致荒谬可笑的结论出现。——译者注

4. 魔法：来自异国宗教、幻想以及神秘主义的诱惑

“谁想成为百万富翁？”这个大获成功的电视抢答节目，为所有参与者提供了机会：只要正确回答出各类琐碎问题，就可以赢得一百万美元奖励。试问有谁能抵挡得住这种从天而降的财富的诱惑呢？当然，能够参加节目的机会，被选中成为答题者的荣幸，以及正确回答出 15 个越来越难的问题的困难程度，使得最后获奖的机会少之又少。不过也用不着沮丧，除此之外，还有其他方法可以让人获得一百万美元，而且这种机会对任何人都是公平的。

治疗性接触（therapeutic touch）是一种伪科学，但是它在护理行业十分风行。Swenson（1999）指出，进行治疗性接触的治疗师声称运用这种疗法的人已经遍布 75 个国家，超过 48,000 人，而且还有 75 所大学和专业学校在专门教授这一课程，95 种保健设施具备这一治疗功能。那么，到底什么是治疗性接触呢？按照这一疗法的治疗师所说，他们能够通过感觉与操纵“人体能量场”来诊断和治疗疾病。他们通常将手放在病人的身体上方约 10 厘米的地方，并且慢慢移动，以此进行诊治。虽然名为治疗性接触，其实在治疗时完全没有实际的接触，这可以算得上这一疗法诸多神秘之处中的一个。

现在让我们再来看看如何赚得一百万吧。费城批判性思维协会以及 James Randi 教育基金联合举办了一个活动：任何人只要能通过一个简单的测验，就能获得一百万美元奖励。这个测验的内容是：蒙住眼睛，将你的手在某人的胳膊上方约 10 厘米处移动，要你辨别是否此人的胳膊插在一个玻璃纤维的套管里（这种套管仅仅是起到隔热的作用，因为如果不那样做，被试就非常容易觉察到胳膊

的存在)。如果经过一系列测试，你的正确回答率超过了随机猜测的机会率，那么恭喜你，你的一百万已经到手了！再说说治疗性接触，如果那些治疗师声称的令人惊叹的能力是真的，那他们之中任何一个人通过这个测试，赚得这一百万都是易如反掌的。迄今为止，这份奖金还是无人领取，为什么？

相信魔法

纵观人类历史，人们对知识的渴望，总是源于必须在严酷的、不可预知的世界里努力求生存。我们勤奋不已，不断地通过正规和不正规的试验来探索 and 了解世界。例如，在科学方法形成之前很久，人们就掌握了一套方法来追踪和捕获猎物，同时又保护自己不被大型食肉动物捕食；人们根据日照时间的长短判断季节更替，指导农作物的耕种和收获；人们还会根据夜空中星星分布的不同图形，来指导长距离的航行等等。上述每个成就都表明人类具有惊人的智力与创造力，都是人类在领悟自然界的奥妙、并将相应知识应用于实践的过程中前进的一大步。

然而，伴随着这种对现实世界认识的不断发展，同时出现的还有，人们对有关魔力信念的接受。魔力信念可以提供给人们一种控制感（即使这种感觉本身是错误的），一种对于未知世界中可能存在的威胁的控制感。在我们还没有真正掌握和了解所处的社会及自然环境的时候，相信我们能通过魔法拥有知识和力量，至少也是令人感到安慰的一件事。例如，在科学家们尚未揭示雷电形成的原因时，这些自然现象通常就被解释为神在发怒。

魔力思维，或者说相信魔力的存在，其危害就在于我们可能会坚持迷信的观点不放，而对已经能够清楚解释事物真相的现实却置之不理。短时间内相信超自然力量，相对来说，可能不会产生什么害处，但如果它总是阻止我们去探索真实的世界，那么我们就必将尝到苦果。魔力思维对于人类进步毫无作用，仅仅是提供了错误的期望和信心。而且，对于某种超自然现象的强烈信念，可能会使你继续形成其他的错误信念。一旦你部分地摒弃了现实，接受其他关于魔力的信念也就变得更为容易了。

一厢情愿的想法

下面要说的内容听起来有点孩子气，但它的确是令人惊讶的事实：人们通常

会选择性地相信某些事情，其原因仅仅是他们内心里希望它是真的。一般来说，一厢情愿的想法会使人们觉得这个世界更有趣，威胁性也更小。

在过去的几十年里，Andrew Weil，伪科学家中最直率的人之一，就真实展现出了他推销一厢情愿的想法的诀窍。他将完美的感觉与完全无意义的东西错综复杂地组合在一起，然后出版了一系列的书籍，进行了多次公开演讲，赢得了大批的追随者。他的第一本书，《自然之心》（*Natural Mind*）出版于1972年，在书里，Weil对魔力思维多次表示支持，事实上，这种态度也充斥在他的每一本畅销书中。《自然之心》的主题之一就是理智的或理性的思考其实会阻碍我们获取知识。Weil说道：“为了直接地认知现实，一个人迟早都要学会如何抛弃理智，将自己从不停思考状态中解放出来。”不过，他从未清楚地解释过，从思考和理智加工过程中解放出来是如何使我们理解和认识世界的。而在这段话前面的一小节内容，更是具有明显的神秘主义色彩，里面所说的每一点似乎都不甚明了：

隐藏在现象多样性之下的对于统一性的意识体验，是千百年来贤能哲人们和神秘主义者都公认为最令人振奋和喜悦的一种人类体验，也许只需要一会儿工夫的真正自由就可以达到。而要获得这一自由，就必须从感觉信息的洪流中撤出，以一种非常规的方式来看待事物。

关于“隐藏在现象多样性之下的对于统一性的意识体验”究竟是怎么回事，要如何才能达到这一体验，读者从未真正了解。但是，这个说法是多么铿锵有力，多么富有感染力啊。在后面他谈到心灵感应，或者不同个体间直接进行精神交流的能力时，这类极具吸引力的观点还在不断地出现：

我不仅认为我们每个人都可以分享意识，而且我认为所有人都已经在时时刻刻地做着这件事了。并不是非要专门去学习通灵术，只是必须知道我们已经在这么做了：不停地让心灵感应的事件进入我们清醒的意识……心灵感应不是别的，就是当别人想到什么事的同时，我们也出现同样的思考，其实我们所有人时刻都在这么做，只是它处在意识体验的某一特殊水平，大多数人都意识不到而已。只要意识到它，你就可以自动地进行心灵感应了。

作为一个科学家，我完全被心灵感应的存在吸引住了。我很想去弄清楚它到底是如何进行的，而且可以想见在这一领域中，令人兴奋的、值得探索的新研究课题一定层出不穷。下面就是我最先想到的大约12个问题：

- 什么类型的思维可以分享？意识的，非意识的，还是两者皆有？即心灵感应是要求主动的参与，还是说别人可以不顾我们的本意，随意阅读我们的心灵？

- 在思维交流的过程中，是只能交流文字信息，还是可以包含图像？
- 思维共享的内容存在复杂性或容量的限制吗？
- 心灵感应信息的传递距离究竟能达到多远？
- 当我们想要传送或接收一条信息的时候，如何能够确定对方就是我们要联系的个体？毕竟在这个星球上生活着近六十亿人，看起来似乎很容易弄混淆。

- 我们如何以信号表示一条信息的开头和结尾，又要怎样去辨认这种信息的首尾？即我们发送和接收信息的器官如何才能保证是进行了一次恰当的连接？

- 我们要如何来区分一条新近收到的心灵感应信息和幻觉？
- 心灵感应的信息是由大脑的哪一部分进行加工？
- 我们在处理心灵感应信息时，是否还可以进行正常的感觉输入的加工？
- 对于心理感应的信息，我们是每次只能处理一条，还是可以同时传送、接收多条信息，进行多线程的加工？

- 在传送或接收心理感应信息的能力方面，存在多大的个体差异？
- 传送或接收心理感应信息的能力可以通过自我训练得到加强吗？
- 灵长类动物或其他动物也具有某种程度的心灵感应能力吗？

想到这里，我对于 Weil 富有吸引力的观点开始持怀疑论调了，并不是因为我本人对他所说的东西缺乏兴趣，也不是因为它威胁到了我的科学世界观。虽然根据目前所了解的科学原理，我还找不到这些问题的答案，但我对于新的发现与新的证据是持开放态度的。毕竟我是一名心理学家，因此真正的心灵感应术对我来说是十分具有诱惑力的。问题是目前的证据还根本不足以让人信服。在仅仅只有未经证实的个人断言的情况下，就相信一种非凡的奇异现象，例如心灵感应术，恐怕还纯粹是一种一厢情愿的想法。

相信生命力

虽然我们已经了解一厢情愿的想法的出现，是由于人们总是选择性地相信自己愿意相信的事，但现实的情况是这些一厢情愿的想法还在四处传播，并且十分

具有感染力。举例来说，只是含糊说明的、令人很难理解的各种形式的能量，被作为许多伪科学的前提和依据。我们在前面曾提到，治疗性接触的治疗师宣称他们可以在离病人身体 10 厘米的高度，用手掌检测到一种能量场，而且这种能量在诊断和治疗疾病的过程中都要用到。针灸师则相信一种被称为“气”的神秘能量，他们认为气会在周身的经络中不断循环游走，而穴位则是经络中的特殊位置，当用针刺穴位时可以帮助打通经络，或者以一种更佳的方式重新组织能量的流动。还有一种相对更新一点的伪科学——思维场治疗，同样也是基于周身经络内能量的流动阻塞或中断会致病之说，但它的治疗措施是有节律地轻叩指压点，以增强能量，或者改变能量的流动方式，以达到治疗躯体疾病之功效。另外，按摩师所说的“固有点”（the innate），同种疗法师所说的“生命力”（vitalistic energy），都是关于能量场的变相说法。除了这些之外，依靠未经证实的能量形式的存在，来达到其目的的各类伪科学数不胜数。

对于上述各种各样纯粹假设的能量，最根本的问题就是即使用最高精尖的物理测量仪器，也根本检测不到这些能量。根本就没有任何证据证明有这种能量的存在，更不用说有人能检测到它，并判断它是否处于“不平衡的”状态，或者对它进行所谓的“调节”了。这些都是关于魔力的信念。

基于能量的伪科学的拥护者们进而宣称，那些能量并不存在于我们日常体验的那个物理空间，企图以此自圆其说。根据他们的说法，那些能量存在于另一些领域，但那些领域究竟是什么样的，依旧是含糊不清、未经证实的。虽然他们通过这种在论证当中的闪烁其词，解释了为什么科学家们没有检测到这些神秘的能量，但最终还是导致了最大的自相矛盾。也就是说，如果是由于这些能量与我们所处的物理空间不存在交互作用，我们的检测仪器才无法测到的话，它们又怎能与我们患病的身体器官发生相互作用呢？又怎能与那些利用能量来诊治疾病的治疗师发生相互作用呢？实际上，极少有人会因为自己能量场的流动发生阻塞而寻求医疗保健，这是因为我们通常只会为那些自身正在体验的问题，可以察觉到的问题，还有令我们感到痛苦的问题而寻求帮助。但是，可以帮助对这些问题进行诊断的实验室仪器，却根本不能检测到这些神秘的力量。

尽管如此，还是有许多人并未因这种神秘力量之说的内在矛盾之处——选择性地与物理世界相互作用，就放弃这种观念。事实上，对于某些相信魔力的人来说，要与他们进行一次理性的辩论也十分困难，那些逻辑原则经常被他们抛诸脑

后。不过有一点值得庆幸，很容易就可以对那些伪科学家所宣称的奇异力量进行检验，从而让魔力的拥护者与怀疑者双方都感到满意。要做好这一点，首先争论的双方得同意“检验”这一术语，这意味着无论结果如何，他们都必须接受检验的结果。实际上，已经有人做过关于人们是否具备检测出人类能量场的能力这一检验了。值得注意的是，仅仅是一个初中女孩的科学设计作业，就给了那些相信魔力的人最有力的打击。

将“人类能量场”付诸检验

Emily Rosa 是 Linda Rosa 的女儿，而 Linda 是美国国家反健康欺诈委员会之可疑护理从业行为特别小组的协调员。由于治疗性接触在美国护理界已经被广泛接受，这对母女也经常在家讨论有关这方面的话题。Emily 为了完成一份科学设计的作业，针对治疗性接触最基本的假设——能量场可以被检测到，设计并完成了——一个简单但非常精巧的检验任务。自然，如果治疗性接触师没有通过这一检验，那么也不可能再坚持在疾病诊断及治疗性操作过程中继续运用所谓的能量。Emily 的实验在严格控制的实验室条件下进行，要求治疗性接触师蒙住双眼，对另一个人的存在进行检测，具体任务是辨别在其面前的人每次所伸出来的手是左手还是右手。Emily 预先向参与者详细说明了实验程序，每个参与者当时都保证说成功完成这一实验简直易如反掌，然后每人进行了 10 次关于“左手/右手”的判断。根据实验设计，正确答案中的左和右是随机出现的，那么单凭猜测作答的正确率，与每次用抛硬币的结果来选择一样，应当是 50%。然而 Emily 的实验结果是：参与者回答的正确率是 47%，比单凭猜测作答的正确率还稍低一点。1998 年，Emily Rosa 将她的实验结果发表在了《美国医学会杂志》（*Journal of the American Medical Association*）上，成为了历史上在该杂志发表论著的最年轻的作者（Rosa, Rosa, Sarner, Barrett, 1998）。

对精于治疗性接触之术的人来说，这可能是最简单的任务，但那些自封的专家们都栽了跟头，这已经是对能量场存在之说的强有力的驳斥了。很明显，一个人如果连围绕在别人身边的所谓能量场都分辨不出来，就根本谈不上在此基础上诊断或治疗疾病了。Scheiber 与 Selby（2000）在所著的一本书里，列举了许多关于治疗性接触的逻辑分析及实验室检验结果，发现其中没有一个能够支持治疗

性接触的拥趸们所作的那些大胆宣言。而将此事说得更为清楚明白的是在 1997 年 5/6 月号的国家反健康欺诈委员会时事通讯上，来自科罗拉多大学健康科学研究中心的一个科研小组作出的如下结论：“关于这些能量场存在与否，或者其性质如何，事实上完全不存在任何可接受的科学依据。”

就像那些治疗性接触专家在 Emily 的实验中遭遇失败一样，至今无人赢得费城批判性思维协会与 Randi 基金设置的一百万美元的奖励，也同样很能说明问题。从没有任何基于能量的治疗方法的操作者能证明其所说能量的存在。而且，现有的全部逻辑、理论以及研究结果也都一致性地表明，关于人类生命力这类能量的信念是毫无根据的。

迷信与奇特仪式的产生

除了关于神秘能量的信念之外，许多迷信的观点与行为也因为同样的原因流传甚广。目前，根据 Skinner 的行为学习理论，我们已经能对许多迷信观点与奇特仪式得以发展的原因进行相当好的解释了，首位运用“迷信”这一术语的心理学家也是 Skinner 本人（Skinner, 1948）。要理解迷信是如何发生的，首先让我们来看看学习理论的一些基本原则。

当一种行为出现以后就给予强化，那么将来这种行为再次发生的几率就会增加。例如，一个学生学习很用功，并且在考试中取得了好成绩，他的父母就会强化他的这种行为。强化可以是正性的，也可以是负性的。正性强化是给予一些让被强化者感到高兴的东西，也就是进行奖励（比如因为考得好给他一些钱）；而负性强化则是移除某些让被强化者不愉快的东西（比如免掉他所承担的一些家务）。这样，通过仔细地设计强化的性质与时间，就可以塑造出更新的理想行为方式。大象会列成方阵依次行进；狗熊会踩单车；海豚会从水面跃出并用尾部击打悬在空中的球；以及动物表演的其他种种技巧，都是驯兽师运用行为及学习理论进行训练的结果（Breland & Breland, 1961, 1966）。运用上述原理，行为主义治疗师也成功地治愈了许多心理障碍患者（Lazarus, 1971; Skinner, 1953）。人类能够生存下来，与我们拥有从环境中学习的能力是分不开的，因此，这些学习原则会如此有效也就不足为奇了。虽然我们每个人都具有某种与生俱来的或者内在的行为倾向，但作为一个物种来说，人类的成功很大程度上不得不归功于随环境

改变而改变的适应能力，适应能力的基础就是学习机制。

但是，如果没有人来制订计划、树立榜样，或进行强化呢？就像日常生活中经常发生的那样，如果我们是随机地接受强化呢？观察资料表明，一些古怪的行为模式发展起来，往往并不需要刻意的强化（Skinner, 1948）。Skinner 指出，虽然在实验所观察的个体中，每个被试的怪异行为都会有所不同，而且缺乏有计划的强化，这些行为仍旧表现出了显著的持续性特点。例如，把一只鸽子放在一间有一些机关的房间里，它可以自由地操作那些机关，而无论它采取什么行为，都会随机地得到一粒食物。结果表明，即使在鸽子的行为与强化（得到一粒食物）之间并没有任何事先设计的联系，鸽子通常还是会表现为对那几个刚刚给过食物的机关反复进行操作。这种偶然的强化导致了鸽子更多地重复先前的行为，继而增大了在这种行为以后按随机原则得到食物的几率，在得到食物后又会偶然地强化相同的行为。在这一过程反复出现以后，实际上鸽子操作的任何普通行为都会与这种正性强化联系在一起。行为主义学家称这种形式的偶然学习过程为“迷信行为”，而人类的许多想法和行为的形成，都与该过程有着不可思议的类似之处。

的确，人类迷信观念的形成，与实验室里的动物并无二致。例如，祈雨舞蹈就是一种偶然的强化导致的结果。当一些正处于干旱的人们偶然跳起了某种舞蹈，碰巧又真的开始下雨了，那么以后当旱灾再次发生时，人们就会跳起舞蹈来祈求更多的雨水。也许在真的又下雨之前人们要跳上好多天，也许他们会跳一种更为复杂的舞蹈，但总之随着人们的坚持，天又开始下雨了——毕竟下雨原本就是干旱发生以后的最终或必然结局。各种各样的护身符、避邪物、或其他所谓带来好运的幸运符等等，可能都是基于这样的原理和过程，很容易地就确定下来了。在这其中，个人所做的就是总会尽可能地搜集与好运有关的证据，同时对那些运气不好的时候，或者没带幸运符时仍旧出现了好运的证据一概予以忽略。从理论上来说，几乎任何一种观点或行为，都存在充足的机会来形成这种碰巧的联系，也因此催生了各种各样的迷信观点或行为。

作为典型思维方式的迷信

迷信观念常有使人不得不相信的特点，因为它总是基于一些因与果的配对联系，表面上看起来很有道理。许多迷信观点纯粹是反映了“相同的东西总是伴随

出现”的观念 (Potter, 1983), 这被心理学家称为“表象思维”^{*} (Kahneman & Tversky, 1972)。这里有一个例子, 可以说明表象思维的力量有多大。

Linda, 31 岁, 是一位爽朗、聪明的单身女性, 她的专业是哲学。当她还在读书的时候, 她十分关心性别歧视和社会公平问题, 并且还参加了反对核武器的示威游行。下面请你挑选出你认为最可能的选项:

- Linda 是一名银行出纳员。
- Linda 是一名银行出纳员, 并且在女权主义运动中十分活跃。

如果你选了第二个答案, 那么你并不是少数派。在 Kahneman 和 Tversky 进行的这项实验当中, 有将近 90% 的参与者都选择了这一答案, 人们更愿意相信 Linda 除了是银行出纳员之外, 也是非常活跃的女权主义者。但实际上, 第二个选择在逻辑上比第一个选择更不可信。原因很简单, 许多银行出纳员都不热衷于女权运动, 而且银行出纳员当中也只有一部分人是女性, 因此第一个选项更可能是对的。那为什么我们大多数人都会出错呢? 因为第二个选择看起来与我们内心设想的 Linda 的形象更一致。不过, 从逻辑的角度来说, 一个一般性的语句为真的可能性, 总是会高过那些将同样的句子再附加一些细节后形成的语句。

这个例子让我们得到的经验是: 我们很容易就会被那些看上去相互匹配的事物所误导, 尽管事实上它们之间很可能并没有真正的联系。例如, 相似法则是同种疗法的两个基本原则之一, 这一法则由 Samuel Hahnemann 于 18 世纪末首次提出。他认为能导致某种疾病发生的同一种物质, 也可以用于治疗患这种病的人。在关于表象思维的例子中, 很难想象还会有比这更危险的了, 因为对于那些已确实患病的人来说, 再给予他们导致疾病的病原体以治疗疾病, 其实是不可能有帮助的 (不过, 幸亏 Hahnemann 也相信无穷小法则, 因此他是用稀释的药液来治疗疾病, 这对于病人来说, 可谓是不幸中的大幸。关于同种疗法我们在第 7 章中还会详细讨论)。

迷信行为的维持

如果说迷信行为是由于偶然的学习或表象思维产生的, 也许你会对它究竟如

^{*} 表象思维指唤起表象并在想象中对表象进行加工改造的思维活动。——译者注

何持续下去感到困惑。难道我们都不够聪明，不会发现并去除我们这些愚蠢的信念吗？答案是：还存在另一条学习的规律，正是它使得迷信行为十分顽固，难以消灭。如果每次行为之后都跟随着强化措施，那么一旦强化停止，这种行为也会很快停止。例如，一名学生每次取得好成绩时都会得到奖励，在好成绩和奖励之间就会形成明确的联系；而当学习行为仅仅是由于不断的强化才维持下去时，一旦撤去奖励，学习行为就会很快消失。

那么，如何才能合理地运用强化来塑造理想的行为，但同时又避免上面所说的行为迅速消退的问题呢？关键在于不要每次在期待的行为出现后都给予强化，而是运用心理学家们所说的“间歇强化法”。例如，可以选择每隔一次对该学生的好成绩进行强化（固定时间表），也可以选择偶尔强化一次（可变时间表）。虽然实施这两种间歇强化法，比起持续强化，学习速度要慢一些，但最终行为的消失速度会慢很多。因为如果该学生知道自己只是有时会获得奖励，那么有时没得到奖励也并不意味着以后就没有奖励了，因此他还会继续努力学习。毕竟不是每次好成绩都会被强化，这样做才有机会再次获得奖励。只有连续多次取得好成绩后都没有得到奖励，才会导致学习行为的消失。

综上所述，在考虑某种行为是如何发生的问题上，不仅要考虑行为是经由有意的强化还是偶然的强化形成的（后者有时就是迷信行为的成因），强化的方式和时间也是很重要的问题。例如，某人曾试用同种疗法来治疗流行性感冒，碰巧几乎是马上就好了。在这种情况下，即使所用的同种疗法对于康复过程完全没有作用，也完全有可能被当作一次成功的治疗体验，迷信观念就是这样通过偶然的强化形成的。从那以后，这个人治疗感冒就更有可能选择同种疗法了。由于治疗方法本身并没有什么躯体治疗效果，那么再次康复也就是一次偶然事件。在多次感冒的过程中，此人所得到的偶然的强化实际上就是按照前文所说的间歇的、可变的时间表来进行的，这使得其采用同种疗法的行为固定下来，而且对同种疗法疗效的信任很难消除。在第6章中我们还会讨论到，个人推荐和证明，实际上并不能作为有价值的证据，其原因就是我们不可能弄清楚那些被推荐的个人经验中，是不是存在上述迷信行为的学习成分，而唯一能够排除这一可能性的方法，就是进行有严格控制与对照的科学实验。

假科学之名支持神秘主义

那些热衷于神秘主义、白日梦以及异国宗教的人，实际上经常假科学的名义来为其服务。尤其是很多人会利用科学理论中一些违反直觉的部分，来证明其所持的“现实本身是极其复杂的，任何事都有可能”的论调。在上面这句话里，前面半句的总结的确反映了真实的情况：现实世界是非常复杂的。Robert Park (1997) 将相对论、量子机制，以及混沌理论并称为科学界的“三大神秘理论”，这三种理论中的任何一种都揭示了在该理论出现以前从未想到过的自然界的精妙之处。但是，“任何事都有可能”这句话并不客观，它是一种随意的“什么都可以”的态度，反映其在运用科学知识的时候可能是进行了高度选择性地利用。事实上恰恰相反，我们每个人都应当把科学进步看作是在判断什么事可能发生时的一种限制。例如，Robert Park (2000) 曾对热力学的第一、第二定律如何对物理现象产生严格限制的问题作过阐述。他将第一定律总结为“你不可能赢”，意思是对于某个系统而言，你从中得到的能量不可能比已进入其内的能量总量还多。他将第二定律概括为“你甚至不可能收支平衡”，意思是在整个能量流通或转换的过程中，总会有某些能量损耗为热能或其他形式的能量，无一例外。这两条定律是自然界中的既定法则，因此实际上大大限制了物理空间中各种可能性出现的范畴。总的来说，根据科学知识对现实进行评估，其实是缩小了可能性的范围，而“什么都可以”的态度，意味着只是承认完全无知。

我们可以将科学理论看作是对真实世界的一系列成功的估计，而每一次新的领悟都会使我们离真相更近一些。好的科学研究的目的实际上就是排除那些不正确的假设，使得留下来的假设其正确的可能性更高。当然，这并不是说科学家们就不会有惊奇的发现。事实上某些重大的科学发现，如青霉素的疗效、X 光的辅助诊断用途等等，都是在偶然的情况下被发现的。但即使每一个新的发现都开启了一扇新的研究方向之门，它的作用还是便于我们更全面地理解现实世界。而那种“什么都可以”的态度，则恰恰是否认我们所说的要更全面理解的那部分现实。伪科学家们总是希望我们将那些已经出现的重大的困惑置于一旁，这样对于他们的神秘主义观点或未经证实的信念的理性思考，就不会与那些最恰当的研究证据相冲突了。

伪科学家们滥用科学的另一种表现是，宣称自己是拥有先进观点的真正天才，只是目前由于观念太超前而不被接受、甚至被人误会。纵观历史，有许多人曾因为不能使当时的大众接受其观点而惨遭迫害（Gardner, 1952）。现在伪科学家们很时髦的一些类比——“就如人们嘲笑伽利略”或“就如人们嘲笑牛顿”——其实是存在许多逻辑上的漏洞的。关于这一点，Carl Sagan 带有讽刺性的评论十分到位：“人们也嘲笑四肢发达、头脑简单的大汉是粗人。”这句话说的的是一个简单的事实：在那些倡导改革或边缘理论的人中，有着卓越智慧的天才毕竟少之又少，绝大多数人是错误的。许多伪科学家实际上只是将他们自称的受迫害状态作为一个卖点，或荣誉的一种标志。他们明显持有一种幼稚的观点：伽利略和牛顿的理论之所以最终被接受，就是因为他们受到了迫害。而且，他们甚至还希望我们能够相信这样的观点。

从历史来看，科学家实际上是最乐于转变并最乐于接受那些有证据支持的、哪怕是激进的新观念的一群人。科学家从未嘲笑过伽利略或牛顿。爱因斯坦的狭义相对论与广义相对论，实际上揭示了物理领域长期存在且备受推崇的一些理论的诸多不完善，但在科学界很快就得到了承认。而且，由于具备有说服力的科学证据，爱因斯坦的理论和伽利略或牛顿的理论一样，最终得到了社会各界各色人等的承认，无论他们是来自什么样的社会，或者有什么样的宗教和政治背景。

现代科技的魔力

在当今这个有着惊人的科学与技术成就的年代，神秘主义居然还吸引了如此多的民众，不能不说有些讽刺意味。也可能科学的极度成就反而会导致一些魔力思维与信念的产生，因为对于绝大多数人来说，顶尖的现代科技简直就与魔法或魔力无异。

电是如何产生的，计算机是如何操作或交换信息的，微波炉是如何烹调食物的，还有遥控器或手机是如何运转的，对诸如此类的简单问题，我们当中又有多少人能解释得清楚呢？上面提到的技术中，没有一项是现代科学最先进的技术，但是对于绝大多数人来说，还是没有完全掌握，而是仅有一点模糊的理解。

然而我们出于各种实际的目的，越来越依赖那些具有魔力的仪器或装置。一旦我们对事物的运行原理及规律失去理解，我们就更可能去接受一些基于高深莫

测、难以理解的原则的观点。如果没有一些基本的科学知识，我们不可能分辨出哪些魔力是真实的，哪些魔力则纯属幻想。但如果我们自己不能分辨出这些区别，至少应该意识到如果将它作为科学问题考虑的话，需要听取具备一定资格的科学家的意见。

学习科学思维的方式，了解科学的知识基础，其实并非易事。科学需要对复杂的问题进行精巧的方法学设计，同时还要注意伦理、法律以及可操作性等问题，才能得到合理的解答方案。完成这一过程不仅需要创造性的思维来考虑所有可能的理论解释，也需要实际的怀疑主义，以帮助我们将最可能正确的理论从那些赝品中挑选出来。也就是说，一方面你要保持一个开放的头脑，而另一方面又要像老话说的那样，“不要太过开放，以免你的头脑一团糟，各种观点争论不休”。如果我们可以不再考虑科学的要求，遵循最小阻力途径（the path of least resistance）来认识世界，那生活该变得多么容易啊。不过，这仅仅是梦想而已。

当你听见某人在运用“相对论、量子机制、混沌论”等术语，或者其他听上去同样复杂的名词时，你可能就需要问问自己，这个说话的人是否真正懂得这些理论。即使是浮光掠影地了解上述理论中的任何一个，都需要广泛的科学及数学专业知识基础，要掌握它们更是需要花上十年或更长的时间来潜心研究。有时，某人会随意地说出一个他自己并不熟悉的理论的简短版本，那只不过是混淆重要的特征与差异，将神秘主义合法化，或者掩盖他自己的理论与人们熟知的现实状况之间彻底的矛盾而已。

“能量”这一术语的运用也是如此。一名物理学家可以很清楚地告诉你各种特殊形式的能量究竟意味着什么，以一种公正的、可验证的方式演示如何精确地测量能量，解释如何将一种形式的能量转换成另一种形式，或者让这种形式的能量与物理空间中的其他部分发生相互作用。因此，当一名自封的治疗师运用“能量”或“人体能量场”等术语的时候，你完全可以要求他提供实际运用这些术语时的具体事例来证明。而当某人将自己与伽利略、牛顿，或其他受过迫害的天才相提并论时，你就需要考虑，这两者之间的关系实际上是否只是“四肢发达、头脑简单的大汉和粗人”这样的表面联系。

5. 权威：要求盲从

二战期间，有 500 万到 600 万的犹太人在德国纳粹集中营中被杀害。二战以后，心理学家们进行了多方努力，试图找出这种大屠杀暴行的原因所在，特别是如此之多的德国民众是如何接受纳粹哲学的。其中一种观点就是权威主义（authoritarianism），这一人格因素导致了一些人特别容易服从命令。不过，在进行研究验证时，这一假设只得到了部分支持。1960 年，Stanley Milgram 重新考虑了这一问题，不过这次角度有些不同，他提的问题不再是“纳粹为什么和我们有如此大的不同”，而是“我们在哪些地方与纳粹有点类似”。

Milgram 的研究

请设想现在是 1960 年，你住在美国康涅狄格州的纽黑文市，看到报纸上有一个广告，耶鲁大学心理学系正在为一个关于“惩罚在学习过程中的作用”的科学实验召集志愿者，每位志愿者会得到 4.5 美元的报酬，价钱不高但还算公道。你就按广告上的地址回了信，并得到了参与实验的具体时间安排。那天到了实验室以后，你发现还有一个人也来参加实验了。工作人员出来迎接你们，感谢你们的参与，并解释说为了检验惩罚在学习过程中的作用，你们两人中的一个将扮演老师，另一位则扮演学生，当学生答错问题的时候，老师可以通过按钮给学生电击作为惩罚。你们抽了签，抽签的结果是由你扮演老师。研究人员将扮演学生的人带到另一个房间，把他用皮带绑在椅子上，并接上电极。为了让你形成对

电击的感性认识，你本人也接受了一次 45 伏电压的电击。在你和实验人员离开房间去电击操作间之前，扮演学生的志愿者提到他的心脏有点问题，并询问电击是否会对他产生不利。实验人员再次向他保证，虽然电击会让人感到痛苦不适，但并不会产生“永久性的组织损害”。

接下来，实验人员将你带到电击操作间，你可以控制整个学习任务的进行。你的面前是一个宽大的操作台，有电线将电击按钮与隔壁学生身上的电极相连。电击的强度共分 30 个等级，起始强度为 15 伏，然后是 30 伏，再是 45 伏，依次递增，直到 450 伏，而且每一个按钮下方都用文字注明电击强度，15 伏为“轻微”，强度再高一些的被标为“中等”，自 400 伏开始标有“危险：强烈电击”的文字，而最后 450 伏的按钮则被标为“致命”。

当你观察完整个仪器后，实验人员向你说明了你作为教师的职责是什么。首先，你要通过一个通话装置向学生朗读一个很长的单词表，而学生的任务是记住你所读的那些一对对的单词。然后，你就只读一对词中的一个词，并给出四个词让学生选择哪一个是前面听过的词对里与之匹配的另一个。如果学生选错了，你就进行下一个词，但要告诉学生他选错了，随后念出正确答案，并按电钮给予一次电击作为惩罚。这个实验里最令人不安的就是，每次你给予电击的时候，强度必须要比上一次电击至少高 15 伏。也就是说，学生第一次犯错，要接受 15 伏的电击，第二次就必须接受 30 伏的电击，然后依次递增。实验将一直进行到学生完全记住整个词表为止。

你开始实验了，学生记住了一些词对，但偶尔也会犯个错误，你就忠实地按照实验要求给予短暂的电击。但是，当电击强度逐渐加大时，学生由于疼痛开始呻吟、哭喊，最后尖叫。学生反复地要求放开他，说他的心脏有问题，不停地抱怨剧烈的疼痛。你不知道该怎么办，求救似的看着实验人员，而后者只是简单地说“请继续下去”。如果你再表示抗议，得到的回答就是“实验要求你继续下去”。随着痛苦的叫喊声越来越大，学生坚持说必须放他下来，实验必须停止，然而，实验还是继续进行。当电击强度达到 300 伏和 315 伏时，学生仍在叫喊，拒绝再接受电击和进行这个实验；而当强度达到 330 伏时，隔壁房间就再也没有任何声音传过来了，不仅是没有回答问题，连痛苦的叫喊声也没有了，这实在是让人感到焦虑和担忧。尽管如此，实验人员还是说“没有选择，你必须继续下去”，并且要你把没有回答算作答错一次，仍需给予电击。你现在已经开始冒汗，

神经质地笑，而且不时停下来恳求坚定不移、铁面无情的实验人员，不过你还是继续扮演着教师的角色。电击强度逐渐加大到了标有“危险：强烈电击”的程度，而且还更高，一直达到了最后标有“致命”的450伏。当你给了学生这台仪器所能给的最强烈的电击之后，实验人员又命令你对学生所犯的每一个错误都给予一次450伏的电击。在你又这样做了几次之后，实验人员终于告诉你整个实验完成了。你的恶梦也终于结束了。

现在你可能在想，“我可从未做过这样的事，无论有没有实验人员要求我这样做，我都不愿意带给别人伤害”。如果你认为上面描述的一定是一个有施虐狂的老师，与绝大多数正常人完全不同，那么你的想法和大多数人是一样的。在实验之前，Milgram向40名精神病学专家进行过咨询。在他详细描述了整个实验的流程之后，专家们认为在1000个人里只会有一人将教师的角色扮演到底，也就是坚持到最高的电击强度。不过，与实际的实验结果相比，这个预测实在是太保守了。在前面描述的这个实验中，65%的被试完全服从了实验人员的命令，坚持到了实验的最后。而且在35%拒绝执行实验人员命令的被试当中，也只有少部分的人是在电压增加到330伏、学生停止反应之前做出反抗的。这样的结果，用一句双关语来说，实在是太震撼（shocking）了*。

在对这一实验结果进行分析与探讨之前，我们必须先谈谈有关这一研究的伦理学问题。首先，在实验中扮演学生的人并没有接受电击，这一点很重要。事实上，整个实验中唯一的一次电击就是让扮演教师的人所接受的一次强度为45伏的电击，是作为对电击强度的体验，也为了让他相信电击按钮真的在起作用。实验的目的也并不是研究“惩罚在学习过程中的作用”，而是研究对权威的服从行为。开始的抽签只是为了让被试相信而特意设置的步骤，实际上每次抽签的结果总是真正的被试扮演教师，而当学生的却是实验者的助手，他的所有反应都只不过是表演而已。学生的抱怨、抗议等都经过了标准化的设计与训练，然后才正式进行表演，这是为了保证每一名被试都体验到完全相同的情境。一旦实验中止以后，扮演教师的人就会和扮演学生的助手重新见面，以向教师证实学生的确没有受到任何伤害。你可以想象得到，对于被试来说，这会多大程度地缓解他们的焦

* shocking 此处的意思是“令人震惊的”，但这一词也可理解为“令人休克的”，在这里为双关语的用法。——译者注

虑与不安。然后，实验人员还会就这一研究的实质以及隐瞒实验意图的原因，与被试进行更多的探讨。最后，Milgram 本人在全部实验结束后会与所有的被试见面，以确定他们没有因实验而出现遗留的不良心理应激反应。还有一个精神病学专家小组专门评定 40 名被试的心理健康水平，以观察被试在实验过程中受到的短暂、剧烈的应激所带来的影响。在所有的被试中，绝大多数人（84%）报告很高兴能参与这一实验，只有 1 人表示参加了这样一个实验很后悔。许多人都认为虽然在这一过程中有些不安，但自己学到了一些东西，对于人性有了更深入的认识，自己的生活也因这次参与有了变化，开始不再那么盲目地服从权威的指导，而是更多地表达自己或他人的真实意愿。有几个被试甚至表示只要是 Milgram 的实验，他们就愿意再报名当志愿者。

对 Milgram 研究结果的分析

如果你是首次了解 Milgram 的这一研究（毫无疑问是心理学界最著名的实验之一），最常见的反应就是“我本人是决不会一直服从命令到最后的”。但是 Milgram 进行的一系列后续研究表明，选择服从和表示抗议的人并没有什么显著的差异。无论是性别、年龄、职业，还是人格特征因素，都与是否会服从没有显著的关联。而且，如果你认为你至少会在 300 伏的电击之前就表示抗议，那你就很有可能想错了。事实上，并没有什么理由可以说明你和 Milgram 的实验参与者们有什么本质区别，而他们当中有 65% 的人完全服从了指导者的命令。你会那么想很可能是由于你低估了实验情境的影响力。

有许多因素导致了 Milgram 的实验结果——很高比例的完全服从（将实验一直进行到 450 伏的电击强度），要了解这些影响因素，需要比较一下 Milgram 所做的一系列实验的结果，这些实验只是在情境和方法上有细微的差异，实验意图和流程是完全一样的。虽然在每个实验中都会有完全服从者，但实验结果还是揭示了一些重要因素。其中一个重要因素就是被试对实验者的权威性的判断或理解。最初，实验是在一个享有盛誉的世界名牌大学——耶鲁大学做的，后来的重复实验则是在一个租来的仓库里做的，研究者自称是由“桥港研究联合委员会”的成员组成的实验小组，在那个实验中，完全服从率就降到了 48%。第二个重要因素则是权威人物对于参与者的可接近性。当研究人员通过电话来给参与者下

命令时，服从率就降到了 21%。第三个重要因素是教师与学生情感交流的距离远近。当将学生与教师安排在同一房间时，服从率降到了 40%；而当要求教师在实施惩罚时要握着学生贴有电极的那只手时，服从率进一步降到了 30%。

研究表明，服从行为的重要决定因素还包括其他的情境因素。比如在后续实验中，从未严格控制过的一个情境变量就是按等级递增的电压强度大小。每次电击在等级强度上只有细微的差别，可能会导致不服从行为似乎是任意地出现在各个电压点。也就是说，从一个电击等级到另一个电击等级，情境并没有出现明显的变化，使人们（包括被试和实验人员）得以判断什么时候采取拒绝的态度是适当的。例如，当已经给了 150 伏的惩罚后，再拒绝给 165 伏的电击是不是显得有些伪善？如果说 165 伏的电击是残酷的，那么 150 伏也是如此。因此，如果电击的强度等级被拉得更开一些，可能服从行为的比率就会下降一些。为什么那些不服从的行为都出现在同一个电压点——330 伏，唯一的理由就是在 330 伏电击的时候，扮演学生的助手停止反应了。正是这种“电击对学生已造成了可怕的影响或结局”的可能性显著地改变了实验的情境，而仅仅是 15 伏电压的增值却不会有这样的效果。这也可以看作是滑坡谬误造成有害影响的另一个例子。

我们应当在什么情况下服从

Milgram 的实验的最大意义，就是揭示了我们每个人都具有潜在的、明显的服从权威的倾向。从某种程度来说，我们不应当对这一研究结果感到如此惊讶。实际上，绝没有哪个人在一生当中从未服从过权威。我们不可能证实每一件所听到的事，因为我们没有那么多的时间来证实；我们也不可能拥有那么多的专业知识，可以理解和判断各种观点的相关证据究竟是什么含义。我们必须接受一些专家的权威观点，否则就可能永远处于无知与蒙蔽的状态。但问题就在于，我们必须将具有真正的专业知识的合法权威，与仅仅打着专家旗号的不法权威区分开来。因为前者往往是真正关心人类的福祉，而后者可没有如此高尚的情操。

不幸的是，要判断权威的正规性或者合法性，往往十分困难。即使是医生，拥有相当多的专业知识，接受过正规的理论与操作培训，并具有一定的实践经验，也并不是完全可靠。那么，有人会指出他们的问题吗？比如一名内科医生在开处方的时候犯了一个明显的错误，执行处方的护士会质疑他所开的药物吗？很

遗憾，通常都不会。让我们来看看 Cohen 和 Davis 报道的一个“直肠耳痛”的奇怪病例吧。Robert Cialdini，一位影响力和说服术方面的心理学专家，是这样描述这个案例的：

一位内科医生给右耳疼痛并有感染的病人开了一种滴耳用的外用药物。但是在写处方的时候医生没有将“右耳”（right ear）两个字写完整，而是写了个简写“in R ear”（rear 有“屁股”的意思）。拿到处方以后，护士迅速将滴耳用的药剂挤进了病人的肛门。很显然，直肠治疗对于耳痛病人是不会有any效果的。但无论是病人还是护士，没有人对这件事表示疑问（1993）。

毫无疑问，这个小故事背后隐藏着大问题。在对更严重的治疗失误进行的另一项系统研究当中（Hofling, Brotzman, Dairymple, Graves, & Pierce, 1966），实验者要求护士们执行医嘱，给一名病人配药。但是，至少有四个理由让她们对上述医嘱提出质疑：

1. 这一医嘱是通过电话通知的，而这违背了医院医疗工作的常规。
2. 给出该医嘱的人自称是医生，但却没有给出任何能证实其身份的资料。
3. 医嘱中要给的是一种尚在实验阶段的药物，而不是已通过检验允许进入临床使用的药物。
4. 医嘱中的用法明显超量，是临床可用最大剂量的两倍。

这一实验在 22 个不同的病房进行，参与的护士来自外科、内科、儿科以及精神科。在目标病人的病房外面有一名研究者在暗地观察护士的行为。95% 的护士完全按照那个医嘱准备给药（这时，该观察者就会出来干涉，说明这只是一个实验，防止真的出现过量给药的情况）。想到这种盲从可能导致的结果，不由得让人感到后怕。这一实验的结果提醒我们，在执行命令之前，必须仔细地考虑权威人物的意见的可靠性。

靠不住的权威

伪科学界经常有一些令人怀疑的权威人物或自封的专家，他们的主张和判断并不是来自于经过严格控制的科学研究。很多时候，伪科学的宣传者试图混淆真

正的权威与冒牌权威之间的差别。下面我将举例说明一些常见的、容易让人误认为是真正权威的谬误。当然，这些讨论不可能那么完整详实，但还是可以提供一些有用的工具，帮助我们更好地思考这些问题，从而发现事物的真相。

有关“远古智慧”的谬误

在伪科学中，往往有用所谓的“远古智慧”来取代真正的科学依据的倾向。伪科学的拥趸们总是说任何理论或观点，如果能保持千百年，那么一定是有效的。实际上，这种观点可能会导致十分危险的后果。如果仅仅因为年头长短就能判断某种理论或观点的真伪，那么根据性别、种族、是否反犹太主义等等原则进行判断都是可以的了。从另一方面来说，在人类历史的长河中，出现过许许多多相互对立的观点，它们也都曾广为流行过。不断对远古智慧进行质疑，是社会前进的必经步骤，没有任何理由可以认为某种理论能免除这一挑战和质疑的过程。

传统中药的拥护者，如针灸师或中草药师，其实是相当依赖“远古智慧”之说的。但不幸的是，治疗方法的有效性并未因其年代久远而得到完全的保证。事实上，很少有证据能够说明这些古代的疗法是否“经得起时间的考验”，因为几乎没有人对它们做过有效的检验，当然也就谈不上有没有通过检验了。事实上直到现代，才开始出现准确详尽的医疗记录，在此之前，人们很容易主观地得出某种方法治疗有效的结论。但是，在没有健康状况及预后的详细记录的情况下，又如何能真正了解治疗方法是否有效，整个治疗过程是否利大于弊呢。

关于普及性的谬误

伪科学的另一个花招就是夸大他人支持或赞同其理论的程度。许多人都会同意的观点至少具有表面的合理性，看上去更像是权威的意见，这一点 Robert Cialdini (1993) 称之为“社会认可”。1993 年，David Eisenberg 及其同事发表了一项调查的研究结果，这一调查结果目前已成为替代保健领域的文献中引用频率最高的调查报告之一，虽然很少有人真正理解它到底说了些什么。在这项研究中，研究人员询问被调查者在过去的一年里采用了哪些类型的健康保健措施，重点询问的是那些尚未在美国医学院正式教授的课程，也未在医院里广泛使用的非常规治疗方法。得出的结论是：34% 的美国人采用了非常规治疗方法，而且这一数字在被引用时还往往是以“每三个人里就有一个”的形式来进行描述。

在大众新闻媒体反复报道的情况下，很多人已经知道了 34% 这个数字或“每三个人里就有一个”的说法，但实际上这一说法存在一些严重问题。当研究者讨论非常规治疗的时候，列出了同种疗法、针灸、草药治疗等治疗方法，并暗示了这些替代治疗的措施相当普遍。但仔细研究一下他们的数据，你就会发现情况并非如此。实际上只有极少的被试选择了同种疗法（约 1%）、针灸（低于 1%，比以前的调查数据还要低）、草药治疗（约 3%）。在这极少数选择非常规治疗的人当中，主要选择的还是市面上无需处方即可合法出售的同种治疗或草药产品，只有 1.33% 的人真正到治疗师那里接受了治疗。因此，这些常和“三分之一”的统计数字联系在一起的非常规治疗方法，事实上正是因其稀少，或者说缺乏普及性，应当引起我们的注意。

另外，在那项调查中，研究者将许多已经广泛教授或使用的治疗方法都算作了非常规治疗（如商业性的减肥程序、放松训练、生物反馈治疗等）。比如，饮食控制和运动的治疗方法，实际上在每个医学院的课程中都有，而减肥程序的基础就是上述两项内容。因此，当有人再跟你说“每三个美国人里就有一个采用了替代治疗”的时候，你应当认识到这一结论实际上与其数据不符。它造成了一种基于“普及性”的权威的错觉，而实际上对其普及性的估计是大大言过其实了。

关于自封专家的谬误

有一次，我以前的一位同事参加了一个研讨会，其中一个报告的内容是“感知学习模式”，一种认为学生在阅读、听或观察等学习能力上存在个体差异的理论。报告者认为，因为有这种个体差异的存在，必须要对学生的学习模式进行评估，然后再施以与之匹配的教育方法。这一观点看上去十分合理，而且如果它说的情况属实，可能还具有相当重要的实践价值。但如果这一理论有问题的话，也会产生许多负面的影响。因为许多教师总是在试图改进自己的教学方法，其中一些人就有可能根据这个尚未证实的理论，来改变自己正在使用的行之有效的办法，而这样做并不明智。

我的同事决定要了解清楚，他们究竟是用什么方法来区分学生的学习模式。报告者提到的用于评估和区分的问卷是否具有良好的信度和效度？将学生划分成他们所说的那几种学习模式的类型是否恰当？在任何科学的学科领域，提出要提

供参考文献或依据之类的要求都很正常。但是在这里，报告者听到上述问题以后气愤不已，完全不是在讨论相关研究与发现时的应有反应。

我的同事并未因此而被吓退，反而提出了另一个关键性的问题：有什么证据可以证明根据这种学习模式差异理论对教学方法进行改动以后，会增进学习效果？该理论预测学生的学习模式与教学模式之间会存在交互作用，因此在进行了适当的匹配以后，就会比匹配前产生更好的学习效果。在许多关于不同教学模式效果的研究中，老师们通常对所有学生都采取同一种教学模式。那么在这里，问题的关键就是老师是不是应该对不同学生采取不同的教学模式。例如，当通过视觉形式进行教学时，视觉型的学生就会比听觉型的学生学得更好；反之，听觉型的学生也会在他们更喜欢的模式中学得更好。但是，当我的同事询问有关这一问题的研究数据时，专家立刻变得慌乱不安，闪烁其词。她给了我的同事一个参考文献的清单，然后很快就离开了。对她来说，提出提供依据或者相关资料的要求，好像是一种威胁，甚至带有敌意，而不是任何一个值得尊敬的科学家应尽的职责。在她所列的参考文献中，许多都来自于会议报告，只有很少一部分是发表在有同行评议程序的学术期刊上的实验性研究论述，而且没有一篇说明了其最基本的观点——交互效应。当再次被问到这些问题时，报告者肯定了我同事的猜测。事实上，关于运用这种教学与学习模式相匹配的方法以后是否会增进学习效果的问题，从没有人作过直接的研究和检验。因此，除了不断地要求人们根据假设的学习模式调整教学方法之外，报告者给不出任何需要这样做的依据。也就是说，她是在要求教育工作者们根据某种推测来改变自己的观点或实际教学行为。

自封为专家所带来的问题，实际上很普遍，也很令人头痛。要是我们可以简单地通过检查一些个人证明资料（文凭、学术兼职等等），就确定一个人是否是专家的话，我们的生活该变得多么轻松啊！但不幸的是，专家的身份也不一定就能保证其所说内容的真实性。对此，Faust 和 Ziskin（1988）曾发表了一篇颇有争议的论文，根据已有的关于人类易犯错误的研究，他们认为心理学家与精神病学家们很难对专家一词给出合法的明确定义。尤其是有相当多的研究表明，关于专家身份认定的最直观、最明显的标准之一——临床经验，就和其判断及主张的正确性没有直接的相关（参见 Garb 的综述，1989，1999）。我们对一名已有几十年临床经验的心理学家的权威说法，不应当比对一位新近毕业的心理学博士的说法更信任，这听起来多少让人有点不安，但它的确是避免盲从的行之有效的办

法。所谓专家，并不能根据其主张或观点来进行判定，也不应根据是否得到了整个业界的承认来进行判定，唯一判断的标准就是他证明自己的判断有效可信、结论真实准确的能力。

将政治游说作为警示旗

最后一种导致人们误认权威的情况与某些政治行为有关。还有什么方法比得到政府的重视更能让人获得尊敬和树立权威呢？游说别人并不算什么错误，只要游说的理由最终是合理的，而且在伦理学上也站得住脚。不过，基于可靠的证据而游说政府改变政策方针，与出于意识形态立场不同而进行的此类游说，两者有着天壤之别，尽管有时这种差异并不是那么清晰可见。一些重要的社会问题的本质其实就是意识形态问题，但另一些问题则是科学研究的问题。对于后者来说，当一场精心组织的政治游说完全代替了科学研究，而不仅仅是起到其应有的宣传作用时，情况就变得十分令人不安了。

当具体的医疗保健措施也进入政治舞台时，通常就是伪科学的一个警示信号。有效的治疗方法其实是很少需要政治支持的，其坚实的科学依据与成功的健康保健效果就足以说明问题了。Jonas Salk 需要向政府机构游说，让他们支持使用脊髓灰质炎疫苗吗？当然不需要。那些能提供和证明真正疗效的人自然会得到认可与重视。因此，有意为之的努力游说就意味着那些提倡者正在试图绕过正常的渠道——通过提供科研证据来促进健康保健政策的制定，取而代之的是通过立法机关来允许一个尚不可靠的实践。这种策略非常有问题。立法机构不是对病人负责，而是对选民负责；政府官员的出发点往往也不是保障人们基本的健康需要，而是让公众满意。此外，立法者们通常也不具有医学相关的那些技术性的专业知识，但这一点恰恰又是充分理解和恰当评估某种健康保健措施所必需的，因此他们就很容易被这一领域中的错误信息所左右。在 20 世纪 90 年代发生的两场与健康相关的政治论战，就很好地说明了这种建立权威的不法方式所带来的危害。

美国国家补充与替代医学中心

作为一场大型政治运动的结果，美国国家卫生研究院在 1992 年成立了替代医学办公室这一分支机构。到了 1998 年，这一办公室正式更名为美国国家补充

与替代医学中心。据说，成立这一机构的目的是通过投入对替代治疗疗效的研究，来探索常规医疗保健方法的替代性途径。下面是直接从其网站上的“常见问题”栏摘录下来的一段话，清楚地说明了这一目的：

人们有时会问，国家补充与替代医学中心是否也运用与常规医学相同的那些科学标准来进行衡量。补充与替代医学需要用与常规医学同样的科学方法来进行研究。国家补充与替代医学中心鼓励关于补充与替代医学的有效研究成果的发表，而这些研究应当是运用与常规医学同样严格、甚至更为严格的研究方法来进行的。因为这些研究经常会涉及一些新奇的观念及主张，而且需要运用复杂的操作系统，需要系统、清晰明确、全面的知识与技能。

在这一段文字中，对科学研究方法的明显重视，以及想要将其推行至整个健康研究领域的态度，是很值得称赞的。但是如果你注意到替代医学办公室/医学中心以往的记录，就会发现与其公开宣称的使命相互矛盾之处。1998年，《新英格兰医学杂志》的期刊社论回顾了这些记录，并将其总结如下：

直到目前为止，结果都是令人失望的。例如，在该办公室的在线数据库中查到，1993年该办公室拨款资助的30项研究中，有28项都只是提交了一个摘要形式的“结题报告”。在世界上最具权威的临床医学文献数据库——Medline中，检索那些研究者获资助后6年内所发表的文章，发现28项研究中只有9项是将研究成果总结成正式的期刊论文发表了。而且，其中5个发表论文的期刊甚至还不在于Medline的医学图书馆收录的3500种学术期刊之列。剩下的4项研究，没有一项有临床对照实验，因此也不能得出任何关于替代治疗的疗效的确切结论。

说上述情况“令人失望”实在是太过保守的说法。虽然由国家补充与替代医学中心资助的另外一些研究出版物可能还是“按正规期刊的审稿流程”操作的（有专家评审，有修改程序等等），但如果以研究能力的任何一条评估标准来看，都很不合格。就拿那5项仅仅发表在两个无名期刊上的研究来说吧，由于那两个期刊的录用标准极低，在学术界声望极差，所以在Medline这种收录广泛的系统中都没有它们的名字，这其实已经很能说明问题了。而且，没有发表过一项具有良好对照的实验结果，这个事实令人震惊，同时有力地驳斥了某些人宣称已经使用了严格的研究方法的说法。2001年，Green对替代医学办公室/医学中心

成立9年期间资助的所有研究进行了评估，他得出的结论是，“直到现在，也只是进行了极少的基础科学实验和临床研究……而且根据这些研究，你不能对任何一种治疗方法得出有效或无效的结论”。换言之，根据那些研究文献，替代治疗方法的疗效研究既没有阳性结果，也没有阴性结果。

1993年，也就是替代医学办公室运作的第一年，它有200万美元的预算。尽管这一机构没有提供设计精巧的研究报告，或者有实践指导意义的研究结果，这一经费预算仍然不断增加，2005年已达到12310万美元，约是1993年的60倍。直至目前，替代医学办公室/医学中心所获得的经费总额已经累计达7亿美元。对比这一数字和该领域的研究成果，想到其他更有前景的医学研究领域，不禁让人唏嘘，如果它们有限的联邦研究经费能有所增长，会取得多少成就啊！尤其是当详细了解了补充与替代医学中心所取得的成绩有多么少时，更让人感到疑惑。这一机构所做的除了为一系列治疗措施塑造表面的可信性之外，没有其他的作用；而它所支持的那些治疗措施都是有问题的，既没有任何有意义的相关科学研究结果，也从未获得过来自补充与替代医学中心之外的其他任何医学研究机构的支持。

膳食补充剂保健与教育法案

几十年来，美国政府管理机构为了保护消费者的权益，要求所有的药物在上市之前，都必须经过安全性与有效性检验。负责检验的部门有美国食品与药品管理局、美国联邦贸易委员会等，它们都是为了避免不法商人出售无效甚至不安全的产品，为了公众利益而特别增设的机构。纵观历史，江湖郎中的狗皮膏药害得许多本该得到有效医治的人饱受病痛的折磨（相应的经济损失就更不用说了），究其原因就是他们出售的药品事先都未经过正式检验。一个世纪之前，“蛇油销售员”贩卖各种各样的所谓灵丹妙药，其中一些药可能的确对于某类患者有些疗效，但绝大多数都没什么效用，甚至还有害处。由于没有经过安全性与有效性检验，很难说人们从江湖郎中那里买到的究竟是什么。建立相关管理机构，给予它们权力，以取缔不安全或无效产品的销售，并对那些只想在保健医疗市场中大捞一把的人征收罚款和提出犯罪控诉，种种相关措施已经取得了显著的成效。

在几十年来坚持保护消费者权益、为消费者说话以后，至20世纪80年代，美国政府管理机构的执行力度更强了，并且获得了打击保健食品工业中许多欺诈性产品及措施的契机。1984年，美国国会委员会将“欺诈性庸医”定义为“任

何出于利益的目的，企图促进已知是错的或者未经证实的医疗方案、治疗措施的人”（美国众议院，老龄问题特别委员会，1984）。随着上述政府管理工作的执行力度越来越强，保健食品工业感受到的威胁也越来越大，它们必须做出抉择：是更谨慎地只提供有研究证据支持的产品和服务，还是联合起来与政府管理工作作对。不幸的是，它们选择了后者。

一个名叫营养健康联盟的组织进行了大量政治游说努力。虽然这一组织的名字看上去好像代表着值得信任的权威，但实际上只是一些含糊的措辞，主要是用来掩饰该组织的真实用意。美国食品与药品管理局拥有批准新药能否以膳食补充剂的名义进行销售的权利，营养健康联盟要求取消这一权利，实际上是支持保健食品工业彻底违反规定，即以牺牲对消费者权益的保护来换取商业利益。游说活动很大程度上是基于两种关键的错误信息：（1）如果零售商不支持他们所提出的法令，就要被踢出业界；（2）如果消费者不支持他们所提出的法令，美国食品与药品管理局的做法就会使市面上没有维生素可买。这两种观点虽然都是完全错误的，但在煽动人们的行动方面却十分有效。整个游说运动被构架为一个“保持选择自由”的问题，产生了一种持久与强烈的游说效果，最终促使1994年通过了美国《膳食补充剂保健与教育法案》。

根据该法案，任何人销售任何药物都是合法的，只要他认为那种药物是膳食补充剂，而且没有声称其有特殊的医学疗效。那些未经证实的治疗方法（根据国会所给的定义就是“欺诈性医术”）在较早以前经常打着蛇油之类的牌子声称是灵丹妙药，现在已经是打着膳食补充剂的幌子得到官方的承认了。一些人通过含糊其辞的做法打擦边球，还有很多人则公然违反《膳食补充剂保健与教育法案》的规定，使得原本就没有力度的关于膳食补充剂的标准和约束，更是形同虚设了。1998年11月10日，联邦贸易委员会进行了一次引人注目的行动——网络交易侦察日活动，检查在网络广告中是否存在违法的保健说明。结果是，仅仅一天他们就向1200个对膳食补充剂产品做违法保健宣传的公司发出了书面警告。

然而让人感到害怕的是，即使发现了不良反应，上报了打着膳食补充剂幌子的药品名称，政府机构仍然没有对其进行干涉或处理的权力。也就是说，在产品到达消费者手里之前，不会再有安全性检验措施了。现在判断那些膳食补充剂产品是否安全有效，已经变成你我这样的普通人的事情了，而在以前，这一费时费力、存在危险性并且令人厌烦的工作，是由具有专业知识和职业资格的专家来进

行的。健康食品工业本身对于质量控制也是一种装腔作势、反应迟缓的态度。在美国可信卫生知识理事会 1998 年 11/12 月的时事通报上，摘述了《药剂师之信》（*the Pharmacist's Letter*）上的一项研究结果：“草药类产品的成分含量变化非常大。这种植物是如何生长、采摘以及贮藏的，制药选择的是植物的哪个部分，运用了哪些提取程序，以及其他很多因素都有可能对疗效产生很大的影响。而在一些草药类产品的说明中，其有效成分的剂量差异达 10 倍之多。”

这类警示性的提示并不是空穴来风，许多对市售产品进行研究的结果都支持这种观点，而且那些都是已经通过同行评议程序的正规研究。例如，Ko（1998）报道，在 260 种拥有亚洲专利权的药物中，有 83 种包含有一些尚不清楚的药物成分，或者药物剂量已经超标。大约 10% 的产品含有铅，15% 的产品含有砷，另有 15% 的产品含有汞，其中砷和汞的含量平均值已大大超过了美国药典中所列的安全剂量标准。根据这类研究发现，《消费者报道》（*Consumer Reports*）杂志 1999 年 3 月刊上发表了一篇文章，严厉批评了草药补充剂产业的现状。它敦促国会重新考虑膳食补充剂保健与教育法案，理由是“这一法案造成了市场混乱的局面”。同时它也鼓励读者“抱着怀疑的态度检查‘自然的’这个词，因为说起来，毒芹或砷也都算是自然物质”。

随着膳食补充剂保健与教育法案的通过，市面上新的膳食补充剂的数量激增。膳食补充剂原来被限制只能在保健食品商店出售，因此消费者很容易做到避免其影响，而现在不再有这一限制了，在药房和食品杂货店的柜台上也可以找到它们。膳食补充剂与那些已通过检验的治疗性药物放在一起，而且通过各种各样的邮购服务或网上订购方式都能买到。这种合法的自由混战的局面，使得我们必须保持见多识广、博学多闻的头脑，而运用批判性思维的技巧显得尤其重要。

随着美国国家补充与替代医学中心的成立与发展，随着《膳食补充剂保健与教育法案》的通过，所有人都清楚了，一些政策公然反对科学研究，进而损害到消费者权益保护的基本原则。这是一种错误的权威形式，而且是最为危险的一种形式。信息时代的来临，使得误导民众与教育民众一样，变得比以前要容易得多。由于判断权威的合法性，以及分辨他们究竟是站在谁的利益立场上说话，通常都特别困难，因此，我们必须培养自己在这方面的能力，并时刻准备为了自己的利益，对所了解的信息进行仔细思考。不过从这个意义上说，各位读者当然也不应当不假思索，就轻易接受我的观点。

第二篇

自我欺骗

6. 经验：作为证据的局限性

1994 年，Martha M. Christy 出版了一本书，书名十分大胆，语出惊人：《你自身拥有的完美药物：医学尚未揭秘的令人难以置信的自然神奇疗法！》（*Your Own Perfect Medicine: The Incredible Proven Natural Miracle Cure That Medical Science Has Never Revealed*）。David G. Williams 在《替代保健通讯》（*Alternative Health Newsletter*）中对此书极尽赞誉之词：“如果今年你只打算买一本健康方面的书籍，你就应该买这本书，它介绍的治疗方法任何人随时随地都可以使用，而且对已知的疾病几乎都有作用。”你能猜出这种“完美的药物”是什么吗？让我来告诉你，这本书写的是尿液疗法。

人们如何使用尿液进行治疗？一些颇具个人特色的专家，如 Andrew Weil，认为尿液对皮肤有一些治疗效果。Helen Kruger（引自 Gardner, 1999）则写道：

在南部的某些地方，人们用尿液清洗婴儿的脸，认为可以保护皮肤。尿液还被用作漱口药来治疗喉咙痛、痤疮、割伤和其他伤口。法国有一个习俗，人们将长袜浸泡于尿液之中，然后再缠在脖子上，用它来治疗链球菌感染导致的咽喉炎……人们还把尿液洒到眼睛里，以“治疗”白内障，不过这一习惯让我的眼科医生感到有点毛骨悚然，因为他认为可能会使白内障变得更严重。就在最近，我还听说佛罗里达州的一些医生提出了一种新的减肥疗法，他们将孕妇的尿液注射到肥胖患者的体内，认为可以“分解肥胖细胞”。

Christy 在她这本独一无二的书里，还提到了尿液的其他用法。她就喝尿可以治疗癌症、关节炎、肺结核、艾滋病以及不计其数的其他疾病的问题进行了解释与说明。一点也不奇怪，Christy 举出了自己的亲身实例来证明尿液疗法的功效，据她说这一疗法治愈了她的许多疾病。Christy 称自己为“会走路的疾病百科全书”，说她患的疾病之多别人根本想象不到：盆腔炎、溃疡性结肠炎、回肠炎、慢性疲劳综合征、慢性甲状腺炎、单核细胞增多症、严重的慢性肾炎、慢性膀胱炎、严重的念球菌感染、酵母菌感染、严重的肾上腺机能不全、严重的慢性中耳炎、鼻窦炎以及对某些食物的严重过敏等。Christy 医疗费用的自付部分就达到 10 万美元，她花了好多年的时间尝试了营养计划、大剂量维生素疗法、针灸、脊椎指压疗法，以及她所能找到的“所有的草药制剂和非药物的自然健康疗法”，后来，奇迹就发生了。

Christy 因病绝望，有严重的抑郁，甚至想到自杀。她偶然看到一本书，书上推荐人们喝自己的尿，她决定试一试。结果，她的症状“几乎立即减轻”了。她曾经脱落的头发又重新长出来了，“浓密而有光泽”。她的体重增加了，也有了精神。她说她现在可以游泳、远足、骑马。“令我和家人感到十分惊讶的是，我又重新开始工作了。经过 30 年无休止的病痛折磨，我重新拥有了一个丰富、充实的人生。而这一切，都要归功于那种令人难以置信但简单有效的自然药物”（Gardner, 1999）。

这些令人惊骇的描述，很难不让人怀疑是夸大其词或伪造之词，这与伪科学支持者所提供的许多推荐证明也没什么不同。事实上，由于达不到科学证据的标准而限制发表的个人经验报道，实在是太多太多了。

自我欺骗

尽管法律明令禁止宣传缺乏证据支持的保健观点，但却无法阻止口耳相传。由于伪科学没有得到足够的研究证据支持，其鼓吹者几乎毫不例外地依赖个人经验、传闻和推荐来支持他们的大胆断言。这些故事鲜明生动，极富感染力和鼓动作用，许多人都觉得它们不仅引人入胜，有时还有很强的说服力。在健康领域，由于存在着明显的利益驱动，人们会有意识地撒谎，甚至进行彻头彻尾的欺骗。

不过，有时人们与其他人分享个人经验，也只是出于想要帮助他人的愿望，但却不知那可能是错误的经验。有许多因素导致人们对个人经验产生的原因出现明显的误解。即使是最有善意的人所做的真诚推荐也有可能产生误导，本章将对其原因进行探讨。

在心理学史上，最早获得经验的一种方式就是内省。内省旨在了解自己思想或行为的性质和原因，但它其实是非常不可信的。我们都有自我欺骗的本事。要讨论自我欺骗的所有形式及其出现的影响因素，已经超出了本书的范畴。在此，我将重点探讨我们对自身疾病的原因和治療措施产生误解的倾向，因为这样的误解将会产生可怕的后果。同时，针对完全依据个人经验，而无视所持观点的实质的现象，我也会深入地探讨其局限性。

自我评估健康水平的基本问题就是区分“病感”（illness）和“疾病”（disease）之间的不同。在一篇名为《为什么假的治疗方法似乎更有效》（*Why Bogus Therapies Seem to Work*）的论文中，Barry Beyerstein 就（1997）病感/疾病的区别进行了详细的解释：

尽管“疾病”和“病感”往往被等同和通用……但“疾病”指的是由于感染、组织退化、损伤、中毒、癌变等造成的有机体的病理状态，而“病感”则指的是伴随疾病出现的不舒服、疼痛、定向障碍、功能紊乱或其他不适感。被我们称为症状的那些客观感受的主观反应，其实会受到文化和心理因素的影响，如信念、建议、期待、需要特征、自我服务偏差和自我欺骗等。生病的过程也受到许多社会和心理的利益的影响（常常是不知不觉地），那些社会及心理的利益往往伴随着病人角色的获得（由医学专业人员确认或社会承认）而产生。对于某些个体来说，由于病人角色而产生的特权状态以及获益，足以使其在疾病被治愈后仍保持病感的体验，并永久性地保持下去，甚至使其在没有疾病时也出现病感。

竞争性的解释

谨记“疾病”和“病感”的区别之后，让我们来看看是什么原因使人们认为一些无价值的治疗方法是有效的，并且还真心相信自己所说的是事实。一般来说，人们尝试某种治疗方法后感觉好多了，这常常被视为治疗有效的最有说服力

的证据，然而事实并非如此。“我试过这种治疗方法，感觉比以前好些，所以这种方法有效”这句话背后存在着逻辑矛盾。因为还存在许多能使感觉好转的竞争性解释，其中任何一种解释都可能是病情好转的真正原因，而我们在得出治疗有效的结论之前还没有排除这些影响。下面我要介绍 10 种所谓治疗有效的竞争性解释。

安慰剂效应

在与健康相关的所有现象中，最令人惊奇的是仅仅是人们对于康复的期待，就会产生积极的作用。也就是说，仅仅因为暗示作用，人们就会觉得自己正在接受的治疗有效，感觉好很多。用科学术语来说，这是“安慰剂效应”，而且安慰剂效应的作用还很强。举例来说，在抗抑郁药物试验中，大约有 75% 的药物疗效都是来自于安慰剂效应（Kirsch & Sapirstein, 1998; Walach & Kirsch, 2003）。

但是，怎样来判断身体好转是由于暗示的作用？另外，即使能确定有暗示作用的影响，又如何判断这种好转多大程度是治疗的效果，多大程度是由于暗示的作用？要回答上述两个问题，必须进行科学的实验。在征得病人的同意后，将其随机分配到三个不同的实验条件组，然后再比较他们的实验结果（Walach & Kirsch, 2003）。以抗抑郁剂疗效研究为例，第一组病人接受抗抑郁药物治疗（治疗组），第二组则使用糖丸代替抗抑郁药物，但病人不知道（安慰剂组），第三组是等待治疗的病人（实验结束后才开始治疗），因此在实验过程中是未接受任何处理的（空白对照组）。无论是病人还是发放药物的实验助手，都不知道分组的情况，这一点很关键。这样，安慰剂组就和治疗组具有相同的期待，而空白对照组则没有期待。这种双盲（double-blind）处理实际上是从研究设计上就将暗示作用作为一种竞争性解释，以考察是否存在安慰剂效应的影响。那么，实验者要做的就是运用标准化的评估方法，监控所有病人抑郁程度的变化并收集数据，在事先定好的治疗时间完成之后，比较各组病人的结果。

如果治疗组和安慰剂组的疗效相似，而且两组的疗效都优于空白对照组，那么最可能的结论是：暗示作用可以解释所有的治疗效果——真正的药物治疗效果与糖丸无异，或者说在治疗中更重要的是病人的信念，而不是药物治疗。当然，如果治疗组的疗效显著优于空白对照组和安慰剂组，你就完全可以作出“抗抑郁剂治疗有效，且其疗效显著优于暗示作用”的结论。此外，由于有安慰剂组的存

在，你还可以计算出治疗组和空白对照组之间的疗效差异有多少是安慰剂效应，也就是估计出有多大比例的疗效是由于暗示作用产生的。

要对某种药物的疗效进行检验，空白对照组和安慰剂组都很重要，但在某些情况下，治疗中活性成分的副作用会“暴露”实验组，因而破坏了关于暗示作用的实验对照（Greenberg & Fisher, 1997）。如果通过副作用的出现与否，可以推测出谁接受了药物治疗，谁没有接受药物治疗，那么治疗组和安慰剂组之间的比较就被“污染”了。例如，在一篇对当时所有采用了双盲和安慰剂对照实验设计的、关于百忧解*药效研究的综述中（Greenberg, Bornstein, Zborowski, Fisher, & Greenberg, 1994），作者发现在每一项研究里，被试报告有副作用的比例和治疗效果的大小都有很高的正相关，因此，可能是百忧解的副作用暴露了治疗组的许多成员，从而加强了其暗示作用。根据这一综述及其他这类研究，Greenberg 和 Fisher（1997）提出，和药商宣传的正相反，百忧解不仅没有比原有的抗抑郁药物更有效，而且也没有可信的证据表明它比安慰剂更有效。有趣但也极具讽刺意味的是，他们提出，市场宣传比其他同类药物副作用更小的百忧解，可能恰恰是通过产生副作用、引起生理反应来产生疗效的。当然，这种假设也可以通过实验来进行验证：即不给安慰剂组无任何活性成分的糖丸，而是给其含有某些活性成分但没有治疗作用的安慰剂，使该组患者产生和治疗组相似的生理反应。这样，就可以排除前面所说的那个竞争性假设：即由于实验对照不好，药物副作用暴露了治疗组，导致得出治疗有效的错误结论。

如果不进行一个有良好对照的实验，就没有办法确认安慰剂效应对健康改善是否会产生影响。因此，那些基于有限的个人经验，没有进行实验对照研究的推荐，是不可能将真正的治疗效果与暗示作用区分开来的。

自愈

有时候，由于一些我们目前还不甚明了的原因，疾病甚至是癌症，没有经过治疗就自行痊愈了。但是，实际上这种情况非常罕见。Beyerstein（1997）指出，一位肿瘤学家亲眼目睹了在他治疗的大约 6000 名癌症病人中，有 12 例是自行缓解的。毫无疑问，这 12 个幸运者尝试过的任何一种治疗方法，都会将该患者的

* 百忧解，Prozac，一种抗抑郁药。——译者注

痊愈当作是本疗法治疗成功的一个典范。鉴于这种自愈现象的存在，Sagan（1995）对天主教圣地 Lourdes 所谓神秘治疗作用的成功率进行了分析。当时已有近 1 亿人到过法国的 Lourdes，寻求治愈各种疾病的机会。天主教教会拒绝了其中绝大多数人，最终只在 Lourdes 接受了 65 人给予治疗。这就意味着治愈的几率甚至低于百万分之一，这个几率与到 Lourdes 商务旅行遇到飞机失事的几率差不多。而且，从统计数字来看，癌症病人到 Lourdes 的治愈几率实际上比那些呆在家里出现自愈的几率还要低。一般而言在每 1 万个到 10 万个癌症病人中，就会有一个人出现自愈现象，那么在去过 Lourdes 的 1 亿人当中，即使只有 5% 的癌症病人，我们都可以期望有 50 名到 500 名癌症病人会由于自愈现象奇迹般地好转。但是，在天主教教会确认的 65 例非凡的治疗成功案例中，只有 3 例是癌症病人。

疾病发展的自然过程

如果所患疾病并非慢性病或致命性疾病的话，那么经过疾病的自限过程，我们的身体最终是会全面康复的。就像一句老话说的，“无论你是连续七天都喝鸡汤，还是坐等一个星期，感冒都会好”。但是，由于对自愈过程的误解，当疾病最终按照其自然发展过程逐渐消失时，当时所进行的任何治疗方法都会被误认为是有效或成功的。而且，一些医生也借此弄虚作假，将患者自动送上门的信任照单全收，全然不顾实际上是什么原因使得疾病痊愈。例如，Livingston（1998）引用了据说能治疗头痛的磁性项链的使用说明，“一出现头痛就戴上项链，头痛一消失就取下项链”。即使磁性项链实际上没有什么作用，严格遵照该使用说明，我们都会不断地产生这样的印象：磁性项链有助于减轻头痛。问题是，这种治疗方法要求时间上的一致性：直到头痛消失，你才能取下项链。因此头痛消失和取下项链必须同时发生。为了更清楚地认识到“磁性项链能减轻头痛”这一结论的逻辑缺陷，我们不妨设想一下：你也可以保持单腿站立直到头痛消失，然后认为就是单腿站立的方法治愈了头痛。

许多疾病的循环特点

诸如关节炎、多发性硬化、过敏症、抑郁症、焦虑症等慢性病，都会有阶段性变化的倾向，病情时好时坏，不断在症状缓解与加重之间变换。那么，人们什

么时候最可能寻求医疗帮助呢？当然是在病情最糟糕、痛苦程度最高的时候。然后，仅仅是由于正常的疾病循环的特点，治疗过程就很可能与病情好转同步出现。在疾病循环往复的变化过程中，治疗有很多机会适逢病情好转。由于通过间歇和变化的强化过程形成迷信观念的过程十分容易（见第4章），人们也因此很容易对一个毫无价值的治疗方法坚信不已。

误 诊

医生的工作绝不可能做到十全十美，即使有最精密的实验室检查手段，也难免会出现误诊。如果一个人被误诊为有某种疾病而寻求治疗，那么当他后来被确诊为健康状态的时候，人们就有可能认为他接受的那种治疗是有效的。其实起初就没有疾病，但无论是医生和还是病人都对这一点毫不知情，因而得出了错误的结论。我们在第13章里还会举例说明，持伪科学健康观的江湖郎中常常会采用未经确定的诊断性测试，然后再据其结果得出可疑的诊断。

对疾病的过分担忧

目前，可能是由于人们日益关注和强调预防性保健工作的重要性，许多没有疾病的人也会关心健康问题并寻求保健。一些普通的疼痛或者痛苦，伴随着一系列纯粹主观性的担忧（我们在前面曾提到的病感），被人们上升到了疾病状态的层面。对于有上述这样症状表现的人，有一个术语描述是“对疾病过分担忧”。他们的“病”从没有真正治愈，因为他们的抱怨一开始就不是以真正的疾病为基础。伪科学的从业者们使用伪医学的行话，热心地造就了大批的疑病患者。这里有一个刊登在报纸上的关于“结肠治疗”的广告（见 Barrett & Herbert, 1994），上面这样写着：

你会清洁牙齿，为什么不清洗你的肠子？我们提供一种健康疗程帮助你清肠，让那些使你感到疲惫、难受、胀气、抑郁、浮肿的有毒物质一扫而光！让你不再厌食，进而对碳水化合物重新产生渴望！清肠可以让你的便秘、有念球菌感染和寄生虫的状况得以减轻！清肠可以增添精力，纯净心灵，还可以减肥！

说服健康的人，让他们相信自己也需要保健，这是一个多么有创意的办法！

通过吸引那些处于正常健康状态的客户，伪科学的从业者们就可以轻而易举地使其治疗事业兴旺发达，然后舒舒服服地享受治疗成功带来的好处。

症状缓解

把诸如疼痛之类导致病感的症状缓解，和真正的治愈疾病混为一谈，这是一个相对低级或简单的错误。症状的缓解并不意味着疾病的病理学基础得到了治疗和改善。例如，艾滋病患者可以通过服药来减轻某些痛苦，但他们的病根——艾滋病病毒并没有消失。在这里，我要再次重申“病感”和“疾病”是有很区别的。当症状有所缓解时，即使疾病仍继续存在，人们也会觉得病好了一些。仅仅因为症状出现缓解而相信治疗是有效的，会使得病人的注意力从关注疾病病程的发展中转移开来，而事实上，后者才是真正最值得关注的东西。

两面下注的行为

为了保险起见，许多病人在接受伪科学治疗的同时，也会接受有证据支持的科学治疗。1998年5/6月刊的《美国反医疗欺诈委员会通讯》中，报道了一个戏剧化的案例，就是典型的为避风险两面下注的行为：

Anne Frahm 的乳腺癌早在诊断之前就已经广泛转移了。她接受了手术切除、化疗、放疗以及激素治疗等多种方法的治疗。在这些方法都没有治好她的病的情况下，她又进行了实验性的自身骨髓移植，医生告诉她这是最后的治疗方法了。五个星期后，她彻底痊愈了。

奇怪的是，Frahm 并没有将病愈归功于骨髓移植的生理治疗效应，而认为这是她采用极端的灌肠食疗方法的功效——为了“排毒”，两个星期只吃生的水果和蔬菜，喝水果蔬菜汁，不吃一切糖、肉和奶制品。

Frahm 后面提到的治疗癌症的食疗方法，没有一个是经证实有效的，但 Frahm 却将乳腺癌的痊愈完全归功于它们。她非常相信就是这些非传统的治疗方法治好了她的病，以至于她和丈夫还准备在科罗拉多的斯普林斯开办一个服务中心，向人们传授这些治疗方法。如果你面临治疗方法的选择，你会把赌注押在哪里？骨髓移植，还是 Frahm 夫妇的健康服务中心？

衍生出的好处

许多保健行业的从业者是情绪高昂、极具说服力的。他们能让病人产生目标感、意义感、希望以及其他许多心理上的获益。例如 Andrew Weil，在前面有关心灵感应术的章节里我们曾提到过他，他既是一个具有超凡魅力的演讲者，也是一位极具天赋的作家。许多人对他那种将有意义的观点与无意义的废话独特地糅杂在一起的言论非常满意，觉得可以给他们“全新的生命意义”。由于这些由医生提供的心理上的获益，病人开始更加关心自己。他们会更加注意饮食平衡、积极参加锻炼、好好休息、多参加社交。这些副产品或衍生出的好处中的任何一种，都可能使病人的健康状况有所好转。但是，当人们在确认成功治愈疾病的原因时，这些衍生出的好处就不如治疗方法本身那么显而易见，往往会被人们忽略。

对现实的心理歪曲

与健康相关的个人经验可能产生误导的最后一种形式是，对现实的心理歪曲，它涉及了社会心理学中研究最多的一个原理：当我们同时具有两个或两个以上相互矛盾的观念时，就会感到不舒服，这被心理学家称之为“认知失调”现象。我们总是尽量设法去减少这种认知失调，常用的解决办法有改变已有的观念或形成新的观念（Festinger, 1957）。下面我们将通过一些有趣的例子，来看看认知失调是如何被减轻的。

首先，让我们看看犹太裁缝的案例（Ausubel 原创，1984,）：

在一个反犹太分子的小镇上，一个犹太裁缝冒失地在大街上开了一家裁缝店。为了把这个裁缝赶出小镇，一群年轻人天天到店里来，他们站在门口大声喊：“犹太人！犹太人！”

度过了几个不眠之夜后，这个犹太裁缝终于想出了一个办法。当那群人再次来威胁他的时候，他宣布：凡是喊他“犹太人”的人，都可以得到一角钱。然后，他给每人发了一角钱。

这种新的刺激和鼓励措施使那群年轻人兴奋不已、兴高采烈，他们第二天又来了，大喊“犹太人！犹太人！”犹太裁缝微笑着，给每人发了五分钱（他解释说今天只付得起这么多）。那些人满意地走了，毕竟，五分钱也是钱。

第三天，犹太裁缝给每人只发了一分钱，他解释说再也付不起更多的钱了。一分钱可算不上什么刺激，那些人开始抗议。

犹太裁缝回答说：“你们可以拿走这一分钱，也可以不要。”那群人决定不要了。他们大声说：“这个鬼裁缝一定是疯了，他居然以为我们会为了一分钱来喊他‘犹太人’！”

就这样，通过一个巧妙的心理学策略，犹太裁缝让这些年轻人产生了认知失调。这个聪明的裁缝假装很喜欢他们的行为，于是他们接受了他的一角钱，并把大声喊“犹太人！犹太人！”看成了一份工作。当报酬减少以后，他们感到这份工作的收入与自己付出的努力并不相称，因此产生了认知失调。而要减轻这种认知失调，只能放弃工作。

当我们做决定时，也会面临认知失调的压力。在一个经典的“决策后”失调减轻的案例中，Robert Knox 和 James Inkster（1968）评估了人们在赛马场赌马的信心。他们发现，那些在下赌注后 30 秒接受询问的人所表现出的信心，比如下赌注前 30 秒接受询问的人要强得多。一个人对赌博中获胜几率的看法，是如何在这么短的时间内发生改变的呢？答案就在于下赌注的行为使人们产生了认知失调。你在尚不确定赢利机会的时候就把钱押了出去，这会使你产生认知失调：如果赢的可能性很小，还去下赌注，似乎很愚蠢。而解决的办法则是：增强你对获胜几率的信心，这样就可以减轻认知失调了。

同样的道理在许多人评估治疗疗效时也起着作用。如果你自愿采取的治疗措施已耗去了大量的时间、金钱、努力和希望，那么宣布治疗完全失败岂不是反映出你的判断失误？已投入的资源一去不复返了，这是你根本无法改变的事实。那么解决的办法只能是：改变你对治疗效果的看法。你必定会宣称该治疗方法已取得了某些方面的疗效，即使那只是症状层面的。再通过一些轻微的歪曲，治疗效果就可能被拔高成“取得了相当的成功”。就这样，出于减轻认知失调的目的，病人（间接地还有医生）就会从许许多多的失败中揪出一星半点的成功。

底 线

上述 10 个因素中，没有一个是有意欺骗或不诚实，但它们都说明了将个人经验作为证据的不可靠性。对于所感觉到的身体康复，存在有那么多竞争性解释，简直就没有办法知道究竟哪一种解释才是正确的。因此，对于传闻、个案

研究或推荐，最合理的反应就是：作为一个故事来听，它们可能很有趣，而且还提供了进一步研究的可能方向，但是在判断一条基本原理是真是假的问题上，它们则完全没有价值。

鉴于此，即使收集了很多人的推荐意见，也是毫无用处的。前面我们所讨论的那些竞争性解释，没有哪一个能通过大量增加推荐的数目来排除或解决。每一个案例都会面临那些竞争性解释，而且，同样的解释也可以用来说明许多人明显的治疗获益。请设想一下，如果有许多人都在服用一种被认为可以治疗抑郁的草药，但他们不知道这种药其实是没用的，这会怎么样？由于安慰剂效应可以解释抗抑郁药研究所报道的大部分治疗效果，那么许多用过这种草药的人很可能是由于安慰剂效应而获益。事实上，仅仅是相信自己正在接受一种有效的治疗，就确实能使病人的抑郁出现某种程度的减轻。所以说，即使有很多人推荐，还是无法排除这种可能，即一个或多个竞争性解释可以说明所有人觉察到的好处。由于同样的原因，或者是在某一特殊案例中各种原因的综合，会导致许多人在同一时间全部犯同样的错误。要解决这个现象解释的难题，唯一可靠的办法就是科学实验研究，也只有它才可以将真正的治疗效果从所有潜在的混淆因素中系统地分离出来。

概率的基本规律

竞争性解释貌似合理，但其本身存在显著问题，除此之外，从个人推荐中学习经验也是非常危险的，因为仅仅是单个案例可以支持许多观点，无论这些观点有什么样彻头彻尾的错误。例如，《新英格兰医药杂志》(*New England Journal of Medicine*) (Kern, 1991) 上曾刊登了一篇颇有吸引力的报告，说一位 88 岁的老人每天吃 25 个鸡蛋，至少已有 15 年了，可他的胆固醇水平仍然正常。虽然大众报刊常常过分夸大鸡蛋的危害，但实际上吃适量的鸡蛋（平均每天一个）不会增加患心血管疾病的风险（Hu 等，1999），而且可能还比吃其他替代鸡蛋的食品更健康（Vaupel & Graham, 1980）。尽管如此，也没有哪个有责任心的营养学家会根据某人每天吃 25 个鸡蛋仍然能保持身体健康，而建议大家都来每天吃 25 个鸡蛋，事实上，这个病人引起医学家们的注意，并对他的胆固醇新陈代谢过程产生特别的兴趣，恰恰是因为他的身体特性与众不同。要知道个人经验，不论是你自己的还是其他人的，都只是一个单一的个案，而个案往往是非常特殊的。

知识是以一些基本的原理为基础的，而基本原理就是当将其运用于新的案例、事件、病人等的时候，都具有一定准确性的规律。如果案例之间不存在共同性，知识就无法积累。实际上，人类任何一种思想或行为的基本原理的本质都是概率问题，记住这一点至关重要。基本原理揭示了事件之间显著的关联，即在事物间存在一定程度的相互联系的趋势。任何一种概率趋势都会出现例外，健康科学中所有与症状、疾病、诊断和治疗有关的问题也是如此，总会有一些特例出现。但无论如何，仅凭单一的反面例子是不能推翻一个概率规则的。

请考虑这样一个问题：对学业表现进行量化测量，如高中成绩等级和标准化测试分数，是否可以预测学生会在大学里取得成功？如果有人愚蠢地宣称这些测量成绩就能完美地预测在大学阶段的成功，那么仅仅一个特例就足以证明这个论断是错误的。我们可能都会想起一些人，他们在高中的成绩很好，但在大学成绩不好；或者在高中成绩不怎么样，但到了大学则成绩优秀。不过，因为有这样几个人的存在，就意味着成绩等级或测验分数不能用来预测大学阶段的学业成功吗？完全不是那么回事。让我们来看看学生高中学术能力评估考试（Scholastic Assessment Test, SAT）和大学一年级平均成绩（Grade Point Average, GPA）的关系图（见图3）。

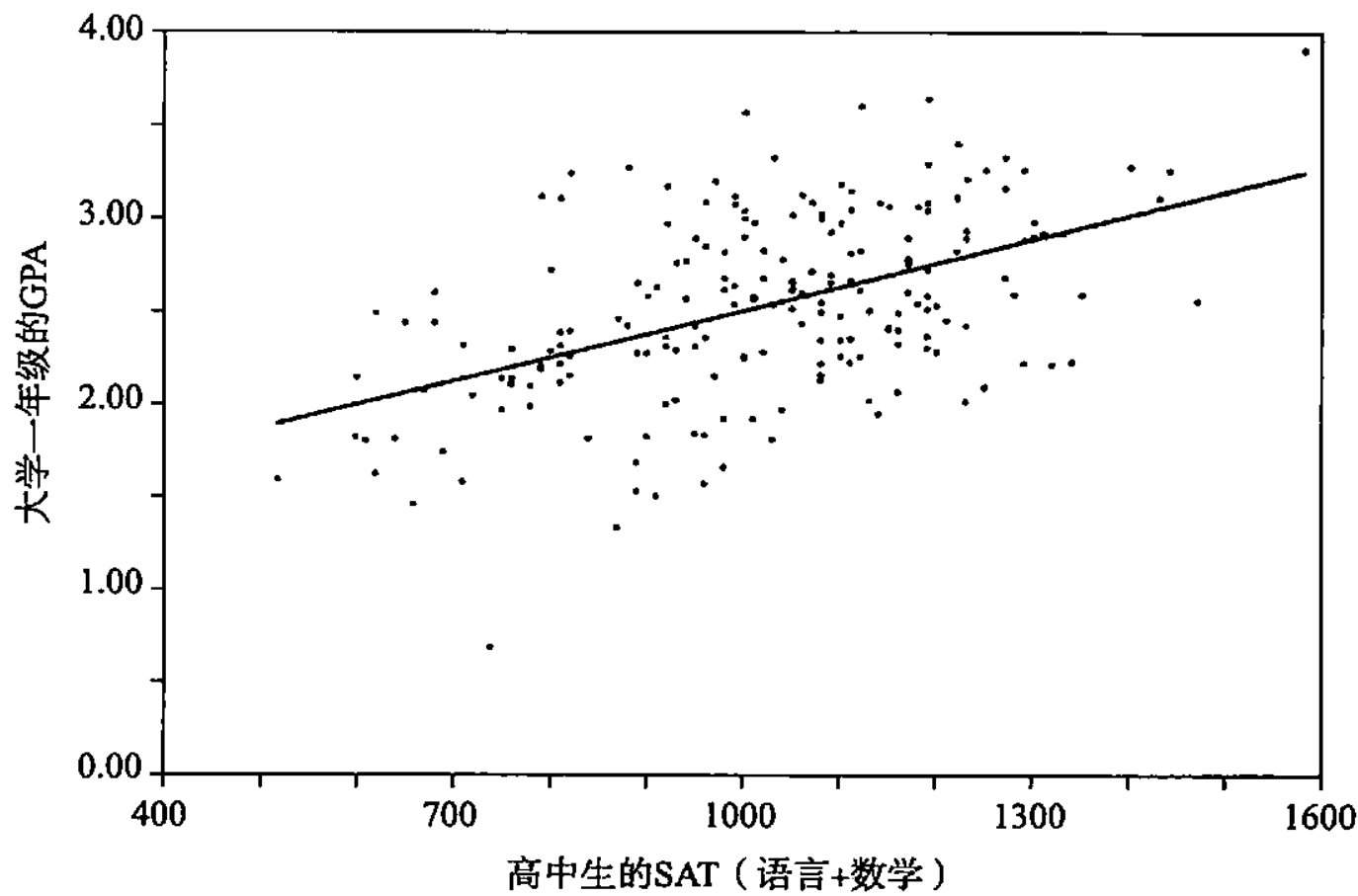


图3 高中生 SAT 分数和大学一年级 GPA 之间的关系。直线表明，虽然有相当程度的变异，但还是存在一个概率普遍规律：SAT 分数较高的人往往大学分数的等级也较高

如图所示，两个变量之间存在一种普遍性的关联，但也有一些特例。图中直线左上方有几个点，表明有几个分数低于 SAT 平均分的学生在大学里是优等生；而右下方也有几个点，表明有几个分数高于 SAT 平均分的学生在大学里被留学察看。但仅仅是这些孤立的案例，并不能否定数据的整个趋向。绝大多数点的位置表明，在一定误差范围内，学生的大学成绩是可以根据他们的 SAT 分数来预测的。因此，研究揭示了一个概率的基本规律：大学入学成绩好的学生，更有可能得到好的大学成绩等级。如果你根据自己认识的几个恰巧是特例的高中生的情况，进行过度概括的话，你就可能会得出错误的结论：以大学入学成绩为录取标准是没有用的。

上面的这个例子证明了通过个人推荐来学习经验的危险性：那些个人推荐可能正好是概率规则中的特例。而在日常生活中，我们常常会选择性地接触到这样的特例。哪个人的故事更可能引起人们的注意？是 SAT 分数达到 1300，在大学里成绩也不错的学生；还是 SAT 分数达到 1300，但在大学里成绩却很糟的学生？特殊案例往往被认为具有更多的新闻价值而广为传播（见第 9 章），而对那些符合概率基本规律的情况，人们则因其理所当然而熟视无睹。

一旦我们对那些属于恰巧治疗有效，但其实是概率基本规律的特例的个人推荐进行了过度概括，我们就可能会形成错误的信念。实际上，我们自己的经验也仅仅等同于一个个人推荐而已，这虽然与直觉相悖，但也许是最重要的一条原则。因此，如果有任何信念系统是鼓励高估自己的个人经验，我们都应该对它保持高度警惕。只有当我们认识到仅凭个人经验无法得出结论，需要系统地观察更多人来收集可靠的证据时，我们的知识才能真正有所积累。

7. 貌似合理：并不是所有信念都具有同样的价值

一天早上，Zeke 醒来的时候，发生了一件非常可怕的事情。他完全动弹不得，却有一种轻飘飘地浮在半空中的奇怪感觉。很快他就觉察到房间里还有陌生人，一切都很清楚：是外星人来了。这些小小的灰色生物，有着不合比例的大脑袋和大眼睛，把 Zeke 劫持到碟形的飞行器中。在飞行器中，对外星人对 Zeke 进行了创伤性的医学实验，自他的身体探查取样，盗取他的精液。当这些可怕的实验结束后，外星人警告 Zeke 要尽快净化地球环境，防止全球性的艾滋病爆发。然后，外星人把 Zeke 送回了他的卧室。那些医学实验并没有在他身上留下可见的伤疤，而正在他身旁安然熟睡的妻子，更是一点儿都没有觉察到他曾离开过。这是一件非常激动人心和不可思议的事情。然而，究竟发生了什么事情呢？

世界上并没有 Zeke 这样一个人，但是许多声称曾被外星人劫持过的人都讲述过类似的故事，仅仅是在细节上有所出入而已（Saga, 1995）。他们十分激动地反复述说自己的故事，看上去完全信以为真。哈佛大学精神病学家 John Mack 相信这些关于外星人劫持的故事，并认为这些故事之间存在的明显一致性可以作为确认其可靠性的依据。但这一证据究竟意味着什么？目前已有的证据是否足以证实那些奇异的想法？科学家常说，“非同寻常的观点就需要非同寻常的证据”。外星人劫持的情节里包含了如此众多的非常元素，因此出于谨慎的考虑，应当要求提供强有力的证据来证明其真实性。

我们已经了解到伪科学的制造者是如何使用各式各样的伎俩使人们相信其想法和做法的，其手段包括欺骗性的语言、魔力思维、一厢情愿的想法以及借助权

威的力量等等。成功的伪科学的标志之一就是表面上的合理性，它们往往有一个十分引人入胜但却虚假不实的主题故事来做幌子。然而在这个故事的背后，隐藏着各种矛盾，并且违背了公认的、以科学证据为基础的有关现实的概念与想法。因此，仔细审视各种观点所提供的主题故事，从知识的角度评价各种观点的真实性，是非常重要的。即使在将各种信念放在严格控制的条件下进行实验验证之前，也能看出有一些想法会比其他想法更可信。周密的合理性或真实性检查，是一种十分有用的初步测量观点可靠性的手段。尤其要考虑的问题是各种观点中所涉及的机制，即所谓的原因是如何导致相应效应出现的。如果没有令人信服的理论或独立的证据来支持这种机制，那么该观点的真实性就值得怀疑。

为了阐述由于主题故事表面的可信性而接受其各种观点的危险性，在这一章里我们将探讨几种广为流传但有疑义的伪科学信念。深入地评估其合理性或真实性，会揭示它们与现实极其脱节，因而也极其不可信的实质。在讨论外星人劫持的例子之前，我们先来研究两种缺乏令人信服的理论基础、与健康相关的伪科学——磁疗法和同种疗法，以及一种垃圾科学（其特点是一些非真实的理论被误认为是支持性的证据）——电磁场与癌症之间的所谓关联。

磁疗法：假说 VS 现实

Sabadell（1998）汇编并列出了一些关于静止磁体与磁场的奇怪观点：

- 细胞依靠磁能生存，而绝大部分的磁能是由水提供的。
- 磁场会改变水的结构，以及水的一些物理特性。
- 泉水是已经磁化了的。但如果你用瓶子装泉水，五天后它就会失去功效。
- 磁化水接触金属就会丧失疗效。
- 80% 的普通疾病产生于缺磁综合征。
- 多喝磁化水。这是最简便易行的办法……这种方法可以治愈脓肿、痤疮、过敏、贫血、关节炎、哮喘、支气管炎、蜂窝组织炎、坐骨神经痛、糖尿病、腹泻、消化不良、性冷淡、阳痿、流感、疱疹、头痛、带状疱疹，还有许许多多其他疾病！
- 磁场增加了体内血浆中氧的饱和度，与血红蛋白一起输送更多的氧气。

- 将治疗性磁体用于疾病领域。北极是治疗疼痛的，南极是用于恢复已失去的能量与活力的。

面对这些观点时，人们也许会首先问一个基本问题：它们的机制是什么？换句话说，磁体是如何产生上述功效的？历史上，关于磁体和磁性曾出现过不少误解。“动物磁力说之父” Franz Anton Mesmer (1734—1815) 是最先使用强磁体治疗重性精神疾病的。到后来，他认为自己可以“磁化”任何事物，如纸、木、水等等，而且不论磁体是什么，都可以取得同样的疗效。最终他得出结论：他已经拥有了真正的力量，并且将他所认为的治愈疾病的效果都归功于他和病人之间的“宇宙磁流”。后来这一疗法被称之为“麦斯麦术”（后也被称为催眠术）(Mesmerism)，在巴黎备受推崇。然而，关于这一疗法也存在许多争议：

1784 年，路易十六成立了皇家委员会，评估动物磁力说的种种观点与主张。该委员会成员包括安东尼·拉瓦锡 (Antoine Lavoisier) 和本杰明·富兰克林 (Benjamin Franklin)。他们做了一系列实验后得出结论：所有观察到的效果都可归功于暗示的力量（安慰剂效应），而且，“磁化行为是逐渐增加患者想象的方法”。委员会报告发布后不久，托马斯·杰弗逊 (Thomas Jefferson) 抵达巴黎，在他的日志上写道：“动物磁力说完了，徒留笑柄而已。” (Livingston, 1998)

遗憾的是，虽然已证实“麦斯麦术”只是起到安慰剂效应，和糖丸或其他无效治疗差不多，但并没有终结“麦斯麦术”的继续推广和应用。事实上，有许多伪科学现象就源于这种理论或信念。例如，注重共情的治疗者发展出通过手势治疗疾病的方法，即现在所谓的治疗性触摸。Mary Baker Eddy 深信是磁体治好了自己的病，后来她又认为祷告的方法不仅能取得同样疗效，而且比磁疗更有效，由此她创建了基督教科学派 (Christian Science)。

上述治疗方法都是建立在安慰剂效应的基础之上，没有一种能够证明其有效性超过了暗示的力量。第 15 章中将会提到：实际上，无论是采用还是倡导任何一种未经验证的治疗方法，都是违反伦理道德的。在第 4 章中，我们讨论并且证实了治疗性触摸是一种缺乏实验支持的治疗方法。第 12 章还将讨论在基督教科学派抚养的儿童中，由于宗教目的而出现的医疗忽视问题。在这里，我们并不会深入探讨那些源于麦斯麦动物磁力说的令人置疑的分支及流派，我们将集中讨论

一种与动物磁力说有关，据说能取得显著治疗效果，因而已发展成为数百万美元的工业产业的现象——磁疗现象。

目前，消费者在市面上可以买到各种各样包含永久性磁铁的系列产品，如项链、手镯、耳环、坐垫、枕头、床垫、腰带等等。而且，有适合身体的各个部位（手、腕、肘、膝、踝和足等）使用的磁性产品，最近磁性鞋垫也特别畅销。Florsheim 鞋业公司就紧紧抓住了这一商机，利用磁疗风潮拓展业务，将磁植入一些鞋类产品中。磁体不仅用来治疗疾病、缓解疼痛，还被用于提高运动成绩；不仅受到了高尔夫玩家的欢迎，还用在了纯种赛马的训练中。不过在这里，有必要再次提出“机制是什么”这个基本问题了。下面我们将根据已知的磁体对水和人体的影响，以及通常所出售的磁体的磁场强度两个方面，对治疗性磁体疗效的基本理论进行讨论。

磁与水

我们前面提到由 Sabadell 收集的许多观点，都涉及了磁体对水的影响，以及磁化水对我们身体的可能影响。但是，水究竟能不能被磁化呢？磁学研究专家 James Livingston 曾简要地介绍了相关的科学知识：

水是抗磁性的，即与磁场微弱相斥。如果你使用一个磁场来进行影响，水分子中电子的运动方向就会有微弱的调整，会产生一个与应用磁场方向相反，且比其强度小 100,000 倍的净磁场。不过，一旦移除应用磁场，电子就会回到原有的轨道。水分子又重新变得没有磁性。（我们也许应该注意到，某些磁疗法的拥护者倡导饮用“磁化水”。但其实你不可能磁化水。虽然水对外界的应用磁场会有微弱的反应，但这种反应在磁场移除后几乎立即就消失了。）（1998）

磁疗法中提到的有关磁体与水的观点，有部分是正确的。磁场确实改变了水的结构，尽管其产生的影响是微弱的、方向相斥的（而不是相吸），而且十分短暂（在磁场移除的瞬间即消失）。不过认为泉水在装瓶后的 5 天内可以保持磁化，则几乎没有这种可能。唯一能喝到磁化水的方法就是在磁场中把水喝掉。

磁与人体

人体接触磁场后，真的会产生所谓的疗效吗？磁疗理论认为，磁场有治疗疾

病的作用，因为磁场会增加达到身体目标部位的血液流量（血中含铁），由此可带来更多的营养成分，并带走新陈代谢所产生的废物。但是，这一理论与我们所知的事实直接矛盾。水与磁场微弱相斥的特性是尤为关键的原因。与绝大多数生命形式一样，人体的主要构成成分就是水，血液、其他身体组织和体液均主要由水构成，因此，人体也是与磁场微弱相斥的。理论上说，足够强大的磁场可以使人，或者水滴、花、蚱蜢和青蛙等物质飘浮在空中。但我们的血红蛋白中仅含有少量的铁，其在外界磁场作用下使血流产生的变化，根本不足以消除数量大得多的水所产生的相斥作用。所以，如果磁场能对人体血液产生任何影响的话（这种影响实际上是极其微小的），那影响也是它会促使血液流出身体的目标部位。

“治疗性”磁体的力量

最后要注意的是，市面上出售的磁疗产品，其磁场强度通常是非常微弱的，以至于人们用不着担心它们会使血液更多地从易受影响的部位流出。一个磁场的强度要取决于几个因素，包括磁体的大小、磁体是有独立的北极和南极（在单极磁体中）还是没有（在多极磁体中）。一般来说，当其他所有条件都一样时，较大的磁体会产生较强的磁场，但南北极的排列方式也是很重要的一个因素。比如，老式马蹄形磁铁（单极磁体）的南北极排列方式，会产生一个较强的磁场。而新式的冰箱磁贴（多极磁体），其南北极是呈交替排列、圆形排列或三角形排列的，其产生的磁场强度则是相对较微弱的。这点实际上很显而易见，因为许多冰箱磁贴仅仅只能夹住一张很薄、很轻的纸。

绝大多数用于治疗磁体的都是多极的，也就是说它们的磁场仅能穿透几毫米，几乎连皮肤以下的组织也不能达到。因此，人们根本不必担心血液会从磁场影响的部位流走。尽管这会令磁疗的拥护者们大失所望，但我们其他人则应当感到十分庆幸：正是因为这样，我们的信用卡、支票保证卡、ATM卡、磁带、录像带、软盘、硬盘，还有其他的磁性装置和设备，才不会因为我们自己或其他人携带的磁体而不分青红皂白地遭到破坏。另一方面，好在治疗性磁体的磁场微弱，在别人采用了错误的方法给我们治疗时，我们才可以免遭大的伤害。因此，静止磁体能对人类健康产生有益影响，这一观点及其机制并没有说服力。既然磁疗方法缺乏理论基础的支持，那么这一大胆假说没有得到实验依据的支持，也就不足为奇了。

同种疗法：空洞的承诺

如果说磁疗法毕竟还是运用了实实在在与人体相互作用的仪器（尽管是以微不足道并且无效的方式在起作用），那么同种疗法与现实的关联就更少了。同种疗法是德国医生 Samuel Hahnemann（1755—1843）创建的一种治疗方法，其治疗假说是基于两条法则，而对每一条法则而言，都存在大量与科学事实矛盾之处。

相似法则

假如你被毒蛇咬伤，来到急诊室求治，你会怎么办：是选用有针对性的解毒剂治疗，还是再注射一点毒液以毒攻毒？这不是无谓的发问，因为在现实生活中，真正相信同种疗法的人就会选择以毒攻毒的方法。

同种疗法的第一条法则被称为相似法则，它是指能导致疾病发生的物质同样也能用于治愈已经患有该病的人。Hahnemann 花费了其职业生涯中的大部分时间来测试各种物质，记录它们导致的种种症状，并给出现了同样症状的人们开含有该物质的处方。这种行为实在是颠倒黑白，因为给予更多相同的病原体或毒素，通常都不是治病的方法。

相似法则看上去与主动免疫的过程是类似的，但两者的效果则完全不同。主动免疫法，是在健康人身上注射少量病原体，以使他的免疫系统产生相应的自身抗体，从而抵御病毒，避免日后感染。而同种疗法所说的是，往已经饱受病原体或毒素折磨的病人身上再增加少量同样的致病物质，使免疫系统不得不更努力地与疾病斗争。有害物质能促进身体健康，而不是有损健康，这一作用机制很难让人感到合理可信。当然，这一荒谬的主张也就只得到了一些关于治愈的小道传闻的支持。

无穷小法则

幸运的是，通常同种疗法也并不会使疾病恶化。这是因为有 Hahnemann 的第二条法则，“无穷小法则”的存在。根据无穷小法则，即将物质多次稀释以增加其效力的方法，同种疗法会将病原体或毒素稀释到物质消失点，再用这种稀释液进行治疗，正是这一方法保护了患者免遭相似法则可能造成的危害。无穷小法则

与目前公认的理论以及现有的大量证据是完全矛盾的。化学研究表明，实际上物质的作用遵循剂量—反应曲线，即增加某种物质的剂量，其效应也更加强烈。每一种药物的曲线陡度不同，而且一个简单的线性函数不能适用于所有的剂量水平。例如，给予极低量的有毒物质，可能不会产生可观测到的效应；但剂量超过初始阈限后，效应就会迅速增加；剂量超过第二个阈限后，所有的剂量水平就都是致命的了。在这种情况下，所形成的剂量—反应曲线是S形的，而不是一条直线。为什么同种疗法的治疗方案与“剂量增加效力增加”这一公认的定律不同，事实上还完全相反，同种疗法的医生们对此从未做出过合理的解释。而且我还得再重申一次，关于同种疗法取得疗效的机制究竟如何，为什么会与相关的科学理论与数据相矛盾，该理论也从未做过说明。

无穷小法则的谬误，其实还远不止它与药物剂量—反应曲线特征的矛盾。许多同种疗法的治疗方案中，活性成分不仅仅是被高度稀释，实际上已经消失了。Hahnemann所提倡的、至今仍被广泛使用的稀释倍数“30X”（10的30次方），最终导致的结果是溶液里甚至没有一个原有物质的分子留存。用化学术语来说，就是超过了溶液的稀释极限。“X”是指以10的倍数稀释：一滴某物质加九滴清水稀释，充分混合之后，再抽取一滴混合液。“30”表示该稀释过程连续进行了30遍。即原有物质的百分比，会从第一滴的100%，经第一轮稀释下降到10%，第二轮稀释后下降至1%，依此类推，直到最终该物质被稀释到 10^{30} 倍。

你也许想起来了，阿伏伽德罗常数（Avogadro's number）代表1摩尔的任何物质所含有的该物质的分子数，大约为 6.02×10^{23} 个。这就意味着，一摩尔物质的分子数还不够进行 10^{30} 倍的稀释。由于分子不能分解，你最终会发现通过这样的稀释过程，到最后原有的物质连一个分子都没有剩下了。

Park（1997）计算出，你要服用大约20亿粒经过 10^{30} 倍稀释的同种疗法药丸，才能获取药物的一个分子。这还不算最糟的，更高、更极端的稀释倍数也时有采用。在某些同种疗法药物的标签上，你会看到“200C”的字样。C代表100倍的稀释，而200则指重复200次，故其稀释倍数实为 10^{400} 次。想想整个宇宙也只有大约 10^{100} 个分子（Park, 1997），你就会明白这一药品说明是多么荒谬了：即使你以世界上所有存在的物质为起点开始稀释，那么早在你完成整个200C的过程之前，就已经达到物质的稀释极限了。

Hahnemann的同种疗法理论发表于1810年，之后一年阿伏伽德罗常数才被

发现。因此，Hahnemann 也许并没有意识到，他所提倡的稀释方法已超过了稀释极限。但是，现在我们确实都知道了这一点。虽然同种疗法几乎从 Hahnemann 时代后就一成不变地传承下来，但有些同种疗法产品的制造商似乎意识到了，超过稀释极限是极其愚蠢的。因此，又出现了两种解决这一与现实脱节问题的方法。不幸的是，每一种方法所引发的问題，都要比其解决的问题更多。

规避方法 1：降低稀释倍数

目前市面上出售的许多同种疗法药物，其稀释的倍数比以前的产品都要低，比如 6X。这一稀释倍数相当于将原有物质稀释到一百万分之一的浓度。但是，正如 Park (1997) 所言，药丸中的杂质都会在百万分之几的水平上，因此这一稀释过程虽然没有超过稀释极限，但仍然太过极端，对于保留下来的任何物质而言都不可能达到一个起效的水平。而且，采用这一策略与同种疗法理论也不再一致了：这样做相当于承认无穷小法则是错误的，即药物的效力不会通过稀释而增加。所以说，与其围绕稀释倍数的问题瞎弄一通，还不如干脆抛弃这一存在致命错误的伪科学。

规避方法 2：“水记忆”

为了试图挽救前面提到的相似法则及无穷小法则，同种疗法的一些拥护者又提出了另一种不可思议的想法。他们宣称，尽管稀释以后原有的成分连一个分子也没有保留下来，但是用于稀释物质的水会“记住”曾经溶解在其中的东西。这种异想天开的说法，在 1988 年引起了媒体的极大关注。当时在权威科学期刊《自然》(Nature) 上就曾登载了一篇论文，认为增加物质的稀释倍数，并不会使药物的效力降低 (Davenas 等)。这一报道使人们一度以为“水记忆”的假说已经得到了科学实验的支持。遗憾的是，该项研究所存在的严重问题，并未像它声称的研究发现那样得到媒体的广泛关注。

Davenas 等的研究所存在的严重问题有：第一，研究结果无法重复。尽管有一些研究小组非常仔细地重复了该研究的实验程序，但没有一个研究小组取得过与其相似的结果。正如我们在第 2 章中所讨论的，可重复性是科学的基石之一。如果有认真负责的实验者采用完全相同的方法却不能取得类似的结果，那么说明原有的研究结果要么是侥幸取得的，要么有意或无意间犯了某种错误。第二，

Davenas 等 1988 年的研究结果和无穷小法则并不一致，无穷小法则认为稀释程度越高效力越大，该研究结果则认为加大稀释倍数，效力没有下降，也没有上升（在剂量反应一曲线表现为该段曲线趋于平坦）。事实上迄今为止，从没有人能够通过实验证明相反的标准剂量—反应曲线。第三，请你想一想，如果水真的有“记忆”，我们的世界会变成什么样子？每天喝那些能“记住”它所流经的所有物质的水，这是多么令人恐怖的事啊。说到这儿，我第一个想起的就是下水道。我并不想让读者不快，但事实上我们日常饮用或做其他用途的水在由水厂供应之前，都必须经过过滤和一些其他的特殊处理，这是因为在那之前它会从许多令人恶心的或有毒的物质上流过。因此，如果水真的有“记忆”，解决上述问题的唯一办法，就是让水具有“选择性记忆”，即只记住我们认为有益的物质。这种推理的方式，这种没有理论或数据支持的猜测之说，只会使其自身情况变得越来越糟，甚至连貌似合理也很难做到了。

由于同种疗法存在许多极其不可信之处，Barrett 和 Herbert（1994）将这一疗法称为“极端伪科学”，Park（1997）则称其为“无药医学”。1999 年，备受内科医生推崇的药物信息来源——《药物与医疗通讯》（*The Medical Letter on Drugs and Therapeutics*），对同种疗法当时的地位进行了简要的总结：

同种疗法产品的化学成分常常无法明确。有些产品稀释程度之高，使它们根本不可能包含任何原有的有效成分。这些产品的有效性从未得到过证实。使用这些产品毫无必要。

电磁场与癌症

你是否听说过电线能致癌？微波炉、警察用的雷达测速器、手机或其他电子设备也会致癌？这些假想的关联都来自一种不符合科学逻辑的理论：电磁场与人体的相互作用会使得癌症发生的风险增高。这种观念的来源令人费解，但更让人不安的是，尽管这一观念与科学知识相悖，它仍然十分成功地占据了人们的头脑，并得以广为流传。在下面的讨论中我们会看到，电磁场致癌说并没有什么合理的机制，相反，可以证明电磁场不致癌的理由倒是不少。Robert Park（2000）讲述过一个有关电磁场的故事，主要是说一个人的阴谋理论如何滚雪球式地发

展，最终演变成了一场关于这一不可信观念的耗资巨大的辩论。我们下面的讨论也将围绕 Park 叙述的故事展开。

Paul Brodeur 是一位冷战研究记者，曾揭露过中央情报局的秘密。20 世纪 60 年代晚期，他转向关注环境问题。遗憾的是，Brodeur 没有相关科学学科的深厚背景，在他的报道中充斥着利益冲突和所谓的阴谋。1976 年，他在《纽约客》(*New Yorker*) 上撰文，将微波炉与包括癌症在内的许多健康问题联系起来，这是他所发出的数个有严重错误的警告中的第一个。正如当时的科学家们都熟知的，微波能够增加组织的动能，从而达到加热的目的，但它不可能切断化学键、改变 DNA 以导致癌症。当科学家、政府机构以及其他具有相关专业知识的个人和组织都质疑 Brodeur 的观点时，他出版了《美国的溃败》(*Zapping of America*) 一书进行回应，在书中他概述了一个大阴谋，而且认为美国国防部、电子工业界和学院派科学家们都身涉其中（有关这一阴谋的更多内容请见第 11 章。）该书一度成为畅销书，这说明即使在电磁场—癌症相关论战的早期，已有许多人乐于接受这一与合理的科学原理相冲突的、纯粹出于猜测的理论。

1979 年，在丹佛进行了一项关于儿童白血病的发生与临近电线之间关系的研究。这项未得到严格控制的研究发现了某种统计学意义上的数据关联。但究竟引起这种关联的原因是什么呢？很多东西都与电线的存在有关。电线在非常贫困的地区随处可见，一些人由此认为，从这些数据中可得出的最合理的结论是：脱离贫困是解决儿童白血病问题的好办法。此项研究发表后，很多地区都出现了传闻之类的证据，支持癌症与电线之间的联系。而电磁场导致癌症之说缺乏合理机制的事实，则遭到了人们的普遍忽视。与这一问题有关的一个基本科学知识就是，电线周围同时有电场和磁场围绕，电场的强度仅仅取决于电压，而磁场的强度则仅仅取决于流过电线的电流量。众所周知，电场是不足为虑的，因为我们的皮肤是天然的保护屏障。而对于磁场而言，磁疗等运用的静止磁体产生的是静态磁场，电线中的交流电产生的则是随时间变化的磁场。这种时变磁场对人体的各种加工过程可以产生影响，但事实上它们是十分安全的，因为电线的电流量是很小的。这正是为什么电线通常都负荷高电压，因为那样它们就不用传送太强的电流。即使是在现代社会存在如此多电线的情况下，由电线所产生的磁场强度也不过是地球磁场强度的约百分之一。

1989 年，Brodeur 再一次无视科学界的批评，在《纽约客》上撰文，创造了

由电磁场 (electromagnetic field) 这一词组中三个单词首字母缩合而成的新词“EMF”，发表并宣扬了电线与癌症有关的警告性观点。该文引发了公众大范围的恐惧情绪，人们担心电线会影响学龄儿童的健康、引发癌症、使财产贬值。吸取了他前一次遭受抨击的教训，Brodeur 对于微波（波长约 100 MHz）与电线（波长约 60 Hz）在物理特性上的差别有意不加区分。而且，在遭到了科学界批评的情况下，Brodeur 成功地将此文又扩充为一本畅销书——《死亡电流》(*Currents of Death*)。1990 年，他在《纽约客》上又撰写了系列文章，举出了许多传闻轶事之类的证据，但完全不理睬科学的理论和数据。不管自己报道中存在的明显偏见受到了多少批评，Brodeur 还是出版了第三本畅销书——《电线的弥天谎言》(*The Great Power-Line Cover-Up*)，再次促进并利用了公众对于电磁场的恐惧心理，更不用说对于美国政府与科学家所策划的所谓杀人大阴谋的恐惧了。

到 1999 年，世界各地开展了许多大型的、严格控制的研究，以验证电磁场和癌症之间的联系。这些研究结果非常清晰、一致：绝对没有任何证据支持电磁场与癌症之间的联系（参见 Park 的综述，2000）。当然 Brodeur 还是继续坚信电磁场的危险性，并从三本畅销书的出版中获利颇丰，但是已经确认的科学理论表明：在电磁场和癌症两者之间建立关联是毫无道理的。我们花费了上千万的研究经费，却没有为疾病的诊断、预防或治疗提供任何新的知识。前面我们曾说过磁疗和同种疗法可能是无效的，但是其费用往往相对较低（参见第 15 章隐性成本的讨论），并且是由那些寻求治疗的患者自己承担。而据白宫科学办公室估计，宣称电磁场导致癌症的垃圾科学可能已经耗费了 250 亿美元，这笔费用则大部分是由美国纳税人来负担的。所以说，对于某种新的观点，不充分考虑它的合理性，尤其是其假定的风险因素的相关机制问题，可能会导致灾难性的后果。

外星人劫持的故事：现代社会的误解

本章开头 Zeke 被外星人劫持的故事在很多方面存在疑点。尽管有许多人都热心地提供了类似的报告，但它们是真的吗？

污染效应

科学家把某一现象的一致性与重复性称之为信度 (reliability)，但信度是真

实性——效度（validity）——的必要条件而非充分条件。比如，只有在称了数次，卫生间里的体重秤都得到一致的（可靠的）读数后，你才能肯定该读数是有效的。但也存在这样一种可能：该秤每次总是给出一个一致（可靠的）的读数，但事实上它却总是错的（无效的）。有信度并不一定就意味着有效度，这一点与是否能将许多看上去一致的故事作为故事真实性的证据直接相关。

为了弄清这个问题，让我们先撇开那些外星人劫持故事之间的不一致，假定这些故事是基本可靠的，也就是说它们的主要情节是一致的。那么接下来要确定的问题之一就是，为什么它们是可靠的。至少存在两种可能性：第一种可能是外星人的确正在劫持大批地球人，那么现有的故事描述对于一般人来说是十分准确的，这当然是被劫持者想让我们相信的。第二种可能则是实际上根本没有劫持活动，但存在某种文化共享的固定印象或脚本，外星人劫持之说就来自于此。现有的证据实际上明显地支持了后一种可能性。

花点时间自己试一试吧。闭上眼睛，想象一个不明飞行物，直到你的脑海中形成一个清晰的图像。一旦出现了这个可描绘的图像，你就不断地用语言对它进行描述。你看到什么了？我试验过这种简单的论证方法，几乎所有参与的学生都说他们看到了一个近似圆盘状的、像飞碟的东西。为什么大家都这么说呢？实际上没有一个学生曾见过这样的飞行器，更不要说被外星人劫持过。那么，这一相同的意象又是从何而来呢？

1947年，民航飞行员 Kenneth Arnold 报告自己在夜空中看见了一群香烟状的物体，它们的运动轨迹就如同掷过水面的碟子一样。当这一描述被一名记者错误地引用为“飞碟”后，一个文化意象诞生了。飞碟一进入大众的想象，许多观察者就开始声称见过它们，而且这一意象还持续至今（Sheaffer, 1997）。所以，在缺乏任何效度的情况下，出现文化共有的意象也是可能的。

那么，如何来解释外星人劫持事件主要出现在北美这一奇怪现象呢？是外星人特别喜欢劫持美国人，还是美国的大众娱乐媒体迷上了外星生物，源源不断地为诱人的外星人劫持之说提供原料？如今的外星人原型，有着不成比例的大而秃的脑袋、大而细长的眼睛，这来源于 1975 年的电影《不明飞行物事件》（*The UFO incident*）。该电影是根据 Betty 和 Barney Hill 的劫持故事改编而成的。此后，不计其数的电影、电视节目、书籍、杂志和网站，使得几乎是同一版本的外星人常规形象得以普及。而且，这些媒体在描述外星人劫持事件时所给出的事件发生

程序也几乎是完全一致的。在美国，还有谁不“知道”外星人劫持事件是如何发生的？谁让我们生活在媒体为导向的文化里呢？即使是纯粹虚构的事件，大家也会很快形成一种共同的期待。

乏味的幻想和矛盾

关于外星人劫持事件的诸多看法既暗藏着矛盾，也暴露了人们想象力的严重匮乏。让我们思考一下对这些看法提出质疑的四个问题：

第一，为什么外星人与我们长得如此相像？想想地球上各种生命形式之间是多么惊人的不同，而我们只是上千万种生命形式中的一种而已。说到来自另一星球、独立进化而成的生物，即使是与我们有一丁点儿类似的可能性，都是微乎其微的。这种缺乏创意的人类中心主义，可以称作“星球大战谬误”：

在各种星球大战电视连续剧和电影中，有不少的外星人种类。我们只要稍微花点时间观察就会发现，它们都是在人类形象上稍加改造而成的。这也许是由于经济上的考虑，只需要一个演员和一个乳胶面具来扮演，但这一点儿不符合进化过程的随机性特点。如果真有外星人，我想它们大部分看起来都会更不像人类（且其发展水平应该是非常不均衡的）。星球大战的摄制人员看样子没有掌握好进化论。（Sagan, 1995）

第二，外星人能在我们毫无察觉的情况下穿墙而入吗？如果不能，它们怎么能避开许多家庭安装的保安系统？难道它们事先知道谁家有这样的系统，然后就不去这些人家的？同样，为什么外星人从没被保安系统的摄像头拍摄下来过？难道它们能选择性地隐身？如果这样的话，它们为什么还会让人类看到呢？

第三，外星人能将身体悬浮在空气中；在没有身体接触的情况下运送人类的躯体；还能在跨越如此遥远的星际距离的情况下，利用其生命供给系统在家乡环境以外的地方生存下来。要做到上述这些，所需要的技术远比我们目前所能达到的复杂得多得多。如果真是这样，那么为什么人们所描述的外星人就生物学特征来说显得那么笨拙和无能？为什么它们还要劫持人类，一次一个地来做“医学实验”？为什么它们不迅速地取一些人体组织的样本，进而克隆人类？这件事对它们来说应该易如反掌，连我们地球人差不多都可以做到了。

第四，为什么外星访客给被劫持者的建议是如此平淡无奇，就好像是针对我

们经受的考验和苦难的产物，反映的都是地球上现代人当前所面临的各种问题。除非 Zeke 是在洞穴中生活了十年甚至更久，否则作为现代人，他一定对环境和艾滋病扩散问题知道得不少。那么，要是外星人想让我们相信它们更聪明或者有更高的智慧，为什么它们不说些它们已证实是正确的、但我们目前由于知识所限尚未得知的东西？Carl Sagan 是一名太空科学家兼科普作家，经常被邀请去对被劫持者进行提问，询问关于他们所接触过的外星人的信息，但他坦言自己并不相信外星人劫持的故事。不过，由于这个特殊的身份，Sagan 总有机会让那些被劫持者去对外星人提出一些问题，令人称奇的是，时常还能收到答复。但当他想要具体的信息，如费尔马最后定理（著名法国数学家费尔马在 1637 年写下的一个看似简单的数学命题，但学者们长期无法解答，直到 1994 年）的简要证明时，他从未收到过回复。不过，当他问道：

诸如“我们应当做善良的人吗”这样的问题，几乎总是会得到答复。任何模棱两可的问题，尤其是涉及世俗道德判断的问题，外星人都极其乐于回答。但是，一旦涉及具体的问题，那些能让我们有机会发现它们是否真的知道人类都还未知晓的事情时，结果却只有沉默。我们可以从这种选择性回答问题的举措上推导出一些东西。（Sagan, 1995）

如果外星人真诚地希望给我们一些关于现实危险的警告，为什么它们不在环境问题出现之前或者艾滋病出现和传播之前，事先提醒我们？那时给我们这种信息会有用得更多。另外，为什么它们不干脆告诉我们如何治疗或预防艾滋病？为什么它们现在不就一些科学家们还没发现的威胁提出警告？所谓的外星生物好像只愿意、或者只能够告诉我们那些人类已经知道的事情。

睡眠麻痹

也许，对外星人劫持故事最致命的质疑是：它与人们常常忽略但却真实存在的一种现象——睡眠麻痹，有着惊人的相似。睡眠麻痹多发生于刚入睡或将醒未醒的临界状态。个体感觉自己完全清醒了，但却根本无法动弹。那是一种让人觉得很难受的体验，人们经常描述说自己的胸口上有东西压着，但同时又觉得全身轻飘飘的，好像浮在半空中，通常还会伴随有幻听或幻视，如亮光、人、外星人、魔鬼和幽灵等等，感到房间里还有陌生人存在。另外，这种体验还常常包含

强烈的性的成分。所有这些足以令我们做那些有极端情绪反应的噩梦了。

在典型的睡眠麻痹与标准的外星人劫持故事之间，显然存在着惊人的相似性（Blackmore, 1998）。据报道，外星人劫持活动总发生于人们即将熟睡或者将醒之际，或者长途驾驶途中即将陷入恍惚状态的时候。被劫持者飘浮在闪着亮光的飞碟里，外星人限制了他们的人身自由，拿他们做实验，还强迫他们发生性行为。所有这一切既令人眩晕又让人痛苦，难以解释。有些人从未听说过睡眠麻痹，但通过流行文化，他们对外星人劫持活动颇有了解，这些人就很有可能曲解了自己的体验。

事实上，对睡眠麻痹做出外星人劫持的曲解，也仅仅是自古以来对这一经历的一系列文化共有错觉中的最新版本而已。例如，这类故事在早期都有着相同的剧情，发生在将睡之际，或者将醒之时，都是用恶魔来进行诠释的。在中世纪，女魔鬼和男人交配以获取精液，然后再变成男魔使妇女受孕。人们认为这种邪恶的交媾所产生的子女都是巫师。这种故事读起来与我们今天描述的外星人劫持活动颇有类似之处，也像是睡眠麻痹的情形。种种明显的联系可作为有说服力的证据，而否定它们会带来棘手的问题。有关外星人的历史记录可以追溯到什么时候？大约只是50多年前，始于错误地引用了飞碟一词的报道。女妖、男魔如今又在哪儿呢？纵观人类历史，神秘生物总是随着当时文化偏见的变化而消失，而为了满足新的文化期待，奇特的新生物又及时地诞生了，两者总是有着惊人的巧合。

今天，当我们回顾中世纪巫术和魔力的种种说法，总是将其当作人类想象力的产物。在中世纪，腐败的政治和宗教机构兴盛于那个极度轻信的社会，他们拟造出来的种种荒唐说法让人信以为真，这似乎十分可怕。但是，外星人劫持故事作为超自然神话的现代版，其实质难道不也是我们自己的幻想、恐惧和渴望吗？究竟是外星人穿越了遥远的星际距离来劫持我们，实施医学和育种实验而不留下一丁点儿证据，还是被劫持者们就像人类千百年来一直所做的一样，用文化共有的期待把一个可怕的经历说得有血有肉，而实际上只是曲解了睡眠麻痹现象？

那些自称被外星人劫持的人所提供的报告，其实并没有揭示有关外星生物的任何信息，他们所说的主要与人类的心理有关。同样地，购买磁疗产品或同种疗法产品也暴露了人类的某些本性。我们总是容易异想天开，并且轻信那些缺乏真实性的理论。然而，只要抛开某种观点表面的东西，对其进行更深的挖掘和探索，就可以发现隐藏于其下的逻辑矛盾或负面证据。

8. 关联：建立和解释相关

《三面夏娃》(*The Three Faces of Eve*) (Thigpen & Cleckley, 1957) 是一本畅销书，后来被改编成了电影。这本书描写三种不同的人格—Eve White、Eve Black 和 Eve Jane 竞相控制同一个身体。其中，Eve White 和 Eve Black 的观点总是针锋相对，行为也截然相反。Eve Jane 是在治疗过程中最后出现的一种人格，更多的是前两种人格的综合。翻开《女巫》(*Sybil*) 这本书 (Schreiber, 1973)，开篇就是纵横交错的“家谱”和“人物特征和出生日期表”，介绍了令人吃惊的一系列人格。最初名叫 Dorsett 的女巫分裂为多种人格：作家、画家、两个木匠（均为男性）、两个狂热的宗教人物、一个青少年、两个儿童、一个婴儿，还有其他 7 种纯粹根据其主导情绪来描绘的人格，总数竟达到了 16 种！这些人格在 1923 年至 1965 年之间出现，每次出现 1~2 种。作者还对这些人格进行了清楚的外貌描写，例如，Mary L. S. Dorsett “丰满的、深棕色的长发从两边垂下”；Vanessa Gail Dorsett 有“高高的个子、身段苗条、红色的头发、浅棕色的眼睛、富于表情的椭圆形脸蛋”，而 Sid Dorsett 则有“浅色皮肤，黑色头发，蓝眼睛”。

被诊断为多重人格障碍 (MPD) 的个体，表现为同时拥有两种或两种以上明显不同的人格，轮流控制其思想和行为。据报道，MPD 个体甚至可以达到成百上千种人格。在美国精神病协会 1994 年出版的《精神障碍诊断与统计手册》(*Diagnostic and Statistical Manual of Mental Disorders*) 最新版本中，已经将多重人格障碍称为分离性身份识别障碍 (dissociative identity disorder, DID)，但本书中，仍会使用 MPD 这个更广为人知的名称。在 Eve 和 Sybil 的案例中，这些“变化

的”人格表现出惊人的不同（比如 Eve White 行为和衣着均很保守，而 Eve Black 行为高调、衣着华丽），有时简直让人不敢相信（比如 Sybil 的一种人格是有着浅棕色眼睛的女人，另一种则是蓝眼睛的男人）。是什么导致了这些极富想象力的信念和行为？一般来说，人们认为 MPD 是由于童年期创伤所致，比如身体创伤或性虐待（如 Putnam, 1989）。MPD 的创伤理论认为，由于创伤经历所产生的应激过于强烈，儿童无法承受其痛苦，便将意识“分离”或分割成不同的人格，从而承受创伤，并将关于该事件的威胁性记忆保留下来。在考察有关多重人格形成的其他理论之前，我们先来评价一下创伤理论。

如果要建立儿童创伤经历和 MPD 之间的真实联系，我们需要通过系统的调查搜集四种数据：（1）报告有儿童期创伤经历且被诊断为 MPD 的人数；（2）报告有儿童期创伤但并未被诊断为 MPD 的人数；（3）没有儿童期创伤，但被诊断为 MPD 的人数；（4）没有儿童期创伤，也未被诊断为 MPD 的人数。这四部分信息可以很容易地用一个 2 × 2 的表格来表示：

		诊断为 MPD	
		有	没有
报告有儿童期创伤经历	有	单元 1	单元 2
	没有	单元 3	单元 4

四种类型的数据缺一不可，否则无法比较。要建立童年期创伤经历和 MPD 之间的相关，有过童年期创伤且患有 MPD 的个体的比率，必须与没有童年期创伤但患有 MPD 的比率不同，这些比率可以用下面的公式计算：

- 比率 1 = 单元 1 / (单元 1 + 单元 2)
= 所有有童年期创伤经历的个体中患有 MPD 的人数比率
- 比率 2 = 单元 3 / (单元 3 + 单元 4)
= 所有无童年期创伤经历的个体中患有 MPD 的人数比率

如果这两个比率相同，说明童年期创伤和 MPD 没有关联：无论有无童年期创伤经历，被诊断为 MPD 的可能性都是相同的。如果这两个比率不相同，则说明童年期创伤经历和 MPD 可能存在某种关联，这又可分为两种情况：

- 如果比率 1 大于比率 2，说明存在正相关（有童年期创伤经历的个体增加，诊断为 MPD 的人也增多）
- 如果比率 1 小于比率 2，说明存在负相关（有童年期创伤经历的个体增

加，诊断为 MPD 的人减少)

若四个单元的数据缺少一个，这个比率就无法计算，也无从得出合理的结论。一旦你了解了方法以及相关的意义，这个比率检验是很简单的，但还是有许多人认识不到进行这一检验和通盘考虑四类数据的必要性。这些人常常做出以下两种错误的推断：

根据很少的例子提出存在相关的假设

第一个错误就是，只根据少数例子就假设有某种相关关系的存在。如果有人注意到有一些 MPD 患者曾有过童年期创伤的经历，并因此认为童年期创伤和 MPD 存在正相关，那么他错误地高估了上表中的“有一有”单元（单元 1）。举个例子来说，假设有一个医生比较熟悉 Eve 和 Sybil 的案例，他记录了有童年期创伤且诊断为 MPD 的病例，却没有注意或记录那些不符合这种期望模式的病例。而在 20 个假设的患者中，可能有 5 个符合这个期望模式（有童年期创伤经历且患有 MPD），另外，各有 5 个分属于其他三种情况（没有童年期创伤经历但患有 MPD，有童年期创伤经历但未患 MPD，没有童年期创伤经历也未患 MPD）。要确定童年期创伤经历和 MPD 是否存在相关关系，比率计算应当如下：

比率 1 = $5 / (5 + 5) = 0.50$ ；比率 2 = $5 / (5 + 5) = 0.50$

因为比率 1 与比率 2 一样，所以根据比率检验的结果可以得出结论，童年期创伤经历和 MPD 不存在相关。因此，虽然看了单元 1 中的几个案例以后，感觉似乎两者具有相关性，但证据并不充分。

只根据少数的例子就得出某种相关的推断，实际上不仅仅是变量间可能并不相关，而且，还有可能使我们得出与实际情况完全相反的推论。假设这个医生记录了 5 个有童年期创伤经历且诊断为 MPD 的病例，但忽略了在其他三个单元中各有 10 个病人。那么，两个比率值应当计算如下：

比率 1 = $5 / (5 + 10) = 0.33$ ；比率 2 = $10 / (10 + 10) = 0.50$

这一次比率 1 小于比率 2，也就是说，两个变量间存在负相关，这个结果恰好与该医生的假设相反。也就是说，尽管记录在案的少数病例既有童年期创伤经历，也患有 MPD，但实际上，相比那些没有受过创伤的个体而言，有童年期创伤经历的个体患 MPD 的可能性更小。

很显然，根据少数样本得出变量相关的结论是危险的。一些为数较少的案例

被视为代表多数人经验的典型案例，这就导致符合预期的信息被高估，而不符合预期的信息则被低估。尤其是在我们的个人经验或一些推荐、虚构的人物生活，以及媒体报道中，这种情况更为严重。许多人读过或看过《三面夏娃》的书或电影，或《女巫》的书，不免会通过少数的一些案例得出某种结论。这实在是一个很难避免的陷阱，除非你总是清楚地记着，要确定某种相关关系，必须了解四个方面的数据。单纯某方面的实例，并不能给你提供所需要的另外三类数据。科学家们总是努力通过系统方法收集数据，并运用客观的统计方法进行比较，最后才能得出结论。

根据单一的比率提出存在相关的假设

前文提及的比率检验包括了两种比率的比较。然而，人们往往只根据一个比率就得出某种相关关系存在的假设。例如，某些人可能深入研究了一些 MPD 的案例，然后发现其中有童年期创伤经历的人所占百分比很高。也就是说，上述四格表的比率 1 很高，因此人们就推断这两个变量呈正相关。其实这种证据是不充分的，原因很简单：我们并不知道其他比率如何。我们所不知道的比率有可能和已知的比率一样高（说明没有相关），或者更高（说明有负相关）。无论是哪一种情况，关于童年期创伤经历与 MPD 之间关系的结论都会有很大的不同。

实际上，对任何种类的心理障碍患者进行个人生活史的调查晤谈，通常都会发现有童年期创伤经历者在其中占有很高的比率。但 Renaud 和 Estess (1961) 发现，即使在经过严格筛选，剔除了任何有心理异常表现者后的 100 名被试中进行调查，童年期创伤经历也是普遍存在的。许多人都或多或少有童年期的创伤经历，可这种经历并不能解释为什么某一特定的个体会出现心理障碍。因此，如果要证明童年期创伤经历和某种特定的心理障碍存在相关，我们必须对心理创伤给出一致的定义，并且用研究数据表明那些有童年期创伤经历的人罹患该种疾病的比率高于没有童年期创伤经历的人。迄今为止，许多童年期创伤经历和精神病理学之间假定的相关关系都没有得到证明，童年期创伤经历和 MPD 之间的关系也是如此 (Lilienfeld 等, 1999; Spanos, 1996)。

解释相关性

我们已经讨论了如何建立真正的相关关系，接下来我们将讨论对于相关关系

的正确理解与诠释。当我们试图了解事实的真相时，通常的目标总是想了解是什么事情导致了其他事情的发生。事物之间的相互关联当然值得研究，但对我们来说，却不如因果关系那么具有理论和实际的价值。举个例子来说，假定童年期创伤经历和 MPD 的诊断呈正相关，如果这种关系是因果关系，那就意味着预防童年期创伤经历能相应地减少 MPD 的发生。但如果这种关系并非因果关系，还有一些其他因素也同样与 MPD 的诊断有关，那么由于对 MPD 发病的真正原因我们还不了解，试图通过减少童年期创伤经历来预防 MPD 的发生就会徒劳无益，在 MPD 的治疗中对创伤性记忆所投入的关注也可能是时间、金钱和精力的浪费。下面我们将简短地讨论一下这种可能性。

正由于因果关系的解释特别有力，而且富于直觉的诱惑，因此即使在信息不足的情况下，人们也会试图得出因果关系的结论。事实上，我们经常会在证据不充分的情况下作出这种跳跃性的结论。让我们再回顾一下第 6 章，人们总倾向于相信这种说法，“我试过这种治疗方法，感觉好些了，因此这种方法对我有效”。实际上，有很多种其他的解释可以取代因果关系的解释，如安慰剂效应、衍生出的好处、症状缓解而非痊愈等等，可我们往往都没有注意到。要避免问题结论的出现，我们必须记住一条：“相关关系并不意味着因果关系。”就如童年期创伤经历并不必然导致 MPD 的发生。对于单一的相关关系而言，还可能存在着许多其他潜在的因果解释。

相关事件间的因果关系

一般说来，两个相关事件（X 和 Y）之间可能存在三种类型的因果关系：

1. X 导致 Y。比如，治疗过程中提到的童年期创伤经历（X）导致 MPD（Y）。

2. Y 导致 X。也有可能是 MPD 的诊断导致了患者报告有童年期创伤经历。想想我们第一章开篇介绍的关于 MPD 治疗的可怕故事。Braun 认为每一个 MPD 患者都在童年时期遭受过性虐待。因此，当 Patricia 被怀疑患有 MPD 时，她的治疗师就假定她也曾遭受过性虐待，并对她施加压力，以挖掘和发现与治疗师的信念相符的记忆。因此，MPD 的诊断可能会导致治疗师去寻找和诱发关于性虐待的不真实记忆，并由此认为有童年期创伤经历和 MPD 呈正相关。

3. Z 导致 X 和 Y。这里的 Z 代表你能想到的任何其他因素，任何一种相关关系通常都存在很多种可能性。比如说，不是童年期的创伤经历导致了 MPD 的发

生，治疗师也没有因 MPD 的诊断去诱导病人对童年期创伤经历的回忆，而是另有一些变量可能同时导致了 MPD 的诊断和创伤性回忆的出现。这个变量会是什么呢？是不是催眠易感性、幻想倾向，或者暗示性？对所有这些变量，治疗师都有可能以一种非正式的方式进行评估，然后根据它们来判断病人是否会接受 MPD 诊断，以及愿意接受记忆提取与恢复程序。实际上，有证据支持这一可能性的存在。Hyman 和 Billings（1995）发现，凡在创造力想象量表（测量催眠易感性）或分离性体验量表（测量意识、思维和记忆的分离）上得分高的人，最容易产生错误回忆。

因此，对于任何一种相关关系而言，仅根据相关本身并不能进行明确的因果关系的解释，还需要很多其他的信息，我们才能做出正确的选择。

因果关系的三个条件

要证明是某个事件导致了另一事件的发生，必须满足以下几个条件：

1. X 和 Y 相关。
2. X 必须发生在 Y 之前。
3. 所有其他的因素（Z）都已被排除。

如果事实上已存在需要解释的相关关系，那么第一个条件已经满足，但是其他两个条件却是不易处理的。

第二个条件常常被称为“方向性问题”，因为确定因果关系的方向十分困难。在 MPD 的创伤理论中，原因（创伤）必须发生在结果（MPD）之前。由于 MPD 患者往往是在没有任何关于童年期创伤经历的证据的情况下就已被确诊了，一些治疗师就有可能出于想要解决这个方向性问题的目的，诱导病人对性虐待的回忆。然后，这些“记忆”又成了“创伤的确发生在 MPD 之前”的证据。

再举一个例子，自尊和学业成绩之间存在着明确的相关（Sykes, 1995），许多教育家由此推测存在“自尊导致学业成绩”的因果关系，并建议通过培养儿童的高自尊来提高其学业成绩。但是，这种因果关系会不会正好是相反的方向呢？也许是学生学习成绩好而导致其自我感觉更好？这种解释和前一种解释看起来同样有道理，但它意味着培养自尊并不会提高学业成绩（而且从某种程度上说，由于它占用了教学时间反而会降低学生未来的成绩）。相反，这意味着教育过程存在着自身的特殊目的，自尊的不断提高只是伴随出现的衍生获益之一。又

或者上述两种解释都有其合理之处，它们之间存在着一种“互为因果”的关系，自尊影响学业成绩，学业成绩反过来也影响自尊。仅仅基于相关性本身，根本无法断定哪种解释是对的。必须有更多的信息，才能弄清楚这个方向性问题。

第三种情况常常称为“第三变量问题”，因为必须要确定是否存在另一个外部因素，即“第三变量”导致了X和Y的同时变化。这一过程往往是非常困难的，由于外部变量为数众多，因此这种问题往往比方向性问题更为复杂。正如前文所述，可能存在某些因素（如催眠易感性、幻想倾向、受暗示性等）既促使治疗师对病人做出MPD的诊断，也促进了记忆提取与恢复程序的开始。在进行因果解释时，这些因素都无法排除。同样，或许不是自尊导致了学业成绩，也不是学业成绩导致了自尊，而是教养方式和家庭环境同时影响了两者。又或者也不是家庭环境，而是同伴影响、个性和智力中的遗传成分，以及其他许多潜在的因素同时对自尊和学业成绩产生了影响。

把相关关系解释为因果关系是非常危险的。仅凭相关关系，怎么能明确因果关系的方向呢？怎么逐一排除可能存在的第三变量呢？即使你自己认为已经排除了所有其他可能的解释，其他人是不是会考虑到你尚未想到的问题呢？

多重相关的检验模式

虽然在只确定了事件间相关性的情况下，很难证明或讨论因果关系，但情况也不总是那么悲观。结合多种相关的模型，可以对因果关系进行讨论。Hill（1965）提出了以下几个标准，来确定相关关系是否可能是因果关系：

1. 一致性。研究结果是否可以在多个测验和研究中重复，从而排除侥幸成功或偶然现象的可能？
2. 相关的程度。相关性是否具有显著的临床意义，而不是微弱的一点联系？
3. 剂量—反应关系。如果预设的原因水平更高，是否导致了预设的结果效应水平也更高，从而排除其他因素引起的巧合现象？
4. 合理性。这种关系能否用可靠的理论合理地进行解释？正如我们在第7章所讨论的，与现实世界已知的原理相悖的解释是不可能正确的。
5. 连贯性。这种解释是否和其他已知的事实相吻合？大量的事实之间的相关，比任何单一的相关关系更具有说服力。

具体如何应用上述标准，以纯粹的相关性数据建立因果关系，我们将以吸烟

是否导致癌症的问题为例加以说明。出于伦理学上的考虑，在研究吸烟是否致癌的问题时，我们不能进行概念上十分简单而又行之有效的实验设计：随机抽取一些人让其吸烟，一些人不吸烟，过一段时间再观察他们的身体状况。这样我们得到的往往是纯粹的相关性数据，难以说明因果关系，而且还要受到第三变量的限制。因为这些自主群体的健康程度差异可能是由其他与吸烟无关的因素导致的，比如喜欢冒险的癖好、与健康有关的人格因素、社会经济地位、药物滥用、酗酒以及对健康饮食和锻炼的态度等等。

关于“吸烟的人比不吸烟的人肺癌发病率更高”的说法，相关证据可以说是已经堆积如山了，很少有人会表示反对。但能否据此推论因果关系呢？让我们来看看确立因果关系的标准：许多大型研究的数据都显示了吸烟和肺癌之间的相关，因此一致性不成问题；两者相关的程度也非常高，例如吸烟的人预期寿命比不吸烟的短；剂量—反应关系也是存在的，吸烟越多的人得肺癌的风险就越大；还有合理性，众所周知，香烟里含有致癌物质，并且是被直接吸入肺部。此外，Abelson（1995）还阐释了这个合乎逻辑的理论，并总结了以下几方面的证据，说明了整个论断的连贯性。

用简单的、非技术性的形式来解释因果联系的假设机制，那就是烟草烟雾中包含了某些物质，当反复多次接触后，可能对人体组织产生毒害，接触越多，毒性越大。那么，根据这一机制究竟能做出什么经验性的推断呢？

1. 一个人吸烟的时间越长，其罹患癌症的风险越大。
2. 在一定的时间内，一个人吸烟越多，其罹患癌症的风险越大。
3. 戒了烟的人癌症发生的几率要低于一直吸烟的人。
4. 烟民的癌症一般是肺癌，且常常是某一特定的类型。
5. 烟民患其他呼吸道疾病的几率也会升高。
6. 由于烟雾并没有吸进去，抽雪茄或烟斗的人口腔癌的发病率常常很高。
7. 吸带过滤嘴的香烟的人比吸其他香烟的人，其罹患癌症的几率要低一些。

在1964年的美国外科医学期刊（*Surgeon General*）及一系列的后续报告中，上述这些推断都得到了中等至较高强度的实验支持。通过比较不同亚群体的癌症发生率，上述推断中的相关关系都得到了确定。这个案例有力地说明了吸烟与癌症之间的因果关系，因为该案例包含了有毒的烟草烟雾所可能导致的种种不良后果，并且每一种机制都得到了充分的证据支持，因而具有

极强的连贯性。

这个例子表明，尽管相关性组合中每一组成部分的说服力都不强，但是其整体形成的因果关系的证明却十分可靠。将那些单一的、不能解释所有证据的解释综合在一起，就构成了一个连贯性论点的有效性。这个观点运用了科学推理的一条重要原则——奥卡姆剃刀（Occam's razor）原则，即在其他条件相等的情况下，最简单的解释优先。因此，根据奥卡姆剃刀原则，吸烟和肺癌的关系可以解释为因果关系，因为这是最简单的解释，并与所有的数据相一致。

小心媒体的误导

在我们努力地建立正确的相关关系，并对其进行合理的解释时，最大的阻碍可能是大众媒体（电视、收音机、报纸、杂志、网络等等）。为什么呢？正如我们在第9章中将要讨论的那样，大众媒体讲述的是关于个体的故事。故事越是生动、具体，个人化情绪色彩越是强烈，就越是容易被大众所接受。吸引大众的眼球永远是传媒人士最关心的事情，至于报道的准确性则是次要的，而且这一标准往往被远远地放在了后面。因此，媒体中总是充斥着种种极具戏剧化色彩但却不准确也没有教益的专家推荐、个案研究以及奇闻轶事。媒体大卖场不断地促使大众根据很少的例子或单一的比率形成某些观念。

有一个关于寿命统计的老套伎俩就总是不断地出现在媒体中。比如有这样一个观点：管弦乐队指挥由于职业具有成就感，因此他们的预期寿命比普通人要长。的确，Atlas 统计数据（1978）表明，管弦乐队指挥的预期寿命是 73.4 岁。由于指挥们基本都是男性，因此将这个数字与 1978 年美国普通男性的寿命（68.5 岁）进行比较，就可以得出关于富有成就感的职业与健康、长寿的关系这一论断。但是，这是一个虚假无效的比较。想想看，管弦乐队的指挥代表了什么人呢？健康的成年人。而用于计算平均寿命的男性包含了各种各样的人，还有在婴儿、儿童或是青少年时期就夭折的个体，这样得到的统计数字当然会偏低一些。Abelson（1995）进行了一次更为适当的比较，比较管弦乐队指挥和活到 32 岁以上男性的预期寿命，两者分别为 73.4 岁和 72.0 岁，几乎没有差异。显然，1.4 岁寿命的差异并不足以作为人们在选择职业时的依据。事实上，如果进行某些更恰当的比较，还可以更大程度地减少这种差别。例如，一般来说白人比其他

人种更长寿，而白人在管弦乐队指挥中占大多数，如果将管弦乐队指挥与成年白人男性的预期寿命进行比较，也许差别会更小。这种比较的伎俩，可能会被用在很多地方。根据这样的比较结果，几乎任何职业都是保持长寿的明智选择。

然而，即使你幸运地从媒体报道中获知了充分的案例或足够的比率，还应当小心不要从相关的事实中得出因果关系。如前所述，最简单的原则也是最重要的原则是：**相关并不意味着因果关系**。要时刻记住我们很容易根据日常事件，主观地错误解释相关的事物。很多时候，这些都是无心的错误，因为新闻报道人可能根本不知道（或者也不在意）相关关系和因果关系的区别。

再举个例子，我们常常听到素食有益于身体健康的报道。被用来支持这一观点的证据往往是素食主义者比普通的杂食者身体更健康的一系列数据，例如寿命更长，或者心血管功能更好等等。但是，记者们有没有考虑到第三变量问题呢？素食主义者是一个自主选择的群体，他们在很多方面都与杂食者不同。最主要的是，他们更有健康意识，因而可能在与吃肉无关的很多其他方面都会特别注意照顾自己（比如更有可能不会吃得太多、不吸烟、没有药物滥用、酗酒或久坐的习惯等）。考虑到第三变量以后，情况就变得很明显了：素食主义与健康指标之间的相关根本说明不了存在因果关系。

要确定哪种饮食方式最好，就需要进行实验来证明。除了要比较的饮食方式——素食和杂食之外，两组被试在其他方面的条件应是完全相同的。随机安排参与者进入不同的实验组，从而确保被试的生活方式、态度、行为等不会存在系统性的差异。也就是说，两组间唯一不同的变量就是是否素食，排除了诸如摄入的总热量、体育锻炼等第三变量。实验者严格监控被试的饮食，以确保他们按照饮食规定进食。另外，当实验结束时，要使用“盲评”来评估被试的健康状况，即评估的医生并不清楚被试属于哪一个实验组，以避免先入为主或者已知信息导致的误差。这种研究得出的结果应当说是很准确，并且有参考价值的。但是如果缺乏相应的设计得当的实验证据，仅仅是比较那些各方面条件差异都很大、而且是自主选择的群体，我们无法得出任何可靠的结论。

此外，还存在某些专业人士故意把相关关系偷换成因果关系来误导大众的情况。一些广告人、政治策略家、某种行业的拥护者和倡导者，往往会从一些相关性数据中得出不应有的因果结论，来达到推销其产品或信念的目的。例如，某制药厂商说他们的药物能比其他厂家的同类药物更快地缓解头痛，但他不会告诉

你，他们用的是超强配方而别人用的只是一般配方。当然就更不会跟你解释，你为什么要购买他们的产品，而不能简单地加大原有药品的剂量，以获得更快的疗效。为了获得金钱、选票、信仰或者其他利益，某些别有用心的专业人士精心编织了一张张相关性之网，实在很容易让人身陷其中。

那么，最好的自我保护措施是什么呢？就是尽可能地依靠实验性数据，而不是相关数据来得出结论。如果实验设计得当，就不会有相关性研究中的方向性问题和第三变量问题。如果由于可操作性问题或伦理学问题导致严格的对照实验无法实施，那么就坚持使用相关性的连贯性模型。如果一系列的相关关系一致地指向同一个结论，那么对预先假设的观点是非常有支持作用的。如果情况不是这样，那么其中就可能存在欺骗的成分。

占星术真能预示什么吗

最后，让我们再迅速浏览一个媒体实例——占星术，借以总结这一章。许多报纸都曾登载过关于太阳星座占星术，一种以个人的太阳星座来进行预测的学说。人们可能会好奇，这种占星术的建议或预测到底有没有用。占星家们将太阳星座和人类事件之间的相关假定为因果关系的性质。天体真的会影响我们的生活吗？下面这个简单的测试方法就可以帮助你得出结论。

找一群被试（越多越好），给他们看前一天的星相预测，但是以随机的顺序排列，并且不告诉被试哪个出生星座对应哪个描述。让被试仔细考虑，思考每一种描述与自己前一天所发生事情的吻合程度。然后，让被试选择自己所属星座的描述。如果星座真能有效预测现实世界所发生的事情，那么被试正确选择自己星座描述的比率就会比较高，比如至少在75%以上。如果占星术的预测完全是无效的，那么这个正确性的比率就会接近猜测的机会率，大约是8%。最好第二天再重复一次这个实验，尤其是当被试人数很少的时候，有些人仅凭运气就能够猜对。占星术若有效，就应该经得起重复的检验。

我在许多班级做过这个测验，没有一次正确率能超过机会率的水平。因此，许多报纸在占星术一栏旁边写的“仅供娱乐”的标识看来是有道理的。除非所有无价值的信息都标注有这样的免责声明，否则我们就必须要为自己进行更深入的思考和更合理的判断。

9. 风险：知觉偏差和媒体悖论*

你认为在美国，每年因坠机残片击中而死亡的人和遭受鲨鱼袭击而死亡的人，哪种更多？死于龙卷风和死于被闪电击中的人，哪种人更多？死于糖尿病和死于谋杀的人，哪种人更多？谁更有可能出现再犯（因先前的罪行被定罪后再次犯同样的罪）：性罪犯还是非性罪犯？当对不同风险的相对危险程度进行评估时，我们很容易忽略所想到的信息样本的两种微妙偏差：第一，大众媒体呈现给我们的是一些细节逼真、精心挑选的非同寻常的事件；第二，那些逼真的、非同寻常的事件很容易被我们想起，而且对我们随后的判断还会产生不成比例的明显影响。这两种偏差加在一起，就产生了所谓的媒体悖论——我们越是依赖大众媒体提供信息，就越容易产生错误的恐惧感。

一项被广为引用的研究结果表明，大学生们将原子能列为 30 种不同的活动和技术中最危险的一种。然而，当风险评估专家们在做同样的风险排序时，他们把原子能列为第 20 名，甚至认为还没有骑自行车危险（Slovic, Fischhoff, & Lichtenstein, 1979）。罗斯（1995）在回顾了一些关于风险知觉的严重错误后，提出了下面的批判性问题：“我们熄灭了一根火柴的火焰，就背离了狂怒的地狱吗？”

* 第 9 章节选自“危险的商业：逼真性、可获得性和媒体悖论”（Risky Business: Vividness, Availability, and the Media Paradox）。作者：约翰·鲁西欧，2000，《怀疑论者》。

逼真性

我们中的大多数人都依赖大众媒体（电视、收音机、报纸、杂志和因特网等）来获取日常信息，帮助我们在所处的世界中规避风险。然而有些信息来源，并不能提供有代表性的事件样本。由于各种原因，其中包括在某种媒体渠道之内或不同媒体渠道之间争夺受众的激烈竞争的需要，媒体必须对潜在新闻进行严格的筛选，以确定其捕捉受众注意力的能力。寻常、平凡的事件没有机会上黄金档的新闻。那些经过艰难的选择、得以上黄金档的新闻故事，还得在故事具体的、个人的和情感的内容上进行精心的构思以增加故事的逼真性，这样才会引起我们注意，并且让我们有可能记住这一信息（Nisbett & Ross, 1980）。

不管是轶闻趣事，还是实验性的证据，都可以用来证明逼真性所造成的影响。请设想你要在汽车市场上选购一部新车，于是你到《消费者报告》（*Consumer Reports*）杂志上寻求建议（Nisbett, Borgida, Crandall, & Reed, 1976）。杂志上巧妙地总结了来自几百个消费者的个人经历，外加汽车专家的个人评点和道路测试结果，让你相信从价格和可靠性来说，本田思域是满足你个人交通需要的最佳选择。但是在正式购买前，你碰巧和一个朋友提到你的这一决定，你的朋友大感震惊地对你说：“本田？你一定在开玩笑！我的一个朋友买过一部，简直麻烦不断！一开始是燃油喷射系统坏掉，然后刹车坏了，最后传动装置也完蛋了！不出三年，他就被迫把这部老爷车当废品卖了！”这种警告的逼真性非常有说服力。我们中有多少人会公平、认真地权衡从几百个消费者和汽车专家的观点中得来的有利证据，和一个通过二手转述的方式得来的不利证据，然后以一种恰当、理性的方式来对待这种警告呢？

当然，我并不会试图用一个逼真的例子，来让你相信逼真性的重大影响。实验调查数据更能确定地说明逼真性信息的影响。举个例子，Borgida 和 Nisbett (1977) 让选修心理学导论课的学生评定他们对参加 10 门高级心理学课程的兴趣。为了使评定更加科学，这些学生被随机地分配到三种不同的信息条件组中。基础组的学生读到一份对课程评估的统计报告，这些报告是由上个学期参加过这门课程的所有学生完成的。面对面组的学生则与一个小组的高年级学生直接接触，这些高年级学生与他们分享自己关于这 10 门课的观点；而在分享观点之前，

高年级学生会先报告与基础组所看的统计报告中相同的评估数据。还有一组是对照组，对照组的学生没有收到有关这 10 门课程的任何信息。

实验结果表明，和对照组相比，读了统计报告的学生（基础组），对所推荐的 10 门课表现出的兴趣呈轻度的增加。值得注意的是，直接与高年级讨论的学生（面对面组），对被推荐的课程表现出了十分明显的兴趣。也就是说，比起来自大量已选课学生实际反映的统计归纳数字，面对面的信息交流对于学生的选择有着更显著的影响。而且对于最近决定主修心理学的学生和那些未主修心理学的学生来说，对前者的影响更大。因此，信息的个人相关性可以放大逼真性的力量。

大众媒体善于在多种渠道上利用这种力量。例如，对于一个涉及全州几千个家庭的福利改革的问题，为什么 90% 的报道都要包含对某一个受影响者的个人访谈呢？为什么用“超越统计，走上街头”的逻辑来调查该问题，对观众、听众和读者来说更具有说服力呢？那是因为制片人很清楚，科学的分析与某人的个人经历（当然是经过精心选择的）相比，在情感说服力上是远远不够的。为什么在报道当天早些时候的一个重要判决时，电视新闻记者要站在法院门口呢？为什么气象通讯员在镜头前报道一场正在逼近的暴风雪时，非得站在户外忍受着严寒、冰雹和呼啸的大风呢？原因很简单：即使是非常表面的背景要素，都会使一个故事看起来更加真实、具体。

可获得性

如前所述，通过大众媒体渠道，我们总是被有偏差的、逼真的信息样本所包围。然而，这究竟对我们随后的决策会产生什么样的影响呢？心理学家已经发现，我们所能想象、回忆一个事例的难易程度，对有关频率和可能性的判断会有很大的影响。请你考虑一下某研究文献提到的两个问题：

1. 如果要从一个 10 人的小组中选一些人组成委员会，那么组成 2 人委员会的不同组合方式，是不是比组成 5 人委员会的方式更多？
2. 如果你随机地从一篇课文里选取一个英文单词，那么字母“K”出现在单词首位的几率大，还是出现在单词第三个字母处的几率大？

Tversky 和 Kahneman (1973) 发现，他们这项研究的参加者在回答上面两个

问题时，是以一种系统的、可预期的方式在犯错误。在第一个问题中，也许在人们的想象中，应该是更容易组成较小的委员会。因为较小的委员会看起来互相之间更容易不同，而较大的委员看起来会在很大程度上有重叠。实际上，组成2人委员会只有45种不同的组合方法，而组成5人委员会却有252种不同的方法。在第二个问题中，人们更容易想起以字母“K”开头的单词来。但大量的单词计数研究发现，字母“K”出现在单词第三个字母处的几率，几乎是以“K”为首字母的两倍。事实上，很少有人会注意到，在这一章的英文原文里，字母“K”出现在单词的第三个位置上的单词就有12个，而以“K”为首字母的单词只有一个，而且还只是Kahneman这个固定人名，可能还不能计算在内。

上面的研究表明，我们的判断实际上会被我们所能想象、所回忆信息的难易程度，或者说信息的可获得性所左右。在日常生活中，使一件事能比其他事情更容易自记忆中获得的原因通常是什么呢？一个至关重要的因素就是逼真性。当我们在拼命地搜索记忆，以便进行一个有关频率或可能性的判断时，最容易获得的往往是那些具体的、个人的、情感的事例。例如，选课实验中的大学生，更倾向于记住那些高年级学生小组成员表达的观点，而不是枯燥的统计学结论。同样的道理，新闻报道也会通过某个个体的亲身经历来给观众留下更持久的印象，而不是在“纯统计”的基础上提供科学的论据。

由于我们的判断会受到回忆某具体事例的难易程度的影响，而不是对所有逻辑上可能发生的事件的实际频率进行仔细评估，因此会很简单地因为某种记忆的存在或其他记忆的缺失，而阻碍人们进行详尽、理性的评估。我们很少会注意到不发生，例如犯罪或意外事故的不发生（Fazio, Sheman, & Herr, 1982; Nisbett & Ross, 1980），大众媒体就更不会报道事件的不发生了。当然，当不发生本身代表现实状况的戏剧性转变时，那又另当别论了。例如，一个大城市自杀率突然出现显著下降，这就有了新闻报道价值。Bach（1973）曾经描写过一对年轻人在其第一次飞机旅行时所感到的恐惧：

风，发动机的轰鸣，大地的震颤，在我们的脚下变得越来越小的东西，我观察着来自威斯康星的老友和他的女友的变化。尽管他们还在笑，但他们已经对飞机感到恐惧了。他们对飞行仅有的一点知识都来自于报纸的头条，那些对撞机、坠毁和死亡的描述。他们从来没有读到过一条这样的新闻：一架小小的飞机起飞了，飞过天空，安全地降落在陆地上。他们只能相信，尽

管所有的报纸报道都是有关事故的，但那只是一种可能性，并且相信他们正在拿自己的三美金和性命在下赌注。

媒体悖论

Bach 对那对年轻人恐惧飞行的描写，突出地表明了逼真性和可获得性的结合对于人们判断的影响，而这种影响最终会导致媒体悖论。我们可能都曾经听过或者读过这样的观点：按照每公里旅程计算，飞行其实比开车更安全。但是，媒体对这一事实置之不理，他们对飞机空难的报道将我们引到了一个错误的方向上：乘坐飞机比驾车更加危险。事实上，有关死亡事件的发生频率判断的一系列经典研究，均强调了媒体的新闻报道对于风险知觉所造成的强大影响。例如，让大学生和女性选民同盟的成员对每年在美国发生的 41 种死亡原因的频率进行估计和判断（Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978）。研究者分析他们的判断结果时，发现了两种系统性的偏差。第一，对最不常见的死因的频率估计过高，而对最常见的死因的频率估计过低。而且在绝大多数人的判断中，判断的偏高程度超过了 10%。第二，根据被试判断的结果和 41 种死因的实际发生频率之间的总体关系，可得到一条最佳拟合曲线，而据此人们发现那些受到更多媒体报道的死因，存在着被高估的倾向。例如，相对于最佳拟合曲线来说，摩托车事故和自杀的致死率存在着被高估的情况，而天花接种和糖尿病的致死率则被低估了。

在得到风险知觉和媒体之间存在联系的证据之后，Combs 和 Slovic 又仔细地审查了当地两份报纸实际所报道的死亡事件。他们在一年中每隔一个月进行一次统计，查找 41 种死因中每一种的相关报道，记录了所有提到死因的，报道出了事故的以及报道了死亡事件的文章的数量。这两份报纸对死亡事件的报道相当一致。有趣的是，尽管两份报纸的读者群大不一样，但这两份报纸都认为某几种死亡原因比其他死因更具有新闻价值。这种现象并不是由实际事件发生的频率造成的，如自杀身亡的报道率是疾病所致死亡报道率的 3 倍，但因病死亡的人数实际上是自杀人数的 100 倍。更为重要的是，在前述研究中大学生等参与者所判断的各种死亡事件的频率，和报纸报道的因各种原因致死的频率之间存在惊人的一致（这样的高相关，并不是说人们判断的死亡频率和报道的死亡频率，两者与实际

的死亡频率之间存在着共同的联系。事实上，当实际死亡频率的统计数字保持不变时，人们判断的死亡频率与两份报纸所报道的死亡频率的相关度还更高。)

上面的证据看起来令人十分感兴趣，但我们必须谨慎，不要单单根据这种相关的证据，就得出结论说，媒体应该为我们在风险方面的知觉偏差负责任。也许实际情况正好相反：媒体工作人员只是对我们实际存在的恐惧和兴趣简单地进行反应，报道了他们认为对我们来说有新闻价值的东西。或许还有第三变量在起作用，例如不同死因在可控程度上的差异，或者不同死因在灾难性上的差异，都可能导致某些特定的死因同时会引起我们的恐惧和媒体的关注。对媒体报道和我们信念之间可观察到的相关所给出的上述解释，以及其他各种解释，实际上都可以找到例证。由于相关所涉及的因素非常多，其间无疑存在着复杂的交互作用。

不过幸运的是，我们还可以运用相关关系以外的证据，来评估因素之间的关联。一些进行了严格控制的实验研究其结果表明，逼真的信息与我们的判断之间是一种因果关系。因此尽管还存在其他的影响因素，逼真性的确是作为一种原因影响到我们判断的结果。而且，可获得性与我们的判断也存在因果关系，这一点提示正是通过可获得性的中介机制，逼真性才得以发挥它的影响。

上面的这些证据，让我们很难否认媒体对我们的恐惧所产生的因果影响。讨论这种影响的强度，或者考察这种影响与我们的反应之间是否存在循环的相互作用，是绝对有必要的。但我们也不得不承认：在媒体报道中任何与客观现实的系统性背离，都可能是我们自身信念的映照。

重新分析坠机残片

我除了在早先一部电影《楚门的世界》(*The Truman Show*)中曾看到坠机残片的描述，再未听说过其他关于这方面的事故，无论是科幻的或是真实的。我的学生曾经告诉我，他们听过一些关于人们找到坠毁飞机的残片，但并未造成实际伤亡的事件。而另一方面，鲨鱼袭击这一死因却很容易让人想象得到，并且也被媒体广泛报道过。在我所能想到的第一部有关的电影——史蒂文·斯皮尔博格(Steven Spielberg)的《大白鲨》中，鲨鱼的确造成了几个人的死亡。所以当了解到在美国一年中因坠机残片致死的人数，是被鲨鱼袭击致死的人数的30倍

时，不得不感到有些惊讶（“Death Odds”，1990）。

说到这里，在本章开始部分所提出的问题，答案已经变得清晰起来。对于每一对死亡原因来说，媒体总会倾向于报道其中某个死因的频率更高，而报道另一个死因的频率更低。普劳斯（1993）对本章开头的前三个问题所给出的正确答案为：是坠毁飞机残片（而不是鲨鱼袭击），闪电（而不是龙卷风），糖尿病（而不是谋杀）造成更多美国人的死亡。另外，尽管刚从监狱释放的性罪犯如果再次骚扰儿童或犯下强奸罪，会被广泛宣传并被媒体详尽报道，但整体来说，性罪犯实际上比非性罪犯更不可能再犯（Beck & Shipley, 1989）。

因此可以简单地说，媒体悖论是这样形成的：事件必须存在某些特别之处才会被认为有值得媒体曝光的价值，但恰恰是这种曝光使得我们对它们的发生频率产生了过高的估计；然后我们就会相信这些原本比较稀奇的事件是常见的，还相应地对这些不太可能出现的危险采取更多的防御措施，却忽略了更为明显和重大的风险。很多时候，我们的注意力会被关于某些特殊风险的警告所吸引，而实际上这些风险对我们的健康和安全很少造成严重的威胁。例如，Gilovich（1991）曾对广为媒体关注的、有关异性之间传播 HIV 病毒的机会进行过讨论。他引用了美国著名脱口秀主持人 Oprah 说过的话，“研究显示，五分之一的异性恋可能在三年之后，死于艾滋病，我是说，到 1990 年，五个人中就有一个。相信我，艾滋病已经不再仅仅是同性恋的专属”。很明显，现在已经证实这一估计过于夸张。虽然通过异性性交传播艾滋病的确算是一个严重的公共健康问题，但首先将危险的程度考虑在内，仍然是非常重要的。

激进的推测

让我们再回到本章开头罗斯（1995）的问题。有时候我们的确可能会在熄灭了一根火柴的火焰时，忽视了一个狂怒的地狱。但我还想提出一些更为激进的建议。我认为留心那些警示性的忠告，反而会让我们对其实相对安全的区域提高警惕。也就是说，一些媒体报道的本意是保护我们远离危险，但经过我们的“深思熟虑”之后，实际上却可能起到了完全相反的预警作用。

当我还住在波士顿郊区时，当地有家快餐连锁店因为违反了卫生法规而临时关闭了，原因是几个顾客吃了沙拉吧里变质的食物生了病。当地媒体当然在这方

面大做文章。当这家餐馆重新开门时，至少有一家当地报纸发表了跟踪报道，提醒当地居民去那家餐馆就餐的风险。而我可以大胆地推测，在当时所有为顾客提供餐饮的餐馆中，这家餐馆其实是最卫生的地方之一，至少在短期内会如此。自从上次事件所招致的灾难之后，餐馆的经理肯定会严防出现第二次事故。经理知道，无论是顾客也好，卫生督察员也好，都会对食品卫生非常在意。这一点肯定可以提高他们对卫生问题的关注程度。

如果你觉得上面那个快餐店的例子很平常，那么就再想想下面这个更复杂、影响更深远的例子吧。1979年3月28日，宾夕法尼亚州米德尔顿的三里岛上的核电站发生了一次意外事故。在那次悲剧的20周年纪念日到来之前，宾夕法尼亚中南部有大量的媒体报道了此事，而那些报道无疑会引发人们对过去的回忆和恐惧。就在那个时候，在我工作的大学里有一个空缺教席的候选人，就因为其配偶拒绝住在三里岛附近而拒绝了该职位。但是，当时不仅风险评估专家认为三里岛核电站是非常安全的，我也认为在全世界所有的核电站当中，那里的核电站是最安全的，住在那里的任何人都无法忍受再来一次事故。而且在事故发生后的二十年里，三里岛核电站的所有操作都特别小心。当然，他们一定还会继续这样做，已经得到广泛关注的跟踪报道也迫使他们必须这样做。

媒体从业人员往往有一种强烈的倾向，就是把个体的不幸事件戏剧化，而不管这些事件是纯属意外还是恶意所为。当一个事故引起大众的关注以后，就很容易忽略这样一个事实：所谓的作恶者本身也会对这一问题非常敏感，从而提高警惕不要再犯。因此，我们在准备根据事情的表面价值接受媒体的警告时，必须持非常谨慎的态度。更加谨慎的怀疑态度，甚至有可能会指引我们采取与媒体所讲的逼真故事方向正好相反的行动。

对媒体报道的谨慎关注

一个聪明又好奇的人，他完全依赖于大众媒体来获取信息……他不加批判地接受某个传说或谎言的可能性，几乎是冷静、均衡地对其进行评估的可能性的成百上千倍（Sagan, 1995）。

很明显，要在信息的媒体迷宫中导航，必须持有一种健康的怀疑态度。我们可以经常问问自己下面这些简单的问题，“为什么要采用这样的程序宣传这件

事？”或者“这些问题或做法被描述的频率有多高？”通过这些问题，我们可以发现媒体所报道信息的无力、缺陷，甚至是矛盾之处。

Gilovich (1991) 还列出了几条其他的策略，以使我们避免不必要的恐慌。第一，考虑信息的来源。这个信息来源的声誉怎么样？这个信息是基于科学研究的证据吗？是不是存在明显的利用你的动机？譬如这故事和广告收入有直接的联系吗？第二，不要相信预测。历史中充斥着错误的预测。恐慌制造者往往等着你作出不明智的购买决定，然后从中获益，就像在 Y2K 计算机病毒传播期间某些无耻的恐慌散布者所表现的那样。第三，如前所述，要对个人推荐或个案研究持谨慎态度。每种新的风靡一时的伪科学都会有那么一群虔诚的拥护者，他们发自内心地坚信该观念所称的惊人的好处是真的。记住，就算是老实、善良的人，也可能在毫无价值或者有害的治疗方法上犯致命的错误，而且人们犯错的方式多得数不胜数。

下面的例子可以证明上述的一些原则。20 世纪 80 年代早期，对世界上所有的精神病学文献进行检索，你会发现被报道的多重人格障碍患者仅仅有 200 例 (Bliss, 1980)。几年以后，Coons (1986) 估计，仅仅在北美就有 6000 人被诊断为多重人格障碍。之后 Ross (1991) 又推测说，5% ~ 10% 的成年人都患有此病，也就是说，全世界大约有 1 亿人患有多重人格障碍。那么，我们怎么解释多重人格障碍发病几率的爆炸式增长呢？我们可能可以追溯到书和电影《女巫》、《三面夏娃》，正因为这些书和电影，多重人格障碍开始广为人们熟知。尽管影片中的人物因其不典型与独特之处显得极端戏剧化，但很多人就是通过这些故事，了解到多重人格障碍和它的表面病因。这些不寻常个体的戏剧化的故事已广为人知，证明了媒体对传播逼真的个人推荐意见的热情，以及受众对这些不确定的观点不加批判地全盘接受后，所导致的知觉偏差、古怪的信念和令人质疑的行为。

最后，就让我以在电子邮件中广泛传播的一个故事，来结束这一章的内容：

4 月 26 日，鹰岩中学的一名学生，在泛爱达荷瀑布地区科学展 (the Greater Idaho Falls Science Fair) 中获得一等奖。他试图向人们展示，我们已经被那些垃圾科学的制造者和在我们周围散布恐怖谣言的大惊小怪的家伙们，弄得有些条件反射化了。在他的计划中，他敦促人们签署一个请愿书，要求对一种“二氢一氧化物”的化学物质实行严格的控制，或者完全禁用。

下面是许多支持的证据：

1. 这种物质可以引起过度的出汗和呕吐。
2. 这种物质是酸雨的一种主要成分。
3. 这种物质在气态的情况下可以引起严重的灼伤。
4. 意外吸入这种物质会引起死亡。
5. 这种物质会引起地表侵蚀。
6. 这种物质会降低汽车刹车的功效。
7. 在临终的癌症患者身上已经找到了这种物质。

他问了 50 个人，问他们是不是支持禁用这种化学物质。43 人回答“是”，6 人未作决定，只有 1 人知道，这种物质就是水。而该学生获奖项目的题目是：“我们有多么容易被欺骗？”在他看来，结论是不言而喻的。

10. 信念：验证性意见、事后解释和过度自信*

假设你刚完成了一个心理测试，测试结果是得到一个人格剖面图及解释，大意如下：

你有很强烈的寻求他人喜欢、赞美的需求。你有时外向，亲切，善于交际；有时内向，警觉，有所保留。你还有大量精力尚未转化成你的优势。对于你的某些人格弱点，你通常会从其他方面进行补偿。一定程度上，你喜欢变化与多样性，当你陷入束缚或限制之中时，你会感到难过。你为你独立的思考而感到骄傲，不会接受没有合理证据的其他观点。你还有自我挑剔的倾向。另外，你的一些渴望是非常不切实际的。

这段话对于你这个独特的个体来说有多相符？如果你重新读一遍这段话，你会发现其中有些描述的确能和你的某些人格特点精确吻合。那么，这一人格剖面图到底有多准确呢？运用上面这段描述所做的研究发现，绝大部分人都说这段描述非常准确（McKelvie, 1990）。

但这又怎么可能呢？每个人的性格是不一样的，而同一个剖面图不可能适合每一个独特的个体。Forer（1949）写道，这个剖面图被特意写成有通用性，可适用于每一个人，但大多数人仍会把它看成是对自己独特个性的恰当描述。根据马戏团组织者 P. T. Barnum 曾经许诺的“为每个人提供一些同样的东西”这句

* 第10章节选自“事后解释的风险”（Perils of Post-Hockery），作者：约翰·鲁西欧，1998，《怀疑论者》。

话，Meehl（1956）创造出了巴能效应（*Barnum effect*）这个词，来指称非特异性的人格描述。人们通常不仅相信通用性的人格描述会适合自己，而且还相信上述那个虚假的人格剖面图解释比真正有效的心理学测验结果更准确（Dickson & Kelly, 1985; Standing & Keays, 1986）。更令人不安的是，很多心理学家（以及他们的病人）也开始相信，对一系列墨迹的反应（即洛夏墨迹测验，见本章后面内容）具有诊断性的价值。那些效度较低的测验，还有那些完全是通用性的个性素描，为什么会有如此的说服力呢？

在得不到科学和逻辑证据支持的情况下，人们仍会顽固地坚持某些信念。不幸的是，所有具有活跃的想象力，总是持续地接触到模棱两可的数据，并且自愿放弃批判性思维去“发现”和“证实”变量之间关系的人，都比较容易犯这样的错误，甚至在根本没有某种特定关系存在的情况下，也会出现这种错误（Gergen, Hepburn & Fisher, 1986）。问题的关键是，人们总是注重对所谓的现象进行事后的解释（发生在事实之后），也就是有选择性地考虑现存的证据。这种思维方式称为事后解释（post-hockery）。

发现 VS 证实：预测的重要性

事后解释的方式，其实是利用了人类判断过程中有相互联系的几个长处和缺点，比如我们常常弄不清产生假设和验证假设之间的关键区别。在探索发现领域，我们出色的模式识别和推理能力是必不可少的。我们可以检测出潜在的相关信息，并对潜在的因果关系形成复杂的假设。然而，我们在核实验证领域却很少运用同样的独立判断过程。我们一般都不会把自己持有的信念诉诸严密的检验，而是把所能得到的第一个解释当作事实。特别是当这种解释很有趣、很有吸引力时，我们的信心就更加不可动摇了。

事后解释必定具备一个特征，即它不能对未来发生的事件进行预测。尽管成功的预测并不一定意味着深刻的理解，但不能预测则意味着解释的不当和无力。如果个体真正理解了事件之间的关系，那么就可以利用已有的知识进行可靠的预测，但事后解释者却只能解释过去的事件。下面有两个例子，可以说明预测作用在检验对知识是否真正理解这一点中的重要性。

圣经密码

1997 年，Michael Drosnin 出版了《圣经密码》（*The Bible Code*）这本极富争议的书。在书中他宣称发现了隐藏在《圣经》中的上帝的信息。之所以说是隐藏的信息，是因为这些信息并不是以连续的单词出现在《圣经》中的。Drosnin 又是如何知道这些隐藏的信息呢？他编了一个计算机程序，并用这个程序执行了无数次、无数种的文本搜索。结果他发现，如果从文本中一个特定的“n”开始，然后每隔 4772 个字母开始计数，接下来的三个字母将会是“a”、“z”和“i”，这样你就在隐藏的圣经密码中拼出了“Nazi”这个单词*。Drosnin 发现的并不止于此。他还发现如果给每 4772 个字母标上坐标，不仅会找到“Nazi”，还可以找到单词“Hitler”——如果你纵向搜索字母，并和“Nazi”共享字母“i”的话。除此之外，Drosnin 还展示了许多对这样的单词，包括“Roswell”和“UFO”**。Drosnin 解释说，这些不太可能只是巧合，因此在《圣经》中肯定隐藏着上帝的信息。他还进一步宣称这些隐藏的“信息”只能在《圣经》中找到，在其他的文字作品中就不可能找得到。

这个所谓的发现便是事后解释的一个典型例子。Thomas（1997，1998）极好地证明了发现这些“信息”的计算方法并不合理。Drosnin 忽略了这样一个事实，他的计算机程序是以无数种不同的方式来搜索《圣经》的。事实上，如果你能进行足够多这样的搜索，即使纯粹出于偶然，也一定会让你发现一些东西。Drosnin 在计算他所发现的特定配对单词的概率时，就像是他只做了这些搜索一样，完全忽略了还有数不胜数的更多次的搜索并没有任何发现这样一个事实。问题在于，任何事情在发生以后，都会被认为是极不可能的。例如，2000 年 6 月，数学家 John Allen Paulos 在 ABC 新闻频道的一个“谁在计数？”节目中指出，“如果发一次桥牌，任何一手桥牌的组合都有六千亿分之一一的可能性。然而如果认为发给某人的 13 张牌是一组奇迹组合，那就不合理了”。假如你要在某地寻找一个特定的数字，然后就开始数你遇到的所有东西，作为支持你预感的证据，那么你肯定会在并没有什么特别的地方，发现某种特殊的“意义”。

* Nazi 指纳粹党，纳粹分子。——译者注

** 1947 年 7 月 4 日晚，美国新墨西哥州罗斯韦尔（Roswell）上空突然出现发光物体，然后一不明物体坠毁在罗斯韦尔北部，这就是著名的 Roswell UFO 事件。——译者注

对于 Drosnin 所宣称的“隐藏的信息只能在圣经中找到”的发现，后来已经证明，那些所谓的隐藏信息可以以统计上可预测的规律，在很多其他的文字作品中找到，其中就包括 Scopes 的《猴子审判》（*Monkey trial*）手稿（1925 年法庭曾就 Scopes 在公立学校讲授进化论这一事件进行过辩论，而这份手稿的确不像是上帝隐藏信息的地方）。事实上，Thomas 已经证明，如果把不同字母在文本中的出现频率也考虑在内的话，使用 Drosnin 的搜索方法，我们可以相当精确地预测一些单词或词组被找到的概率。而这些单词或词组在任何文字作品中都可以找得到，包括宗教的、文学的和其他一些通俗作品。因此，Drosnin 在《圣经密码》中所谓的发现“隐藏信息”，并不局限于《圣经》，它们其实是巧合的、普遍存在的现象。我可以假设，如果你在美国税务代码中，把段落的编码和数字去掉，Drosnin 的搜索程序也会在其中找到“隐藏的信息”，或任何事先选定的其他信息。

在这里，我们还需要讨论一下关于《圣经密码》的一个更加微妙的问题。Drosnin 宣称，重要的信息是以配对的方式显现出来的，如 Hitler 和 Nazi，Roswell 和 UFO。那么，我要在这里向那些 Drosnin 的支持者们提出一个挑战：进行预测。《圣经密码》的支持者宣称它可以预测未来，那么就让他们预测一下给我们看。不幸的是，Drosnin 的程序只能是先选定一个单词或词组之后，再对其进行搜索。因此，原则上这个程序不能预测任何尚未发生的事情，它只有在事件发生以后才能找到其“意义”，这一事实表明：不能预测任何东西的信息肯定没有隐藏着什么信息。

篮球比赛中的“顺手现象”

关于事后解释的第二个例子是被人们广泛认同的篮球运动员连续投篮得分的现象。如果在整场电视转播的篮球比赛中，没有一次听说有某某人出现顺手的情况，你一定会觉得很纳闷。理解“顺手”有两种完全不同的方法。“顺手”可以简单地指一个运动员的连续投篮得分，恐怕没人会反对用“顺手”这一描述性的说法。然而，对于那些参加比赛的运动员来说，“顺手”一词有着不同的意义。

运动员对于“顺手”这个词的使用和理解含有“可预测”的意味。91% 的球迷、教练和运动员都报告说，“一个运动员在刚刚投篮命中了两次或三次之后，就更有可能会再次命中；而在刚刚投篮不中两三次之后，更有可能不中”。84% 的

人则报告说，“把球传给那个已经连续多次（两次，三次，或四次）投篮得分的人，是很重要的”（Gilovich, Vallone, & Tversky, 1985）。这些话含有两个可检验的假设：（1）一系列命中（或不命中）是经常发生的；（2）“成功滋生成功”。为了检验这些假设，Gilovich, Vallone 和 Tversky 分析了费城 76 人队整个赛季所有的中距投篮记录，以及波士顿凯尔特人队的所有罚球记录，还有关于“顺手”这一现象的另外一些实验研究。研究的结果十分清楚一致：

1. 和随机的机会率相比，有没有更多次数，或更长时间的连续投篮命中（不中）呢？

答案：没有。

研究者在分析连续命中（不中）的次数时，发现观察到的次数和纯粹机会率的命中（不中）次数非常吻合。

2. 与刚刚失球相比，运动员更有可能在刚刚得分后再次得分吗？

答案：不是。

统计在刚刚得一球、两球、三球，刚刚失一球、两球、三球后的命中概率发现，在失三球后的命中概率是 56%，失两球后为 53%，失一球后为 54%，得一球后为 51%，得两球后为 50%，得三球后为 46%。上述统计结果并不支持问题中的假设。如果说的确存在某种趋势的话，那么运动员在刚刚投篮不中后，比刚刚投篮命中后，稍微地更有可能命中。

3. 当次投篮的结果是否和前一次投篮的结果呈正相关呢？

答案：不是。

和第二个问题所提到的趋势一致，在两次投篮之间有微弱的负相关。

4. 运动员是不是有时出乎意料地顺手，有时又不顺手呢？

答案：不是。

分析表明，运动员在不同赛次中表现出的变化和统计学上正常的浮动相吻合。

防守策略的变化可能对“顺手”假设起着相反的作用——对方队员对那些刚刚投篮命中的运动员会严加防守，而对投篮不中的人则可能疏于防守。因此，研究者们也考察了与防守因素无关的情景。例如：

5. 运动员在第一次罚球成功后再次罚球，是否比第一次罚球失败后更

容易罚进呢？

答案：不是。

实际上两者的相对频率是一样的。

最后，研究者们还考察了自信心对运动员在球场上的表现的作用。

6. 运动员能够预感到他们什么时候会出现“顺手”吗？

答案：不能。

康奈尔大学的运动代表队和中学运动代表队队员站在一个距篮筐同等距离的位置进行投篮。在投篮之前，他们对每一次投篮做预测（表示信心指数）。就像预期的那样，他们似乎相信“顺手”会起作用：他们在投篮命中之后信心会增加，在不中之后信心会降低。但是，他们的预测结果和实际投篮的命中率毫无关系。因此，运动员所感觉到的“顺手”与否（用他们的信心指数来测量），并不能预测任何东西。

除了来自 Gilovich 等人的证据以外，Adams（1992）也做了一个研究，以检验对“时间依赖”的假设。这一假设是这样的：成功后跟随成功和失败后跟随成功，二者相比，前者在时间上是更加接近的（这也是为什么教练认为通过喊“暂停”，可以打断对方气势的原因）。

7. 有没有“气势”这样一种效应，即成功后跟随成功和失败后跟随成功，二者相比，前者在时间上更加接近呢？

答案：没有。

两次连续投篮命中的时间间隔，和一次命中、一次不中之间的时间间隔是一样的。而且，无论这种比较是在所有的 83 名队员之间进行，还是在每个队员的不同投篮之间进行，得到的结果都没有显著差异。

因此，对“顺手”信念的支持者来说，很不幸的是，“顺手”并没有任何预测价值。上述的 7 条证据显示，“顺手”标志着过去的成功，但对未来的表现没有预测作用。而且迄今为止，也并没有任何人提出与此相反的证据。这不是说，投篮的结果只是由偶然所决定的。事实上，有很多的因素可以导致成功，例如球员的基本水平，投篮的难度，防守的力度等等……但是研究结果揭示了一个非常简单的事实：如果你列出所有能够决定一次投篮结果的因素清单，那么前一次投篮的结果一定不在其中。尽管很多球员、球迷、教练都对“顺手”的假设坚信

不移，但这只是一个错觉。

那么，怎么会有那么多的人都如此坚信这样一个共同的错觉呢？因为我们都是组织模式的高手，以至于有时会把规律强加在本来的无序当中。在球员几次连续投篮命中之后，我们总觉得会有些东西可以总结。球迷、教练和球员们都觉得有必要为此创造一种因果解释，而不是简单地接受这样一个平凡的统计现实：因为随机概率的问题，连胜（或连败）的现象比我们能想象到的次数更多。对模棱两可事件的选择性解释也会支持这样的信念。例如，一个出现“顺手”的球员投了一个球，这个球在篮框周围转了一圈，但最后还是掉到了篮筐的外面，这样的球会被认为是接近成功的；而同样的一个球，被一个尚未“顺手”的球员投出，会被认为是接近失败。这一分析还提示我们，不仅是在投篮行为中存在着观察者对连胜或连败的知觉偏差，在其他运动项目中，人们对运动员表现的评价也同样存在偏差。

尽管“顺手”假设使我们的生活看起来更有趣，但如果继续坚持这个已被研究证明是错误的信念，则有可能会犯愚蠢的决策错误（例如，把防守集中于出现“顺手”的球员，浪费暂停的机会来打断对方并不存在的气势等）。在任何运动项目中，运动员或教练都会形成自己的信念，并依此调整他们的策略。如果他们的信念是基于真正可以预测成功的因素，那么和那些坚持错误信念的人相比，他们就更有可能会赢。

上述这些例子说明，发现错误的关联是很容易的。如果我们对于发现领域和证实领域之间的区分很清楚，也就不会犯这么多的错误了。然而，有关人类判断的研究表明，我们检验自身信念的决心是多么的微弱，即使我们真的去做了检验也是如此；而与此相反的是，我们总是倾向于寻找微弱无力的、经验性的证据来支持我们的假设，而不是追求更加合理的、有说服力的证据。而且，我们的许多思维捷径，也会给我们的这类信念造成一种正确的假象。

验证性意见

我们总是对支持我们观点的信息（而不是反对我们观点的信息）进行选择性的搜索，以免放弃自身已经持有的信念。只要有足够多的模糊信息，我们总可以找到支持我们观点的证据。例如在本章开篇部分提出的问题：怎么会有那么多人相信一个可通用的人格描述就是对他自身人格的个性化描述呢？这其中的诀窍

就是使用非常模糊的描述，而这些描述几乎可以应用于任何人。这个描述确实捕捉了几乎每个人的个性特征，但这同时也是它最大的缺点：它是通用的，而不是个性化的。假如我们深挖自己的记忆，那么每个人都可以找到一些事情来支持这样的描述。但是我们恰恰忽略了，这个个性素描与任何随机选定的个体相吻合的可能性是很高的。在这一方面，旨在寻求证实的偏差超过了理智的思考。

洛夏墨迹测验

有关验证性意见的另外一个证据是，我们会选用一个具有双重标准的证据来支持自己。我们将支持既有信念的证据赋予很大的权重，同时忽略那些威胁我们信念的证据。洛夏墨迹测验解释的发展历史，和这项测验效度的证据，都向我们展示了验证性意见的危险性。

洛夏墨迹测验是一种投射测验。通过你对一些模棱两可的刺激（一系列墨迹）的反应，可以反映出你的人格特征。对于如何实施测验、计分以及解释结果都存在很大的争议，使得这项测验看起来很有问题。对测验本身的理解缺乏一致性意味着不同的测验者会对你的测验反应给出不同解释。如果就连“专家们”都不能对测验反应有一致的看法，这样的测验又怎么能说是有效的呢？实际上，几乎没有证据可以证明洛夏墨迹测验是有效的人格测量工具，特别是在考虑“增加效度”（与那些更为经济简便、已经证明有效的测量工具相比，在增添有效新信息方面的贡献）和临床有用性（根据任何合理的定义，所产生的对执业临床医生日常工作的用处）时，更是如此（Hunsley & Bailey, 1999）。洛夏测验作为《心理测量年鉴》中主要的心理测验工具之一，几十年来一直受到高度的批评和质疑。例如，Arthur Jensen 在 1965 年版的年鉴中写到：

许多心理学家对洛夏测验进行了评价，他们都同意说 40 年来在洛夏测验技术上浪费了大量的精力。至少，在发展一套有效的心理学测验工具这项工作上，这些精力被证明是无效的……。在临床心理学中科学进步的程度，可以用彻底抛弃洛夏测验的速度和程度来衡量。

那么，临床心理学界在“抛弃”洛夏测验方面做得怎样呢？在 40 年后的今天，洛夏测验依然是心理测验中最流行的工具之一。临床心理学家们有没有意识到自己是将洛夏测验作为微弱无力的、经验性的支持证据呢？显然没有。洛夏墨

迹测验依然流行而且充满争议。测量领域的领先杂志——《心理评估》(*Psychological Assessment*) 杂志，曾花了整整两期的篇幅（1999 年第 9 期和 2001 年第 12 期，分三部分连载了洛夏测验的支持者和反对者就这个话题进行的争论。那些一直在使用洛夏测验的支持者，在大量个人经验的基础上为其进行了辩护。他们在接受自身个人经验的同时，其他人的经验以及证明洛夏测验无效的科学证据，统统都被他们忽略或低估了。

再来讨论一下画人测验。画人测验也是一个投射测验，Chapman 和 Chapman (1982) 曾引述了一些执业治疗师在描述画人测验时的伪科学态度：“一个临床心理学家说，‘和某些期刊上的文献相比，我更相信我自己的直觉’；另外一个心理学家说，‘我知道在实验室里妄想狂的患者不会画大的眼睛，但在我的办公室里可不是这样’。”这些临床心理学家已经成为了验证性意见的受害者，全然不顾反面的信息，只能看到他们想要看到的东西。

Gilovich 和 Savitsky (1996) 曾经描述了验证性意见是如何影响对洛夏墨迹测验的解释，其中有代表性的一句话是：“相同的东西总是伴随出现。”例如，在洛夏墨迹测验中看到眼睛，被解释为多疑或是偏执狂；看到大的头颅，被解释为对自己的智力缺乏信心；看到妖怪或非人的生物，被解释为同性恋等等。这些表面的、事先拥有的观念，使临床心理学家对洛夏测验的解释产生了强烈的偏差 (Chapman & Chapman, 1967)。因此，洛夏墨迹测验本质上是一种伪科学的测量方法，正是由于验证性意见的存在，才使它表面上看起来很有效 (Garb & Boyle, 2003; Hunsley, Lee, & Wood, 2003)。

笔迹学：书法的分析

许多欧洲国家和以色列的雇主，在选聘雇员之前，会付钱给笔迹学家，让他们对候选人员的笔迹进行分析 (Schmidt & Hunter, 1998)。有没有科学证据支持笔迹学呢？没有。就像洛夏墨迹测验那样，书法样本提供的是一系列模糊的信息，专家们再从这些模糊信息中抽选出一部分来形成他们的解释。在这种情况下，验证性意见很容易发生。有关笔迹学的科学研究结果也清楚地表明，专家们的判断很大程度上会受到验证性意见的影响。

首先，和普通人相比，专家在从书法样本推论书写者的人格方面，并不会更加出色 (Bar-Hillel & Ben-Shakhar, 1986)。因此对于这样的任务来说，根本没有

必要花大价钱雇请笔迹学家。第二，尽管书法的内容对判断书写者的人格特质还有点儿用，但书法的形式却没有丝毫用处。在让书写者自由决定书写内容的时候，笔迹学家们猜测他们的人格特征还有一定的准确度，但在让书写者抄写一段既定的文字时，他们就不能预测到任何东西了。因此，你书写的形式和特点，对表现你的人格没有一点效度（Neter & Ben-Shakhar, 1989）。笔迹学家能够猜到你的人格特点，完全是基于你写了什么。他们会强调你书法的形式和特点的某些方面，同时忽略不一致的信息，以此来证实自己的解释。也正是因为这样，笔迹学家们相信自己可以从书法的形式和特点上猜测书写者的人格，这与临床心理学家们相信洛夏墨迹测验是同一个原因——验证性意见。

多通道生理记录仪：测谎

关于验证性意见的最后一个例子是常用的多通道生理记录仪，也就是俗称的测谎仪。要确定一个人是否在说谎是非常困难的，而且也的确没有一种绝对可靠的方法来作出评判。在许多人的心目当中，测谎仪是当前可用的最科学有效的测谎程序，而实际上有很多人人为的因素参与在其中，因此，多通道生理记录仪比人们所想的更容易受到验证性意见的影响。在测谎过程中，最常使用的是控制问题技术（*control question technique, CQT*），即让被测者回答一系列的问题，同时用仪器持续地记录被测者的生理过程（如心跳、皮肤电阻等）。这一系列问题是由两类问题组成：一类是目标问题，如“你是不是从你的老板那儿偷走了 2000 美元？”另一类是控制问题，如“你为了摆脱麻烦，是不是撒过谎呢？”如果被测者在回答目标问题时的生理反应比回答控制问题时有所提高，那么测试者就断定：被测者不诚实。

多通道生理记录仪存在几个严重缺陷，大大地限制了它的使用。例如：控制问题不能控制的一点是，老实人在被问到目标问题时，也会感到紧张或受到威胁，因此生理反应也会升高。更为合适的做法是：将一些有关捏造的罪行的问题作为控制问题，然后比较被测者在回答这类问题与真正的目标问题时的生理反应。如果在回答目标问题时的生理反应比回答捏造问题时更高的话，那么就能更肯定地认为此人不诚实。多通道生理记录仪的另一个问题是，撒谎者可能会通过学习一些技巧与对策，来抑制自己的生理反应，因此这种情况下测得的生理反应就没有假定的那么有效和精确。因为大多数的相关研究都存在方法学上的缺陷，

所以有关多通道生理记录仪的效度很难确切地估计（美国国家科学研究委员会，2003）。一个较合理的估计是：测谎仪对不诚实者的正确检测率为85%，对诚实者的正确检测率为60%（Lykken, 1998）。在第13章里，我们还会看到实际操作的例证，表明当多通道生理记录仪检测的人很多，而其中说谎者比例又极低时，它的正确检测率会大大下降，而另一项鲜为人知的技术却可以达到更高的正确率。

更为重要的是，多通道生理记录仪也有其心理上的局限性。Ben-Shakhar (1991) 提出了一个四因素的模型，重点考察在多通道生理记录仪测试中验证性意见所起的作用。首先，多通道生理记录仪给测试者提供了大量的复杂信息。通常，测谎程序至少包括三个测试序列，由10个或更多问题组成，同时记录的生理指标大约是三个。要整合所得到的如此多的信息，是一项十分艰难的任务。第二，测试结果会被其他来源的信息所“污染”。不管是谁雇用了测试者，都会提供一些相关的信息，其中甚至还包括了归罪信息或疑犯信息，在测谎之前还会对被测者进行大量的审问。第三，测试者会形成一个有关被测者是否诚实的假设，这一假设就来源于被测者的测谎数据和案件相关“污染”信息的某种程度的综合。第四，多通道生理记录仪的结果被用来检验假设，而这种检验很大程度上受到验证性意见的影响。当测试者对多通道生理记录仪的生理读数进行检验时，可能会有选择性地寻找支持的证据，而忽略矛盾的信息，以此把大量的复杂信息整合起来证实最初的假设。因此，多通道生理记录仪的测试结果并不一定是为最后的决定提供信息，而是为了把支持性的证据整合起来，以提供看似客观的数据形式，最终证实原本的假设。在1986年的“60分钟”^{*} 节目中播出过一个有趣的调查，为 Ben-Shakhar 的这个模型提供了支持性的证据：

三家多通道生理记录仪公司分别应邀对一起盗窃案的嫌疑分子——摄影杂志社办公室的4名员工——进行测试，他们被告知办公室一部相机和一些透镜被窃，而事实上这个办公室里什么也没有丢。测试者还被告知：盗窃人只可能是这个办公室里4个员工中的一个，而且“某人可能是嫌疑犯”，被指认的嫌疑犯分别为不同的员工（都是很微弱、不太确定的指认）。结果，多通道生理记录仪测试者们都检测出“被指认的”员工撒了谎，其他员工

^{*} 60分钟，是美国哥伦比亚广播公司的王牌电视新闻节目。——译者注

是清白的；而且所有的测试者都对他们的测试结果非常有信心。这件事说明，不仅是多通道生理记录仪的测试者会犯错误，而且他们的判断和决策过程会受到一种系统、强大的偏差的影响，这种偏差就来自于“污染”（Shakhar, 1991）。

关于多通道生理记录仪有效性的争论一直持续了几十年。只要发现一个安全漏洞，政府官员们就开始呼吁增加多通道记录仪的使用，以此来解决问题（Brooks, 1985）。例如，20世纪80年代早期，里根总统曾提出要在政府官员中广泛使用多通道记录仪，幸运的是，Saxe, Dougherty 和 Cross（1985）在回顾了相关的科学证据后，提出了科学的反对意见，这项行政指令就此搁置了。1998年，联邦政府通过了一项法律，禁止私人公司使用多通道记录仪，但允许警察、军方、国家安全机构和其他政府机关使用。由于出现了“9·11”恐怖袭击，以及其他一些机密信息泄漏给国外机构的事件，美国能源部要求国家科学院重新考虑多通道记录仪在政府部门现有员工的筛选和未来人员选拔方面的作用，以此作为增强国家安全性的必要措施。

国家科学院的研究机构——国家科学委员会，在综合了所有有关多通道记录仪的现有理论和研究之后，得出的结论是否定的（美国国家科学研究委员会，2003）。该研究指出，多通道记录仪的理论基础不堪一击，大多数研究存在设计和实施上的缺陷，而且所得到的支持性证据也非常模糊和微弱。此外，研究还指出，很难确定多通道记录仪的威慑作用（即潜在的间谍会因为惧怕接受测试而不敢从事谍报活动）是否掩盖了测试本身的一些问题，如测试的效度低；由于错误的测试结果使得员工的事业受到严重损害，从而降低了政府机关的整体士气；又或者让一个真正的间谍通过了测试，反而产生了虚假的安全感，而其他的调查手段却因此被放弃，等等。

不过，目前美国能源部仍然不顾大量研究得到的数据，继续使用多通道记录仪作为筛查工具之一。看起来，联邦政府官员们也非常急于忽略多通道记录仪的严重局限性，要求对政府雇员和科学家进行测试，尽管在这种测试中，测试者的最初怀疑比实际的科学数据更能左右最终的结论。就像“60分钟”节目所展示的那样，随意的暗示可以对多通道记录仪的实施和解释产生戏剧性的影响。不可否认，测谎工作是一项极为重要的现实问题，但是像多通道生理记录仪这样的伪科学方法，受到验证性意见的严重“污染”，绝不是关于测谎问题的明智的解决方案。

缺乏反馈或误导性的反馈

人类判断过程的另一个局限，是我们对矫正式反馈的使用愿望，我们需要这样的反馈来评估和修正我们的观点。可是，因为认知与注意的偏差（如验证性意见），以及我们自身寻求并建立的环境的性质，使我们很少接收或注意到反馈的信息。实际上，我们得到的反馈信息总是严重偏向于肯定的方向。也因为这样，我们的观点很少会受到挑战。

通常，即使我们收到过负面的、可用于验证自身信念的客观反馈信息，那也是为数很少的。假设一个临床心理学家非常依赖洛夏墨迹测验来辅助诊断，那么测验实施了，反应解释了，诊断也就出来了。对于一些临床从业者来说，诊断错误对于治疗过程可能根本没什么影响。例如，当所有的病人都接受相似的治疗时，根本不可能发现诊断的错误。也正是因为缺乏矫正性的反馈，人们对洛夏墨迹测验的信念变得坚不可摧。

再举另一个例子：假设成功的治疗取决于正确的诊断，那么临床心理学家会不会接受矫正式反馈（即使他已经做出了错误的诊断和无效的治疗）？鉴于病人寻求咨询的一般原因以及病人与治疗师之间的动态关系，病人会倾向于接受和同意治疗师所给出的诊断、建议，这一点也不让人感到奇怪。即使被误诊的病人，也会相信他们正在接受最好的治疗，治疗的失败因此变得不那么明显，最后也许永远也不会受到怀疑。而如果治疗没有取得成功，治疗师和病人也并不会去质疑原有诊断的正确性，更不会去质疑洛夏墨迹测验本身的效度。他们还有很多可以利用的借口，比如“或许这是个特别的案例”、“或许治疗开始得不够早”、“或许病人对治疗师存在阻抗（或者正好与此相反：治疗师对病人有阻抗）”。我们可以找到无数的借口为失败的结果做辩护，但从来都不会去质疑我们所持的信念。所以临床心理学家根本无法从经验中得知，洛夏墨迹测验不能帮助他们做出正确的诊断。

通常，我们得到的反馈总是存在着系统性的偏差。许多反馈本身就是正性的，或者很容易被歪曲成正性的。我们怎么评估教师的教学水平呢？一些教师认为标准化的评估一文不值，他们更倾向于从更直接的经验中获取信息：哪位教师的前任学生还依然和他保持联系？哪位教师的学生请求教师帮他写推荐信？如果

一个学生去修某个教师的另一门课，那么显然是这个教师的授课水平不错（不管他是不是教必修课）；而如果有学生没修这门课，又正好被注意到了，那肯定是因为其他课和这门课时间安排上有冲突。就这样，当这门课结束时，教师自前任学生那里总结的直接经验，只会来自于那些相信他教得很好并表示感激的学生，或是那些为了以后的学分或推荐信想要得到一个好成绩，因此去恭维老师的学生。

我们总是选择性地将自己置身于一个只包含有利信息的环境；我们倾向于和与自己有同样想法的人呆在一起；我们喜欢读那些支持自身观点的书和文章而不会理会观点相反的东西；我们乐意参加与我们兴趣相投的俱乐部和组织，等等。如果已经有了那么多的防御性措施，那么矫正性的反馈又从何而来呢？

后见之明偏差和过分自信

对于个人来说，一旦我们发现了事情之间存在有意义的关联，挖掘出了证实性的证据，并且收到了恭维式的反馈信息，后见之明就会阻挡我们的视线。我们开始觉得自己“一直就知道”，并通过“渐进性决定论”（*creeping determinism*）的方式，用确定代替了猜想（Fischhoff, 1982），我们开始对自身的信念持有过分的自信。要弄清楚是否大多数人大多数时候都存在着过分的自信，最好的方法就是检验一下你自己的自信。

Russo 和 Shoemaker（1989）设计了一个简单的测验，通过常识性知识测验来评估自信水平。对于下面的 10 个问题，你要分别设立一个区间，你有 90% 的信心确信正确答案就在区间的范围之内。例如下面这个问题，“人脑要正常运作的话，大概需要多大的功率？”假如你有 90% 的信心相信正确答案在 30 到 90 瓦特之间，这就是你的置信区间（不过这个区间里并没有包括正确答案，正确答案是 15 瓦）。这个测验的关键并不是你是否知道正确答案——尽管某些问题你的确知道——而是你要设立一个有 90% 的信心认为包含正确答案的置信区间。好，下面就是这 10 个问题：

1. 马丁·路得·金逝世时，年龄是多少？
2. 尼罗河长多少英里？
3. OPEC 有多少个加盟国？

4. 旧约全书包括多少本书？
5. 月球的直径有多少英里？
6. 一架空的波音 747 飞机重多少磅？
7. 莫扎特生于哪一年？
8. 一头亚洲象的妊娠期有多少天？
9. 从伦敦到东京的空中距离有多少英里？
10. 大洋中最深的点有多少英尺？

这些问题的答案在本章的最后。如果你答对了所有的问题（都在 90% 的置信区间中），那么你很可能是没有自信的，不用说是令人惊异的极少数人中的一员了；如果你刚好答对了 9 个问题，你的自信很可能和你的实际水平恰好相符，这样的人也是非常少的；如果你答对的问题在 9 个以下，那么你就像大多数人一样，可能是过分自信了。这些问题看上去微不足道，但却是衡量我们自信心的一种方法。过分自信的问题实际上对于人们所持的许多类型的信念来说都会出现，而且它在许多任务和情境下都普遍存在（Alpert & Raiffa, 1982; Oskamp, 1965; Ruscio, 2000）。这一点无论是对专家，还是对外行人来说都是如此。

阻碍成功的事后解释

在好奇和怀疑之间会有一个合适的平衡点，找到这个平衡点可以有效地防御事后解释的情况出现。我们需要在发现的领域表现出强烈的好奇心。要有效减少事后解释所引起的判断缺陷，最好的办法就是考虑多种假设（Arkes, 1991）。因为必须要寻找强有力的证据来驳倒那些竞争性的理论，所以多个假设的同时存在可以适当地减少对某个特定假设的过分自信。而且事实上，正确的假设也更可能存在于一个更大的可能性空间里。

而在证实的领域，就需要我们表现出更多的怀疑了。不仅是要求别人这样做，我们最好把自身的信念也付诸于严格的检验。用最严格的标准来选择证据，可以使你获得并保持最正确的信念。你观察到的结果有可能只是由于概率问题出现的随机波动，并非什么真正有意义的发现，你一定要持开放的态度来对待。Abelson（1995）在其《作为原则性论据的统计结果》（*Statistics as Principled Argument*）一书中，敦促读者“给随机现象一个机会”，就很好地总结了这种观点。

当你准备宣布一个你认为十分有意义的重要发现时，请再问问你自己，根据同样的证据，你是否还能接受其他不同的观点。

最后，如果你只能通过回顾性的方式，在事情发生以后看出某种问题或者意义，就说明你受到了事后检验的影响。只有当你能根据某种现象进行可验证的预测时，才是真正地理解了这一现象。当有人宣称发现了某种知识，那么作为寻求证据的方式，你可以问他“你可以预测什么呢？”让这个简单的习惯成为你批判性思维的技巧之一，它可以帮助你清楚地区分发现的领域和验证的领域，可以在你自己或别人论证的过程中，清楚地认识到事后解释问题的存在，这是一种特殊的能力，绝对值得拥有。

信心问题的答案

(1) 39 岁。(2) 4, 187 英里。(3) 13 个国家。(4) 39 本书。(5) 2, 160 英里。(6) 390, 000 英镑。(7) 1756。(8) 645 天。(9) 5, 959 英里。(10) 36, 198 英尺。

11. 图式：大阴谋理论的诱惑力

以“我们沉默的世纪”为题，讽刺性作品《洋葱头》的出版人在一系列虚构的报纸头版新闻上，针对发生在20世纪的几乎每件事进行了评论。他们在1963年11月22日的报纸头条中，围绕许多关于肯尼迪总统遇刺的阴谋理论作了一番取笑，“肯尼迪是被美国中情局，黑手党，卡斯特罗（Castro），LBJ*，Teamsters**成员，互济会会员一起杀死的；总统被129发子弹从43个不同的角度打中了”。人们曾经提出过很多阴谋理论来解释总统的遇刺，到现在已经成为谈笑的话题。但问题是，这么多的阴谋理论是如何发展起来的呢？为什么阴谋理论如此盛行，而且经久不衰，使人不得不信服呢？另外，我们要如何来评估一个阴谋理论是否真实呢？

在一个阴谋理论中的三种角色

理解阴谋理论的第一步，是区分三种角色的不同作用。任何一个阴谋理论都涉及三种主要角色：阴谋者、拯救者和被蒙蔽的人。

* LBJ, Lyndon Baines Johnson, 美国第36任总统。1963年11月22日肯尼迪遇刺，时任副总统的LBJ在空军一号上宣誓就职，1964年正式当选。——译者注

** 美国卡车司机协会。——译者注

阴谋者

阴谋者是指那些被指控犯了阴谋罪的人。这个角色很少是自愿的，大多是被那些提出阴谋理论的人强加上去的。让一个阴谋成为“大”阴谋的因素是：时间跨度长，参与阴谋的人数多。如果任何一个关于肯尼迪遇刺的阴谋理论是正确的话，那么阴谋者已经成功地掩盖他们的阴谋长达40多年。在这期间，已经有半打以上的总统宣誓就职。因此，我们完全可以说，联邦政府具有一项迄今为止尚不为人知的能力，那就是保持秘密的能力。

拯救者

拯救者是指那些相信他们自己发现了阴谋的证据，而且想要把阴谋公之于众的人。他们宣称发现了精心掩饰的表象下隐藏的疑点或丑闻，而且想让所有人都知道这一切。这个角色是自愿的，通常还非常热衷于扮演这一角色。拯救者相信，他们拥有把其他人从邪恶的罪行或者阴谋计划中拯救出来的钥匙。当我们考虑到拯救者这一角色所带来的心理获益时，就不难理解阴谋理论是怎样被制造出来，以及它为什么可以维持下去了。

拯救者的角色在两个方面拥有特权：第一，他拥有秘密的信息。我们中有谁会愿意错过拥有一些可供分享的有趣新闻的机会？第二，拯救者拥有的独一无二的信息，最终可以铲除阴谋者的谎言和欺骗，让真相大白于天下。难道你不想成为一个真理的使者，一个腐败谎言的粉碎者吗？拯救者通过揭露阴谋中的危险，给公众带来安全。这个目标还可以达到更高的境界：拯救者愿意帮助任何缺乏敏锐的观察能力、缺乏秘密信息的来源渠道，或者缺乏对事情进行合理分析能力的人。任何不能发现阴谋的潜在危险的人都需要帮助，而拯救者就是第一个冲在前面的人。

试想一下，对于一个单调的生命来说，拯救者这一角色是多么有意义啊！投入到与巨大的阴谋作战这一高尚的事业，又是如何能激励人啊！因此，有些人总可以在任何地方发现阴谋，这有什么奇怪呢？当某个拯救者厌倦了陈腐的阴谋故事或者无效的抗争行为，总还会有新人出来继续与阴谋战斗下去，其实这也没有什么奇怪的。这只是反映了人们对特殊地位的渴望，或者对更大的人生目标的追求而已。

被蒙蔽的人

被蒙蔽的人不是那些被指认为阴谋者的人，拯救者相信他们并没有参与策划什么阴谋，但他们也并不相信有拯救者所说的阴谋存在。他们在拯救者眼中是不知情的、容易轻信的，在阴谋者的蒙骗下只顾享受安逸生活的人。很明显，他们十分需要拯救者的启迪，以认识到阴谋的危险。被蒙蔽的人对拯救者所说的危险通常是持怀疑态度，挑战所谓阴谋的证据，甚至会把整个阴谋理论看成纯粹是时间和精力上的浪费。通常，他们在数量上远远大于拯救者，这会让拯救者有更强烈的动机去继续他们的使命——启迪被蒙蔽的人们。

大阴谋理论：“第 51 区”

尽管大众文化中充斥着阴谋理论，但很少有阴谋可以称得上是“大”阴谋。大阴谋意味着涉及的人很多、持续的时间也很长。不过确实也有这样的阴谋理论，让人觉得不仅很有说服力，而且经得住时间的考验。一个广泛流传而且深受人们信任的大阴谋理论是：1947 年，外星人的宇宙飞船在新墨西哥州的 Roswell 附近坠落，随后这艘飞碟和几具外星人的尸体被运送到一个秘密的美国政府机构——“第 51 区”，用于进行秘密研究。

在这个阴谋理论里，阴谋者是成百上千的政府机构人员和科学家，他们对这件事守口如瓶，而且已经超过了 50 年；拯救者是那些坚信美国政府正在实行集体隐瞒的人们；而与拯救者恰好相反，被蒙蔽的人，始终不相信有这么一个飞碟坠落在 Roswell，或者有什么外星人的技术及尸体被藏匿在一个神秘的军方基地。

媒体的角色

在研究外星来访方面，Phillip Klass 是公认的、可信的科学家之一。他花了几十年时间考察政府记录，其中很多记录是最近才解密的。最后他的结论是：真正的阴谋隐藏者是新闻媒体。他不是指美国有线电视新闻网在守卫着“第 51 区”，他已经收集了足够多的证据证明，在 Roswell 根本就没有飞行器坠毁，根本就没有什么外星人造访，整个所谓的“第 51 区”阴谋理论纯属虚构。

然而，是大众新闻媒体延续了这个虚构的故事，因为它非常有趣。Klass 详

细地记录了有关 UFO 或外星人的有高度偏差的、单方面的新闻报道。在他的一些书中，我们可以发现，那些支持神秘的“第 51 区”的证据，是一些值得怀疑的观点、被公众所接受的随意推测，或者未经证实的假设。例如到现在为止，媒体大篇幅报道的一些支持外星人造访的证据，就已被证明是伪造的：

- “麦田怪圈”或空地上出现的更为复杂的图案。在第 2 章里我们已经讨论过人类制造神秘的麦田怪圈的问题。

- “MJ-12”档案的照片。这个档案的附录中提到了令人惊异的外星科技。MJ-12 档案是真的吗？一段尚未完成的影片神秘地出现在一个相信外星人造访和政府阴谋的电影制片人的邮箱里。这个照片档案应是几十年前拍下的，那么它们是在哪儿被发现的？具体是什么时间拍下来的？事实上，并没有什么令人信服的证据可以证明照片档案确实存在过，摄影排版和文字排版分析也显示这个档案是伪造的。特别令人奇怪的是，摄影者没有拍到一点儿关于档案附录的真实信息，我们所拥有的仅仅是未经证实的参考资料。总而言之，MJ-12 档案看起来缺乏证据，而且相对容易伪造。

- 福克斯电视台重复播放的外星尸体解剖影片。同样有很多理由可以怀疑这些证据的真实性。令人十分不解的是，为什么现在高清晰度的彩色电影随处可见，而这些影片却被拍成画质很差的黑白影片呢？影片的所有者拒绝提供任何可以确认拍摄日期的影片样本，而他们提供的空白片段（在胶片前面的一段空白胶卷）可以是任何地方截取的，如任何一个旧新闻的胶卷。看过这段影片的人，都可以证明它的画面质量很差；拍摄的所谓外星人，是采用远景拍摄，并不能提供任何近距离的细节；人们解剖外星人尸体的动作，也显得紧迫而又奇怪。当考虑到发现一个真正的外星人是多么难得、多么重要时，所有的这些就显得更加让人不解了。

在读这段文字之前，你是否听过或看过有关神秘的麦田怪圈或 MJ-12 档案的报道，或是外星人尸体解剖的电影？依我的经验来看，大多数人都看过或听过。你曾经对此提出过质疑吗？我个人的经验表明，很少有人会这么做。而这些都是大众新闻和娱乐媒体不加批判地报道完全不可信的证据的例证（更多关于这一话题及其他有争议的观点，参见 Klass, 1983, 1997a, 1997b, 2000）。

关于“第 51 区”的批判性思考

对阴谋理论进行批判性的思考，有两点需要考虑。第一，回忆一下第 7 章的

内容，大多数人把 UFO 想象成类似于飞碟的东西，就像所谓在 Roswell 坠毁的那个一样。然而，根据 Sheaffer (1997) 的解释，从来就没有飞碟这样一种飞行器存在过。一旦一个错误的消息报道开了，其他有关飞碟形状的 UFO 的报道和照片就会层出不穷。在这方面我们可以总结的教训是：人们总是容易看到他们想要看到或期望看到的東西。其实，原始报道描述的是一个雪茄形状的飞行器，像飞碟一样掠过水面。如果事实是这样的话，后来所谓发现飞碟的真实性就可见一斑了。

第二，如果所谓制造阴谋的两方——政府和科学家，对外宣称有外星人的存在，而不是采取隐瞒的态度，这样不是对他们更有利吗？政府每年出钱资助大量的科学研究项目，为什么他们会隐瞒这个巨大的科学发现，不让其他的科学家知道，阻碍科学发展的进程呢？譬如，外星人的发现，可以让生物学家直接研究一个独立进化的物种。而且不难想象，研究外星人很可能带来极大的技术突破。如果外星人可以制造一个飞行器，飞越星际间浩瀚的宇宙，他们肯定在技术上远远超过我们，美国国家航空航天局（NASA）用这一理由一定可以申请到不少基金的支持。既然 NASA 总是在为它的预算发愁，那么这样的好事他们为什么不愿意做呢？

整件事情中最为矛盾的地方，也是“第 51 区”理论最大的漏洞就是：为什么美国政府要对存在外星人这件事进行隐瞒和压制呢？毕竟对任何一个国家的政府来说，最基本的功能就是保护国民。没有外部的威胁，政府就很难保持它的权力（同时征收巨额的税款）。政府总是擅长通过唤起民众的恐惧，甚至有时夸大和制造来自敌人的威胁来召集民众。试想，如果确实有强大的外星人可以到达我们的星球，那政府就会有很充足的理由增加防御经费，提高税收，或者把民众联合起来抗击共同的敌人，这不正是政府梦寐以求的理由吗？我相信我们的政府，还有其他国家的政府，都会非常兴奋地宣布外星人的存在，如果有任何证据可以证明的话。

不幸的是，尽管有这些逻辑的分析，也有 Klass 等人收集到的大量的证据，有关“第 51 区”的幻想依然受到大众媒体的青睐，而且继续在我们的社会中广泛流传。例如，1997 年的夏天，在 Roswell 市公开举行了发现飞碟坠毁事件的 50 周年庆祝活动。而且，电影《独立日》（Independence Day）把“第 51 区”的阴谋描述成一个既定的事实，也让这个阴谋理论看起来更加可信了。

大阴谋理论存在的问题

从“第51区”现象我们可以发现，任何大阴谋理论都需要面对一些问题。第一，所谓的阴谋方，必须有组织地操纵阴谋。这听起来让人觉得有些难以置信。是否真的存在一个包含众多人、众多机构的联盟呢？这听起来可信吗？举个例子来说，大阴谋理论的一个常见的目标是“政府”。在美国可信赖健康资讯理事会1999年3/4月刊的时事通讯中，曾发表了William Jarvis的一篇题为“政府不是一个人”（The Government Is Not a Person）的短篇通讯，通讯的开头是这样的：

尽管美国政府经常被戏称为“山姆大叔”，但美国政府并不是一个人，也不具有道德方面的人格特征。美国政府是一个包括行政、立法和司法机构在内的复杂的网络。这些机构之间经常存在着矛盾冲突。

既然美国政府中存在着内部冲突，那么难道就不会有系统内部的人将所谓的阴谋泄漏出来？难道就不会有联邦政府或州政府的部门，可能从揭露所谓的“第51区”阴谋中获益？难道就不会有官员将这么有趣的秘密偷偷地告诉家人或朋友？我们能够保证所有人都不会与其他人分享这则消息吗？能够保证听到这则消息的人都可以保守秘密吗？

上面所说的这个问题与大阴谋理论的第二个问题是直接相关的。使大阴谋理论如此有趣的特征——规模宏大，同时也是大阴谋理论最致命的缺陷。想象一下，让成百上千的能接触到“大阴谋”的人都坚决保守这个秘密，这种可能性究竟有多大？大阴谋理论的规模使它从一开始就注定了是不可能的。如拯救者所说的那样，设法掩盖“第51区”的真相，也许会有可怕的危险，难道没有一个人有勇气将这样的信息匿名泄漏给新闻媒体吗？难道我们真的相信，能接触到所谓政府机密的所有人都愿意并且能够保守这个机密吗？

第三，必须考虑涉嫌所谓大阴谋的各方的自身利益。必须要问的一点是，所谓的阴谋者有没有足够的理由去实施这样的阴谋？他们会不会从所谓的阴谋中获得利益？我们只要仔细想一想就会发现，所谓的阴谋者并没有一个像样的理由来从事这项阴谋。所以，这一点本身就是一个强有力的证据，可以证明大阴谋理论

的不合理。如前所述，如果国防部将外星人入侵这样的消息散布出去，将会有巨大的收益，而国防部却对此事保持沉默，这一事实本身就说明了一切。

“隐匿疗法”的阴谋

有了批判性的思维武装头脑，我们再来看看下面这个卫生领域的大阴谋理论。耶鲁大学医学院的研究员 Steve Novella 将此称作“隐匿疗法理论”。支持这一理论的伪科学家们宣称，因为有一个“医疗机构”禁止“隐匿疗法”，因此这些疗法只能存在于科学医学领域之外。例如，一些伪科学家声称有人发现了治愈艾滋病、癌症或其他不治之症的方法，但这些方法被贪婪的科学家和医师隐藏了起来，以谋取自己的个人利益。如果进行仔细的审查，这些说法还会可信吗？我认为不可信，至少有以下四条理由。

激烈的竞争被忽视了

这个大阴谋理论的第一个主要问题是，根本就没有这样一个统一的医疗机构。阴谋理论的支持者们相信，美国医学协会、医学院、医院、制药公司、保险公司、医生、国家卫生研究所、国家癌症研究所等机构，统统都联合起来维持现状。大阴谋理论家所忽视的事实是，这些机构和这些机构中的个人之间很少进行联系，因而这个庞大的网络很难组织（更别说成功地组织了）这样一个大型的阴谋。就像所有其他的大阴谋理论所面临的问题一样，这个大阴谋涉及太多人，以至于根本就没有办法组织起来。

这些机构不仅仅是独立运作的单位，而且它们之间还存在着激烈的竞争。在科学机构内部，总是存在着对基金、知名度、声誉以及实验场地与设备等资源的持续争夺。任何一个独立的研究者及其雇员想要保持竞争优势，就必须取得科研成果。没有任何科学家或科学领域可以从保守现状和无所作为中获益。所有的资源——可见的和不可见的——都集中在那些可以证明自己科研成果的人手中。

搜寻新想法，并将它们付诸严格的检验，是研究者的既得利益所在。科学卫生保健系统的奖励机制也旨在重视和鼓励开发新想法，无论它们是何来源。因此，新的疾病理论、诊断工具、治疗方法不断在替代旧的理论、工具和方法。这一点和许多伪科学家的论断正好相反，他们总是鼓吹他们的做法千百年保持不

变，其实质就是没有改善和进步（详见第5章“关于远古智慧的谬误”部分）。任何鼓吹存在一个阴谋科学机构联盟的人，似乎都缺乏对于现实的科学运作机制的基本理解。

所谓的利益驱动毫无道理

相信所谓“隐匿疗法”阴谋的人认为，医疗机构因为害怕新疗法会降低他们的利润，因此将所谓的新疗法隐藏起来。这个观点初看起来似乎有些合理之处，但是请想想看，究竟谁会从一种新疗法中获益，谁又会从中受损呢？执业医师大部分时间在为病人治疗，如果一个新疗法被发现，他们也许会感到受威胁；而研究者，他们的大部分时间是在实验室里验证他们的假设（一般他们不会同时也是执业医师），就不太可能因为新疗法的发现而感到受威胁了。事实上，假如一个研究团队发现了一种治疗艾滋病或癌症的新方法，很多很多的奖励和名誉会纷至沓来，无数的好工作和基金支持在等着他们，另外还有不计其数的专业津贴。因此，研究者们完全没有任何理由去隐匿医学突破，事实上，他们可以从获益无数。

此外，我们还可以从另一个角度来考察个人利益的问题。医疗从业人员愿意拿自己的事业冒险吗？那么医生宣誓过的希波克拉底誓言——要尽力为患者减轻痛苦——又意味着什么呢？如果医生没有使用最有效的治疗方法，病人可以针对医生的治疗失当进行诉讼，医生们愿意承担昂贵的诉讼费用吗？隐瞒一个有效的治疗方法，会给医生们带来道德和法律两方面的巨大挑战，谁又愿意冒这个风险呢？

科学永远也不能治愈所有的疾病

如果医生们不隐瞒“隐匿疗法”，而是将这种神奇的疗法公布出来，那么他们的事业是不是就要面临破产呢？回顾一下医学发展的历史，我们就不难发现这种想法是极其愚蠢的。在过去，当疫苗的发明使得人们对致命的疾病不再束手无策，医学科学事业有没有因此而轰然倒塌呢？当天花和脊髓灰质炎被消灭，执业医师们有没有因此而面临困境呢？当然没有。随着对抗疾病的重大进步，的确某些专业人员会受到一定的影响，但正在从事这些专业，而且全部工作也只是治疗一种疾病的人毕竟很少很少。即使真是这样，他们也仍然可以因需改变，重新选

择自己的专业。这种事情其实经常发生，并不仅仅是卫生服务领域的特点，其他不断进步的学科也是如此。错误地认为取得了某种医学发现就可以治愈所有的疾病，持有这种观点的人真可以说是天真得有些可悲。

极端无情的假设

“隐匿疗法”阴谋理论的最后一个是，支持者们会将所谓的阴谋者看成是极端无情的人。隐瞒一个可以缓解患者痛苦的有效治疗方法，可以说是既无情又不人道（同时也违反了医师的职业道德）。那些在所谓的阴谋医疗机构工作的人，也有家人和朋友，他们的家人和朋友也有可能患上这些不治之症，这些医生难道会如此狠心，以至于连自己所爱的人的生命都不顾？认为阴谋者会如此极端的无情，本身就是对人性的一种极端扭曲。现在你可以自己再来做一个判断，看看这种论断是否站得住脚。

为什么“隐匿疗法”理论依然存在

现在看来，根本没有充足的证据证明，存在那样一个统一联合的阴谋医疗机构。这个想法当中有太多的逻辑错误了。再进一步说，如果真的存在这样一群阴谋者，他们也还是缺乏足够的动机去隐匿所谓的医学奇迹。人类本性的一个基本原则是：我们的行为大部分是目标取向的。发现并推广一种新疗法，绝对符合卫生领域研究者的切身利益，因此必须有更充足的理由来说明他们为什么要隐瞒一种有效的治疗方法，问题是这样有说服力的理由一直都没有出现。

那么，为什么还会有人坚持宣称医疗机构隐匿了新疗法呢？问题的答案可能就是前面我们提到过的拯救者角色所带来的心理获益。或许，“知道真相”的特权、能够拯救他人的特殊地位，就足以使某些人产生强大的动机了。

拯救者角色的心理获益可能可以解释大阴谋理论，但还存在一个问题：为什么所谓的阴谋医疗机构从来就没有涉及伪科学的从业者呢？我曾经多次听过“隐匿疗法”理论，而且听过许多不同的版本，其内容不外乎是指控科学家或政府隐瞒一种神奇的、只有一两个曾被人忽视的医师所知道的治疗方法，而那些孤独的人往往有一些十分有问题的想法。正如我们曾在第4章所讨论过的，这就足以让我们怀疑，这些极少数人可能实际上只是一些被忽视的骗子，而并非少数被忽视的天才。

那么，也许问题的答案和许多鼓吹伪科学的骗子所持有的反科学态度密切相关。把科学家描述成残忍的人，说他们故意隐瞒新的有效疗法，就可以为那些伪科学家们提供新的商机。伪科学家们把人们的注意从合理的证据上引开，把对卫生保健服务的选择说成是道德问题，并且说服别人相信他们处于所谓的弱势地位。许多人的确有过一些不愉快的就医经历：也许是等得太久，也许经历过尴尬的或痛苦的侵入性检查，也许被误诊过，也许遇到过粗鲁的或自以为是的医生。这些问题以及其他一些问题确实是存在的，而且其中有许多的是医疗服务人员的错。因此可以理解，在这种情况下人们经常会有一些不满，然后会通过其他途径来寻求治疗。

虽然上述问题的确会使人另有想法，或者干脆换个医生，但它并不意味着应该彻底地放弃医学科学。毕竟卫生保健专业人员和我们一样，难免会有犯错的时候。而基于研究证据的、科学的医疗服务，仍然是我们寻求健康保健措施的最佳选择。伪科学家深谙如何激起人们对医生的失望和不满，再加上人们对医学科学缺乏了解、对不确定性抱有恐惧心理，有求医需求的患者就可能因此放弃寻求科学的医疗方法，转而求助于伪科学的方法。所以说，可能正是那些庸医和骗子们出于要在充满竞争的医疗保健领域占有一席之地的企图，将“隐匿疗法”理论作为了有利可图的工具。

质疑阴谋理论

阴谋理论也许表面看上去很吸引人，但只要细细考察，你会发现它们是非常不可信的。如果你看到或听到一个大阴谋理论，请试着问问自己三个基本的问题：第一，所谓的阴谋者真的是出于共同的利益而联合在一起吗？留心，也许所谓的阴谋者会被冠以一个笼统的名称，如“政府”或“医疗机构”，因而让阴谋理论看起来颇为可信，但这些名称很可能只是反映了一个并不存在的个人或机构联合体。事实上，所谓的阴谋者之间往往还存在着激烈的竞争和冲突。尤其是涉及科学家的大阴谋理论，更是非常不可能的，因为在科研领域的奖励机制中，处处都存在着竞争。

第二个要考虑的问题是，阴谋理论的规模是否已经反映了它的不可信？有没有可能许多人持续长时间地实施一个阴谋？考虑一下阴谋理论所意味的道德败

坏，就算是阴谋者有合理的动机，也不见得所有的相关人员都会从中受益。另外，在一个大阴谋网络中很容易出现那么一个有社会责任心的人，并不同意其他人的做法，并匿名地将档案材料泄漏给法律机构或是新闻媒体。现在请你再综合判断一下，究竟有没有可能如此多的人如此不遗余力地实施一个阴谋，而且大家都如此意志薄弱，如此道德败坏，以至于虽然对阴谋有所不满，但仍然保持沉默、守口如瓶？

最后，如果前面两个问题使你开始怀疑大阴谋理论可能是虚构的，那么你就要问一下，为什么有人想让你相信这个理论？这只是单纯为了娱乐，还是为了使生活更加有趣？是不是有什么人会从拯救者的角色中得到心理获益，使他想要传播威胁性的信息，试图拯救他人？又或者，是不是还有一些更加见不得人的动机：有没有可能这个人是一个伪科学家，他想通过声称自己的想法遭到排挤来博取信任？那些被主流所遗弃的人，可能会虚构一些阴谋理论来解释自己为什么不为主流所接受，而我们已经了解了这些手法的愚蠢之处。当有人向你推销某个大阴谋理论，但阴谋的存在既没有什么证据，又缺乏极有说服力的论证时，你就应当意识到，前面所提到的这些问题都有可能成为其不可告人的动机。

12. 错觉：控制感

为什么大多数彩票都让你自己选择号码？不是所有的随机号码组合都同样可能赢得大奖吗？事实上，不管选择什么号码，中奖的可能性都是一样的，但人们还是喜欢自由选择号码。这种自由有什么好处呢？让我们来看一个有趣的实验。

在这个实验中（Langer, 1975），实验者邀请办公室的员工每人购买一张 1 美元的彩票，再从中抽取出一个幸运者。需要 20 人参加这个游戏，有一个幸运者会得到 25 美元（实验中的彩票不同于商业彩票，在这个实验中，实验者每实验一次就要付出 5 美元）。一半的参与者是随机得到彩票号码，而另一半的参与者则可以自己选择号码。因为这种彩票胜算相对较大，所以 20 张彩票全部卖完了。这时，实验者开始逐个询问彩票持有者：“有其他人想要参与这个游戏，但已经没有多余的彩票了，你愿意以多少钱的价格转让自己手中的彩票？”从参与者的回答中，实验者得到了两个非常有趣的发现：

第一，平均来说，人们索取的价格超过了他们付出的价钱（1 美元），甚至超过了彩票的客观价值（根据彩票的期望价值计算是 25 美元/20 人，每张彩票值 1.25 美元）。实际上，参与者提出的转让彩票的平均价格大概是 5 美元。经济学家把这种人们过高估计自有财产价值的倾向叫做赋予效应（endowment effect）。换句话说，就是我们常常对自身拥有的东西赋予更高的价值。

第二，这一实验的发现非常新奇。那些随机获得彩票的人，仅仅向实验者收取了大约 2 美元，稍稍高出彩票的客观价值 1.25 美元；而那些自己选择了号码

的人，平均向实验者收取了大约 8 美元。仅仅是自由选择的号码，就使彩票的预期价值增加了 4 倍，这实在是令人吃惊，因为彩票的客观价值是一样的——中奖的概率依然是二十分之一，每张彩票的期望价值依然是 1.25 美元。很明显，那些自由选择号码的人认为他们中奖的概率要高于随机水平。因此，Langer 创造了控制错觉（illusion of control）这个词，来指人们对纯粹随机事件的控制的误解。

控制错觉

当我们错误地理解周围世界中的随机现象时，就会产生控制错觉。随机事件是不确定的，因此会使人感到畏惧。即使是假想不确定性会减少，也会使可预测以及理解的感觉得以上升，这就像我们有时对周围的社会和物理世界会产生迷信或奇怪的信念一样（参见第 4 章）。即使真的是单纯的随机事件，我们同样会倾向于在混乱中强加秩序，从而获得控制感，尽管那只是个错觉而已。我们往往会模糊两种不同情况之间的区别，比如需要技巧的体育竞技和只需要运气的彩票。造成这种模糊的原因是运气情况具有类似于技巧情况的特点。举例来说，技巧情况的一个特点是可选择性，而在运气情况下，可选择性则导致控制错觉。

Langer（1975）还发现了促使控制错觉产生的其他几种特征。例如，在一个只靠运气的纸牌游戏中，参加者在面对一个穿着体面、看起来很时髦的玩家时，下的赌注更少；而在面对一个穿得很差、看起来很卑微的玩家时，下的赌注更多。虽然对手的个人特征和纯粹机会性的纸牌游戏毫无关系，但参加者在卑微的人面前会比在体面的人面前感受到更多的控制感。可见，这种赌博的方式是完全不理智的。同样的情况还包括：尽管中奖的机会均等，参与者还是愿意付更多的钱买上面有字母的彩票，而不是上面有数字的彩票。在另外一个实验中，Langer（1975）发现，尽管练习不能提高机会概率，但人们还是觉得练习提高了控制感。此外，个人经验和时间本身也都会提高控制感。由此看来，我们很想控制周围的环境，尽管从理论上和客观上来说，我们都不可能超越机会概率。

运气情况和技巧情况有时很难区分。在技巧情况下，感知到可控制感是无可厚非的，而且还是一种优势。在运气情况下，是否感知到可控制感有时是无关紧要的，有时甚至还会带来伤害。例如，一个赌徒想要大点数，他就会用力掷骰

子，如果他想要小点数，就会轻轻地掷骰子。这种类型的迷信行为，源于一种稳定的、可预期和可控制的世界观，可能还无伤大雅。但有时在运气情况下的控制错觉却会造成伤害。例如，在 Langer (1975) 的一个彩票实验中，处于控制错觉中的参与者，拒绝用自己的彩票交换更可能赢的彩票。首先让我们来看一个现象，该现象和不确定性的减少有关，之后我们再来讨论控制错觉可能带来的风险。

确定效应

因为不确定性会让人感到不安全，所以在我们的决策过程中，一个很强烈的动机就是选择确定性、避免不确定性。和减少同样程度的风险相比（从 15% 到 5%），我们会更注重完全消除风险（从 10% 到 0%），这一点都不奇怪。现在请考虑下面的情况，以检验你自己对确定性的偏好程度：

假设你被迫要玩一轮标准的俄罗斯轮盘赌，你将拿着一支装有一颗子弹的左轮手枪，对准自己的脑袋扣动一次扳机，那么你愿意出多少钱把这颗子弹从手枪里拿掉呢？现在还是设想你要玩一次俄罗斯轮盘赌，但是枪膛里面有四颗子弹。那么你愿意出多少钱把四颗子弹中的一颗拿掉呢？

假如你和这个实验（Tversky & Kahneman, 1981）的参与者们反应一样，你就会愿意出较多的钱把枪膛中唯一的一颗子弹拿掉，而出较少的钱把四颗子弹中的一颗拿掉。客观地来说，这两种条件都是拿掉一颗子弹，都会把死亡的风险降低六分之一，但人们总是认为，把死亡的风险从六分之一降低为零，比从六分之四降低为六分之三更有价值。

实际上，保险公司就是利用这种确定效应（certainty effect）赚了大把的钱。保险之所以有利可图，就是因为人们宁愿多花钱来消除风险，而不是少花钱来过没有保障的生活。就像人们不大可能赢回花在购买彩票上面的钱一样，买保险也不大可能拿回花在保费上的钱。但保险仍然很吸引人，因为没有保险可能会让你面对无法承受的巨大损失。因此，保险公司和赌场赚钱的道理其实是一样的：概率对他们有利。而且，正是由于我们对巨大损失的极端讨厌和对确定性的偏爱，使得保险具有更大的吸引力。

确定效应消除了不确定性，在混乱中创造秩序，给人们以最大限度的控制感。但是，如果我们明知这种控制感只是幻觉时还依旧自我欺骗，那就是在开始玩火了。

控制错觉的危险

拥有确定感和控制感的确不错，因为它们减少了生活中的不确定性，会使我们感觉更强大。但是，控制错觉也可能会产生不良的后果，如不适当的责怪，或无助的人生观等等。

相信“公平的世界”，责怪受害者

通过 Lerner (1980) 称之为“基本错觉”的过程，我们有时会把身处的社会简化为一个“公平的世界”。这种信念把人和结果之间简单地联系在一起：人们会得到他们应该得到的，而他们得到的也是原本该得的。假如一个陌生人发生了什么可怕的事情，我们会假定那是事出有因，这样就可以从心理上保护我们远离悲剧、免受相似的伤害。尽管这样做让我们感觉比较舒服，但其实质是错误地否定了生活中出现不可预测状况和坎坎坷坷的可能。和我们所想的相反，坏事也可能发生在好人身上，因为我们所处的环境并不是完全可以控制的。坚信公平世界的存在，会诱使我们产生一种安全和可控的错觉。

有时不幸的确发生在好人身上。当面临这样的事情时，我们会感到原有的“混乱中终有秩序存在，这是一个公平的世界”的信念受到了严重的威胁，并感到十分害怕。我们试图解决这一问题，但往往要付出高昂的代价。例如，坚信公平世界的存在会导致错误地责备受害者或者自我谴责。那些有过创伤经历的人，如强奸、乱伦的受害者，受虐者，严重的疾病患者，居丧者，意外伤害以及天灾的受害者等等，需要的是我们的同情和支持，而不是谴责和批判。虽然这些事件基本上是随机发生的，但这些无辜的受害者还是会让我们感到自己的脆弱。消除这种危险的不可控性的一个途径，就是把他们的不幸怪罪在他们自己身上，把那些灾难看成是他们应得的，以此让自己远离灾难。我们可以这样自我安慰：因为我们没有过错，所以灾难不会降临到我们头上。

在第 10 章中提到过的事后偏差，也是我们责怪受害者的一个原因。就如社会心理学家和创伤研究专家 Ronnie Janoff-Bulman (1992) 解释的那样：

有了后见之明，我们就可以煞有介事地因受害者受害之前的行为来责怪他们，一旦不幸发生了，我们就责怪受害者为什么没有想到事情的后果。在

一系列的研究中，我和我的学生都发现，被强奸的妇女往往因为她之前的行为而受到责备，而如果同样的行为后面跟随的是中性或正性的结果，就不会被认为是问题的或值得谴责的。但问题是，这些强奸的受害者在事前根本不知道这些行为的后果。一个男人在约会之后又来到一个女人的住处，这不足为奇，但如果这个女人被强奸了，她就会因为允许这个男人来到她家而受到谴责。因为后见之明，尤其是意识到存在牺牲和欺骗时，普通的行为也会突然变得有特别的含义。

你可能会想，不管怎么说，那些特别聪明、善良的人不至于如此愚蠢，因为受害者的不幸而责怪受害者。但令人吃惊的是，这种错误逻辑的发生非常频繁。请看下面这段让人感到非常不安的话：

Eileen Gardner 曾经在里根政府里做过一段时间教育部部长 William Bennett 的助手。她曾经针对残障人士写道，“以为变化难料的生活是随机地惩罚了他们，那就错了。事实并不是这样。事情发生总有其原因，在人生的某个发展阶段发生，有果必有因”。她还说，“表面上看，这是不公平的，但一个人的外部环境总是和他内在的精神发展水平相适应的”。这样的观点，在美国教育部这样的层次上无论如何是不应该存在的（Gilovich, 1991）。

自 责

当我们评价自己的不幸时，也会犯类似的推理错误。一旦我们成为不可预期事件的牺牲品，我们就倾向于责怪自己。尽管承认自己受害是应得的，会让人很痛苦，但同时也会为未来提供一种安全的感觉：如果我改变了自己，就不会再受害了。例如，强奸受害者可能因为受害而自责，从此小心谨慎地避免再次受害的情境和行为。尽管这种谨慎是合理的，但仍会产生一种虚假的安全感。重要的是，一定要认识到，就算我们的设想和意图再好，这个世界还是充满了不确定因素。如果我们认识不到这一点，单单改变自己的行为，从而产生新的强大与安全的感觉。结果，我们可能忽视了其他明智的防范措施，悲剧也因此可能重演。

Stanovich（1998）总结过 Dawes（1991）讲述的一个故事，故事描述的就是这种虚幻的经验学习可能导致问题的重复发生。下面是这个故事的一部分：

Dawes 所供职的大学，在监狱里搞了一个教育项目。Dawes 在监狱里遇

到了一个犯人，他是一个酒吧里的男招待，同时兼职帮人讨赌债（必要时要用枪）。一天，他被派到另一个州——俄勒冈州收一笔赌债。到了那里以后，他逮住了欠钱的人并用枪指着他。就在这时，警察突然出现，当场抓住了他。告诉了 Dawes 这个故事之后，这名犯人非常自信地说，他再也不会进监狱了。Dawes 问他，他怎么知道自己再也不会进监狱了。他说：“因为我再也不会踏足俄勒冈州了！”

习得性无助

控制错觉除了会导致责怪受害者、自责和虚假的安全感，还会导致一种心理学家称之为习得性无助（learned helplessness）的现象。在一个经典的实验中，Martin Seligman（1975）首先训练狗从一个笼子的一侧跳到另一侧，以逃避痛苦的电击。然后他装了一个栅栏，使狗不能跳开以逃避电击。在尝试了很多次以后，最终狗不再逃避电击了。这时，实验者拿掉了栅栏，结果令人吃惊：虽然狗现在完全可以逃跑，但它们大部分选择留在原地接受不必要的电击。它们已经认识到：它们是无助的。

控制错觉的内在危险是制造了一种虚假的有效性。当有控制错觉的人遭遇到应激或创伤事件，控制错觉的肥皂泡就会破灭。也就是说，他们会认识到自己的行动是无效的，因此进入了一种习得性无助的状态，就像 Seligman 实验中的狗一样。Seligman 发现，习得性无助与抑郁之间存在明显的相关。当一个人已经知道他的行动是无效的，就会产生无助感，这种无助感会引起极端的被动和冷漠，结果就算是有机会，这个人也不会再尝试更加有建设性和适应性的行为了。

心理意象的负面作用

伪科学，特别是那些打着“心身医学”和“整体健康”旗帜的伪科学，经常假定我们可以有意识地控制原本不可控制的过程。回忆一下前面提到过的那个直言不讳的伪科学家 Andrew Weil，他说心灵感应只是一种选择——如果我们选择了心灵感应，我们就会自动地感应到（详见第4章）。大部分的伪科学，都建立在神秘的“精神超越物质”的哲学之上。也许，相信精神可以治疗疾病的强大信念，主要指的就是这种心理意象的操作。

心理意象的支持者们宣称，我们仅仅借助想象就可以有效地对抗疾病。其原理是：我们可以通过正性思维来调动机体的自然恢复过程。不过，不仅没有任何可靠的证据可以支持这样的如意算盘，而且他们还忽略了一点，就是关于健康的思维也完全有可能对我们产生直接的、严重的负面影响。支持者明显是想要利用想象的好处——正性思维可以缓解疾病，却忽略了重要的危险性：负性思维也可以引起或加剧疾病的进程。Gilovich（1991）指出，我们既可以想到好的东西，也很容易想到不好的东西，接下来，他继续讨论了负性思维可能导致的危险：

认为人的免疫系统很容易受到一个人的想法和情绪的影响，这种观点存在很大的问题。更理想的情况是，不管一个人的心境如何，人的免疫系统都正常工作……这样，在看过一场悲剧电影，面对一群挑剔的观众进行演讲，或是得知自己的狗死掉了以后，才不会有特别的风险。就我个人来说，情愿相信不管我在想什么，都不会影响我的健康。事实上如果你像我一样，一想到自己所想的东西会影响健康状况，就会想到掉头发、心率失常、晚期癌症，那么，按照许多整体论者的观点，就会产生非常可怕的后果。这种现象和一个简单的思维实验的结果非常相似：当要求想象有人可以“读你的思维”或“听到你内心的对话”时，许多人报告说他们忍不住会想到最见不得人的冲动。同样，如果有可靠的结果证明我们的健康会受到所想内容的左右，我怀疑大多数人都会不由自主地想到生病或衰老。

持整体论的医疗保健人员，显然对心理意象的负面作用缺乏认识。他们片面地企图通过“水记忆”理论（参见第7章）来拯救同种疗法。试想，如果水真的有记忆，我们因此可以获得水接触过的所有物质的有效剂量的治疗，那将会发生多么可怕的事情。同样，再想一想如果真有心灵感应（如果别人能够读懂我们的思维，这是多么可怕），以及心理意象（如果我们想到一种疾病就会得这种病，那又是多么危险），世界将会怎样。持有魔力信念的人们常常有选择性地考虑事物的正性作用，而不切实际地对负面效果一概忽略不计（Radford, 2000）。

“选择健康”的谬误

许多所谓的心身医学或整体论疗法都基于这样一个观念：所有的患者都需要

主动地“选择健康”。虽然正性思维短期内有利于免疫系统的改善和主观痛苦感（病感）的缓解，但是还没有确切的证据表明正性思维可以治愈疾病。而且，这种谬论还有另外一个负面作用：它暗示那些患者，正是他们自己选择得病的，这也是伪科学竭力掩饰的“责怪受害者”的邪恶特征。

伪科学从业者所遵循的最基本的原则之一，就是必须考虑“什么样的人 would 得这种病，而不是这个病人得了什么样的病”，这一原则和责怪受害者的倾向直接相关。关心“什么样的人 would 得这种病”，暗示着病人的某些想法和做法是导致这种病的源泉。Gilovich（1991）为这种残忍的心态提供了一些例证：

- 一本颇有影响力的“整体护理”教科书的作者声明，“在人们没有充分发挥自己潜能的时候，疾病就会发生”。
- 新世纪信仰治疗师 Elizabeth Stratton 宣称，“疾病只不过是人们可能意识不到的深层心理问题的症状，……我要寻找的是，为什么人们会制造出疾病，以及为什么他们会让疾病持续下去”。

我们能够从这些论断中了解到些什么呢？那就让我们再来看看整体论吧。请设想有一个妇女得了乳腺癌。是什么样的人会得这种癌症呢？那些犯了错误的人，那些没有发挥出全部潜能的人；她一定有某种深层的心理问题，那才是她得乳腺癌的根源。也就是说，在治疗癌症之前，必须先治好她的心理问题。反过来再想想科学的医学又是怎样的观点：也许她遗传了家族的易感因素，也许她存在一些风险行为，不过大多数的疾病依然是不可预测地发生在个人的身上，因此这个妇女完全没有必要为自己的癌症而自责。这种观点消除了病人因为患病而受到的谴责，把注意力放在了治疗疾病这件事本身。从这样的观点出发，我们应该考虑的是，“这个妇女得了什么样的疾病”。这种观点不仅和循证医学相吻合，而且也更加人道。

现在我们的不难看出，关注“治疗整个人”会导致对患者的责怪；同时又很难看出，这种思维方式会有什么样的帮助或益处。更有甚者，一些病人不仅不会对失败的治疗说“不”，而且还会在治疗师的引导下学会自我责备。产生这种现象的关键就在于这些治疗师认为，他们可以对根本无法控制的疾病进行个人控制。Gilovich（1991）引用了心理意象治疗癌症的先锋人物——Carl Simonton 和 Stephanie Simonton 的一些令人十分难过的案例，而 Simonton 夫妇却似乎没有因那些将要死去的病人所采取的自责态度感到任何不安：

一些早期的患者认为我们给了他们治愈癌症的钥匙，并对自己说：“对，我可以做到！”随后，如果他们沒有好起来，他们就会感到自责和内疚……最终，他们的家人会给我们带来患者的临终遗言：“告诉 Carl 和 Stephanie，他们的方法还有效。”或者是，“告诉 Carl 和 Stephanie，这不是他们的错。”（Simonton, Matthews-Simonton, & Greighton, 1978）。

听了这些话以后，很难想象是不是还有更加让人难过的临终遗言了。科学的医学治疗方法至少可以让患者减轻症状，为他们赢得一段宝贵的时间。而放弃医学科学，去尝试一种根本没有任何证据可证明其有效性的心理意象疗法，实在是自欺欺人。这种治疗方法注定是会失败的，于是患者转而责怪自己。这些患者临终时说的那些话，表面上是对毫无作用的心理意象疗法的支持，实际上表达的是人们对于获得控制感的渴望。

祈祷和宗教信仰下的医疗忽视

也许在所有的医疗悲剧中，儿童遭受的因宗教信仰所导致的医疗忽视最让人痛心。造成这种医学愚昧的是一种执著的信仰，尽管它完全缺乏证据。信仰者坚信，虔诚的宗教信仰，特别是祈祷，足以让人保持最佳的健康状态。当一个宗教团体宣称祈祷可以治愈疾病时，相当于提出了一个可以用科学方法来检验的经验性观点。那么我们现在就开始吧，看看这种观点到底有多少道理：

有那么一类祈祷，祈求上帝干预人类的历史，纠正一些真实的或想象的不公平的事，或者减少自然灾害。例如，美国西部的一个主教向上帝祈祷，请求结束一场灾难性的旱灾。为什么主教要做这种祈祷？上帝难道不知道有旱灾发生吗？上帝难道不知道主教的教徒们所面临的威胁吗？无所不能、无所不知的上帝也有不知道的吗？这究竟意味着什么？主教还要求他的教徒也进行祈祷。那么，和只有少数人祈祷相比，上帝是不是有可能因为有更多的人祈求，就赐予怜悯和公平呢？（Sagan, 1995）

在我看来，一个会计算祈祷人数有多少的上帝，不免有点儿失身份，就像他那儿是一个大型的投票站一样。当考虑到下面这些影响深远的问题时，就更令人担心了：全世界毕竟存在那么多信仰，每种都不一样，而且还常常互不相容，那

么相应产生的问题又要如何解决呢？

- 死后的世界（例如，有的认为死后会升入天堂或罚入地狱；有的则信奉转世投胎）
- 世界的起源（例如，有人相信是造物主的创造；而中国人、芬兰人、希腊人、印度人、日本人、波斯人和萨摩亚人所信仰的是蛋形混沌宇宙的杰作）
- 历史事件的记载和解释（例如，基于某个流域的河水泛滥或者山洪暴发的故事，如诺亚舟的故事所塑造的文化，与其他文化的比较）
- 道德和社会禁忌（例如，有的主张一夫一妻制；有的主张一夫多妻制）
- 宗教信仰的传统仪式（例如，在仪式上需不需要献祭牛羊等动物的差异）
- 一神还是多神（例如，犹太—基督教信仰一神论；有的信仰众神论）

总之，应该如何祈祷才是正确的方式呢？什么是可以接受的，什么又是必须禁止的呢？祈祷者究竟应该向谁祈祷呢？除此以外，还有一大堆各种各样的难题。这些难题并不是在玩文字游戏，而是针对祈祷有效论的内部逻辑所提出的实实在在的 challenge。重要的是，祈祷的所谓治疗效果到底有什么样的实证性证据呢？Sagan（1995）指出：

维多利亚时期的统计学家 Francis Galton 曾指出，如果所有其他条件都是相同的，那么英国的君主应该非常长寿，因为每天全世界有无数人在诚心诚意地吟诵“上帝佑我女皇（或君主）”。但事实上，他们并没有比其他富有、奢侈的贵族活得长久。在古埃及，几乎每个人也都在向上帝祈祷让法老“生命永驻”，但很明显，这些集体祈祷都失败了。而这些失败就是实实在在的数据。

关于祈祷对健康的效果，究竟有没有系统的研究呢？当然有。例如在附录“评估信息来源的质量”中我们会讲到，Harris 等人（1999）对 990 个心血管重症监护室的住院病人研究发现，祈祷的结果甚至比单纯的随机机会率还要低。祈祷只是提高了患者对“Swan-Ganz 导管*”的评价，而对其他的研究结果变量，如肺炎、大手术、心跳骤停和死亡的发生率，则毫无作用。这项严格控制的实验，为祈祷作用的研究提供了最可靠的结果。当祈祷有效论的支持者还企图引证

* Swan-Ganz 导管是一种医用心内导管。——译者注

此研究作为他们的支持证据时，恰恰说明了他们的论点是多么不堪一击。

虽然并没有证据证明祈祷是有效的，但很多宗教团体依然选择祈祷来治疗儿童的疾病，而不是选择科学的医疗方法。Asser 和 Swan (1998) 研究过 172 例儿童的死亡，他们的父母以与他们的宗教信仰冲突为由，拒绝接受医学治疗。尤其是信奉正统基督教的教派成员，更崇拜信仰治疗，拒绝医学治疗。其中，基督教科学派和卫礼教派的儿童，其死亡率显著高于研究中的其他群体。研究者们还对儿童的死因进行了详细的回顾性调查，最后的结论是：172 名儿童中有 140 名（占 81%），如果接受医学治疗的话，至少可达 90% 的存活率，另外，还有 18 名儿童如果接受治疗，可以有 50% 的存活率。这些悲剧之所以会发生，是因为一些州的法律规定，可以因为家长的宗教信仰，而免除医疗忽视致死的法律责任。美国国家反健康欺诈委员会 1998 年 3/4 月的一篇通讯曾指出：

很奇怪，尽管美国人非常同情儿童的遭遇，但宗教信仰下的医疗忽视导致儿童死亡的案例，却极少受到媒体的关注，即使在医学界也鲜有报道。

Asser 和 Swan 的研究表明，由于宗教信仰下的医疗忽视引起的儿童死亡，不仅是可以避免的，而且还是可预期的。但如果善良的人们不采取措施去制止这种行为的话，这类悲剧将持续发生下去。事实上，美国最高法院就是否强制父母接受对子女的医学治疗有明确的立法条文，条文如下：

“宗教信仰自由的权力并不包括将群众或儿童暴露于传染性疾病之中，或者威胁儿童健康及生命的自由……父母自己有成为殉教者的自由，但即使在同样的情境下，他们也没有让未达到法定成年年龄的子女成为殉教者的自由。”（Prince v. Massachusetts, 321 U. S. 158, 1944）

也许，比起那些受到伪科学毒害而深深自责的受害者，命运更为悲惨的就是那些由于父母愚昧的宗教信仰而身受其害的儿童，而那些父母总是极其固执地坚信控制错觉。正如我在整本书中不断强调的，对于伪科学的毒害，最佳的解毒剂就是根据实验性的证据形成信念，并把判断和决策过程置于科学推理的基础之上。而摆脱错觉最可靠的方法就是保持一种开放、怀疑的态度。

第四篇

决策与伦理学问题

13. 评估：经典决策理论

请看图 4 中三个症状自陈量表的摘选部分，这些内容都来自于常用的症状自陈量表，你是否会将所描述的特征与自己联系起来呢？如果是，你认为意味着什么？大部分特征看起来模糊不清，让人难以确定“是”或“否”的答案是否意味着有问题存在。许多题目还同时包括了症状谱的两个极端，意味着其中任何一端都是有问题的。此外，这些内容所描述的大部分特征都很常见，几乎每个人都能发现符合自己的内容。产生了以上种种疑问之后，你可能会对这些量表原本的编制用途感到很奇怪，因为它们的编制者认为：如果你发现自己有与这些表述相似的感觉，就意味着你可能压抑了在童年曾受过性虐待的记忆，而否认这种压抑是没有意义的。“如果在你的身上存在一些预警信号，那么你可能是真的有被压抑的记忆。无论你是否否认，都逃不出这一规律。”（Fredrickson, 1992）。压抑记忆运动的两位领军人物，Ellen Bass 和 Laura Davis（1988），在《治疗的勇气》（*The Courage to Heal*）一书中补充说：“如果你认为你被虐待过，而且你的生活中出现了这些症状，那么你可能真的被虐待过。”

其他研究儿童性虐待记忆的治疗师对于自己的观察能力也极其自信，即使是在患者坚持认为没有发生过这种事件的情况下也是如此：

治疗师经常说，他们具有在当事人不知情的情况下发现虐待记忆的能力，并将这作为证明其具有最高智慧水平的直觉和诊断能力的证据。例如，一些治疗师吹嘘自己可以在治疗伊始就区分出患者是否有隐瞒的被虐待史。“通常在人们进门后不到 10 分钟，早在他们自己意识到之前，我就可以发现这些”，这是家庭咨询师 Brenda Wade 在美国国家广播有线电视公司的节目

“真实人性”中谈及她能够识别有童年虐待史的受害者时的说法，“缺乏信任……必定有某种途径可以让人们表现自我，必然有某种身体语言用于表达”。(Ofshe & Watters, 1996)*

1. 你是否具有下述的许多特征？(Blume, 1990)
 - 害怕在黑暗中独处……做噩梦，害怕黑夜
 - 身体有异样感……身体意象差
 - 有胃肠道不适、妇科疾病、头痛、关节炎或关节疼痛
 - 饮食障碍，药物或酒精滥用（或绝对禁戒）
 - 恐惧
 - 有自杀念头、自杀企图或强迫观念
 - 抑郁；无缘无故哭泣
 - 有高度冒险的行为（“挑战命运”）；不敢冒险
 - 罪恶感、羞耻感、低自尊、无价值感；对别人的小恩小惠感激涕零
 - 被抛弃感
 - 遗忘某些早年时期的记忆（尤其是1~12岁的），或者忘记特定的人或地点
 - 体验到疯狂；体验到差异；感觉自己是不真实的而其他所有人都是真实的，或者感觉其他所有人都是不真实的，而自己才是真实的；创造出虚幻的世界、人际关系或身份
 - 性问题
2. 对于下列“预警信号”你是否感到熟悉？(Bass & Davis, 1988)
 - 你是否觉得自己与众不同？
 - 你害怕成功吗？
 - 你觉得很依赖自己在乎的人吗？
 - 你是否感觉与人疏远或是孤独？
 - 你是否感觉难以被激励？
 - 你是否很难相信自己的直觉？
3. 这些描述是否符合你的真实情况？(Fredrickson, 1992)
 - 我容易受惊吓
 - 我过度地做一些事情
 - 我的大脑常充斥着性意念
 - 我经常发呆、开小差或做白日梦
 - 我不敢去看牙科
 - 我经常做噩梦
 - 某些特定的食物或味道令我感到害怕或觉得恶心
 - 我有厄运当头的感觉，就像我的人生将在悲剧或灾难中结束

图4 三个常用症状自陈量表的摘选部分

* 摘自《制造怪物：错误记忆、心理治疗和性癡症》(Making Monsters: False Memories, Psychotherapy, and Sexual Hysteria), Richard Ofshe 和 Ethan Watters 著。

对上述症状评估的实用性提出质疑是明智之举，但我们还需要回过头来再讨论“压抑的记忆”这一主题。首先，我们将以一个每天都会被使用的决策过程为例，来看看个体如何评估这一决策方法的准确性。

多通道记录仪测试

正如我们在第 10 章中所提到的，在犯罪调查中采用控制问题技术（CQT）进行多通道记录仪或“测谎仪”测试，其正确识别说谎行为的敏感度约为 85%，特异度约为 60%。也就是说，在所有确实有罪的嫌疑人中，CQT 可以正确探知 85% 的说谎行为，但会遗漏 15% 的说谎行为；而在实际上并没有犯罪的嫌疑人中，CQT 可以正确判断其诚实的概率为 60%，而将其错误判断为说谎的概率是 40%。假设在所有接受审问的犯罪嫌疑人中，每五个人就有一个人是真正的罪犯（20%），将多通道记录仪测试应用于该人群，并采用“测试结果为说谎就意味着有罪，测试结果为没有说谎（诚实）即意味着清白”的判断规则，那么 CQT 技术能否为嫌疑人犯罪与否提供强有力的证据呢？要正确解释该测试的结果，你首先需要了解两个彼此密切相关的问题：

1. 如果多通道记录仪测试结果显示某位嫌疑人说谎，那么这位嫌疑人真的有罪的概率是多少？
2. 如果多通道记录仪测试结果显示嫌疑人没有说谎（他是诚实的），那么这位嫌疑人真的无罪的概率是多少？

花一点时间思考这两个问题，请明确你的最佳猜想答案以后再继续阅读下去。

如果你的反应与大部分人一样，那么你可能会在估计这些概率时犯下一些预测性的错误。首先，这两个问题的答案是不一样的。许多人认为，诊断性测试（例如 CQT 多通道记录仪检测）有单一的整体效度，该效度可以决定所有测试结果的可靠性，而实际上，需要考虑多种因素，每个因素都可能导致这两个问题的答案不同。其次，不仅上述两个问题的答案存在差异，而且每个答案都会比许多人所想象的更极端——“说谎”的结果能正确识别罪犯的概率只有 35%，而“诚实”的结果能正确识别无罪者的概率则为 94%。

要解答上述两个问题，需要将基本比率（20% 的嫌疑人是有罪的）、敏感度（85% 的有罪者会得到“说谎”的测试结果）和特异度（60% 的无罪者会得到

“诚实”测试结果) 综合起来考虑。幸运的是, 的确有这么一种直接的方法可以整合这些信息, 回答这两个问题。

经典决策理论和频率树

当得出测试结果时, 存在四种可能。首先, 结果可能是阳性 (显示说谎) 或阴性 (显示诚实)。请时刻注意: 词汇“阳性”和“阴性”并不是有价值的判断结果, 它们只是可以方便地被用以描述不同结果的标签而已。无论是阳性还是阴性, 都不存在对和错, 也不意味着应该如何回答问题, “阳性”仅仅意味着该测试所要探测的状况 (这里指“说谎”) 被探测到了, “阴性”则意味着这种状况没有出现, 而结果可能是“真” (正确), 也可能是“假” (不正确)。将阳性/阴性和真/假组合起来, 就可以得到四种可能结果: 当结果是“说谎”时, 这一阳性测试结果可能是正确的 (真阳性, true positive, 简称为 TP; 即嫌疑人确实是有罪的), 也可能是不正确的 (假阳性, false positive, 简称为 FP; 即嫌疑人实际上是无辜的)。当结果是“诚实”时, 这一阴性测试结果可能是正确的 (真阴性, true negative, 简称为 TN; 即嫌疑人确实是无辜的), 也可能是不正确的 (假阴性, false negative, 简称为 FN; 即嫌疑人实际上是有罪的)。之前所提的问题 1 是指在所有阳性结果中判断正确的比例, 也就是测试的阳性预测效力: $TP / (TP + FP)$; 而问题 2 是指所有阴性结果中判断正确的比例, 即测试的阴性预测效力: $TN / (TN + FN)$ 。

我们在这里介绍了许多新名词, 让我们来看看该如何应用它们。相对简单的一个技术就是构建一个频率树, 也就是根据假定人群的期望频率将人群划分为亚群的一种简图 (Sedlmeier & Gigerenzer, 2001)。研究的人群的大小并不会影响结果, 因为百分比是在最终阶段计算的。因此假设我们研究的是包含 1000 名犯罪嫌疑人的人群, 这一总人数构成了频率树的最高层次, 下面还要将其再划分为亚群 (见图 5)。

将全部嫌疑人划分为有罪和无罪, 就建立了频率树的第二层次。我们怎么知道分别有多少嫌疑人各属于这两个亚群呢? 根据该人群的基本比率, 每五个嫌疑人就有一人 (20%) 是有罪的。因此有罪者共有 200 个 ($1000 \times 0.20 = 200$), 剩下的 800 人为无罪 ($1000 - 200 = 800$), 这两群人构成了频率树图的第二层次。

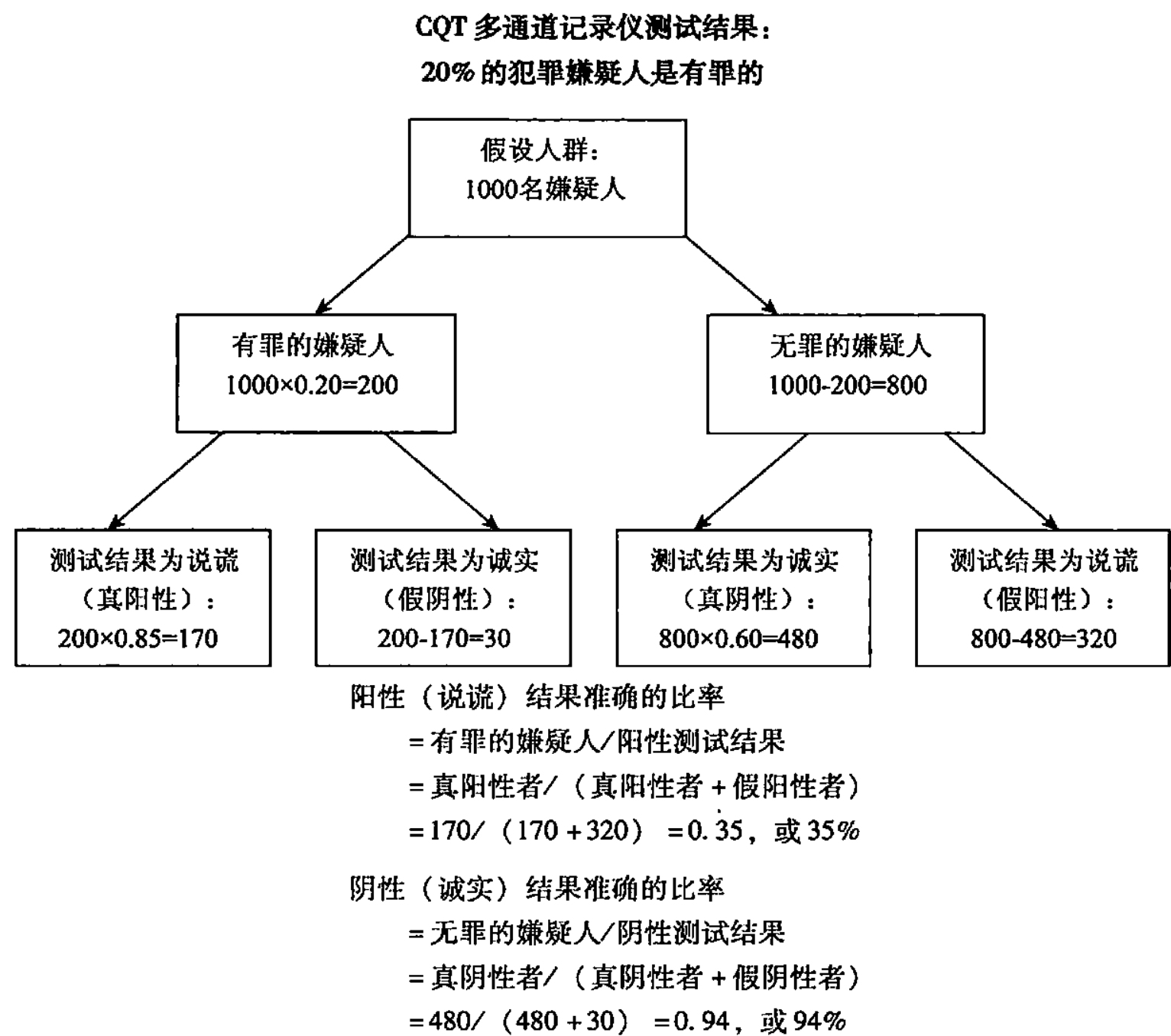


图5 对于 1000 名犯罪嫌疑人（其中 20% 有罪），CQT 多通道记录仪测试结果的准确性

分别再将这两个亚群（有罪和无罪）划分为测试阳性组（说谎）和测试阴性组（诚实），这就建立了频率树的第三层次，也是最后一个层次。我们怎样才能知道每种测试类型各包含多少嫌疑人呢？这里就需要用到敏感度和特异度了。敏感度指的是测试结果呈阳性的有罪嫌疑人占全部有罪嫌疑人的百分比，在这里是 85%，因此在 200 名有罪嫌疑人中，有 $200 \times 0.85 = 170$ 人的测试结果会是阳性（TPs），而其他 $200 - 170 = 30$ 人的测试结果会是阴性（FNs）。特异度指的是测试结果呈阴性的无罪嫌疑人占所有无罪嫌疑人的百分比，在这里特异度是 60%，因此在 800 名无罪嫌疑人中，有 $800 \times 0.60 = 480$ 人的测试结果会是真阴性（TNs），而其他 $800 - 480 = 320$ 人的测试结果会是假阳性（FPs）。这四个类型就构成了频率树的第三层次。

在完成频率树后，再来回答我们前面所提的问题就相对简单了。首先，如果 CQT 结果为阳性（显示说谎），嫌疑人确实有罪的概率是多少？将正确的阳性测试结果数目（TPs：170）除以阳性测试结果的总数目（TPs + FPs = 170 + 320 = 490）得出： $170/490 = 0.35$ ，或 35%，这就是 CQT 多通道记录仪在犯罪嫌疑人中的阳性预测效力。也就是说，当该测试显示说谎并据此做出该嫌疑人有罪的判断时，这一判断只有 35% 的可能是正确的，同时有 65% 的可能是错误的。

其次，如果 CQT 结果是阴性（显示诚实），嫌疑人确实无罪的概率是多少？将正确的阴性测试结果数目（TNs：480）除以阴性测试结果的总数目（TNs + FNs = 480 + 30 = 510）即可回答这个问题： $480/510 = 0.94$ ，或 94%。这就是 CQT 多通道记录仪在犯罪嫌疑人中的阴性预测效力：当测试显示诚实并据此做出该嫌疑人无罪的判断时，这一判断正确的可能性是 94%，而错误的可能性是 6%。

我们已经使用频率树技术演示了一个案例，接下来，我们将探讨对于任一决策过程来说都存在的、影响阳性和阴性预测效力的三个因素。

基本比率：试图探测的结果有多罕见

先前之所以得到令人诧异的结果是因为有罪的基本比率非常低（20%）。一般来说，当相应人群的基本比率接近 50% 时，测试结果更有意义（Meehl & Tosen, 1955）。因此，当基本比率较低时（即很少一部分嫌疑人是真正有罪的），需要特别强有力的证据将原本很低的有罪概率（20%）提高到较高水平，再进行相应的测试。也就是说，一个测试必须有极高的效度才能比较有把握成功地探测出罕见的事件。这一点也反映在 CQT 多通道记录仪的阳性预测效力较低（35%）的方面。因为许多重要的预测都涉及一些罕见事件（例如暴力行为、自杀意图，或者患病），所以这里所讨论的决策时应注意的问题，可以适用于很多重要情境。

这个例子中相对较高的阴性预测效力表明，相比较而言，普通事件可以更好地被预测。在上述由犯罪嫌疑人组成的人群中，大多数人是无辜的，因此阴性（诚实）测试结果很容易从原本就高的无罪概率（80%）上升到更高的水平（94%）。一条普遍的原则就是，当目标状况的基本比率降低时，阳性预测效力会减弱，而阴性预测效力则有所加强。

为了更好地说明基本比率对预测效力的影响，我们假设将 CQT 多通道记录仪检测应用于有一半人确实有罪的犯罪嫌疑人群体中，即这个群体犯罪的基本比率是 50%，而且假设这一群体也共有 1000 名犯罪嫌疑人。我们重新构建该群体的频率树，以进一步分析基本比率是如何影响预测效力的。在 1000 位嫌疑人中，有 500 人（50%）是有罪的，其余 500 人是无辜的。在真实有罪的嫌疑人中，多通道记录仪结果显示说谎的有 500×0.85 （敏感度）= 425 位（TPs），而结果显示诚实的则有 $500 - 425 = 75$ 位（FNs）。在真实无辜的嫌疑人中，多通道记录仪结果显示诚实的有 500×0.60 （特异度）= 300 位（TNs），而显示说谎的则有 $500 - 300 = 200$ 位（FPs）。图 6 是相应的频率树。

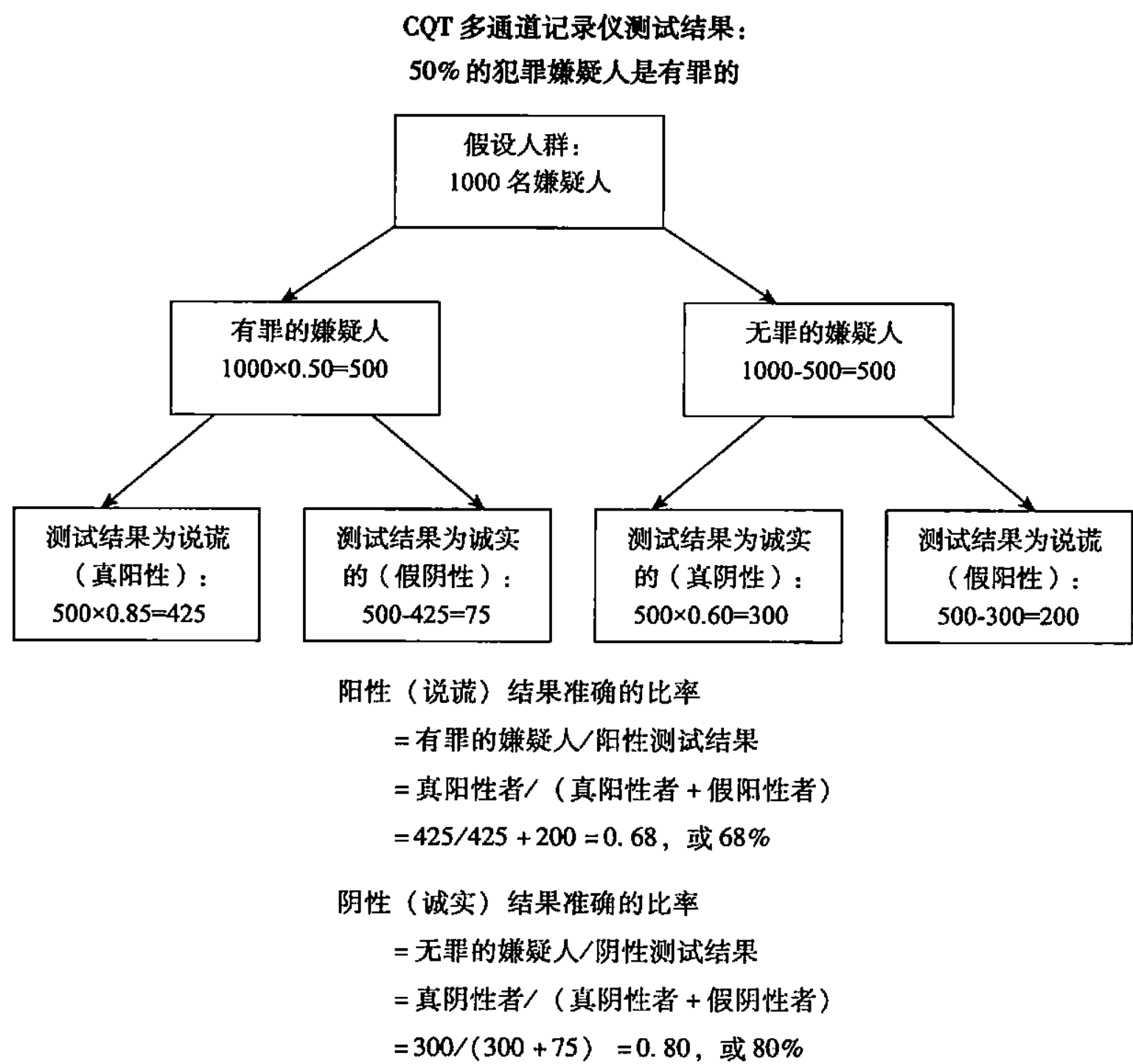


图 6 对于 1000 名犯罪嫌疑人（其中 50% 有罪），CQT 多通道记录仪测试结果的准确性

根据这些，我们可以重新计算阳性和阴性预测效力。首先，阳性预测效力等于正确的阳性（说谎）测试结果（TPs）占有所有阳性测试结果（TPs + FPs）的百分比： $425 / (425 + 200) = 0.68$ （68%），与犯罪基本比率为20%时的阳性预测效力35%相比，有所提高。也就是说，当犯罪概率增加时（新的基本比率为50%），阳性测试结果的准确度也相应提高。阴性预测效力等于正确的阴性（诚实）测试结果（TNs）占有所有阴性测试结果（TNs + FNs）的百分比： $300 / (300 + 75) = 0.80$ （80%），与犯罪基本比率为20%时的阴性预测效力94%相比，有所降低。这一结果与设想是一致的：犯罪基本比率的升高会提高阳性预测效力，而降低阴性预测效力，这是基本比率改变时必然出现的交互作用。

效度：证据的强度如何

除了基本比率之外，效度是影响所有决策过程的预测效力的另一个重要因素。毫无疑问，敏感度或特异度提高了，预测效力也会随之提高。为了深入探讨这个问题，我们先来了解一下多通道记录仪的一种相当特殊的形式——犯罪知识测试（guilty knowledge test, GKT）（Lykken, 1998），该方法也用于调查特殊的违法行为。

CQT 技术尝试应用目标问题和“控制”问题（control question）来区分说谎与诚实，而 GKT 技术则是试图明确被试是否拥有那些只有犯罪个体才会掌握的知识。GKT 模式包含一整套源于真实案例的多项选择题，一般每道题有六个选项，正确答案随机分布于后五个选项中，第一个选项则不记分，目的是消除首位效应（primacy effects）。GKT 测试并不要求被试选择正确答案——犯罪嫌疑人通常也都是假装自己不知道正确答案——只须大声、反复地读出每个选项，同时用标准的多通道记录仪设备同步记录个体的各项生理指标，像 CQT 测试一样。重复大声读问题有助于防止犯罪嫌疑人使用反测谎对策。因为对被试来说，无论在何种情况下，要有意地提高自身对所有错误选项的生理反应，或压抑对正确选项的生理反应，都是很难做到的；而要频繁地控制自身的生理唤醒度且不被测试者所察觉，那更是极其困难了。

根据 CQT 技术的理论基础，有某些特定的生理指标可以明确地显示存在说谎行为，只有犯罪个体会对目标问题表现出比控制问题更多的生理唤醒，然而这一点并不是那么确定。GKT 应答的理论基础则是人们熟知且得到公认的朝向反应

(orienting response) 原则，即个体会对熟悉的刺激物自动产生应答，而且伴随相应的生理反应 (Lynn, 1966)。关于朝向反应最典型的例子就是鸡尾酒会效应 (cocktail party effect)：即使在人很多的嘈杂房间里，只要别人谈话中提到了你的名字，你也很容易会注意到。这种朝向反应是很难被抑制的，因为它在很大程度上不受、或者完全不受意识的支配。因此，如果一个被试持续地对 GKT 中的正确选项有最明显的生理反应，则提示该被试拥有相应的“犯罪知识”。从另一方面来说，如果对正确选项显示最明显反应的频率与其他错误选项类似，约为五分之一（因为有五个记分选项），则表明被试并不具备相应的犯罪知识。

GKT 多通道记录仪测试的效度很大程度上取决于 GKT 问卷中好问题的多少。所谓好问题，是指该问题的正确选项仅仅只有那些确实有罪的嫌疑人才知道（即在犯罪场景中会被注意并被记住的某个突出方面），而对无罪的嫌疑人来说，所有选项的真实度看上去是差不多的。一个 GKT 测试中包含的好问题越多，该测试就越有效，这和包含合适题目较多的学习测验其有效性高于合适题目较少的测验是同样的道理：增加测验中有效问题的数量可以消除运气等因素对测试结果的影响。Lykken (1998) 展示了一个有 10 个题目的 GKT 样本，这是在著名的橄榄球明星辛普森 (O. J. Simpson) 杀妻案中所用的，目的是判断辛普森是否掌握了 Nicole Brown Simpson 和 Ron Goldman 被谋杀一案的犯罪知识。下面是其中的两个问题：

- 当人们发现 Nicole 时，她穿了一件以单色调为主的衣服。她穿的衣服是什么颜色？白色？黑色？粉红色？蓝色？绿色？褐色？
- Goldman 被杀时，手里拿着一样东西，后来人们在他的尸体附近发现了这样东西。他拿的是什么呢？一个购物袋？一个密封的白信封？一个灰色的皮革钥匙包？一个无线电话？一个酒瓶？一个黑色的小通讯簿？

GKT 测试应该在媒体对犯罪现场的很多细节进行报道之前使用，这一点十分重要。不过可能媒体也没兴趣对这类细节进行详细报道。对于是否每次都能编制出一个好的 GKT，研究者们和执法官员之间总是存在争论。就像搜集指纹证据或 DNA 样本需要仔细地寻找、取证一样，详细观察、记录犯罪现场的相关细节并进行存档，通常会十分有助于编制有价值的 GKT 题目。试想，如果根据某案件，我们可以编出 10 个好的 GKT 题目（“好”题目的标准是：确实有罪的嫌疑人中有 80% 的个体对正确选项反应最强烈，而在无罪的嫌疑人中只有 20% 的个体对正确选项反应最强烈），那么这个测试的敏感度就在 97% 左右，特异度则高达

99%。这一效度明显高于 CQT 的效度（敏感度 = 85%，特异度 = 60%）。现在让我们再用最初假设的犯罪嫌疑人群（其中 20% 为有罪）为例，来比较一下 CQT 技术与 GKT 技术的效度。

还是假设有一个包含 1000 名犯罪嫌疑人的群体，并将其划分为有罪（ $1000 \times 0.20 = 200$ ）与无罪（ $1000 - 200 = 800$ ）两个群体。根据 GKT 的效度再进一步划分这两个群体。在有罪者中， $200 \times 0.97 = 194$ 人被判定为有犯罪知识（TPs），剩下的 $200 - 194 = 6$ 人被判定为无犯罪知识（FNs）。在无罪者中， $800 \times 0.99 = 792$ 人被判定为无犯罪知识（TNs），剩下的 $800 - 792 = 8$ 人被判定为有犯罪知识（FPs）。将这些数值加到图 7 的频率树中，再计算阳性预测效力为 $194 / (194 + 8) = 0.96$ ；阴性预测效力为 $792 / (792 + 6) = 0.99$ 。GKT 在阳性和阴性预测效力两方面均较 CQT 有很明显的提高。因此，除了基本比率与决策过程的实际效力存在相关，而且还十分重要之外，决策方法的效度也是一个非常关键的因素。

阈值：用于决策的分界线要定在哪里

第三个影响阳性和阴性预测效力的因素是用于做出决策的阈值。到目前为止我们所讨论的，无论是 CQT 还是 GKT 的测试结果，都是以这样一种形式出现：最终结局为两种结果中的某一个，CQT 是说谎或诚实，GKT 则为有犯罪知识或无犯罪知识。而实际上，事情远比这种情况复杂得多，这两种技术都会得到一个数据连续谱，而结果可能落在这一连续谱上的任一点。就 CQT 而言，将目标问题及控制问题加以配对并进行评分，结果可以说是说谎、诚实或两者均不是（答案不确定）。如果测试的结果表明说谎或诚实的可能性更大，则要再根据一个 3 级评分量表对已有证据在多大程度上支持测试的推断进行评定。因此，对每一对问题而言，CQT 多通道记录表格的得分可以是 -3 到 +3 之间的任何数值。将所有问题的分数加起来就可以得到整个多通道记录测验的总分。因为在典型的 CQT 检测过程中，有许多对这样的目标/控制问题，且每一系列的问题都要被重复询问多次，所以总分的变化范围很大。我们要利用总分下结论的时候，就需要有对应的阈值：若总分大于阈值可判定为说谎，若小于阈值则可判定为诚实。例如当阈值为 0 时，所有大于 0 的分数被视为说谎的证据，而所有等于或小于 0 的分数则被视为诚实的证据。

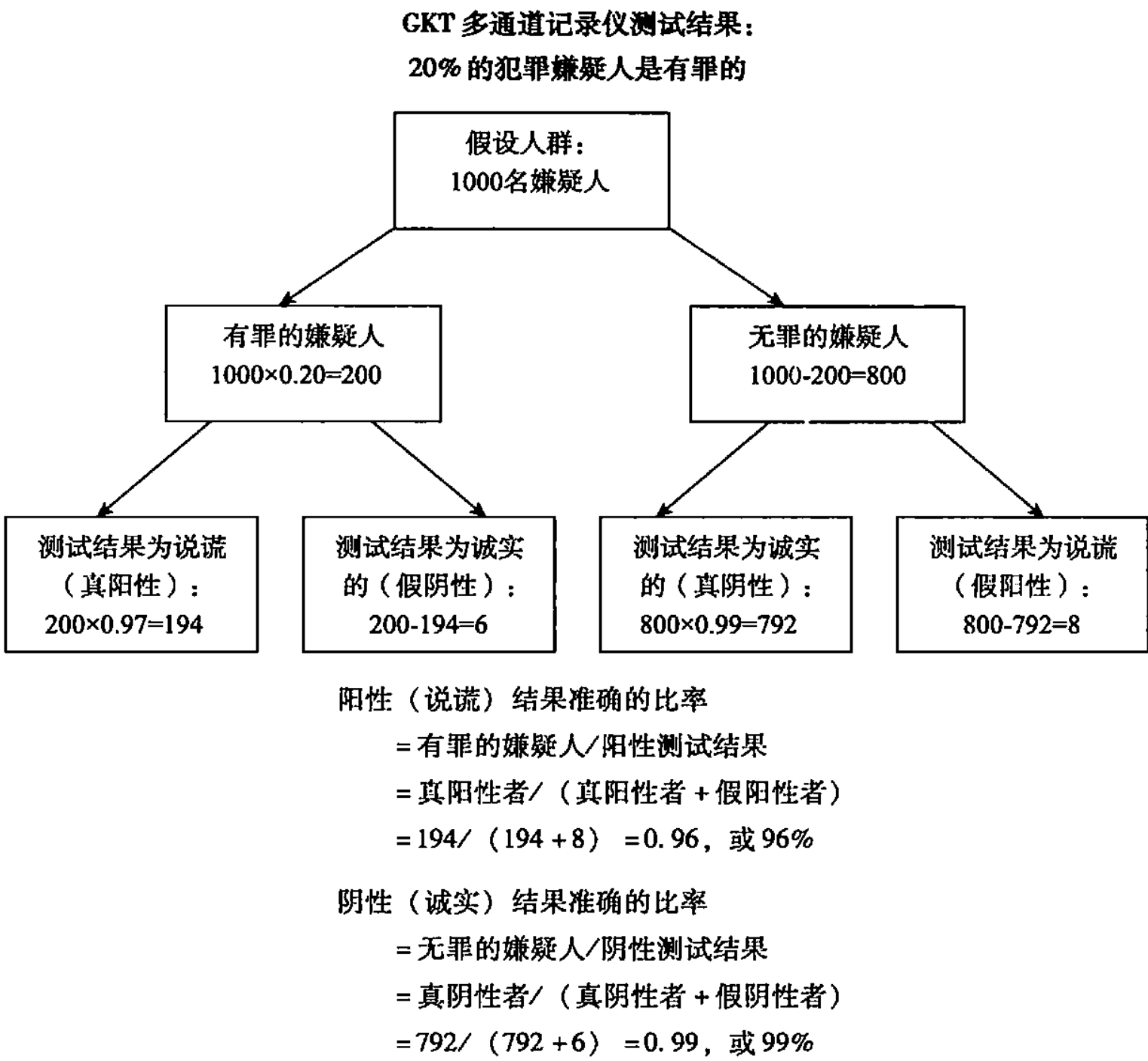


图7 对于 1000 名犯罪嫌疑人（其中 20% 有罪），GKT 多通道记录仪测试结果的准确性

同样，GKT 的结果也是得到一个总分的分布范围。GKT 有不同的记分方法，其中一个简便的方法就是计算被试对正确选项有最强生理反应的项目数。如果 GKT 有 10 个题目，则被试的总分会在 0 至 10 之间。因此在这里也必须设定阈值，并根据阈值来判断谁有犯罪知识或谁没有犯罪知识。假设阈值是 5 分，那么总分的数值大于 5 分就认为该嫌疑人拥有相应的犯罪知识，而数值等于或小于 5 分则认为该嫌疑人没有犯罪知识。

现在你可能已经意识到了，设定阈值可以从根本上影响预测效力。如果将多通道记录仪技术的阈值设置得很高，那么只有很少的犯罪嫌疑人得到的分数会高于阈值而被怀疑有罪。也就是说，假阳性错误或 FP（无罪的嫌疑人被判为说谎）

值很低，而假阴性错误或 FN（有罪的嫌疑人被判为诚实）值则很高。而如果将阈值降得很低，那么只有很少的嫌疑人得分会低于阈值而被判为无罪。也就是说，假阴性错误或 FN 值将会很低，而假阳性错误或 FP 值则会很高。因此，高阈值会提高阳性预测效力并降低阴性预测效力，低阈值则有相反的作用。如果想要平衡这两种类型的错误，那么将阈值设定在合适的水平是很关键的。

实验室效度与领域效用的比较

许多人无法正确理解本章所论述的决策理论的基本原则，其中还包括很多医生（例如，Casscells, Schoenberger, & Grayboys, 1978; Eddy, 1982）。这里最重要的一个原则就是：在实际应用中决策过程的准确度（预测效力，或在该领域的实用性）不等于实验室研究中所测量到的效度（敏感度和特异度）。例如，多通道记录仪技术的敏感度其意义在于一个确实有罪的嫌疑人测试结果为阳性的可能性是多少，但这并不能回答警方调查者在实际工作中提出的问题，因为警方并不会这么问：“如果这个嫌疑人是有罪的，他在测试中得到‘说谎’结果的概率是多少？”如果已经知道某嫌疑人是有罪的（或无罪的），那就没有必要进行多通道记录仪测试了。实际上警方调查者想知道的是，在还不清楚犯罪嫌疑人究竟是有罪还是无罪的情况下，多通道记录仪测试结果正确的可能性有多大，也就是预测效力。阳性预测效力可以告诉调查者，获得“说谎”测试结果的个体有多大可能是真正有罪的；而阴性预测效力则告诉调查者，获得“诚实”测试结果的个体有多大可能是真正无罪的。

那么，这是否意味着敏感度、特异度与预测效力之间没什么关系呢？事实并非如此。敏感度和特异度是明确预测效力的必要组成部分，而且它们是建立在细致的实验室研究基础之上的。为了弄清敏感度和特异度的大小，在评价新的测验或技术时，一开始必须使用已知群体进行测试，计算该测验能在多大程度上准确地探测出已知的事实。例如要评价一个新的 HIV 血液检测技术的效度，就分别将该测试应用于两组人群，一组为 HIV 阳性者，一组为 HIV 阴性者，然后就可以计算出这个新技术的效度。检测结果为阳性的 HIV 阳性者占有所有 HIV 阳性个体的比例就是敏感度，而检测结果为阴性的 HIV 阴性者占有所有 HIV 阴性个体的比例就是特异度。

对于一项新的 HIV 检测技术来说，通过研究获得的效度估计值，是其在实际应用过程中预测效力大小的一个重要影响因素。我们还需要了解该检测技术的使用者是否采用了研究者所采用的阈值，因为设定更高或更低的阈值都将改变 FP 与 FN 之间的平衡，进而影响阳性和阴性预测效力。此外，我们必须知道 HIV 阳性在目标人群中的基本比率是多少。例如将此项检测技术用于全美大学生群体的 HIV 筛查，因为该群体的 HIV 发生率极低，即使该检测技术十分有效（敏感性与特异度高），其阳性预测效力也可能很低。但如果将该检测技术应用于美沙酮*维持治疗诊所，由于海洛因成瘾而进行脱毒治疗的人群常同时伴有 HIV 感染，因此在这类人群中应用 HIV 检测技术的阳性预测效力会较前一种情况高得多。

将所有因素都加在一起

效度、基本比率和阈值是影响预测效力的三个因素，三者结合所产生的共同影响可以用图表加以说明。我们以证据（例如，多通道记录仪测试分数）为 x 轴，以每一等级的人数为纵轴来做直方图，那么每组（如无罪组和有罪组）均包括一系列相互独立的条块，这样就可以清楚地看出每组在各个得分等级上人数的分布情况。图 8 为 CQT 和 GKT 多通道记录仪测试结果的直方图示例，其中白色条块表示无罪嫌疑人的分数，黑色条块表示有罪嫌疑人的分数，点状垂直线表示阈值。CQT 分数从 -30 到 +30 分，说明 CQT 有 10 对目标/控制问题；GKT 分数从 0 到 10 分，代表 GKT 有 10 个题目。

首先，两组分数的分布差异表明了两种技术的效度有所不同。不难发现，在 CQT 图中无罪和有罪嫌疑人的分数存在一些差异，但这两个系列的条块存在较多的重叠；而在 GKT 图中，无罪和有罪两组条块之间的重叠则少得多。两组条块分布的差异越大，就说明决策过程越有效。这两个图都是根据本章前文提到的敏感性与特异度所绘制的，因此它们相当准确地表现出了 CQT 与 GKT 之间的效度差异。

其次，两组的基本比率是通过白色与黑色条块各自的总频率来表示的。在 CQT 与 GKT 图表上，你都可以发现人群中的无罪嫌疑人远远多于有罪嫌疑人。

* 美沙酮是阿片受体激动剂，属麻醉镇痛药，也是最常用于海洛因等药物成瘾的替代、脱毒治疗药物。在吸毒者中开展美沙酮维持治疗是目前国际上证明能够有效控制艾滋病流行的重要手段之一。——译者注

事实上，这两张图都是根据有罪嫌疑人的基本比率为 20% 的情况绘制而成的。

第三，阈值用垂直线来表示，根据该垂直分数线可将测试结果区分为阳性（大于阈值）和阴性（小于阈值）。在 CQT 图中可以看到，大部分有罪嫌疑人的分值落在阈值之上（显示说谎），但有许多无罪嫌疑人的分值也在阈值之上。事实上，因为无罪嫌疑人占多数（无罪嫌疑人有 800 名，有罪嫌疑人只有 200 名），所以即使在阈值以上，无罪嫌疑人数也比有罪嫌疑人数多，这就是为什么其阳性预测效力如此低（只有 35%）的原因。从图中也可以看出，在所有高于阈值的个体中，只有三分之一的个体确实有罪。此外，只有极少数的有罪嫌疑人的分值在阈值之下（显示诚实），而无罪嫌疑人的分值绝大多数都在阈值之下，这就是为什么阴性预测效力如此之高（达到 94%）的原因。在 GKT 图中，几乎所有的有罪嫌疑人都在阈值之上，而几乎所有的无罪嫌疑人都在阈值之下，因此这个技术的阳性与阴性预测效力都很高（96% 和 99%）。

图 8 所示的这类图表有助于确定用于决策的合适的阈值。在 GKT 图表中，阈值线应该划在哪个位置是很清楚的：如果阈值设置得过高或过低，都会增加错误几率。提高阈值，会遗漏更多的有罪嫌疑人，但因此增加的被正确判为无罪嫌疑人的数量并不大，因而会丧失一定的准确度；同样，降低阈值会将更多的无罪嫌疑人错判为有罪，但因此增加的被正确判为有罪嫌疑人的数量也不大，最终也导致了准确度的丧失。

但是对于 CQT 来说情况却并非如此。在图 8 的 CQT 图中，有很多的无罪嫌疑人都被错判为说谎，所以通过提高阈值可以明显降低错判率。一般来说，具有最大区分准确度的阈值位于两组的分割交叉点处。GKT 图中的阈值本来就在这个交叉点上，但对 CQT 来说，需要提高阈值才能使其有最好的区分度。如果设定一个新阈值，将 CQT 分数大于 +15 判为“说谎”，而等于或低于 +15 判为“诚实”，那么会有更多的无罪嫌疑人被正确地判定为诚实，但代价是也有更多的有罪嫌疑人被错判为诚实。此时阳性预测效力从原先的 35% 提高到了 52%，但与此同时阴性预测效力却从 94% 下降到了 76%。因此，CQT 多通道记录技术的效度存在明显的局限性，不可能做到既保护无罪者（具有很强的阴性预测效力，可以证明无罪嫌疑人的清白），又指证有罪者（具有很强的阳性预测效力，可以提供有说服力的证据来证明有罪嫌疑人的罪行）。也就是说，你可以通过改变阈值来达到上述两个目标中的其中一个，但无法同时兼顾。

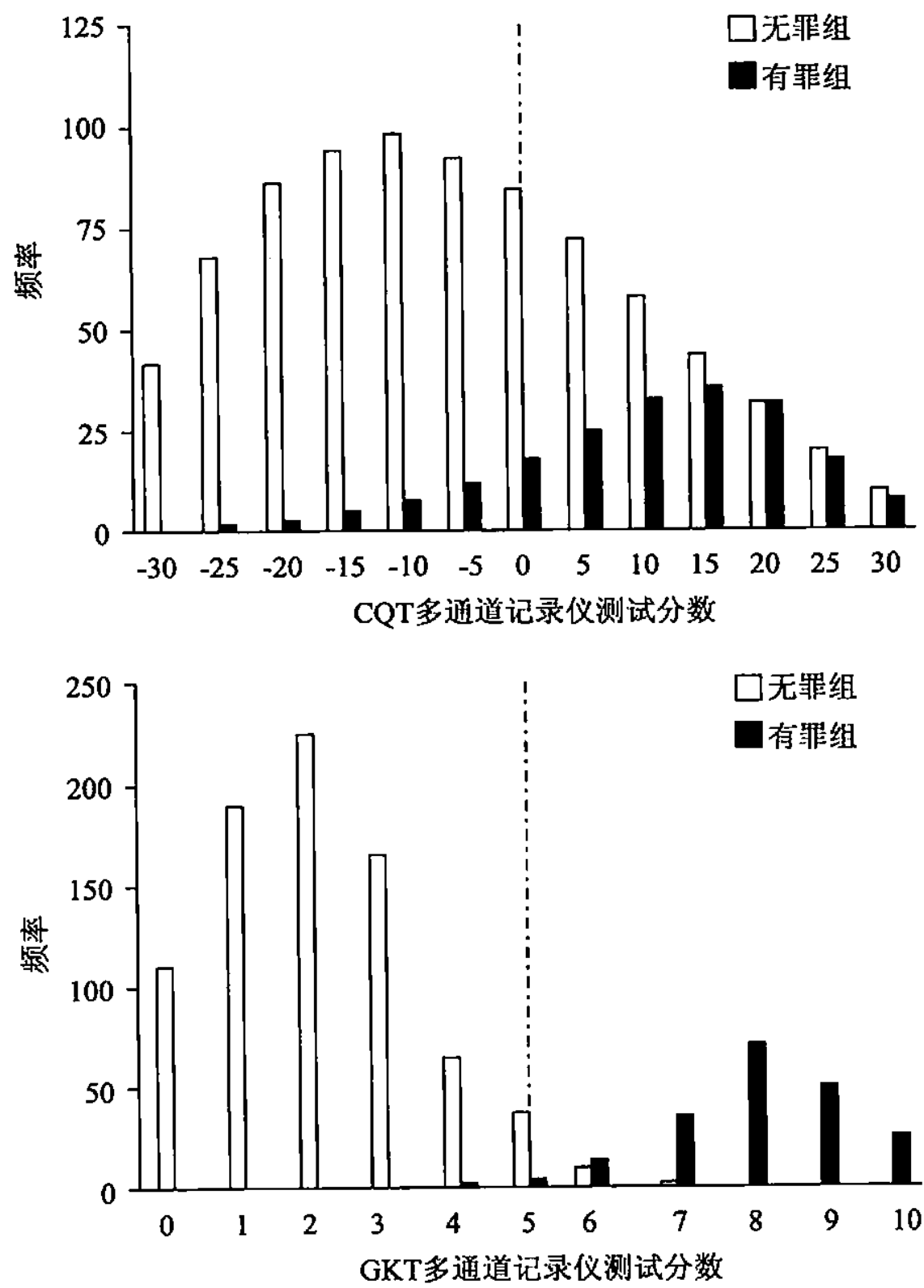


图8 假设人群（1000名犯罪嫌疑人，其中20%有罪）的CQT（上图）和GKT（下图）测试结果分布图

如果没有正确地理解本章所描述的数学关系及图8所示内容的话，那么对决策过程的效用问题进行理性的探讨就会非常困难。为了能够明智地做出判断，我们必须综合考虑四种可能结果（TPs，FPs，TNs和FNs）出现的频率、阳性和阴性预测效力，以及这些数据在现实生活中造成的后果。

例如，多通道记录仪测试的真阳性结果（TP）有助于惩罚真正有罪的嫌疑人，真阴性结果（TN）则可以替真正无罪的嫌疑人洗清冤屈。这两种结果显然都是我们所期待的，但实际上它们在CQT测试结果中所占的比例分别只是17%

和 48%，总共才 65%。相反，假阳性结果（FP）为真正无罪的嫌疑人提供了虚假的犯罪证据，而假阴性结果（FN）则帮助真正有罪的嫌疑人逃过法律的制裁，无论哪一种结果都是我们很不愿意见到的，但它们在 CQT 测试结果中所占的比例却达到了 32% 和 3%，总共有 35%。尽管从数字上来看有利结果还是两倍于不利结果，但实际生活中，我们的司法系统十分强调要保护无罪者的利益，也就是说，FPs 会被视为一个很麻烦的问题。如果在假设人群中应用 CQT 多通道记录仪，其检测结果的 FPs 竟然高达 32%，我们就完全有理由相信不应当选择 CQT 技术来判定谁真正有罪。换言之，人们可能会因为 CQT 的阳性预测效力很低（35%），而认为多通道记录仪的“说谎”结果不能作为司法程序中的证据（因为它们太没说服力了，很有可能使所取的证据产生偏差）。同时，由于 CQT 的阴性预测效力很高（94%），人们会支持将“诚实”结果纳为司法程序中无罪的证据（因为它们更具有结论性的意义）。

另一方面，当同样采用上述标准来评价 GKT 时，可以发现 GKT 的阳性和阴性预测效力都很高（分别是 96% 和 99%），因此它的两种结果都是可以被接受的。综上所述，对于假设的犯罪嫌疑人群进行多通道记录仪评估时所获得的证据，并不意味着就是说服性的证据，只是表明：在全面地判断某种特殊的程序是否可以用于现实生活中的具体情境时，必须谨慎地评估不同结果的出现频率以及可能性，还有它们在现实世界中可能导致的后果。基于上述考虑，心理生理学家们已经提出，CQT 技术无法达到法律规定的从科学性上可采纳的依据标准（Saxe & Ben-Shakhar, 1999），而 GKT 技术则可能需要更多的现场测试才能符合此标准（Ben-Shakhar, Bar-Hillel, & Kremnitzer, 2002；见第 2 章中对法律上可采纳的科学证据进行讨论的部分）。

借助多通道记录仪这个例子，我们详细地探讨了决策理论以及决策过程的具体影响因素。接下来让我们一起来看看这一理论框架如何帮助我们理解一些伪科学的实例。

有关营养品的骗局

影响预测效力的三个因素中的每一个因素，都曾被那些营养保健产品的从业者有意地滥用过，他们的最终目标就是提高假阳性错误（FPs）。他们要么说你患

了一种事实上根本没有的疾病，这一疾病可以通过膳食补充剂加以“治疗”；要么就将你误诊为另一种疾病，这一疾病据称更适合膳食补充剂“治疗”、而不适用于其他有科学依据的疗法。就这样，通过增加 FPs，伪科学家们大大地扩张了他们的产品市场。

效 度

因为实施营养品骗局的人要么就没有什么科学常识，并出于某些理由反对科学，要么就是热衷于那些很少、甚至根本就没有科学依据的特殊理论，所以他们倾向于使用那些效度尚不明确，或者效度不佳的检测手段。例如，在《虚假的测试》(*Phony Tests*)一书中有一个很精彩的章节，其中 Barrett 和 Herbert (1994) 谈到几种可疑的诊断性测试方法，例如头发分析。有许多专业实验室提供这类服务，通过头发分析的结果来帮助诊断各种疾病。在 1983 年和 1984 年，Barrett 将两位健康青少年的头发样本送到 18 个商业性的头发分析实验室，要求进行检测。对于头发中各种矿物质的含量是否正常，不同的实验室采用了不同的判断阈值，所报告的结果当然也不尽相同，甚至在同一个实验室、对同一个女孩的头发样本进行的分析，结果也不一样。然而，几乎所有的实验室都给出了针对所谓“营养缺乏”的治疗性建议，“这些建议是补充维生素、矿物质、非必需的食物成分、酶，以及选择性地摄入动物内脏等的奇怪组合。有一份报告甚至诊断了 23 种‘可能或很可能出现的状况’，包括动脉粥样硬化、肾功能衰竭等，继而建议每天服用 56 种膳食补充剂”(Barrett & Herbert, 1994)。

使用无效检测手段会增加决策过程中的错误率。而且，类似于“可能或很可能出现的状况”这样的说法会显著增加 FPs 的数量，将实际没有病的人判断为有病。很明显，实施欺骗的人是故意地将这种说法解释为某种疾病的征兆。同时，如果情况并不像他们所想的那样发展，因而面临误诊的法律诉讼时，这种模糊的说法又可以帮助那些诈骗者巧妙地逃避法律责任。事实上，对于使用无效检测手段的诈骗者，我们最担心的也就是他们随意给出有关健康状况的诊断，但又不承担相应的重大法律责任。在美国国家反健康欺诈委员会 1998 年 5/6 月的时事通报上，曾报道过 Edwin E. Kokes 的案例。Edwin E. Kokes 拥有内布拉斯加州格兰德岛上的一个独立检测实验室，并是该实验室的实际运营者。自 1989 年以来，Kokes 一直作一些子虚乌有的诊断并且出售非法药物。他自称是内科医生，但却没有执

照。在前后七年中，他通过检测头发或指甲样本来诊断艾滋病或癌症等，并从中获利将近 100 万美元。在案件调查过程中，美国食品与药品监督局也送了几份健康成年人的样本给 Kokes，而 Kokes 则把这些健康者诊断为患有过敏性疾病、毒血症，以及腺体、生殖器官、肝脏和肾脏有病变。更为可笑的是，将几内亚猪毛的标本送给他进行检测，他竟然可以诊断出一系列的人类疾病。下了各种诊断后，Kokes 又推荐了自制的非法药物，其中有一种药物被发现含有稀释了的硫磺酸（一种毒素），该药还被卖到每四盎司* 300 美元。1997 年 8 月，Kokes 被判入狱两年半，罚款 5000 美元，并且赔偿 80,000 美元给相关的受害者。

基本比率

伪科学家们提高 FPs 数量的一个简单方法，就是将焦点置于一些据称极其罕见的疾病上，而那些基本比率极低的疾病是几乎不存在的。例如，环境疾病——也被称为“多种化学物质过敏”或“对任何物质过敏”，是一种所谓的对普通食物和化学物质都高度过敏的状态。那些将自己标榜为“临床生态学家”的人声称，环境疾病源于超负荷的生理和心理压力。目前甚至没有相应的证据可以证明这种假想疾病的存在，但令人费解的是，这些所谓的临床生态学家对所有的人都下这个诊断，并且这是他们所提供的唯一诊断，这不由得让人想起一句谚语，“当你手中只有一把锤子时，所有的东西看起来都像是钉子”。实施“营养品骗局”的人总是回避谈及有效的医学诊断，这样他们就可以更容易用那些由他们自己编造出来的疾病来下诊断。

阈 值

伪科学的从业者常用的另一种手段就是，把据称可以评估是否患有营养缺乏的阈值尽量设低，特别是要设到所有人的分数都超过阈值，那么就可以建议所有的人都来购买营养补充剂。Barrett 和 Herbert（1994）曾复印了一家大型百货公司专柜宣传册中的一个量表，内容如下：

1. 你是否在高压情境下工作？
2. 你是否有躯体或情感劳累感？

* 1 盎司 = 28.3 克。——译者注

3. 你经常旅行吗？
4. 你是否常常觉得有挫败感、紧张或焦虑？
5. 你是否变得容易厌倦、烦乱或抑郁？
6. 你是否摄入过多的咖啡、酒精、药品或烟草，以至于损害了你的健康？
7. 你感觉身体充满活力吗？你经常锻炼吗？
8. 近期你的生活有重大的变化吗？

只要对上述问题有一个肯定回答，就意味着你的身体里那些对于保持健康很重要的营养成分可能正在耗竭。

在这个“压力调查”问卷中，实际上并没有任何问题与营养成分相关，何况没有经过细致的实验室检查，也根本无法判断是否存在营养缺乏。因此，像这样的问卷在评估营养成分方面的效度，即使有的话也是很低的。更重要的是，他们故意将阈值设置得极低。对上述问题的回答只要有一个是肯定的，就意味着存在营养缺乏，这实在令人难以置信。

举一个更极端的例子，让我们看看摘选自 Maharishi Ayur-Veda* 产品手册上的“自由基问卷”中的两个问题：

1. 食物质量：你最常吃的是——
 - a. 加工过的、包装好的，以及剩余食物
 - b. a 与 c 的混合
 - c. 有机食物和新鲜食物
2. 肉类：你最常吃的是——
 - a. 红肉（如猪、牛、羊肉等），汉堡包，香肠等
 - b. 白肉（如鸡肉，鱼肉等）
 - c. 素食（无肉）

这两个问题来源于该产品手册的第一部分“有毒的化学物质/放射物”。很明显，该测试的作者是尝试将特定的某些食物定义为好的食物，而另一些定义为有害的食物，但你看后可能会产生和我一样的疑问：究竟哪个选项是属于“有毒

* Maharishi Ayur-Veda 是一个行销印度药草的企业品牌。——译者注

的化学物质”或“放射物”？该测试的另外两个部分分别是“穿着与眼泪”和“情绪压力”，整个问卷的18个问题最多也只是与营养成分有部分联系。在问卷的最后，你可以看到：

如何评价你的分数：

每个“A”都代表一个极高度的危险因素。

每个“B”代表一个明显的危险因素。

每个“C”代表一个中等危险因素。

因此，无论你给出的是什么回答，你必然会得到从“中等”到“极高度”的危险级别。在这个记分系统下方特别提到：

对所有人来说，关于自由基（free radical）* 这一危险因素的问题已经是前所未有的严重了，因为污染物、放射物以及有毒的化学物质已经进入到我们每天摄入的水和食物中，还有其他很多我们个人无法控制的因素。每一个人都需要尽量避免受到过多的自由基损害。

最后一句话非常清楚地表明了作者的期望效应。如果一个人判断每个人都需要保护，那么还不厌其烦地给别人做量表的唯一理由就是，想说服他的潜在客户，让他们意识到自己是多么脆弱，需要立即采取措施来保护自己。

使用无效的检测方法，诊断基本比率极低（或者根本不存在）的疾病，并且设定过低的决策阈值，这三种方式产生的综合效应就是使假阳性错误大幅度增加，从而创造出对营养补充剂的新的巨大需求，最终获利的只是那些出售营养补充剂的人。

防止专业误导

不仅仅是绝大多数一般人没有认识到有效的心理学评估手段的基本原理，很多专业人员也没有这方面的意识。要保护自己不受那些错误决策的影响，最好的

* 自由基是含有一个不成对电子的原子团。体内活性氧自由基具有一定的免疫和信号传导功能，但过多的自由基会导致人体正常细胞和组织的损坏以及疾病的发生。阳光辐射、空气污染、吸烟、农药及重金属污染、高油高脂食物等都可能使人体产生更多活性氧自由基。——译者注

办法就是自己了解一些经典决策理论的基础知识。

首先，你应该警惕无效的测验以及其他无效操作，这些检测方法总是很勉强地将你纳入某一类问题中，就如同试图将方钉钉入圆孔。那些试图在诊断和治疗过程中采用千篇一律、一成不变的方法的人，常常会将表面客观的“自陈量表”（就像本章开头所提到的那些）作为自己的保护伞。而那些测试没有一个是经过科学验证的，也没有什么证据可以将其所谓的“症状”与任何一种压抑记忆联系在一起，更别提什么儿童性虐待的记忆了。研究证据表明，这些测试的条目都利用了第10章所描述的巴能效应（Barnum effect），即通过某种仅有表面效度的测量方式来获得“可适合每个人的东西”（Emery & Lilienfeld, 2004）。此外，正如在第15章中我们还将要讨论到的，对于压抑记忆的真实性一直就存在着争议。这些自陈量表的编制者将常见的症状统统包括进去，并且将阈值设置得很低，继而声称几乎每个人都存在压抑记忆。因为人们只能猜测这类自陈量表的预测效力，所以有关“测试的合理应用”的伦理指导条例（美国心理学会，1999）应该规定禁止他们滥用测试。不幸的是，这一方法并不一定能阻止那些狂热地相信绝大多数人都存在压抑记忆的人，去使用并且极力推荐其他人使用这些量表。

此外，一些不法治疗者可能会滥用评估技术（事实上已经存在这种情况），对每个人都给出同样的诊断，提供同样的治疗方案，尽管他们的病人可能从来都没有意识到这个事实。你应该十分小心地避开那些伪科学家们，他们总是极力推荐某种诊断或治疗方案，而且认为每个人都适合该模式。避免其影响的方法之一就是询问有多少病人被推荐给其他专家，以寻求更恰当的评估、诊断或治疗措施。那些通常不转诊病人的治疗者都可能是倾向于只提供某一种治疗方案，而全然不顾该方案在有效性方面的局限（如果有效的話）。

你完全有权接受最好、最合适的诊断和治疗，这就要求你不要盲目相信专家提供给你的信息，诸如检测方法的效度、某疾病在相关人群中的基本比率，以及用于制定决策的阈值等等，这样你就能够再次检验相关的概率，最终获得最关键的结论。只要你能提出这些问题，就足以使那些治疗者意识到，你是不愿成为那个硬被塞进圆孔的方钉的。

14. 决策：临床方法与统计学方法的比较

Robyn Dawes (1994) 最后一次实施洛夏墨迹测验的经历，对他的整个职业生涯产生了巨大的影响。他评估的对象是一个被父母强制入院的 16 岁女孩，原因是她正在跟一位比她年长 10 岁的男士约会。被关起来之后，女孩自然会表现得心烦意乱，但相对于她的年龄来说，她既聪明又成熟。这位女孩的 IQ 分为 126，在人群当中百分等级为 95（智力水平较 95% 的人高，较 5% 的人差）。她的洛夏墨迹测验有 40 个题目（总共 41 个题目）的反应被评为“良好模式”，这意味着她对图片的描述可以被测验的施测者所理解。唯一的“异常模式”反应是对第八张图片的回答：“看起来就像一只熊，就这样了。我也解释不清楚。”虽然我不能在这里呈现洛夏测验的这张图片，但它看起来的确不像一只熊。在随后的工作会议上，Dawes 与同事对这一测试结果进行了讨论，当时的情况是这样的：

为首的心理学家向周围每个人展示了第八张卡片，并用夸张的语气问道：“你觉得这看起来像只熊吗？”没有人这么认为。于是他开始“解释”为什么这个女孩会说它像只熊：她一直被幻觉困扰，是一个“假性神经症型精神分裂症患者”。尽管她经常处于幻觉状态，她还是知道如何表现得正常……因此，大家应该忽略她没有表现出精神病症状这一事实……。当我指出，41 张卡片中有 40 张都得到了良好模式的反应，而且这一比例事实上比大部分正常人的反应结果还高时，这位领头的心理学家就说我只了解统计学，但并不了解人类，而且我尤其不懂为何“统计学不适用于这个个

体”……这名病人最终被诊断为精神“分裂症”，因为我所供职的这个部门并不能治疗分裂症，所以她被送回她父母的家中。而且，大家不顾我的反对，进一步决定，如果她的父母还送她回来，那就要把她直接送入附近的州立医院。根据我所知道的情况，她极有可能就这样被判为精神分裂症并关进疯人院，而其依据仅仅只是洛夏测验中的一项异常反应。我不能很肯定这一结果，但我也再不想进一步去了解，我退出。

进行决策的两种方法

进行决策有两种不同的方法，按照经典的区分标准，可以分为临床方法和统计学方法（Meehl, 1954）。临床方法就是指应用非常专业的、独立的人类评判能力，对所获得的信息进行评估，并在此基础上作出决策。这一技术并不一定非要由专业的临床医师来完成。事实上，临床方法一词的应用非常广泛，任何需要某人作出预测的时候，均有可能用到这个词。例如，前面我们提到的那位领头的心理学家，根据洛夏测验的结果将被试诊断为精神分裂症，就是依照临床方法进行决策的。相反，统计学方法则指根据数据信息进行公式运算，并以此来作出决策。Dawes 提议大家应该考虑那些极其正常的个体做出“异常模式”反应的百分比，他就是试图按照统计学方法进行预测。

最终，无论是何决策过程，都会使用这两种方法中的一种。虽然人们普遍认为，任何理性的决策者都应当同时使用这两种方法（如 Kleinmuntz, 1990），但事实上这是不可能的。如果两种方法得到的结果是一致的，那还没什么关系；如果这两种方法所获得的决策结果不一致，比如得出两个有矛盾的诊断结果，那你就只能选择其一了。进行决策时最公平而又符合伦理要求的方法就是使用其中预测效度更好的那个方法，运用预测效度较差的方法无异于欺骗那些受该决策影响的个体。

采用临床方法进行预测主要是期望得到因果关系，而不是期望得到基于概率的解释（Dawes, 1991, 1994, 1995）；同时寻找一种叙述性的真相——一个看起来真实的“好故事”，而不是纯粹的事实报告或历史记录（Spence, 1982）。将所有信息整合为一系列事件，会让这些事件看起来彼此一致并似乎有些道理，这被看作是一种更加危险的预测方法，因为采用临床方法的人拒绝接受统计技术所固

有的误差，一味追求完美预测（Einhorn, 1986）。长期以来，对这两种方法的相对有效性就存在着争论，但四十余年来的研究始终表明：采用统计学公式进行预测的准确性与人类判断的准确性是相等的，甚至还更高（Dawes, Faust, & Meehl, 1989; Grove, Zald, Lebow, Snitz, & Nelson, 2000）。在预测学业成绩、企业破产、寿命、军事训练成绩、心肌梗塞的发生、神经心理诊断结果、违背假释条例的行为、精神病的诊断以及暴力行为等等许多方面都是如此。（Dawes, Faust, & Meehl, 1993）这些结果显示，事件之间的概率性联系比起因果关系，更容易被获得也更容易被证实；历史的真相比叙述性的事实更加准确；接受固定存在的误差实际上意味着可以减少错误决策的次数。

有一点特别重要，就是专门对证据进行回顾和总结。Sawyer（1966）对临床方法和统计学（机械式）方法两者在资料收集和资料整合阶段对决策的影响作用进行了比较。临床数据的收集法往往是运用缺乏常模的一些测验、问卷和访谈方法，因此反对进行个体反应之间差异的比较（如洛夏测验）。相反，机械式数据收集法所运用的是一些在形式、施测过程和记分方法等方面都标准化了的测试和评估工具（如 IQ 测试）。Sawyer 曾写过一篇综述，对四种可能的预测策略（临床收集和临床整合、临床收集和机械整合、机械收集和临床整合，以及机械收集和机械整合）的效度进行了验证，结果清楚地说明：机械式方法在决策过程中的资料收集和资料整合两个阶段，都比临床方法更加优越。无论数据是如何被整合，用机械式方法收集的数据都比临床方法收集的数据更有用。此外，与本章的论点一致的是，无论数据是如何被收集的，通过机械式方法整合信息总会比人为的判断更有效。

Paul Meehl，一位统计学方法的热心倡导者，1954 年发表了一篇有关临床决策方法和统计学决策方法的优劣的综述，该文章被奉为经典，此后他一直敦促决策者们要多关注那些发现。到 1986 年，他已经找到大量可引述的研究结果来支持他最初的观点：

在社会科学领域，从来没有哪个争论像临床决策方法和统计学决策方法的争论一样，吸引了如此多高质量的、多样化的研究，而得出的结果又如此一致。当你综合 90 项研究（在 1992 年的一篇综述中称，研究数目达到了 175；Meehl, 1992），进行从足球比赛结果到肝病诊断等不同事件的预测时，

几乎找不到任何支持、甚至倾向于支持临床方法的研究证据，那么我们就可以得出一个实事求是的结论了。

为什么统计学方法更加优越

相对而言，基于统计学方法的决策具有更高的预测效力，主要是因为统计学方法具有符合我们需要的良好功能，而人为判断则包含许多不恰当的认知偏差及局限性。与容易犯错的人为推理过程不同，统计公式包含了许多有用的数学功能，这对于探测纷繁多变的事件之间的关系是极其有效的。另外，人类思维的局限性，又进一步扩大了统计学方法与临床方法之间的差距。Goldberg（1991）总结了这两种方法在五个方面的区别，可以更清楚地说明统计学决策方法的优越性。

效 度

统计学方法在对不同类型的信息资料进行检测并做出相应调整的方面具有明显优势（Dawes, 1971, 1979）。对于决策过程而言，与重要结果有关的信息比无关信息更有价值，相比之下，人为判断不仅不能像这样恰当地对信息进行权衡，还会在预测过程中带有不同的偏向（Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980; Turk & Salovey, 1988）。尤其是，信息加工过程经常依赖于我们在前面章节中所讨论过的、判断过程中启发式推理方法的三原则——锚定法则（anchoring）、可利用性法则（availability）和代表性法则（representativeness），而且有时也会被诸如迷信观念、控制错觉、随机性感知错误以及验证性意见等问题所影响。在决策者权衡信息时，有诸多因素会导致决策正确性的降低，上述种种因素还仅仅只是其中的一部分而已。

让我们一起来看看作出一个假释决定所涉及的重要判断过程，并以此为例来说明统计学方法比人为判断方法更能恰当地权衡信息。Quinsey 和 Maguire（1986）发现，在对一个高度警戒（最高安全级别）的精神病院里的病人进行危险性评估时，有经验的临床法医尤其注重病人在住院期间出现攻击行为的频率，这在他们的决策原则中占了相当大的比例。虽然这一资料凭直觉看来很有说服力，但经过 11 年甚至更长时间的追踪观察，人们发现病人在住院时的攻击性与复发（出狱后再次犯罪）并没有相关性，这说明临床法医对于危险性评估是无

效的。而与此相反，统计预测系统可以更恰当地对这一信息以及其他资料进行权衡与评估。研究结果也表明，统计学方法预测的准确度虽然没有达到百分之百，但仍明显地超过了临床方法（Quinsey & Maguire, 1986; Wormith & Goldstone, 1984）。

测量单位

运用统计技术，可以很容易地综合分析那些用不同标准的工具所测量的信息，而对临床方法的决策者而言这是一个棘手的问题，因为他们的判断过程会受到心算的局限。对于人类来说，在判断时很难一下子将不同方法所测得的信息综合起来，而且总是倾向于过度使用那些与结果的形式一致的信息（Slovic & MacPhillamy, 1974）。例如，在一项研究中，人们向被试展示了一个按等级排列的线索（根据能力测试的成绩对学生进行排列）和一个按类别排列的线索（将学生分为擅长课外活动或很少课外活动的两组），随机要求被试根据等级排列或类别排列来预测学生的大学入学考试成绩（Tversky, Sattath, & Slovic, 1988）。而在另一实验条件下，将给予信息的形式反过来：课外活动作为等级排列的线索，而能力测试的结果则作为类别排列的线索。结果发现，无论在哪一种条件下，那些被要求根据等级进行预测的被试会更加重视等级线索，而与此相反，当要求根据类别线索进行判断时，被试就更加重视类别线索。

现在让我们思考一个更为重要、也更为真实的决策情境：办案人员根据控告被虐待或被忽视的陈述来决定儿童的命运。而在有关儿童福利的决策过程中，可用的信息非常之多，而且各不相同。许多类型的信息本身是严格的分类变量（如儿童的性别、种族、被控告的犯罪嫌疑人与儿童的关系等），而其他类型的信息则更像是连续变量（如儿童的年龄、被虐待事件的持续时间和发生次数等）。由于缺乏统一的度量标准，要通过临床判断有效地将如此多类型的信息综合起来，的确十分困难。因此，通过临床判断对儿童福利所作的决策其准确度欠佳，也是意料之中的事情（Ruscio, 1998）。

信 度

在进行预测的过程中，人们有一种独特的、无法替代的能力，可以分辨出哪些信息应当被列入考虑之列，但是，他们并不能在不同情境下始终如一地运用这

一预测方式 (Goldberg, 1986)。决策策略应用的一致性就是其判断的信度。信度有多种指标, 可以是应用某一决策策略所得到的判断结果的一致性程度, 也可以是多种决策策略对同一个案的判断结果之间的一致性程度。如果运用同一策略进行多次判断其结果不一致, 或者各种策略的判断结果彼此间有不一致, 那么意味着该决策策略的信度较低, 其效度也会随之下降。

举例来说, 假设某人某次决定将一个受到虐待的委托儿童安置在保育院, 随后又被秘密告知该儿童返家后仍然有受虐的情况。那么如果此人第二次决定将该儿童留在家中, 则意味着其决策的信度很低。因为在前后两次作出的决定当中, 只有一个是最佳决定, 这个决策者就好像是通过抛硬币来决定这个儿童的命运。同样, 如果两个不同的决策者对这同一个案有不同的决定, 他们两人的决策结果至多也只有一个“正确”的, 因此其中必有一个决策者, 甚至同时是他们两个人, 决策的效度比较低。

低信度 (包括依据某种策略所做的多次判断, 以及依据多种策略所做的判断之间的一致性) 可能是有关儿童福利的决策领域所面临的重大难题。研究者认为出现这个问题的主要原因, 一是相关制度缺乏明确定义的目标, 二是在缺乏关于预测的理论或实验基础的情况下, 所做的决策具有极大的主观性 (如 Alter, 1985; DePanfilis & Scannapieco, 1994; Gleeson, 1987; Schuerman & Vogel, 1986)。这样的局限性使得判断之间的一致性很差, 即使是同一领域的专家之间也是如此 (Mandel, Lehman, & Yuille, 1994)。Lindsey (1992) 曾对目前广泛使用的大规模干预的理论基础提出过严重质疑, 认为这些干预措施缺乏足够的科学基础, 其部分依据就是: 目前儿童福利判决的信度非常低。

与此相反, 根据统计技术进行的预测则具有良好的信度。在不同场合下对相同的资料进行统计并作出决定, 都会得到同样的结果。这种一致性可以解释判断引导程序 (judgmental bootstrapping) 技术大获成功的原因, 该项技术首先对人类的判断过程进行方程建模, 然后用其进行预测 (Camerer, 1981; Dawes, 1971, 1979; Dawes & Corrigan, 1974; Goldberg, 1965, 1970)。人类具有在判断时挑选出重要的线索, 并将其纳入预测过程的一系列能力。统计模型充分借鉴和利用了这种能力, 在运用这种能力的过程中又保持了良好的信度。一旦信度提高了, 决策的准确性也就随之提高。

重 复

在运用统计方程时，资料的不同部分之间存在的重复或多余信息也会进入计算过程。如果将一个新变量加入到方程中，而这一变量与其他变量有很高程度的重复，那么该变量所得到的权重，就会明显低于那些提供独特的或新的信息的变量。而在用临床方法进行决策时，人们不太可能在头脑中对所得的各部分信息之间的相互关系进行十分全面的考虑，而且研究证据表明，人们还更倾向于偏好重复信息，甚至主动地寻求重复信息，以支持自己的决定，使自己更有信心（Kahneman & Tversky, 1973; Slovic, 1966）。

例如，一些教育工作者相信，应当在教学中评估多元智力（multiple intelligence），以帮助制定班级教学策略。Chen 和 Gardner（1997）为此提出了七种“智力”——语言智力，音乐智力，逻辑数学智力，视觉空间智力，身体运动智力，自我认知智力，以及人际关系智力。但是，额外增加的对所谓特殊智力的测量真的能够提供新的信息，继而改善决策的准确性吗？虽然多元智力的评估结果是可信的，但其效度却存在一定的问题（Plucker, Callahan, & Tomchin, 1996）。更为重要的是，“总量表”或总智商（IQ）已经可以很好地预测出有意义的结果了，其他类型的 IQ 至多也只能使预测效力略微有点提高（Glutting, Youngstrom, Ward, & Hale, 1997）。即使是在预测阅读能力时，总量表 IQ 分上再增加一个语言智力测试的结果也并不能增强其预测作用。所以说，由于各种智力成分彼此之间存在高度的重复信息，即使真的存在多元智力，从实际的目的出发，单一的总智力测试也已经足够了（Bray, Kehle, & Hintze, 1998）。额外补充的多元智力并不会提高评估预测过程的信度，只会浪费更多的时间与费用（Schmidt & Hunter, 1998）。

回归效应

最后，临床方法和统计学方法之间的比较要涉及一个更微妙的问题，那就是均值回归。请设想有一个学生参加了一次 GRE 考试*（Graduate Record Examina-

* GRE 是美国教育考试服务中心提供的研究生入学资格考试，分为普通考试与专业考试。——译者注

tion)，得了 1400 分。在没有经过更多的学习、练习或其他准备的情况下，这位学生一个月后又参加了 GRE 考试。你预计他这一次会得多少分呢？如果你猜大概在 1400 分左右，甚至还要更高一点，那么你和大多数人猜的结果一样，因为该学生原来所得的分数显示他能力很强。但是，在实际生活中，一般都不会出现这种预期的结果，因为测试成绩包括两个部分——实际能力和随机的情境因素（运气、疲劳等等）。得到高分往往是高能力与幸运两者结合的结果。一个人的能力是稳定的，但运气是随机出现的。那些在某次考试非常幸运的人，就不太可能在随后的一次考试中仍然那么幸运。因此，如果那些得过高分的人再次参加考试，一般来说，他们的成绩将倾向于有所下降，他们不太可能再像第一次那么幸运。同样的道理，那些因为运气糟糕而在某次考试中考砸了的人，一般来说，在重考时就很有可能考得好一些。

尽管结果（如第二次考试的分数）不可能被精确地预测到，但它会有向均值回归的趋势。也就是说，如果第一次出现了极端的分数，第二次的分数就会倾向于不那么极端。如果没有掌握这一原理，我们就有可能作出许多不符合现实的预测。

运动员上了《体育画报》（*Sports Illustrated*）封面后的一周之内，运动员的成绩就会不如之前那么优秀，这就是所谓的“体育画报恶咒”迷信，这样的迷信已经使得一些明星运动员拒绝上该杂志的封面。如果对这一问题进行更深入的探讨，马上就可以发现这里存在错误的推理。是什么原因使运动员可以上体育画报的封面呢？当然，最主要是因为他们具有优异的能力，但其中也与一点好运有关。一位优秀的运动员必须表现得极其突出才能成为画报的封面人物，这意味着必须同时具备优异的能力和好运气。但实际上在比赛中运动员的运气总是时好时坏的，所以大部分运动员在登上画报封面后的一周内，表现很可能就不再像之前那么好了。

无法全面考量影响成绩或表现的运气成分，这是人类进行推理时常犯的一个基本性错误。由于我们的预测太过极端，往往导致其准确度不高。相反，统计公式可以根据预测程序和结果之间统计联系的强弱程度来降低预测的极端值，从而得到精确的回归预测。也正是由于这个原因，大部分常用的预测公式被称作回归方程（regression equation）。由于统计学预测往往比临床预测更少出现极端结果，因而所获得的决策结果也就更加准确。

坚持使用临床方法的原因

尽管有大量的理论和经验性证据支持基于统计学的预测比基于临床方法的预测更好，但许多领域的工作者还是继续应用临床方法进行决策。调查发现，大多数神经心理学家更喜欢用临床方法对智力缺陷进行预测，而不是运用已经被反复证明具有高效度的统计学公式（Guilmette, Faust, Hart, & Arkes, 1990）。Meehl（1986）曾对这一问题进行了探讨，认为存在以下几方面的原因导致大家毫无依据地坚持应用临床方法（进一步了解及更多讨论请见 Grove & Meehl, 1996）。

知识的欠缺

至少有三个方面的知识欠缺，使一些工作者拒绝使用统计学策略进行决策。首先，他们没有意识到许多偏见会影响人们的判断。任何稍微熟悉这个领域的人都会理解，对决策结果的期望会影响决策者的临床判断（Arkes, 1991; Faust, 1986; Kleinmuntz, 1990; Ruscio, 1998）。其次，即使了解到临床判断的不足，人们可能也没有意识到统计学技术是一种可以用于辅助进行决策的方法。第三，在那些确实了解统计学技术的人中，还是有些人并不知道如何应用自己的统计学知识。当然，这些理由还不能完全解释为什么在进行重要决策时，人们会如此坚持使用临床方法。

害怕计算机

运用统计学方法进行决策通常会用到计算机，那些不太熟悉计算机操作的工作者，就会倾向于回避使用这一方法。虽然曾经很多人有计算机恐惧症，但随着时间的推移，这种现象越来越少，甚至人们开始担心相反的问题——计算机依赖了。目前，计算机在工作场所的应用越来越广泛，熟练使用计算机已经成为许多工作的基本要求，尤其是许多办公室的任务。一些现代计算机程序虽然十分复杂，但操作界面却很人性化，不过，正是因为它们简单易用，人们又很可能会过于依赖它们，这也是很危险的。依赖界面友好、但效度未经论证的计算机统计决策，与应用效度未经论证（或低劣）的临床判断都是很愚蠢的（Matarazzo, 1986）。必须记住，一种决策策略的实际效用和伦理接受程度是由经过论证的效

度所决定的，而不是简单地由它的易用性来决定。在选择预测方法时，如果首先考虑的是如何让自己感觉舒适，那么对于那些生活会受到你的决策影响的人来说，这是非常不公平的。

害怕失业

还有许多人逃避使用统计学方法进行决策，是因为他们害怕受到技术性失业的威胁。他们担心执行专业功能的某个公式可能会取代专业人员，甚至可能就是他们自己。这是一种毫无根据的恐惧，尤其是在那些存在许多社会约束、不利于决策过程的场合，如缺乏必要的信息，限期完成或有过多文字工作等行政压力，很多待处理的案件，害怕合法的报复，以及缺少必要的训练和经验等等，这种恐惧更是明显（Murdach, 1994）。尽管面临这些额外挑战的工作人员们特别不愿意放弃临床预测法，但正因为有这些因素的存在，才更需要他们采用统计学方法。工作者的时间、精力和竞争能力都是有限的，这更加突显了提高决策效率的重要性，应用统计学方法进行决策正好可以达到这个目的。花费较少的时间在人类相对做得较差的事情上，例如将信息整合起来为决策服务，也就可以花更多时间在那些非常重要的事情上，如收集相关信息和提供实际的服务。因此，技术性失业并不会成为一个大问题。

相信人为判断的效力

许多人坚信自己的判断是有效的，这也成为阻碍统计学技术应用的另一个常见问题。毫无疑问，统计学方法与临床方法之间的持续“论战”有一部分是源于人们不愿意承认公式可以比专家更好地执行某些专业功能。例如，在一些研究中，研究人员将参与者的预测结果与统计方程的预测结果进行比较，且将按统计学方法得到的预测结果反馈给参与者（Goldberg, 1968, 1970; Sawyer, 1966）。对于该反馈信息，人们最简单的应用方式就是复制，这就使得参与者预测的正确性与那些公式预测的准确性相等了。不过，人们似乎总是强迫性地试图“减少异常值”，即对那些统计学预测结果进行修正，去掉他们不认同的部分，以使自己的预测比统计学方法有更多的命中、更少的失误。然而，他们的这种尝试几乎无一例外地导致了自己预测准确度的下降。

决策者们经常强调他们曾受过良好的训练，强调他们拥有丰富的经验，以此

作为其决策质量的支持证据。已有许多研究对这一问题进行了探讨，即考察针对某一特定的决策任务，接受过大量训练并积累了相应经验的个体，是否能够比统计方法有更好的预测结果。Meehl (1954, 1959, 1967) 曾提出，具有某些特定的任务特征，尤其是变量之间存在着极其复杂的关系的任务，更加适合由经过高度训练且有经验的临床医师来进行决策。Meehl 还认为，复杂的决策任务就需要有复杂思维能力的专业人士。但在此之后，所有的研究（寻求个体比统计方法预测得更好的证据）都是徒劳的（Goldberg, 1991; Slovic & Lichtenstein, 1971; Stewart, 1988）。大量研究表明，训练或者经验并不能显著影响临床判断的质量（Berman & Norton, 1985; Dawes, 1989, 1994; Faust, 1986; Faust & Ziskin, 1988; Garb, 1989, 1999; Garb & Boyle, 2003; Goldberg, 1959; Oskamp, 1965; Ziskin & Faust, 1988）。因此，期望通过正规训练和临床经验等有关证据来支持基于临床方法的预测，按第2章中曾提过的 Dawes 的描述来说，就是“自真空中产生的论证”。

理论认同

那些极其相信某种理论方法的专家，更有可能放弃基于其他理论的决策策略。这其实是不明智的，一种决策策略的价值只取决于它可观测到的准确度，而不是取决于它是否恰好符合某种理论假设。当然，具有很高的理论认同感和强烈的信念会提高决策者的自信，但自信因为两个方面的原因有可能产生负面影响，其本身也还需要仔细考察（Ruscio, 2000）。首先，过度自信对于决策者来说是有害的，因为一个过度自信的专家不太可能会努力去改善其自身的表现。如果你确信自己已经做得足够好了，你还怎么会有动力去做得更好呢？其次，过度自信对其他人也会产生害处，因为人们往往错误地将自信与准确度混淆在一起。例如，研究显示，当某人以自信的方式陈述法庭证据时，不管它的实际效度怎样，所陈述的证据都会对法官或陪审团产生更大的影响（Faust & Ziskin, 1988）。一项大样本的研究也证实，基于临床的个人判断不仅比基于统计方法的判断预测效力更低，而且人们往往还对自己的判断拥有过高的信心，全然不顾其效度实际上较低的真实情况（Dawes, Faust, & Meehl, 1989; Faust & Ziskin, 1988; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982; Ruscio, 2000; Ziskin & Faust, 1988）。

Fischhoff (1988) 指出, 人们更愿意运用临床方法来进行决策, 原因是临床方法看起来似乎可以获得精确的预测。如果选择统计方法进行决策的话, 为了显著提高预测效力, 个体必须先接受“一定比例的决策是错误的”这一前提。而许多决策者似乎并不愿意这样做, 他们宁愿接受那种由自己的判断所造成的固有错误, 至于这种错误的比例有多高, 他们自己也不清楚, 很有可能比统计学方法出错的几率更高。

从这个问题的本质上说, 专家的过度自信是源于他们完全一厢情愿的想法 (Dawes, 1991, 1994)。大家都喜欢看到结果是可以预测的, 而且也乐于相信人们, 尤其是那些受过正规训练、有经验的专家们, 在重大问题上可以做出很好的决策。不幸的是, “最简单的道理就是: 过去的行为是对未来行为的最佳预测者, 但从过去的实际情况来看, 那些专家们的预测并不怎么高明” (Dawes, 1991)。一厢情愿的想法可以夸大专家决策的质量, 但却不能改变它们的实际质量水平。最终唯一的受害者, 还是那些生活会实际受到低劣决策影响的人们。

统计公式给人以“非人性”的感觉

许多人错误地认为, 对资料的统计学处理忽略了个体的独特性, 把人看作是“纯粹的数字”。但实际上, 统计方程同样可以精确地利用那些临床决策者所获得的信息。因此两种方法的差异与个体独特性毫无关系, 但是从以最可靠且最有效的方式来组织信息的角度来看, 统计学方法具有极大的优势。另外, 统计公式可以开放给民众监督, 可以被反复修正、精炼和改良。相比而言, 人为的判断仅限于个人的思考过程, 并存在潜在的偏见, 研究也表明, 专业培训及丰富的临床经验都不太可能对上述局限有多少改善作用。因此, 由于担心那些已证实有效的决策方程有些“非人性化”而产生的不安感, 通常只是一种误解, 这种观念与我们下面将要讨论的与伦理有关的错误概念有一定的关系。

与伦理有关的错误概念

可能是因为前面讨论的那些因素, 临床方法的支持者们总是声称用统计学方法进行预测或决策是不符合伦理的。当仔细加以验证时, 你会发现这种伦理概念似乎包含了一系列的逻辑错误。比如, 人们通常都会同意, 在伦理方面最站得住脚的行动, 就是那种具有可证实的最高效度的行动, 也就是说, 正确决策方法所

能得到的最好记录，就是错误最少的记录。使用一种众所周知的比统计学方法系统误差更大的决策策略，这对于那些决策所涉及的实际个体来说，是非常不公平的。Meehl（1986）曾这样说：

如果我用无效而不是有效的方法对有关大学生、罪犯或抑郁患者的一些重要问题进行预测，同时还索取 10 倍于可以达到更好预测效度的方法的价格，这是一个完全不符合伦理学的行动。原本进行这项重要的工作是因为我会自我感觉更好，也更受人欢迎，而如果这样做的话，它就彻头彻尾地成了一个卑鄙的借口。

接受统计学决策方法

当需要进行决策时，对于大部分任务来说，都不太可能达到精确预测的理想状况。许多重要的决策是以预测为基础的，而预测总是包含了许多不确定的因素，而且我们并不是无所不知的，即使是通过统计学方法也不可能达到最好的效果。因此要提高预测的质量，我们就必须接受必然会犯错这一事实，并将注意力放在如何尽可能减少犯错次数的问题上。尽管我们的大脑不可能达到像计算机那样的计算能力，但在推理过程中我们仍可以很好地模仿统计学的决策策略，最简单的方法就是建立一个“支持或反对观点”的清单。

列出支持或反对观点的清单

当面临很多选择时，你可以用一种老式的方法来改进你的决策，即对每种选择都列出所有“支持”和“反对”的理由。有一项关于不同预测策略之准确度的研究，其结果十分令人惊讶：预测者对每条信息所给予的权重，对于决策准确度来说并无多大的影响（例如，Dawes & Corrigan, 1974）。这就意味着，你只需简单地用每个选项中支持理由的数量减去反对理由的数量，然后选择得分最高的那个选项就可以了。

例如，我们可能采用这种方法到市场去挑一辆新车。假设你已经将范围缩小到几辆车，但仅剩下一点时间就要最后拍板决定，那你就可以模仿统计决策过程，即确定你认为重要的几个指标，如操控性、车型大小、价格、安全性能以及

耗油量等，然后为每辆车列一个清单。然后根据这些指标对每辆车进行正性（支持）或负性（反对）的评定，最后选择有最多正性评分的那辆车。如果你愿意的话，还可以运用一个更复杂的评分方法，那就是使用等级评分而不是单纯的正性/负性评定，或者对不同的车所具有的某些特征给予更大的权重。不过，无论你采用何种方法，你都会惊讶地发现，对于大部分案例而言结果是一样的。统计学决策的成功之处，并不是它对细节的精确把握（虽然在这方面它确实无懈可击），而是它始终聚焦于事实，以事实为根据。你只需通过简单的列表和计算，就可以轻松地享受到统计学决策策略带来的好处。

专业训练

要将统计学决策策略应用到个人生活中，除了简单的支持或反对清单法之外，你也可以使用其他类似的策略来大大提高你的决策质量，甚至达到专业水平。假设你是一个研究生入学资格的审批官员，你必须决定哪些申请者可以录取。每个申请者的档案都包含了大量的信息，包括平均学分、GRE 分数、个人声明、推荐信、完整的大学成绩单，以及社会活动清单等等。无论你用什么方法来进行判断，显然这都是一个庞大而艰巨的任务。

那么，你怎么知道如何对这些资料的每一部分进行恰当的权衡，最后做出明智的决策呢？你可能会高估一些实际上与录取研究生的要求无关的信息，也可能会低估一些本可以提高预测准确度的信息。通过使用统计学程序，根据已有各项资料的效度对其进行权衡和评估，你就可以提高预测的准确性。

如何将每个申请者的平均学分（分值在 0 到 4 之间）与其 GRE 分数（分值在 400 到 1600 之间）综合起来进行分析，才是对申请者最公平的方式呢？在其他条件都相同的情况下，哪个学生才是最有能力的呢？是平均学分为 3.20、GRE 为 1200 分的学生？还是平均学分为 3.40、而 GRE 为 1100 分的学生？如果还需要考虑其他额外的因素，这个任务就更加复杂了。不过，尽管这些任务对于人脑而言十分困难，但统计学方法却可以轻易地解决这些问题。

假设有人将你已经看过的一份资料拿走，然后又将它放回你的文件夹里，而你并没有将其识别出来，那么你有多大把握会对这份资料作出同样的评定结果？我们很容易被暂时的情境性因素，如分心或疲劳所影响。总之，出于各种各样的原因，我们的判断并不能始终保持一致，这些因素的存在使得我们做出决策的准

确度难以得到保证。而与此相反，统计学预测则具有近乎完美的信度。

如果你读了一封推荐信，你可能还想再读一下其他的推荐信，以明确这些推荐是否彼此一致。这类推荐信通常会比较一致地勾勒出申请者的大致特点，因此它们之间至少有一部分内容是重复的，不能被当作完全独立的信息变量。每封推荐信的权重都应该比上一封更小一些，因为它提供的独特信息比原先的那封会更少。运用统计学方法，就可以自动地对这类重复信息进行调整。

最后，如果你还能意识到极端值（无论是好还是坏），都是能力和运气的共同产物，那么你可能也会明智地调整你的预测结果。GRE 分数极高而平均学分中等的人，可能并不比那些 GRE 和平均学分均为中等偏上的人更优秀。尽管某个极高的分数的确会吸引你的注意力，但需要对其进行均值回归才能提高你的预测效度。另外，辅助使用统计学方法——自动地根据均值回归进行调整——可以帮助你做出更准确的决策，犯的错误也更少。

专业人员应该尽可能地利用统计学决策策略的优势。那如何才能得到统计学预测公式呢？幸运的研究者可能通过详细查阅相关的研究文献，就能得到经过效度验证的预测公式。即使没有直接找到公式，也可以通过查阅文献得知各种变量对于评定是否重要，从而有所侧重地将其纳入到决策过程中。有了这些信息，我们就可以轻松地建立起一个预测系统。万一对于要评估的潜在变量，我们的确没有任何的背景资料，那么只能从头开始创建一个统计学预测公式，而这也并不是特别困难（Ruscio, 1998）。

统计学的确可以应用于个体

在第2章中，我们曾驳斥了“几率与单一个体无关”的谬论。令人痛心的是，有很多人，往往还是那些掌握权力的人，一直持有这种错误的信念，本章开头我们所提到的那位领头的心理学家就是如此。不过，不知道在俄罗斯轮盘赌中，这位心理学家是愿意要弹膛中有一颗子弹的手枪，还是有五颗子弹的手枪呢？请记住这样一个重要的事实，即统计学的确可以应用于个体，而否认这一事实，坚持使用已被证明为较差的临床方法来进行决策，那才是违背伦理的。

15. 伦理学问题：使用并推广未经证实的治疗方法

在每年寻求心理治疗的数百万人中，焦虑和抑郁是两种最常见的症状。你可能也曾经成为焦虑或抑郁者中的一员，可能刚刚经历了分居或离婚、失业或者失恋。现在假设你决定寻求心理健康专业人员的帮助，来减轻自己的抑郁和焦虑症状，并学习如何更好地处理未来将要面临的压力，那么我希望首次晤谈是什么样的呢？毫无疑问，你会解释是什么原因让你来寻求治疗，描述你的症状，并希望治疗师可以提供一些支持和建议。但是，如果你的治疗师对你的陈述一点也不感兴趣，甚至在你见面之前，就已经认定，导致问题出现的“真实”原因就是你在童年时期受过性虐待，你在崇拜恶魔的宗教仪式上曾受过伤害，或是你曾被外星人劫持，那怎么办呢？如果你的治疗师相信就是这些假设的创伤事件（尽管你自己完全没有一点印象）导致了你的人格分裂，你并非真的受焦虑或抑郁所困扰，而是多重人格障碍（MPD）的问题，那又怎么办呢？

忽略显而易见的事实，赞成不太可能出现的情况

在你看来，虽然会觉得这种情况不可思议，但它确确实实发生在许多患者的身上。一些治疗师的确是对明显的事实视而不见，而一味凭想象判断。就像下面这位患者所描述的：

我的心理治疗师不断地询问我童年时的情况，而且似乎非常肯定我在童年曾受过性虐待。因为我想要出院，所以我会尽量配合他。但我实在不明

白，为什么我的治疗师没有任何兴趣来听听导致我近期如此抑郁的原因，如痛苦的离婚事件、工作中的性骚扰问题、再次开始一段新的婚姻所承受的压力、近期生活方式的巨大改变等等。不过我还是信任他，如果他坚信我所有的问题都源自于那些似乎已经被我忘却的童年时期的创伤性事件，我想他一定是对的，他毕竟是心理专家（Gavigan, 1992）。

对于这位患者以及本书第一章开头所讲故事中的 Patricia Burgus 来说，她们现有的症状以及相关的压力源都被有意地忽略了。Bennett Braun 和他的同事们非常主观地忽视了 Patricia 明显的抑郁症状及其背后的原因，却十分努力地去寻找那些能够支持他们所偏爱的 MPD 诊断的证据。同样，他们不对 Patricia 的抑郁症进行治疗，却总是坚持她应该恢复那些在崇拜恶魔的宗教仪式上受虐的记忆。试想一下，如果她接受了短程的、有实验支持的认知行为治疗，或者适当的抗抑郁药物治疗，而不是花费 10 年的时间去接受那些未经证实的、将她推进人性苦难深渊的治疗方法，那她的生活会和现在有多么不同啊。

当我们对这些有问题的治疗实践进行探讨时，首先应该想一想这种情况所涉及的道德伦理问题。下面我们将讨论关于压抑记忆和多重人格的问题，讨论的焦点集中在相关的伦理问题上。

压抑记忆与多重人格

目前，在有关心理健康的伪科学中，最恶劣且最具有破坏力的是与创伤相关的两种心理治疗方法（Lilienfeld & Lynn, 2003）。其一，那些专注于“恢复记忆”的治疗家们声称：当我们遭遇创伤性事件时，大脑会自动地压抑创伤性事件，并将它与我们的意识觉察隔离开来，这是一种对无法应对的情境的适应性反应。若干年甚至几十年后，这些压抑的记忆就会以心理紊乱和心理障碍的方式显现出来，而这些心理问题只有在个体对压抑记忆有充分认识后才能有所减轻。因此他们认为，恢复记忆是治疗的核心部分。

其二，专注于诊断和治疗 MPD 的治疗师认为：当我们面对创伤性事件时，我们的意识觉察会出现分裂或者分离，就产生了“转换性人格”，并藉由这种人格来存储创伤性经历，保护自己不受伤害。如果我们此后表现出了某些心理紊乱的症状，那就是受到了分裂的人格的影响。

实际上，恢复压抑记忆和诱发转换性人格的治疗技术是相似的。在上面两个案例中，治疗师都使用了高度暗示和诱导式提问的方式来询问患者过去的情况。他们还劝告患者不要再怀疑创伤性记忆与 MPD 的存在，因为大部分患者开始时往往否认那些稀奇古怪的创伤性记忆的存在，或者不同意 MPD 治疗师的观点。许多治疗师在治疗中鼓励患者积极地想象，完全不受意识支配地任由幻想驰骋 (Poole, Lindsay, Memon, & Bull, 1995)。这些治疗者会建议患者将他们的想法、感受以及梦境等都详细地记录下来，而这些内容随后都会成为证据用以解释压抑的记忆或闯入意识的人格。异戊巴比妥钠、美索比妥钠，或者硫喷妥钠等所谓的“真实血清”药物，还有催眠技术，都是十分常用的治疗手段。最终，治疗师总是成功地引出了被压抑的记忆或转换性人格，而且与其最初所怀疑的极其一致。

现在你可能已有所怀疑，这两种治疗方法以及它们所依据的理论基础都存在着很多逻辑性及经验性的问题。下面，我将总结一下记忆恢复或 MPD 的诊治中所存在的一些主要问题 (Dawes, 1994; Lilienfeld et al., 1999; Lilienfeld & Lynn, 2003; Loftus, 1997; Ofshe & Watters, 1996; Pendergrast, 1995; Piper, 1997, 1998; Sagan, 1995; Spanos, 1996)。

1. 自 19 世纪弗洛伊德提出了“压抑” (repression) 这个术语以来，还没有任何实验性的证据可以证明其相应的心理机制的存在。当然，遗忘信息，甚至抑制 (suppression) 信息，有意地将它与意识觉察隔离开来，这的确是有可能的，然而，大脑似乎并没有压抑或者自动地封闭信息。在弗洛伊德后期的著作中，他自己也否定了压抑机制，并承认他曾经强迫患者产生了一些幻想 (Ofshe & Watters, 1996)。

2. 在那些声称出现了压抑记忆的案例中，有一种奇怪的矛盾现象：没有创伤回忆或记录的个体被认为存在压抑记忆，而那些有创伤历史记录个体则总是不能忘却痛苦的创伤性事件。

相关案例：对于那些以往在少数民族聚居区、货车车厢以及德国纳粹集中营里遭受过难以言表的虐待的儿童而言……并没有证据表明他们发展成了 MPD (Bower, 1994; Des Pres, 1976; Eitinger, 1980; Krystal, 1991; Sofsky, 1997)，或者其中有任何人分离了或压抑了他们的创伤性记忆 (Eisen, 1988; Wagenaar & Groeneweg, 1990)。同样的，对亲眼目睹父母被谋杀的儿童 (Eth & Pynoos, 1994; Malmquist, 1986)、被绑架的儿童 (Terr, 1979,

1983)、被虐待的儿童 (Gold, Hughes, & Hohnacker, 1994) 以及其他类型的儿童 (Chodoff, 1963; Pynoos & Nader, 1989; Strom, Refsum, & Eitinger, 1962) 的研究也得到了同样的结果。受害者们既没有采取压抑创伤性事件的方式而将其遗忘, 也没有发展为 MPD (Piper, 1998)。

事实上, 那些遭遇过创伤性事件的人不仅不倾向于压抑对该事件的记忆, 而且恰恰存在相反的问题: 侵入性的回忆、痛苦的闪回以及创伤再体验, 这些都是创伤后应激障碍 (post-traumatic stress disorder) 的主要特征 (1994)。那么, 为什么只有那些所谓压抑了创伤性记忆的人, 正好没有任何创伤性记录呢?

3. 非常值得注意的一点是, 那些接受记忆恢复治疗的患者, 仅仅只恢复了与治疗师的信念一致的那些记忆。在考虑心理紊乱的创伤来源时, 记忆恢复治疗师总是倾向于赞成以下三种理论中的某一个: 儿童性虐待、崇拜恶魔仪式上的虐待, 以及外星人绑架。每种类型的创伤都有治疗师专门进行研究。但是, 如果相信恶魔仪式虐待的治疗师恰好只恢复了患者在崇拜恶魔仪式上被虐事件的有关回忆, 这难道不令人感到震惊吗? 同样, 相信外星人绑架的治疗师只诱导出患者有关被外星人绑架的回忆, 而那些相信儿童性虐待是大部分心理障碍之原因的治疗师刚好诱导出关于童年性虐待的记忆 (Sagan 1995), 都不禁让我们感到困惑与震惊。在治疗师的信念与被恢复的记忆之间存在不可思议的吻合与一致, 这不得不使我们对这种巧合的可能性提出质疑。

4. 从来没有人否认, 儿童性虐待是一个严重的问题, 毕竟这是一种极其不好的行为, 而且会产生潜在的、长期的负面影响。不过, 我们有充足的理由怀疑外星人绑架或恶魔仪式虐待是否真的伤害过什么人, 因为从来就没有什么证据可以证明的确有这类事件发生过。在第7章里, 我们已经对“外星人造访地球”或宣称“被外星人绑架”的案例进行了回顾与探讨, 同时对其进行了驳斥。此外, 包括崇拜恶魔仪式在内的大阴谋理论 (还往往是跨国界的、跨种族的大阴谋), 也是很可疑的, 原因详见第11章。在宗教运动盛行的时代和地方, 关于崇拜恶魔活动的主张在文化和历史记录中经常可以见到, 这可能是因为特定的宗教信仰, 使得关于恶魔存在的信念成为必要。而且, 无论执法专家是拥护刑法还是《十诫》* (Ten Commandments), 他们都倾向于魔鬼之说。但是, 这些令人难以

* 十诫是犹太教的经典, 又是希伯来国家的基本法律文献。——译者注

置信的主张也是完全未经证实的。对成百（Lanning, 1989, 1991）或上千（Goodman, Qin, Bottoms, & Shaver, 1994）所谓崇拜恶魔仪式上的虐待进行调查，结果表明，所谓在仪式中有动物或人（婴儿或成人）被杀害、吃掉，或活埋等等，并没有什么客观的证据。

5. 与一些治疗师的信念恰恰相反，我们的记忆并不是对过去事件的精确回忆，而是对我们既往经验的重构（Loftus, 1997）。我们倾向于只储存事件的“线索”，因此当我们回忆时就会填充进去许多可疑的细节。以下是一个简单的证明：回忆一件你曾体验过的快乐事件，什么都可以，但必须是你亲身经历过的，而且你可以百分之百准确地想象出这件事。现在闭上你的眼睛并且回忆出这样的场景，在脑子里出现相应的画面以后回答以下问题：在这个场景中你看见自己了吗？大部分人的回答是肯定的。如果你也是这样，那么它就不是真实的记忆而是对过去场景的重组，因为当你在储存原始记忆时，你不可能与你的身体分离。记忆线索无疑是准确的，但你是靠训练有素的猜测或别人事后传达给你的信息，对场景的其他部分进行了填充。因此，认为记忆绝对可靠，以及“我们应该赞成，所有恢复的记忆都是有效的”，这些观点都是错误的。

6. 虽然这一点可能让人感到不安，但目前的确没有什么十分简单的方法，可以用于确定我们记忆的有效性。闪光灯记忆（flashbulb memories）是指对极重大事件的记忆，例如对肯尼迪被刺，“挑战者”号航天飞机爆炸，或者“9·11”恐怖袭击的记忆等。有关闪光灯记忆的研究已经证明，无论是其生动逼真的细节还是情感的力量，都与我们回忆的准确度无关（Neisser & Harsch, 1992）。例如，以下是一位大学新生在“挑战者”号航天飞机爆炸之后，对爆炸那天自己所经历事件的记录：

爆炸的当时我正在上宗教课，这时有一些人走了进来并且开始谈论这件事。除了知道“挑战者”号爆炸了，以及让我感到特别难过的那件事——那名教师的学生全都在现场观看到了那悲惨的一幕，我不知道任何其他细节。

同样是这位学生，三年以后，当她已经是大四的学生时，她对那段经历的描述已经相当不同了：

当我第一次听说“挑战者”号爆炸时，我和室友正坐在新生宿舍里看

电视。这时插播了一个简短的新闻报道，我们都感到十分震惊。我觉得很不安，于是上楼去和我的一个同学谈了一会儿，然后又给我父母打电话。

后一个陈述是不真实的，尽管它包含了大量的细节、丰富的情感，并且陈述者对此很有自信。最近有一项研究让学生们回忆第一次听说“9·11”恐怖事件时的情境，并同时评估他们对该记忆属实的信心，从而进一步阐明了闪光灯记忆的本质 (Talarico & Rubin, 2003)。所有的被试都在 2001 年 9 月 12 日第一次进行了对“9·11”事件的回忆，同时也记录了对当天日常生活的回忆，然后在那之后的第 1 周、第 6 周与第 32 周又分别进行了回忆。以下是该研究者对于其发现的总结：

闪光灯记忆与日常记忆在回忆的一致性上并无差异，它们的一致性都会随时间而下降。然而，只是在日常记忆中，对记忆的生动逼真程度、对信息的回忆程度以及对记忆准确性的信心出现了评分的下降……就像之前所说的，闪光灯记忆在记忆的准确度上并无特别之处，它的特殊之处表现在“自我感知”的准确度上。

7. 就像世界上不存在完美的“测谎仪”一样，也没有完美的“真实血清”药物。这些药物只是降低我们的抑制力，并使我们处于更容易受暗示的状态。因此，在药物的影响下所提取的记忆并不一定准确 (Ofshe & Watters, 1996)。

8. 催眠也带有很强的暗示性。虽然通常我们能够在催眠引导程序的作用下提取出更多的记忆，但真实记忆与虚假记忆之间的比例是不变的 (Orne & Dinges, 1989)。此外，一些治疗师使用的催眠性的“年龄退行”技术 (Hypnotic age regression) 要求患者重新体验早年生活的某个时期，他们在使用该技术时只是鼓励患者参与其中进行角色扮演。使用催眠法促使人们回到过去的生活或想象未来的生活 (年龄推进: age progression) 要比退回到儿童时期容易一些，而且这些记忆都与在年龄退行中获得的记忆保持了相当的一致。因为人们并不能令人信服地证实过去某些生活的存在，或者自身对未来的预见能力，所以个体对这些记忆的过分自信，使得催眠引导出的记忆的效度受到了质疑 (Spanos, Weekes, Menary, & Bertrand, 1986; Spanos, Menary, Gabora, DuBreuil, & Dewhirst, 1991)。

9. 正如第 2 章所讲的，通过一系列暗示性和诱导式提问，尤其是混杂着前述的那些提问技术，可以将错误记忆植入到我们的记忆中 (Loftus, 1997)。那些记忆表面看起来使人不得不相信，但实际上却是错误的 Ceci 和 Bruck (1995)。

报告称，当受人尊敬的、地位很高的治疗师与病人反复进行多次面谈，且面谈过程中存在着强大的验证性意见的影响时，病人出现错误记忆的风险就会大大增加。因此，即使在许多周、许多个月甚至是若干年之后我们还发现，对于那些急切信任治疗师的智慧以求解脱的患者来说，治疗师产生了巨大的影响，甚至导致了戏剧性的效果，也就不足为奇了。

10. 在实验室条件下，也可以观察到多重人格的设定情况。完全正常，健康的个体也可以理解社会线索，并据此表现出不同的人格，而且倾向于根据当时的情境表现出所需的人格特点（Spanos, Weekes, & Bertrand, 1985）。

11. 多重身份的设定是不同文化背景中的普遍现象（Spanos, 1996）。纵观人类历史，那些有恶魔附体，或是歇斯底里大发作的个体，都倾向于表现出与他们所处社会的文化期望相一致的行为。现代的所谓 MPD 病人也是如此。实际上，与文化期望相吻合的多重身份可以通过不同的途径诱发出来，如各种文化团体的宗教仪式和传统，或者心理研究者在实验室的简单引导，而不一定与创伤、转换性人格或其他心理障碍有什么关系。

12. 如同我们在第 8 章所指出的，目前尚没有证据表明那些记录在案的既往创伤经历与 MPD 的诊断存在相关，更不能说创伤是 MPD 的原因了（Lilienfeld et al., 1999; Lilienfeld & Lynn, 2003; Ofshe & Watters, 1996; Piper, 1997）。不幸的是，由于《女巫》和《三面夏娃》这两个真实性未经证实的案例，藉由媒体的作用已经家喻户晓，导致许多人都十分相信“创伤会导致 MPD”这个完全没有根据的结论。

13. 在过去二十多年里，被诊断为 MPD 的患者比例急剧上升。我们在第 9 章曾提到，在媒体大幅报道 MPD 之前，全世界相关文献中报道患有 MPD 的成年人大约只有 200 多人，而在此之后被诊断为 MPD 的患者比例急剧上升，高达 5% ~ 10%（Ross, 1991）。难道是因为 MPD 这种疾病突然之间变得相当普遍了吗？为什么在没有接受心理治疗的人中，没有发现这种疾病的蔓延呢？

14. MPD 几乎只在美国和欧洲人群中被发现，而且相对而言，较少有临床医师发现 MPD 的存在，这一现象又该如何来看待呢（Thigpen & Cleckley, 1984）？一项研究对瑞士所有的精神病学家进行了调查，结果显示，90% 的精神病学家从来没有见过一例 MPD，而有三位精神病学家分别给二十多位病人下过 MPD 的诊断。研究的统计结果是：1% 的瑞士精神病学家（6/655）给出了该国三分之二的

MPD 诊断 (Modestin, 1992)。因此, 对这种数据模式最合理的解释似乎只能是: 少数专门研究创伤的治疗师, 表现出了非常明显的验证性意见, 假定患者存在一些实际上未经证实的创伤记忆, 并且完全没有意识到他自己在创造患者多重身份方面所起的作用。

15. 根据治疗师的诊断, 患者形成第一种转换性人格的平均年龄正在下降, 即个体首次出现“分裂人格”的年龄在下降。20 世纪期间, 该平均年龄从 22 岁降到了 11 岁 (Goff & Simms, 1993)。Coons, Bowman 及 Milstein (1988) 的估计则只有 7 岁。那么, 为什么没有一个存在多重人格的儿童案例呢? 同样的, 在那些没有接受心理治疗的人中, 为什么没有出现 MPD 呢? 既然说 MPD 是以个体某些时候的怪异行为为特征, 那么对于整天有许多成年人照顾和观察的儿童而言, 出现了 MPD 的相应症状后, 要被识别出来肯定不是什么难事。

16. 上述第 13 条和第 15 条已经表明, 在治疗开始之前, MPD 从来就没有被真正地观察到。相反, 总是在专门研究创伤的治疗师对患者进行治疗的数月或数年后, 患者被诊断为 MPD (Ofshe & Watters, 1996; Piper, 1997)。事实上, 个体往往是在已接受了六到七年的治疗后才被诊断为 MPD。在如此长的时间里, 无论是治疗师、家庭成员、朋友、同事, 还是其他与 MPD 患者交往的人, 即使是其中最细致的观察者也没有注意到 MPD 患者存在那么戏剧化的表现, 这又怎么可能呢? 还有, 为什么总是在治疗开始几年以后, 个体才会显现出典型的转换性人格和一些古怪的行为呢?

所有这些未能回答的问题和有关的研究证据都清晰地告诉我们: 压抑记忆和转换性人格很可能是社会期望和治疗性暗示的产物, 在这一过程中, 许多评估方法和治疗行为都存在很多问题。Spanos (1994) 曾总结了 MPD 的社会结构属性: “MPD 是一种受情境制约、目标导向以及受重要人物影响的社会行为, 而且, MPD 的特征会根据期望的改变而在不同时间有不同的改变。”这一描述在许多方面同样适用于记忆恢复。因此, 对这些治疗行为来说, 最重要的评价就是它们均是未经证实的治疗方法, 而最危险的就在于它们代表了专业的误导。

尽管证明其安全性及有效性的证据十分匮乏, 伪科学治疗仍然大受欢迎, 这意味着许多人——包括治疗师和病人, 都放松了在证据方面的标准。虽然我们对周遭的事物可以自由选择信任与否, 但涉及卫生保健问题时, 我们必须遵从已得到广泛认同的伦理原则。下面我们还要探讨两个特殊的原则: 首先是治疗必须有

净收益的要求，即治疗必须利大于弊，也就是伦理学家通常所讲的更好的投入—收益比。其次，应该对知情同意权和患者自主权这两个紧密相关的问题加以评估。基于上述分析我认为，推行未经证实的治疗方法不仅是不符合伦理学要求的，即使只是促进或赞成这种方法的实施，也是有悖伦理的。

投入—收益比

伪科学的支持者们有时会提倡使用一些没有充足实验依据支持的治疗方法。当我们对这类未经证实的治疗方法进行伦理学考察时，必须要问的一个问题就是，该治疗的收益是否比投入大。然而，伪科学家经常试图回避对投入—收益比的估计，转而声称他们的治疗是无害的，可能对患者帮助很大，因此患者不会有任何损失。换句话说，就是“为什么不试试看呢”。这种不负责任的询问其实意味着一种“谁知道呢，任何事情都可能发生”的态度。但是，这一随意的方法存在三个伦理学上的问题——伪科学的支持者们高估了获益的可能性，同时低估了潜在伤害的可能性，并且因为上述两个原因，根本没有考虑放弃已经证实有效的药物治疗可能造成的损失。

利益与伤害的曲解

即使以创伤为中心的治疗有其有利的一面，它的好处也常常被它的实施者所高估。如前所述，并没有证据表明，记忆恢复或 MPD 的相关治疗对患者会有什么好处。对这一问题的对立观点双方进行调查，也没有任何科学数据显示这些技术具有治疗性的价值（Lilienfeld et al., 1999; Ofshe & Watters, 1996; Piper, 1997）。

与此同时，尽管有证据表明这些治疗实际上可能会对患者造成严重的危害，但其危害的可能性却总是被低估。Loftus 曾引述了 Parr 的一项研究（Parr, 1996a, 1996b），该研究对 30 位在治疗中恢复了创伤性记忆的患者进行了追踪：

在记忆恢复之前，只有 3 人（10%）想过或尝试过自杀；在记忆恢复之后，人数增加到了 20 人（67%）。在记忆恢复之前，只有 2 人（7%）住过院（大概是因为精神疾病）；记忆恢复后这一数字增加到 11 人（37%）。在记忆恢复之前，只有一个女性有自残行为；而在记忆恢复之后自残人数增加到 8 人（27%）……在治疗之前，大部分患者（25/30 = 83%）有工作，但

在接受了3年的治疗后，只有3人（ $3/30 = 10\%$ ）仍有工作。在治疗之前，23人（ $23/30 = 77\%$ ）已婚；而在3年的治疗期间，这23人中有11人（ $11/23 = 48\%$ ）分居或是离婚。共有21位有孩子的患者，在治疗后，有三分之一的人丧失了监护权。此外，所有的30位患者都疏远了他们的家人。

如果将这些患者与另一组有类似症状但接受了已证实有效的治疗方法的患者相比，那结果可能会更有意义。无论如何，上述这些结果肯定不是大部分接受治疗的人所期望的结局。我们需要进行控制更严格的研究，来帮助我们判断是否记忆恢复治疗真的像Parr的研究所说的那么危险。

那么，那些被诊断为MPD并且接受了MPD相关治疗的个体又怎么样呢？Piper（1998）总结了一些关于MPD治疗的资料：

MPD患者在治疗过程中常常出现病情的进一步恶化。关于这一症状的主流认知是：MPD心理治疗“导致了患者在治疗情境之外的生活受到严重的破坏”，而且被诊断为MPD的患者几周之后可能出现自杀企图（Putnam, 1989）。在MPD心理治疗过程中，患者可能变得更退缩、更焦虑，或更抑郁（Braun, 1989）；接受治疗的时间越长，他们的所谓童年虐待故事就越精彩、越复杂，也越离谱（Ganaway, 1995；Spanos, 1996）。这种恶化导致患者需要长时间住院治疗，有些甚至花费了数百万美元（Frontline, 1995；Piper, 1994），而这种情况所涉及的患者，通常都有医疗保险，其治疗师也往往是MPD的狂热支持者。此外，在患者被诊断为MPD后，他们住院治疗的次数也越发多了（Piper, 1994；Ross & Dua, 1993）。

被忽略了的机会成本

除了高估好处且低估危害性之外，伪科学治疗的支持者们还完全忽略了经济学家所说的决策中的机会成本。当选择某一种方案时，我们就会错过使用另一方案可能带来的好处。也就是说，如果一个未经证实的治疗是已被选择的“机会”，那么其成本就包括丧失有科学依据支持的治疗方法所带来的收益、病情未受控制而出现恶化，或者治疗本身造成病情加重等等。

例如，在《新英格兰医学杂志》（*New England Journal of Medicine*）上，Coppes、Anderson、Egeler和Wolff（1998）曾描述了这样一个案例：一位15岁

的男孩患有何杰金氏病（Hodgkin's disease），他没有接受医生建议的化疗而选择了草药治疗，两周以后病情恶化了。等到该男孩改变主意要进行化疗时，癌症已经发展到需要用更大剂量的化疗药物才能控制的地步。在这个案例中，草药的机会成本就是病情的恶化。

对于记忆恢复治疗或 MPD 的诊疗而言，实际的机会成本就是放弃了原先要给予患者的有实验依据支持的治疗方案。如前所述，焦虑和抑郁是心理疾病患者最常见的两个症状，针对其中任何一种症状都有好几种已知有效的治疗方案，而选择其他治疗方案的机会成本就是放弃这些治疗方案（Chambless et al., 1996; Dobson & Craig, 1998）。

在所有的机会成本中，也许最沉重的代价就是我们在第 12 章中所讨论过的——宗教信仰导致的儿童医疗忽视。每年都会有许多儿童无辜地丧失生命，原因仅仅是他们的父母拒绝药物治疗，却寄希望于宗教信仰能治愈疾病。

如果把未经证实的治疗方案其获益的不确定性、遭到伤害的可能性，以及其他机会成本综合起来考虑的话，这些治疗的投入—收益比往往是负值。这就意味着，总体而言，伪科学造成的危害很可能远远大于从中获得的收益。而且，所有鼓吹伪科学的人总是鼓动别人去选择一种更有可能受到伤害而非获益的方式。因此无论从哪个标准来看，这样的行为都是不符合伦理要求的。

真正的知情同意权以及患者自主权

另一个重要的伦理原则是：患者必须在治疗之前就知情与同意。这样做是为了保障患者的自主权并避免治疗师的个人独断。侵犯一个人的自主权就是损害了他的个人自由。而充分告知患者他可获得的每种治疗方法的潜在益处和风险，以及每种益处和风险发生的可能性大小，是所有医务工作者的专业职责。病人也只有在得知了所有的这些信息之后，才可能作出真正的知情选择。

如果将病情告知重症疾病的患者，他们有可能变得非常绝望，这时治疗师就陷入了伦理的两难境地。众所周知，当希望破灭时，癌症患者有时会盲目地寻求各种稀奇古怪的治疗方法。当一个绝症患者被告知已经没有什么有效的医疗方法可以用于他的疾病时，他又会多么的悲伤和焦虑，这一点我们完全可以想象。病人总是相信一定还可以做些什么的。一些人会拒绝接受已经没有有效治疗方法这

一事实，因为科学已经取得了如此巨大的成就。令人惊叹的突破、疾病被攻克，以及痛苦有所缓解等等，早已经司空见惯，所以他们根本就无法相信竟然没有希望重获健康。然而，的确有很多疾病是没有有效的治疗方案的。假装有治疗的可能，那就是在利用这些易受伤害的病人。

请看下面两个癌症治疗案例：

- Donald MacNay 向癌症患者提供了一种癌症治疗方法——芦荟静脉注射疗法，并从中获利 12,000 美元，但他的这一疗法既没有被美国食品与药品管理局批准，也没有得到国家癌症组织的认可。在 3 名癌症患者因这一治疗方法死亡之后，弗吉尼亚州卫生管理局吊销了 MacNay 的执照。

- Sandor Olah 将 Essiac 草药茶注射到 54 岁的病人 Petra Hall 的静脉中，来治疗其癌症，但 Hall “在接受这一治疗后开始出现了呼吸急促及其他症状。尽管如此，Olah 还是让 Hall 回家而不予任何处理。第二天，Hall 住进了密歇根州的 Hackley 医院，并被查出已有多器官系统衰竭，11 天后抢救无效死亡。导致她死亡的主要原因是维他命治疗所致的呼吸窘迫综合征”（美国国家反健康欺诈委员会时事通讯 1997 年 5/6 月刊）。

尽管上述患者在接受这些治疗前都签署了同意书，但那并不是真正的“知情”同意。这些案例证实了，对于请求病人同意的特定治疗方法，必须向病人详细说明和解释其所有可能的益处及可预见的风险，这样做有非常重要的意义。

在以恢复压抑记忆或诱导出多重人格为目标的治疗中，患者的自主权通常会受到侵犯。例如，在这些治疗过程中，不允许患者对治疗师的判断提出异议。所谓压抑记忆的症状之一就是缺乏受虐待的记忆。MPD 运动的领军人物（如 Putnam, 1989）声称，即使患者否认患有 MPD 并且没有表现出该疾病的症状，也应该对他们进行这方面的治疗。Kluft（1983）甚至暗示说，并非一定要获得患者的同意才能做 MPD 治疗，“催眠工作应该首先获得知情同意，只要这一点在现实中行得通”。就像 Piper（1997）所评论的，“人们不禁有所疑问，Putnam 和 Kluft 是否意识到了，医生不屑于做那些有关知情同意的繁琐杂事，伦理学和法律的权威们对此都非常反感”。

这种态度与那些在中世纪追杀女巫的牧师和法官的态度相比，又有什么不一样呢？一些治疗师显然就像中世纪的牧师一样坚信自己必然正确，主张患者应该

因他们的压抑记忆或 MPD 接受治疗，无论他们自己愿意与否。在中世纪，人们把被指控的女巫投入湖内，如果她们沉下去就表示她们是无辜的，如果浮上来则表示她们是有罪的，要被烧死。这两种结果看起来都不妙。如果治疗师将创伤性记忆或转换性人格作为患者当前疾病的主要原因，因现时存在一些具体操作的困难而拒不考虑其他任何解释，那么他们的病人所面临的处境也就与被投入湖中的女巫差不多：按治疗师的方法治疗会有更大的危害，而不接受治疗病情也会恶化。

当医生采用未经证实的治疗方法向病人提供虚假的希望，并误导他们的时候，很显然是违背了知情同意的原则。如果不能向患者展示支持该治疗方法的证据，也没有向患者介绍其他可以采用的、已有研究证实其功效的治疗，或者没有向患者充分地解释所建议治疗方案的已知风险，那么就不是真正的知情同意。伪科学家们在他们的广告以及专业实践中总是违背这一伦理学原则。这些事情所致的危险状况不禁让人想起第 5 章所描述的处于自由混战状态的“膳食补充剂”市场。在本章的结语部分，还将讨论一些提议的立法法案，它们会导致这一状况出现更为严重的恶化。

“健康自由”的谬论

《医疗法》（Medical Treatment Act）提案与《美国膳食补充剂保健与教育法案》的情况颇为类似。《医疗法》也是反对在药品市场实施消费者保护措施，而且这一新提案的影响更加广泛，给消费者带来了巨大的威胁，因为它允许治疗师在获得患者同意的前提下，采用任何未经证实的治疗方法。治疗师是否要对治疗的收益或伤害的可能性进行衡量，对此没有任何强制性的措施，而仅仅只有一种含糊的警告：“这种食物、药物或仪器的安全性和有效性，尚未得到联邦政府的认可，任何使用这种食物、药物或仪器的人需自行承担风险。”这其实是给那些不符合伦理要求的治疗方法，披上“让消费者拥有选择的自由”这样的合法化外衣。在美国国家反健康欺诈委员会时事通讯的 1997 年 9/10 月刊上，对此是这样解释的：

显然，该法只会给那些在可疑的替代/补充药物之间进行选择的医疗保健人员以自由。人们肯定不希望被骗，也不希望被伤害，这是首要的原则。

如果要他们接受未经证实的治疗方案，唯一的理由就是有人告诉他们“这个治疗是安全而有效的”。重病之中的绝望患者，往往会愿意尝试医疗保健人员所说的任何“可能有帮助的”治疗方法。

离间对于患者的行为也起到了重要的作用。那些非常规的医疗从业者经常说服他们的患者，使他们相信是一个压制性的、腐败的官僚机构在阻止这些治疗方法进入市场。即使是“获得联邦政府……认可”这样的用词也比“在研究证据的基础上判定”的说法更具偏见。《医疗法》这一提案是江湖医生们梦寐以求的，因为它诱导患者签字放弃了他们自身的权益。

如果法律通过了像《医疗法》之类的提案，那么不仅你我这样的普通人要对不计其数的药物（包括草药和同种疗法的药物等）的安全性和有效性完全负责，而且我们还必须自己来判断治疗师所建议的治疗方法是否安全有效。要做到这些，前提是我们能接触到相关的医疗科技期刊，能阅读医疗专业的文献，而且还能综合分析那些分散的、甚至相互矛盾的研究结果。另外，进行这项极其耗时的、对科学研究结果进行总结的任务其前提条件是，要有那些相关科学研究资料的存在，而这样的研究肯定是不被新法案实施后的新医疗系统所支持与鼓励的。如果研究证据不再是医疗专业领域中的通用标准（原本这一针对治疗方法有效性的评价标准是由治疗者、保险公司以及相关法律来共同进行适当判断的），那么相应研究的质量和数量会受到怎样的冲击啊，更不用说对医疗服务质量会产生什么样的影响了。

任何企图允许治疗者们提供无研究证据的治疗方法的行为，都是极其危险的。只有伪科学家才希望这样的自由越多越好。《医疗法》将对我们苦心建立的科学大厦造成严重的冲击，它会将那道阻止开展不合伦理的医疗活动的“防洪闸”打得更开，甚至引发洪水泛滥。难道要做到对关于我们健康的问题进行知情选择，真的有那么难吗？一旦患者的知情同意是在一种公开地对医疗欺诈行为采取宽容态度的系统中进行，我们就会很容易地成为不法医疗活动的目标。最终，医疗责任只会由我们自己来承担，因为正是我们自己为自己的健康和生命做了一个危险的选择，尽管这一选择是在缺乏消费者安全保障体系（任何文明社会都会为其公民提供的保障体系）的情况下做出的。难道这是我们期待的未来吗？

16. 工具：关于批判性思维的建议

当 Patricia Burgus 从 Bennett Braun 的治疗中脱离出来以后，这位“女祭司”关于崇拜恶魔仪式活动的记忆就开始衰退。她在治疗中建立的虚幻世界开始变得越来越不真实，她对虚幻世界的信念也开始动摇并最终完全倒塌。尽管造成 Patricia 抑郁的原因并不是那么简单或者容易解决，但终究是由明确的问题导致的。然而，在她寻求治疗以后的许多年里，她的生活被彻底打乱了。在错误的治疗过程中，她和两个儿子还曾遭受过据称有正当理由的侮辱。后来，Patricia 对 Braun 及其同事提出了民事诉讼。1998 年 10 月，案件以判处赔偿 Patricia 1, 060 万美元宣告结束。再多的钱也无法弥补 Braun 对 Patricia 及其家庭所造成的伤害，不过，这一判决结果的确是给他们自 1982 年就开始的可怕经历划上了一个句号。

1998 年 8 月，在 Patricia 案公开判决前不久，伊利诺伊州职业管理部门正式控告 Bennett Braun 及他的两位同事——心理学家 Roberta Sachs 和儿童心理学家 Elva Poznanski。在长达 23 页的控告书中，称 Braun 有“严重疏忽”、“不光彩、不合伦理及违反职业要求的行为”、“在治疗任何心理或生理的疾病及相关问题时，只根据自己的意思命令患者接受所推荐的药物、治疗或矫正方法，并对所采用治疗方法的技术、效果以及价值等进行了错误的或者误导性的说明”、“治疗行为已被证明是无效或者不当的”。以下是 Bloomberg（1999）引述其中“严重疏忽”控诉部分的一些内容（其中“P. B.”代表 Patricia Burgus）：

- 在指定 P. B. 不同“人格”的过程中，使用了一个非正规的治疗方案。

- 在实际上没有“多种人格”存在的情况下，被告（Braun）仍错误地执行上述治疗方案，鼓励并帮助 P. B. 发展和建立了被认为是“转换性”人格的“多种人格”。

- 错误地执行上述治疗方案，在这一过程中，使用了暗示技术和强制技术鼓励和帮助 P. B. “回忆起”受虐待的事件。

- 告诉 P. B. 及其丈夫，在她和她孩子治疗的过程中发现了“压抑记忆”，而那些记忆是对确实发生过的历史事件的真实回忆。

- 告诉 P. B. 及其丈夫，的确有一个有组织的、跨越世代的、国际性的、崇拜恶魔的邪教组织存在，它进行了大量的谋杀行动、拷打和折磨，在崇拜恶魔仪式上进行虐待、用人类作为祭品，还有其他类似的活动，而且关于这个邪教组织及其从事不法活动的事情已是尽人皆知了。

- 反复说服 P. B. 及其丈夫并且使他们确信，无论他们当中谁对崇拜恶魔仪式上的虐待行为或者关于这类虐待行为的记忆的存在表示怀疑，他们都是唯一提出质疑的人。

- 告诉 P. B.，她曾经对她未成年的孩子有过性虐待行为。

- 告诉 P. B. 及其丈夫，P. B. 让他们未成年的孩子多次参与了不同类型的崇拜恶魔仪式活动，而且在仪式过程中有将人和动物当祭品、吃人以及各种拷打和折磨人的行为。

- 没有充分告知 P. B. 及其丈夫，在其治疗中使用的技术可能存在产生虚假记忆的风险。这类虚假记忆所涉及的事件，虽然实际上从未发生过，但患者却会感觉像是真的。

其他指控则针对一些有问题的信念与治疗行为——那些恢复了 P. B. 压抑记忆并诱导出其转换人格的治疗师们所持有的信念或进行治疗：

- 没有告知 P. B. 及其丈夫，压抑记忆理论缺乏科学的效度，在科学界也没有被广泛接受。

- 没有充分告知 P. B. 及其丈夫，多重人格障碍的诊断是颇具争议的，该诊断在心理卫生领域还没有被广泛接受，因此做出这个诊断是存在一定风险的。

- 没有告知 P. B. 及其丈夫，心理卫生领域对于多重人格障碍还存在更

深的质疑——这一诊断可能被过度使用了。许多被诊断为多重人格障碍的人后来被发现并没有这一疾病。

- 没有告知 P. B. 及其丈夫，不恰当的治疗方法有引起“多重人格障碍”表现的可能性。

Patricia 对 Sachs 和 Poznanski 也有类似的指控。但当 Poznanski 同意作证，出来指控 Braun（主管所有被起诉的治疗工作的主要被告）后，Patricia 撤销了对她的指控。可能是因为出现了这一转折点，Braun 在审讯正式开始前不久，就其涉及的案件及起诉开始进行调解和协商。最后他的医疗执照被吊销，被处以罚款，赔偿额度达到了 1,060 万美元，并且强制接受医疗再教育。这一针对 Braun 的民事诉讼案判决，以及其他患者因类似治疗遭遇而成功地控告了治疗师的案件审理结果，给了那些使用这类不可靠技术的治疗师们一个强烈的警告。1996 年，Ofshe 和 Watters 进行了如下的描述：

我们很高兴地发现，形势已经发生了改变。1991 年，当我们开始报道记忆恢复治疗的时候，这一理论和治疗行为几乎将要被制度化，作为常规的治疗；而现在，仅仅是几年之后，人们就已习惯对这一疗法提出质疑，而且很快废除了这种治疗方法。在 1996 年初，许多记忆恢复治疗的支持者都忙着推翻自己“发现”了一个国际性的崇拜恶魔宗教组织、这一组织可能虐待和“策划谋害”大量儿童的说法。公众对于在心理治疗过程中恢复记忆之说，也逐渐从接受到不确定，再到怀疑，直至最终彻底摒弃这一观念。在此之前，许多治疗师争先恐后地将记忆恢复理论应用到治疗中，将其市场化，并撰文对其大唱赞歌，称这一疗法为“治疗学上一次革命性的飞跃”，而现在他们早已偃旗息鼓、铩羽而归了*。

在医疗保健行业中伪科学方法的危险性

虽然上述的伪科学治疗方法正逐渐地从治疗实践中退出，这让我们感觉安全

* 摘自《制造怪物：错误记忆、心理治疗和性癡症》（Making Monsters: False Memories, Psychotherapy, and Sexual Hysteria），Richard Ofshe 和 Ethan Watters 著。

了一些，但是谨记这些事件带给我们的经验教训是十分重要的。所有从事医疗保健工作的人，以及那些渴望成为这一领域专业人员的人，都应当特别注意，一旦放弃了科学推理的思维，会产生什么样的危害。目前，有一种很普遍但却很危险的现象，许多希望从事医疗科学实践工作（治疗身体或心理疾病）的人，倾向于把这一工作看作是一门艺术而非科学。当学生甚至专业人员都不愿进行艰难的科学思考，一味追求相对轻松的任务，沉迷于那些听起来不错、让人感觉很好的东西，那么一些有问题的治疗方法又会在专业领域重新占有一席之地。

健康医学常常被划分为两个部分，分别反映了两种不同的实践与操作方式：“研究取向”和“实践取向”。那些强调研究的治疗者将他们的治疗性干预建立在实验的证据上，而那些较少关注研究的治疗者在实践时经常以一些传闻轶事、推荐与证明等等个人经验作为其工作的基础，而这种取向的危险性正是本书的主要关注点。研究取向的工作者很少会认可伪科学的论点，但实践取向的工作者却往往很热心地提出和推广一些可疑的“发现”。

例如，我们在第4章中提到过的思维场治疗，是一种轻叩身体、奇迹般治疗心理问题的方法。尽管该疗法在真实性和证据支持方面仍存在很多问题，但它却以十分迅猛的速度发展起来并风靡一时（Gaudiano & Herbert, 2000; Lohr, Hooke, Gist, & Tolin, 2003; Swenson, 1999）。这一治疗理论认为，有节律地轻叩指压点会以某种特定的方式对人体能量场产生影响。但如前所述，所有这类基于能量的治疗理论最终都被证明是不真实的。而且没有可靠的证据证明，这种疗法会给患者带来什么真正的益处。更令人奇怪的是，该疗法的创始人只将他们的方法展示给那些对这种疗法感兴趣并参与他们工作组（这需要花费100,000美元）的人，同时还要他们承诺不会将这种疗法透露给其他人。同样，一些临床科学家（Lilienfeld, 1996; Lohr, Hooke, Gist, & Tolin, 2003）对眼动脱敏与再加工疗法也提出了质疑，在这种治疗过程中，治疗者们只是让患者不停转移注视点，然后不知何故就缓解了其心理障碍。尽管这种可疑的治疗方法其支持性证据十分无力，但它在创伤治疗领域却赢得了众多坚定的追随者。看来只有时间可以告诉我们，诸如此类未经证实的治疗方法，是否会取代压抑记忆治疗与MPD的诊治，成为心理治疗领域中的伪科学潮流。

伪科学仍将继续骗取治疗者和患者的信任。如果忽略了相关的研究证据，在自己的研究中又没有严格地进行影响因素的控制，仅仅依靠道听途说，并且跌入

认知偏差的陷阱，那么，治疗师就会相信那些伪科学技术是有效的，并毫不犹豫地将其应用到医疗实践中。而这一行为的后果最终要由接受治疗的病人承担。

进行批判性思维

很明显，Patricia Burgus 是错误治疗的受害者，她及其家庭成员所遭受的违法治疗是不符合伦理学要求的，也是不可原谅的。然而，如果说 Patricia 本人有什么错的话，那可能就是她没有进行批判性思维。换作另一个更喜欢提问和质疑的人，也许会对治疗师所告知的一些荒谬理论和建议提出异议。科学心理学为我们提供了科学思考问题的方法，使我们可以在感到惊叹的自然感受与健康怀疑论之间取得适当的协调与平衡。虽然需要多加练习，才能将批判性思维的技能加入到我们的推理能力中，但这样做可以让我们获得更加理性的信念系统，并且降低被骗的风险。下面，我要对这本书的所有内容进行一次概括，对如何在判断与决策过程中避免常见偏差的策略进行总结。

运用多种方法进行概念重建

许多问题最终可以被构架为获益或者损失，当面临获益时，我们倾向于反对冒险，而当面临损失时，我们则倾向于寻求冒险。为了避免受到问题架构的影响，请试着用你自己的思维框架来进行思考。举例来说，如果你面临的是要支持一项惊人的“减税”措施，还是批评一项可怕的“增税”措施，你得先问问自己，从支出和获得之间平衡的角度来说，这到底意味着什么。此外，不要被他人的欺骗性语言所影响。请试着超越“选择优先”或“生命优先”这类模棱两可的词语，去思考隐藏于其背后的根本问题，从而避免被一些无意义的词语或其他的语言陷阱所蒙蔽。

警惕一厢情愿的想法

人们对于未知事物的处理至少有两种方法，其中一种就是简单地想象自己希望世界应该如何，并假设所有最令人快乐或欢欣鼓舞的事情都恰好是真的。然而，这种一厢情愿的想法，只是对于理解的一种错误感知，并且还会阻碍知识的继续增长。一旦在你的脑海中已经有那么一个“解释”存在，无论它是多么不

真实，你都不太可能再积极地寻求新的信息或者修正原有的信念。因此，一厢情愿的想法是轻率、不顾及后果的，你应当尽可能地避免产生这类想法。你可以换成另一种处理未知事物的方法——意识到当前知识的局限性。运用这种方法来理解和阐述以数据为基础构建的理论与假设，是十分恰当，或者说是很有价值的。但是在这一过程中，务必要仔细地将未经检验的假设与基于科学证据的信念区分开来。

考察权威的合法性

如果没有合法的专家提供给我们的参考或建议，我们无法在这样一个复杂而且快速变化的世界中生存下来。但是，我们服从权威的倾向有时比较盲目，而一些自称的权威又并不如他们看上去那么专业。因此，对权威或专家进行仔细考察是很重要的。要尽量从一个专家有多少与其自身利益无关的、真正的知识和经验来考虑。当与那些想要利用你的人打交道时，你就应该特别小心。

寻求有风险的检验，而非无力的证实

我们总是倾向于寻求那些与自身信念相一致的信息，并将那些模棱两可的信息优先解释为可以支持自身信念的证据。这种思维方式只会给我们的想法提供十分微弱无力、没有多少信息含量的证实，却不能提供任何验证、摒弃或修正错误观念的机会。应将我们的想法付诸于真正的检验，虽然检验的结果可能会否定我们原有的想法，但唯有这样，才能清楚地判断我们的观点是否真的值得保留。也就是说，只有那些能通过有风险的检验的信念才是有价值的，而其他未能通过检验的信念则应该被抛弃。

不要被个人推荐所误导

当某种可观测的行为或事件存在多种竞争性解释的时候，往往不太可能确认到底哪个解释才是正确的，这就是将推荐、个案研究以及传闻作为证据的最基本的问题。一些人可能基于自身的特定目的而杜撰出一个故事，或者夸大某个事实，此外，诚实、有良好出发点的人也不是没有可能犯错。例如，对于某种疾病症状缓解的问题，除了疾病是因治疗而好转的解释以外，至少还存在 10 种以上的竞争性解释，而且我们根本无法了解究竟哪种解释才是正确的。此外，个人推

荐还可能恰巧成为概率基本规则的特例。因此，个人推荐的过度推广与泛化也可能导致我们形成与事实完全相反的信念。虽然推荐、个案研究以及传闻可以给我们提供一些有趣的线索，但它们无法对假设进行验证，也无法为某种信念或理论提供良好的证据支持。因此，对于个人推荐、个案研究以及传闻等信息来源，我们均应十分小心，谨慎对待。

不要脱离现实

记住，从知识的角度来评价某种观念的合理性，是判断该观念真实性的一个有用指标。在评估一种理论的合理性时，特别重要的是评估它如何阐述某一假定的原因导致可观测效应产生的可能机制。例如，暴露在 X 射线下可以诱发癌症，这与相关科学理论和研究结果是高度一致的，X 射线与癌症之间存在着一种因果关系的机制。但是，假设电磁场可以诱发癌症，就与相关的理论及研究数据不太吻合了，因为这其中的具体机制尚不清楚（还有大量的理论显示电磁场并不会诱发癌症）。如果某种观点缺乏合理的作用机制，而且与那些已得到科学研究支持的观念相矛盾，那么就不值得你关注，更不用说将其作为你的信念了。

相关不代表因果关系

相关并不意味着因果关系，这个原则非常重要，但却很容易被人忽略。在我们身边充斥着大量的相关文献，尤其在医疗领域更是如此。例如，我们总是被告知采用某种特定的饮食方案或某种特定锻炼方法的人，会比其他人更健康或更不健康。但这样的证据并不能证明那些饮食或锻炼方式会对健康水平产生什么直接的效应，因为选择了那些特定方式的人属于生活方式的自选群体，也就是说他们的生活方式可能在很多方面都与其他人不同。仅仅基于相关性的研究证据就得到因果关系的结论，这种现象很容易出现，但其实相当欠考虑。如果你只有相关的数据，却要进行有关金钱、选票或信念之类的重要选择，那就一定要坚持在得到一组指向同一结论、彼此协调一致的相关数据的情况下再做选择。

小心媒体悖论

避免被你所看到、听到或读到的大众媒体信息过分地影响。时刻牢记，新闻和娱乐信息总是会精心挑选一些不同寻常的特殊事件来吸引受众的兴趣，而且以

一种十分引人注目的方式展现出来。这是很危险的，因为生动逼真的描写（通常仅仅是一些个人的经验与推荐）可能使其更容易自我们的记忆中获得，从而导致我们对某一事件的实际发生率、某种冒险的实际危险程度、某种知识性观点真正的合法性等等出现判断偏差。一些科学知识的信息来源，如有同行评议程序的期刊等，一般提供的是比较可靠的信息。你越是依赖大众媒体所提供的信息，你就可能越忘记不了你的恐惧。

建立多种假设

避免过早接受错误观念的最好方法之一，就是尽可能地多考虑其他可能的选项。尤其是在避免广泛存在的推理陷阱时，一个行之有效的方法就是经常地问问自己，为什么可能出错。这一方法可以使你更清晰地鉴别出观念中的缺陷或薄弱之处，并鼓励你去考虑其他可能更好的观念。

询问可以预测什么

我们都是“解释”过去事件的专家，但其中有很多解释实际上是毫无价值的，只是编得比较圆满的故事而已。为避免后见之明偏差、事后解释及过度自信等情况，我们必须能对未来事件进行预测，并对预测的准确性进行评估。如果要考察一种观点代表的究竟是对现实的真实理解，还是关于过去事件的夸张故事，这是最好的方法。

质疑阴谋理论

有时，阴谋理论仅仅有点似是而非的合理性，如果深入了解，就会发现其夸张的外表之下，其实不堪一击，尤其是那些大阴谋理论。我们完全有理由怀疑是否真的存在足够多的共同利益，让那些所谓的阴谋家们相互勾结。此外，考虑一下阴谋理论在逻辑上是否说得通（例如按该理论所称的规模——时间跨度及涉及的绝对人数，这一阴谋是否现实可行），也能为我们提供一些有用的信息，帮助我们判断这一理论是否更有可能只是一种幻想。按照这样的思路，我们只需提出一些简单的问题，就会发现几乎没有哪种阴谋理论仍然能维持其合理性。

警惕控制错觉

为了减少不确定性的威胁，我们总是倾向于获得一种可控制的感觉，即使这

种控制感仅仅只是错觉。此外，我们常常会低估在可观测事件中机会概率的作用，混淆需要技巧决定的状况和只需要运气决定的状况。最终的结果可能就是产生一种控制错觉，于人于己都会造成一定的伤害，比如在反复失败、达不到目标以后发展出的习得性无助。因此，当我们试着去理解这个世界时，一定要记住 Abelson 的告诫：“给随机现象一个机会。”

注意不要责怪受害者

我们保持安全感的另一种方法就是，相信人们会得到他们应该得到的，而他们已得到的也是原本该得的。这种信念有可能被过度使用，比如将它强加于那些遭遇了无法预知的灾难的受害者身上，认为一定是他们自己做了些什么才导致那些苦难的发生。当你发现自己在为受害者的不幸而责怪他们时，一定要注意避免后见之明偏差。如果你不是已经知道某人后来遭受了某种不幸，你还会认为他的某种特定行为必须受到责备吗？请记住，这个世界并不像我们所希望的那样完全能够预测或完全可以控制。

同时考虑某一观点的积极与消极后果

对于任何理论而言，只考虑到它积极的一面是十分危险的。例如，一些坚信“思维主宰健康”的人只强调积极思维的正性作用，却只字不提消极思维的潜在危害。同样，“水记忆”的拥护者们也没有意识到，如果水真有记忆，那么除了对所接触到的有益的材料有记忆之外，还会对危险和肮脏的东西有记忆，那将会是多么可怕的一件事。请记住，一厢情愿的想法往往只是选择性地呈现该想法所具有的积极特征，而将其实际导致的全部后果综合来看的话，却总是不如它表面看起来那么令人期待。因此，在对任何理论的积极意义进行评估时，都必须同时评估其消极意义。

注意基本比率

某一特定事件发生的简单频率是一种极有价值的信息，但却总是未被充分利用。我们往往倾向于只考虑在某一特例中证据的强度，而忽略了各种结果的基线概率。举例来说，即使是很有效的检测方法，在对某种相对罕见的疾病或事件进行诊断和预测时，出错的几率也很大。有一句简单的谚语可以帮助你牢记要重视

基本比率：“当你听见蹄声时，应当想到那是马蹄声，而不是斑马的蹄声。”

为了减少误差而允许某种程度的错误存在

尽善尽美这个不现实的目标往往是提高我们决策质量的绊脚石。一味地追求对人类行为进行完美的预测，这是非常愚蠢的，因为所有的行为都会包含一些不可预测的成分。与此相反，意识到每种方法都可能存在缺陷，最佳的解决办法就是有最多的优点和最少的缺点，这样就可以极大程度地简化决策过程。优先采用临床方法或直觉的方法进行决策的部分原因，就是想要达到完美预测这个不可能实现的目标，但实际上，其决策的质量却比那些运用统计学方法进行的决策更差。统计学决策方法之所以更好，就在于它始终坚持“会存在一定数量的错误，完美的预测不可能达到”的理念。

利用统计学决策的效力

统计学决策是相当简单却又非常有效的技术。说它简单是因为从根本上来看它只是计算的过程。当你面临难以抉择的状况时，就可以在自己的思考过程中利用这一技术，列出支持和反对观点的简明清单。对每一选项所赋予的权重大小并不重要，因为即使没有权重这一信息，制定列表的过程本身就已经很清楚地揭示了你的真正偏好。如果你是专业人员，那么不仅要使用这个方法，还要使用预测公式（如果有的话）。寻找相关的研究文献，至少会得知哪些变量与特定的决策任务有关，而且还很可能发现某个已有的预测公式，正好可以使用或修改后使用。

不要曲解向均值的回归

请记住，在极端值出现之后就更有可能会出现相对不太极端的值。我们往往容易忽略这样一个事实：真正杰出的表现实际上是一种罕见的情况，它缘于个人杰出的能力与异常的好运气二者同时出现。因为运气不太可能总是那么异乎寻常的好，所以在特别优异的表现以后，表现就可能不那么优异了。如果你试图理解为什么多次重复总是倾向于更平均的结果，那么就想想那个毫无根据的“体育画报恶咒”吧，它就很好地说明了即使是那些能力最强的人也不可能总交好运。

同时考虑成本与收益

为了做出合乎伦理的决策，你必须公正地权衡所有的成本与收益。通常很容易犯的一个错误就是，我们往往高估了收益的可能性，同时低估了损失的可能性。还有更容易出现的一个错误，就是忽略了选择任何特别行为的过程中的机会成本，也就是说，选择了某种方式就意味着对其他可选方式的放弃。例如，把钱放在家里的机会成本就是把它放在银行可能获得的利息收益。你进行的选择所涉及的后果越严重，考虑机会成本也就越发重要。最符合你个人利益的决策就是对所有相关的成本与收益进行公正地权衡后所做的决策，其中还必须包括那些不太明显的机会成本。

练习批判性思维

要在思想开放与怀疑论之间取得微妙的平衡，其显著的特征就是进行科学的推理。这种批判性的思考过程可以保护我们免受那些错误观念的影响，并且让我们不至于被那些没有最大限度考虑我们的利益的人所误导。除了前面我所提到的那些建议——始终坚持充分的科学推理之外，以下六条标准可以作为批判性思维的进一步补充指导：

- 证明某一知识性的观点具有潜在的可证伪性。也就是说，如果该观点所称的确是错的，那么一定有某种方法来证明其错误。不可被证伪的观点是不能被验证的，因此也就没有什么科学效用。
- 检查某一观点的逻辑性，包括检查它每一个前提的真实性，以及基于这些前提所得到的结论的有效性。
- 检查某一观点是否具有综合性，也就是说，它必须要能解释所有可获得的相关证据，而不仅仅是选择性地说明其中一些零散的部分。
- 诚实可靠地评估某一观点。即使是你最坚定的信念，也要与其他观点一样接受有风险的检验，而且要使用同样严格的证据标准，要像你在辩论过程中对另一方的要求一样，对你自己的主张也提出同样的证据标准。
- 坚持各种观点必须以可重复的研究证据作为基础，也就是说，多个彼此独立的检验过程均支持这一观点。要求确认这类证据，你就可以排除该观点仅仅是侥幸成功或偶然事件的可能性。
- 在你相信某一观点之前，一定要有充足的证据。请记住，提出某一观点

的人有义务提供相应的证据（驳斥或推翻它并不是你的工作）；特别的观点需要特别的证据；仅仅有来自权威的证据，是不足以令人信服的。

一厢情愿的结束性思考

我诚挚地期望，本书中对于人类推理过程的心理学方面的探讨，会有助于你在人生各个领域作出更明智的选择。人类大脑最杰出的功能并不是心灵感应，也不是可以有意识地控制健康，而是做出理智推理、正确判断，并最终获得个人最满意的决策。我们生来就具有令人惊叹的进行复杂性思考和批判性思维的能力，但需要努力才能充分发挥这一方面的潜能。数以百万的人每天要花费大量的时间锻炼身体，但有多少人会严格锻炼他们的大脑呢？

扩展大脑功能的第一步就是要学习推理过程，包括对那些经常使用的思维捷径有所认识。那些思维捷径可以提高我们的思维效率，但有时也会导致一些可预知的错误类型。在前面的章节我已经介绍了许多这样的思维捷径和启发式推理方法。

扩展大脑功能的第二步就是要学习何时需要克服以及如何克服这些启发式推理方法的缺陷。批判性思维这一工具对于任何想要学习它的人来说都是开放的。当你使用批判性思维并积累了一定经验后，你就会发现，要摒弃那些愚蠢的观点以及想以表面依据来说服你的企图，是很容易的，这样你就可以很快付诸一笑，而不会受到毫无意义的东西的影响。此外，批判性思维可以应用到你生活的方方面面。利用批判性思维，你可以更清楚明白地表达自己的观点；可以自然地产生和编排合理的想法，以将演讲准备得更好；还可以构建出在逻辑上无懈可击、更具有说服力的论证过程。

批判性思维是很有用的，也是容易上瘾的。一旦你开始发现荒谬的推理过程，无论是你自己的还是其他人的推理谬误，你就会越来越享受思维的力量。而且，一旦你发现了科学推理的某几条原则的价值，其他原则的价值也很快会得以展现。或许这只是我个人一厢情愿的想法，但我写此书的最大愿望就是，希望能提供一些好的通用的思维工具，帮助你更好地走过人生之路，让你受用终身。

附录：评估信息来源

1999 年 10 月，一项关于祈祷是否存在治疗作用的实验研究结果发表在顶尖的医学期刊——《内科医学档案》（*Archives of Internal Medicine*）上（Harris et al）。这项研究很好地控制了各项影响因素，因而其结果得到了媒体的广泛关注，许多媒体还将它作为引人注目的头条新闻进行了大幅报道。

- “心脏病人在进行神秘的祈祷后病情大有进展”——《多伦多星报》，1999 年 10 月 26 日

- “祈祷的‘医学’价值得到了研究的证实”——《圣地亚哥联合论坛报》，1999 年 11 月 3 日

- “科学家们‘证明了祈祷的神奇力量’”——《伦敦每日邮电报》，1999 年 11 月 11 日

这一被认为是里程碑式的研究吸引了大批民众。无论是广播还是电视、杂志还是报纸、网站还是按邮件发送清单投递的宣传资料，所传递的信息都十分清楚、一致：科学研究的数据支持了“祈祷是一种有效的医疗保健方法”。可能是因为先前许多关于祈祷与健康关系的研究都存在方法学上的严重缺陷，因而人们将这一研究当作证明祈祷的医疗作用的最有力的研究证据，进行了广泛的引证和吹捧。但是，这一研究结果的实质究竟是什么呢？

在这项研究中，研究人员共调查了 990 名心血管重症监护室的住院病人，其中约有一半病人（466 人）被随机地分配至祈祷组，接受为他们的迅速康复所做的祈祷，而其他的病人（524 人）则不接受祈祷。实验记录了所有病人的 35 项有关疾病结果的指标，如肺炎、大外科手术、心跳骤停或者死亡等等。但是，在组间进行的 35 项指标的比较中，祈祷组病人只有一项指标在统计学意义上显著

优于未祈祷组，那就是患者对“Swan-Ganz 导管”的评价。说到这里，问题就已经很明显了：为什么仅仅在对 Swan-Ganz 导管的评价这一项指标上，祈祷显示出了对病情的正性治疗作用，而在其他 34 项测试指标中，祈祷组均未出现更好的治疗效果呢？这个问题非常重要，因为根据统计测试的规律，每进行 20 项测试，就可能有一项是纯粹因为随机几率的作用而出现统计学意义上的显著差异。在这项研究中，35 个测试指标只有一项有组间差异，这一比率甚至比单纯的随机几率还要低。而且，除了这一令人失望的结果之外，两组病人的实际存活时间也没有显著的差异。

不过，研究人员对两组之间缺乏显著差异的事实予以忽略，反而想出了一个合成的“病程分数”（course scores）指标，来掩盖 35 项有关疾病结果的评价指标中有 34 项并不支持祈祷的医疗作用这一现实。可以说，这项研究作为能证明祈祷的医疗作用的最有力证据而被广为引证，并且有着很高的知名度，但事实上它所得的研究结果并没有什么价值，完全不足以证明上述结论。

虽然研究报告的作者很谨慎地宣称“我们并未证实上帝会以实际行动回应祈祷者，甚至连上帝是否存在也没能证实”，但前面我们引用的那些新闻标题说明，媒体普遍没有在意这样的警示信息，而是在忙着宣告祈祷是一种有效的疾病治疗方法了。

强化和均化

很明显，这一实验的实际研究结果被戏剧化地歪曲了——尤其是那些媒体的标题，真正去仔细阅读标题以外的具体内容，这样的人只是很少的一部分。这个例子很好地说明了以讹传讹的道理，也就是说，二手信息源通常总是会改变信息原本的内容。当人们重复某种信息时，他们自身的目的与意图总是会使所述说的内容发生某种程度的改变，比如强调某些内容而忽略另一些内容。

Gilovich（1991）曾描述了人们出于使话题更富娱乐性或更具新闻价值的目的，歪曲所叙述内容的两种方式。一种方式是，我们会强化或者强调一个故事中与我们自身的信念、理论，或者偏见相一致的某些成分；而另一种方式则是，我们会均化或不强调故事中与自身信念等不一致的成分。无论是哪种方式，都会使所叙述的内容与最初的信息资料来源有所偏离，继而导致该故事有更多被强化或

均化的机会，因而信息的歪曲程度又进一步增大了。

举例来说，由于许多人不注重研究数据，一味相信祈祷有效，因此 Harris 等人的研究结果在公之于众的每一个阶段都会得到如此有选择性的反映，也就不足为奇了。

首先，Harris 等人以表格报告他们的数据，但在描述这些数据时明显受到在他们自身观念与期望的影响。他们夸大了研究证据的某些细枝末节的重要性，这些被夸大的内容使得他们的“认为祈祷有效”的信念进一步得到了确认，但与此同时他们却忽略了与其信念相反的大部分实际证据。其次，报纸杂志发布关于这一研究的简要消息，无疑是为了使这一消息更具有新闻价值，因而强化了研究结果中有利于祈祷的成分，同时均化了那些不利于祈祷的成分。再次，新闻播报员（其中很多人可能仅仅读了报纸杂志的简要报道），还要对研究结果进行另一次过滤，出于想制造一个能吸引公众的“好故事”的目的，他们又进一步地进行了强化与均化的过程。最后，如果你像大多数人一样，是从一个看过或听过这一消息的朋友那里听说这个研究的，那么你离了解这一研究的实际结果就又远了一步。

上述研究结果的歪曲过程说明，寻求原始的数据与信息来源并亲自检查其可靠性是多么的重要。当你获得的信息是由媒介来进行强化或均化，以符合他们已说出或者未说出的某种假设时，他们所做结论的可信赖程度就要打个折扣了。实际上，有许多方法可以用于评估信息来源的质量，在下面我将会一一进行介绍。

信息来源质量的评估标准

假设你是一个世界级图书馆的读者，在这个图书馆里保存了历史上所有的文字材料，收藏的内容从史前穴居时代的石刻到现代的网页内容，无奇不有。然而，面对浩如烟海、如此众多的资料，你要如何才能找到适合你阅读的材料呢？你会采用什么样的标准呢？假如你已经留出了专门的时间分别来阅读那些具有娱乐性质或者具有知识参考价值的材料，当然有时也有例外——有些材料会同时具有这两种功能，你要如何在这两个领域选择适合你阅读的材料呢？完全以随机的原则来选择显然是很愚蠢的做法，因为必须得运用某些标准，才能将那些智慧的、有见地的、可信赖的内容，从那些沉闷无聊、缺乏创见、有偏见的材料中区

分出来。现在请思考一下你将用什么方法来完成任务。我想，根据你的目的不同（选择的读物要具有娱乐性质，还是要具有知识参考价值），评定信息来源质量的有效标准也会有极大的差异。

流行性

要达到被广泛阅读的目的，文字作品就必须能够吸引读者。举例来说，如果一本书的定位是供大众消遣娱乐，那么只有当它的确好玩有趣时，它才能流行得起来。因此，对于那些娱乐取向的作品，其相对流行性将是你判断自己是否真的会喜欢阅读的适当指标。可以用来评价文字作品流行性的方法有：读者的阅后评分、书籍销售量以及书籍所获的奖项等等。

你可能会希望通过一些粗略的分类方法，如喜剧文学、传奇文学、戏剧、惊险小说等，来得到一个更个性化的阅读清单，但在某一特定的类别中，也许还是要依据流行性这一大众的评定结果来选择最为吸引你的那些书。因此，虽然流行性这一指标并不是那么完美，但它可能是评定读物娱乐价值的最佳标准了。

然而，从另一方面——对于信息质量的要求来说，流行性并不是一个好的评定指标，理由我们在下面会进行解释。关于事实性信息最可靠的来源就是那些有同行评议程序的学术期刊和学校的教科书。这类出版物的作者总是力求精确地引述他们的证据，清楚地解释他们论证的逻辑方法，严格地基于研究数据来得出结论。

不幸的是，这类作品在学术严谨性方面的品质保证并不是获得商业成功的秘诀。这些学术期刊与教材的读者数量，与许多其他并不那么可靠的读物相比，简直无法相提并论。譬如，当你在任何一家综合性书屋里闲逛时，你会看到些什么呢？几乎所有被摆在“心理学”专柜的书籍，都是由一些自封的、没有多少心理学科学知识的专家所写的“自助性”读物。这类书很少有介绍心理学研究进展的文献或资料，取而代之的是给出一些他们自己认为不错的主意，宣传和介绍提高自信、维持婚姻以及取得事业成功的最好方法。最典型的还是在“科学”这一分类或专柜里所陈列的书（如果那家书屋还保留了这一分类的话），科学书籍与那些介绍神秘题材或超自然现象的书籍，例如关于 UFO、外星人劫持、超感官知觉、意志力、鬼或妖怪等内容的书相比，总是少得可怜。总的来说，能够成为畅销书的书籍通常都会有一些这样的内容：以介绍引人注目的个人传奇故事为

目的的报道，允诺某种轻松获得健康、财富或爱情的捷径。这类书的写作素材除了作者本人的个人经验以外，很少会有其他来源，而且也很少或者根本没有支持作者观点的证据资料（Rosen, Glasgow, & Moore, 2003）。因此我们可以说，虽然用这类畅销书来消遣一个下午的时光的确不错，但从所提供信息的可靠性、质量或者真实性的角度来衡量某种读物的话，流行性通常是一个很差的评价指标。

评论和综述

也许在选择书籍的过程中，你会结合对其有利以及不利两方面的评论来进行挑选。这看起来是一个明智的策略，可是实际操作起来又很难。你要怎样来决定该读哪些评论呢？你要怎么判断一篇评论是否与你的看法一致呢？你又怎么来协调那些相互矛盾的评论所提供的观点呢？事实上，要确定哪一位评论家的作品符合你的口味（如果有的话），需要付出相当多的时间和精力，而你所花费在这一判断过程中的时间，又不如阅读最终所选的读物时那么愉快和有趣。评论家们通常总是彼此互不赞同观点的，有时甚至他们自己的观点也会前后矛盾，毕竟艺术审美的评判是极其主观的一件事。因此，阅读相关评论对于选择娱乐性的读物来说，可能是一个完全无效的方法。

不过，在选择具有知识性参考价值的作品时，利用评论或者综述就是一个比较有效的方法了。所有的专业学科都有一种内部的匿名同行评议系统，可以监控和保证出版物的质量。例如，当一名学者向某学术期刊投稿想要发表论文时，他通常需要接受这一领域中几位专家对该稿件的匿名评议。这些同行评议专家往往会对文章所呈现的研究证据的全面性以及关联性进行批评性的评论，并且严格地检查作者论证的逻辑严密性——包括各种明确说明了的和未说明的假设。根据这些方面的考察结果，每位评议人会就这一文章是否合适在该刊物上发表给出自己的意见。由于是匿名评议，同行评议专家们可以无所顾忌地采取怀疑的态度，并给出直率得近于残酷的反馈意见。绝大多数学术期刊与学术书籍的出版商都只发表那些已通过同行审议过程认可的文章或书稿。因此对于读者来说，评估这类信息来源质量的任务就大大减轻了。

在每一个专业学科中，都有一个清晰的信息来源体系，向读者说明文章质量的评估标准究竟如何。顶级的学术期刊与学术书籍会采用最为严格的评估标准，因而也可以提供最可靠、最值得信赖的信息；二等的出版商则采用略为宽松的标

准，以此类推。比较而言，许多其他种类的信息源，如杂志、报纸以及网络等等，都只有极少的质量控制手段，或者根本没有什么质量控制过程。作为一名读者，你完全可以充分利用这种学术同行评议程序所提供的质量控制结果，将你获取知识的来源锁定在那些有同行评议程序的出版物，并且在同类作品中优先选择顶级的出版物，而不是那些评定等级较低的、很少或没有质量控制程序的材料。

将网络作为一个研究工具

正如我们前面的讨论中已经说明的，网络提供的信息通常没有经过有独立见解的批评者进行评议，因此在接触这类信息的时候，必须要持相当谨慎的态度。网络信息的最大优点同时也是它的一大缺点：几乎什么人都可以向网络上传信息，而上传的信息也几乎是什么都有。全世界范围内信息均可自由获取，并且信息交换丝毫不受限制，这是指引网络发展的主要价值观。然而，在网络提供的无可比拟的言论自由和质量控制的要求之间会产生某种矛盾，因为其中能提供经过同行评议的信息的网页极其稀少，人们就必须自己去判断网页信息的可信性与质量高低。

网站内容

你可以通过一些问题来帮助你评估在线信息资源的可信赖程度，而且，其中绝大多数问题也适用于评估其他的信息源。

- 这个网站是出于商业运作目的而建立的吗？在这个网站上是不是总有广告出现？如果是，那么请考虑一下，是否存在任何潜在的利益冲突，可能导致网站上的信息可靠程度下降？
- 是由谁来负责监控网站信息内容的准确度？信息是由什么人传到网站上的？这些上传信息的人是否提供了明确的联系方式，使你可以对信息提出疑问、进行反馈或者要求给予进一步的信息？如果是匿名发贴或上传资料的网站，肯定不会负责进行反馈，那你就小心，要对这些资料的可靠性持怀疑态度。
- 那些信息是什么时候发到网上的？它们多久更新一次？如果是根据可能已经过时的信息所做的结论，你也要仔细考虑。
- 那些信息通过引用或参考一些科学研究的证据，得到充分证明了吗？要

注意从最初的信息源到你所读到的信息，中间究竟经过了多少个步骤。如果参考文献是未经证实的、当中还包含了一些无关的限制条件（例如“参见 Oprah*”），或者一些未列出具体名称的信息来源（如“首席科学家们均同意……”），那么你也不要相信它。务必要寻求那些允许你检查其结论及解释的科学研究作为参考资料。

- 那些信息是否包含有关效应值的大小，或者侥幸成功的可能性大小的定量评估信息？仔细检查一下作者有没有对实验组间的微不足道的差异夸大其辞，或者将因机会概率的影响得到的一些发现说得言过其实。

- 那些信息中包含对实验不一致的结果的讨论吗？或者有没有承认还有其他可能的解释？尽可能多地考虑针对该信息的各种解释方法，这样才能确定网站所提供的论证其可信程度究竟有多高。一个值得信赖的信息源对于争议是持承认与肯定态度的，而且还会从一个问题的各个方面来讨论所持证据的优劣。

- 该网站是否提供了来自于正当的、质量有保证的信息源的其他信息？如果该网站提供的都是些自身参考性质的链接（如一个网站上所有的页面都只是链接到同一网站上的其他网页），或者提供的是隐蔽性链接（如用按钮或其他图片来掩饰链接），那你就要小心了。

网络搜索

除了对于网站内容评估的指导之外，我还要就网络信息搜索的问题提出一些警告。基于环球网的搜索引擎可以提供给你许多可能有价值的相关标题，事实上，你很快就会找到更多标题，多到你都无法一一追踪下去。但是，千万不要误认为这就是一次全面的搜索。

第一，网络搜索引擎并不会真的就你的主题词去搜索整个环球网。真实的情况是，每个搜索引擎都是在它已编入目录的那一部分网站中进行操作。由于这一目录实际上只是全部网络资源的很小一部分，那么对于一个特定的搜索引擎来说，就可能会漏掉大多数与你感兴趣的内容相关的网页。1999 年有一项研究调查了 11 个最常用的网络搜索引擎，发现它们的搜索覆盖范围各占全部网络资源

* Oprah 指 The Oprah Magazine，即欧普拉杂志，是由美国 ABC 电视网的大牌节目主持人欧普拉·温弗瑞（Oprah Winfrey）和赫斯特杂志集团 2000 年起共同推出的网络杂志。——译者注

的 2.2% 到 16.0% 不等。而且，即使是联合使用这 11 个网络搜索引擎，也只能搜索到全部网络资源的 42 %。自开展那项研究后到现在，搜索引擎的覆盖率也许已经有了某种程度的提高（如通过扩大服务器容量，或运用效率更高的搜索算法等），但也有可能更低了（如网页总数的发展速度超过了搜索引擎的搜索能力的发展速度）。

第二，搜索引擎对于输入的搜索关键词是极其敏感的。以搜索关于“祈祷的治疗功效”的信息为例，如果你按“祈祷”和“健康”（health）这两个词搜索，会得到一系列潜在相关的网页，但那些讨论祈祷的所谓“治疗”（healing）效应的网页可能就会被漏掉。

第三点，也是最重要的一点，网络搜索的结果并不是直接链接到那些最可信的信息源——学术期刊与教科书。在网络上搜索通常得不到原始的信息资料，即使有也非常少。你可能会找到一些参考了学术期刊或教材中某些内容的网页，或者对那些内容进行讨论的一些文章，但不管怎么说，最原始的一手资料往往很少出现在网络上。

为你自己而思考

虽然网络信息几乎没有什么质量控制程序，但还是可以利用许多基于网络的工具对包含高质量学术作品的数据库进行检索。本书最重要的主题就是强调运用批判性思维以及审慎周密的推理，当你着手寻找某种信息的时候，这也可能是你所能运用的最好工具了。记住，唯一能保护你不受误导性的信息、有偏差的解释、愚蠢的信念所影响的方法，就是去回顾那些原始数据或信息源，并对其进行批判性的思考与评价。

本书的目的正是为你完成这一任务提供必要的技能，使你能够经过充分的思考，作出与你个人价值观以及人生目标一致的明智决策。