

哲学导论必读材料整理

2026 秋

Wenhao Zhang

GPT-6 Astra

目录

编写说明与课程全景	x
一、这份材料怎样使用	x
二、全学期板块与必读篇目	x
 第一部分 知识论	 1
知识论板块导读	2
 Feldman 第 2 章：知识的传统分析	 5
F2.I 知识的种类	5
A. 几种主要的知识	5
B. 所有知识都是命题知识吗	5
C. 小结	7
F2.II 知识与真信念	7
A. 知识的两个条件	8
B. 知识等于真信念吗	8
F2.III 知识的传统分析	9
A. 信念 (Belief)	10
B. 真理 (Truth)	10
C. 证成 (Justification)	12
F2.IV 真正的知识与表面上的知识	13
F2.V 本章结论	13
 Feldman 第 3 章：修正知识的传统分析	 15

F3.I 对传统分析的反驳	15
A. 三个反例	15
B. 反例的共同结构	17
F3.II 为传统分析辩护	18
A. 拒绝 JF	18
B. 拒绝 JD	18
F3.III 修正传统分析	19
A. 无虚假根据理论	19
B. 无击败者理论	20
D. 一个审慎的提议	21
F3.IV 本章结论	22
从知识分析到因果联系	22
Goldman 第一章：证成的结构	27
1.1 认识论的概念与问题	27
1.1.1 认识论研究什么	27
1.1.2 信念与信念态度	28
1.1.3 证成是一种认识规范评价	28
1.1.4 真理不等于证成	29
1.1.5 知识与其他概念的关系	29
1.2 认识论倒退问题	30
1.2.1 推论证成的基本结构	30
1.2.2 理由倒退	30
1.2.3 共时理论与历时理论	31
1.3 无穷主义	31
1.3.1 证成从何处产生	31
1.3.2 有限主体不可能拥有无限信念	32
1.3.3 信念的现实活动与倾向	33
1.4 融贯论	33
1.4.1 从循环推理到整体网络	33
1.4.2 一致性不足以构成融贯	34

1.4.3	高度融贯是否产生证成	35
1.4.4	隔离反驳	36
1.4.5	同样融贯的竞争系统	36
1.4.6	序言悖论：合理信念必须共同一致吗	36
1.4.7	同一系统成员地位问题	37
1.5	基础主义与基础信念	37
1.6	基础信念是自我证成的吗	38
1.6.1	第一种理解：相信本身保证内容为真	39
1.6.2	第二种理解：事实本身证成信念	39
1.7	驳斥若干对直接证成的驳斥	40
1.7.1	随意性反驳	40
1.7.2	塞拉斯二难	40
1.7.3	君主身份的类比	40
1.7.4	层级上升与门票类比	41
1.7.5	Sellars 的观察知识要求	41
1.7.6	BonJour 的高阶论证	42
1.8	强度、内容、击败者与增强者	42
1.8.1	基础信念的内容	42
1.8.2	直接性与证成强度	43
1.8.3	初步证成、击败者与增强者	43
1.8.4	829 型飞机案例	43
1.9	从地基走向更高楼层	44
1.9.1	从心理经验到外部世界	44
1.9.2	鹿迹与模拟器	44
1.9.3	解释推理证成原则	45
1.9.4	Berkeley 的竞争解释	46
	第一章总结：证成的来源与传递	46
	第一章课后问题参考思路	46
Goldman	第二章：证据主义、可靠主义与内外在主义	51
2.1	证成与证据	51

2.1.1	传统认识论的共同预设	51
2.1.2	当下时间切片假设	52
2.1.3	信念证成与命题证成	52
2.1.4	心灵主义证据主义	53
2.2	证据主义面临的问题	55
2.2.1	证据共存是必要条件或充分条件吗	55
2.2.2	“证据”指什么	57
2.3	证成与过程可靠性	58
2.3.1	为什么仅仅“基于证据”仍不充分	58
2.3.2	过程可靠主义的核心思想	59
2.4	过程可靠主义的表述	59
2.4.1	从初步原则到两种可靠性	59
2.4.2	历史性原则	60
2.4.3	可靠主义与命题证成	61
2.5	可靠主义的吸引力与困难	61
2.5.1	吸引力一：降低怀疑论门槛	61
2.5.2	吸引力二：自然连接证成与真理	61
2.5.3	元证成要求	61
2.5.4	Wright 的进一步怀疑论挑战	62
2.5.5	一般性问题	62
2.5.6	新恶魔问题	63
2.5.7	认可清单可靠主义	63
2.5.8	Norman 的可靠千里眼	64
2.5.9	Olsson 的归因比较思路	64
2.5.10	历史性与“下游后果”	64
2.6	内在主义与外在主义	65
2.6.1	争论框架	65
2.6.2	规范规则模型	66
2.6.3	新手与观鸟专家	66
2.6.4	可实施配方与无误指导观	67
2.7	内在主义与推论规则	67

2.7.1 INF: 仅要求主体相信前提	67
2.7.2 INF*: 要求前提信念得到证成	68
2.7.3 INF**: 主体看出蕴涵关系	68
2.7.4 INF***: 主体认为存在蕴涵	68
2.7.5 规则正确性与更高层级的证成因素	69
第二章总结: 证据、过程与证成因素	69
第二章课后问题参考思路	72
Feldman 第 6 章: 怀疑论 (I)	77
F6.I 怀疑论的种类	78
A. 全面怀疑论与局部怀疑论	78
B. 怀疑主张的强度	79
F6.II 怀疑论者究竟主张什么	80
A. 存在但不为我们所知的真理	80
B. 怀疑论、真理与证成	81
F6.III 四个怀疑论论证	82
A. 错误可能性论证	82
B. 不可分辨性论证	83
C. 确定性论证	85
D. 传递性论证	86
F6.IV 回应怀疑论	87
A. 怀疑论会自我反驳吗?	87
B. 摩尔式回应	88
C. 可错论	89
F6.V 阶段性结论	96
怀疑论的其他回应	98
知识论板块总结: 从知识分析到怀疑论	103
第二部分 心灵哲学	105
心灵哲学板块导读	106

Heil 第 2 章：笛卡尔二元论 (Cartesian Dualism)	107
2.1 科学与哲学	107
2.2 笛卡尔的二元论	108
2.2.1 空间性与非空间性	108
2.2.2 物理性质与经验性质	108
2.2.3 公共可见性与第一人称私密性	108
2.3 实体、属性与样态	109
2.4 笛卡尔二元论的形而上学结构	110
2.5 心身互动难题	111
2.6 一个因果封闭的宇宙	111
Heil 第 3 章：笛卡尔的遗产 (Descartes's Legacy)	113
3.1 无互动的二元论：平行论	113
3.2 偶因论	114
3.3 因果关系与偶因论	114
3.4 唯心论	115
3.5 心灵与意义	115
3.6 副现象论	116
3.7 尾声：非笛卡尔式二元论	117
Bayne 第 3 章：如何成为物理主义者 (How to Be a Physicalist)	119
3.1 行为主义	119
3.2 心脑同一论	120
3.3 功能主义	121
3.4 取消主义	122
3.5 科学的角色	123
Bayne 第 4 章：知觉 (Perception)	125
4.1 三种知觉理论	125
4.2 意向论与现象特征	126
4.3 支持和反对意向论	127
4.4 知觉允许什么内容	128

4.5 无意识知觉?	128
Bayne 第 5 章: 思想 (Thought)	131
5.1 思想的种类与特征	131
5.2 思想语言假说	132
5.3 思想语言的替代理论	132
5.4 中文房间论证	133
5.5 图灵测试与思想边界	134
Bayne 第 8 章: 意识的形而上学 (The Metaphysics of Consciousness)	135
8.1 蝙蝠、神经科学家与僵尸	135
8.1.1 蝙蝠论证	135
8.1.2 Mary 与知识论证	136
8.1.3 哲学僵尸	136
8.1.4 解释鸿沟与困难问题	136
8.2 A 型物理主义与不可推知性前提	137
8.3 B 型物理主义与桥接前提	138
8.4 幻觉论	139
8.5 二元论与罗素式一元论	139
Bayne 第 10 章: 心灵因果性 (Mental Causation)	141
10.1 为什么需要心灵因果	141
10.2 因果排斥反对	142
10.3 非还原物理主义的回应	143
10.4 外在主义与心灵因果	143
10.5 Libet 的挑战	144
心灵哲学板块总结: 从心身区分到心理解释	146
第三部分 形而上学	147
形而上学板块导读	148

Ney 第 6 章：持存 (Persistence)	149
6.1 变化之谜	149
6.2 关于持存的三种观点	150
6.2.1 持续论 / 三维论	150
6.2.2 延续论 / 虫体论	150
6.2.3 阶段论 / 远续论	150
6.3 物质构成悖论的解决	151
6.4 暂时内在性质问题	152
6.5 阶段论与时间对应物	152
6.6 为三维论辩护	153
Ney 第 7 章：模态 (Modality)	155
7.1 作为真理方式的可能与必然	155
7.2 可能性与必然性的种类	155
7.3 模态逻辑与可能世界分析	156
7.4 Lewis 的模态实在论	157
7.5 替代模态实在论	157
7.5.1 丰富语言之角	158
7.5.2 贫乏语言之角	158
7.6 拒绝可能世界分析	158
7.7 本质主义与反本质主义	159
7.8 当代本质主义	160
7.9 本质与必然性的关系	160
Ney 第 9 章：自由意志 (Free Will)	163
9.1 什么是自由意志	163
9.2 自由意志与决定论问题	163
9.3 决定论有何根据	165
9.4 相容论	165
9.5 自由意志主义	166
9.5.1 行动者因果自由意志主义	166

9.5.2 事件因果自由意志主义	166
9.6 自由意志怀疑论	167
形而上学板块总结：时间、可能性与行动者	169
 第四部分 伦理学	 170
伦理学：板块导读与扩充目录	171
 第五部分 语言哲学	 172
语言哲学：板块导读与扩充目录	173
 附录	 174
附录：中英术语表	175
本版结语：三大板块的共同问题	183

编写说明与课程全景

一、这份材料怎样使用

本书不是将原文压缩成若干结论的“速记”，而是依照每篇原书的章节顺序，把主要段落的论点、例子和回应译述为可以连续阅读的中文教材。编排遵循五个层次：先交代作者正在回答的问题，再翻译并解释关键论证；原书的案例尽量完整保留；较长的推理以公式或步骤重构；最后标明该节在整章中的作用。个别章节与原书纸本页码对照，便于回查英文文本。

译述以准确传达论证为首要目标，不追求逐词直译，亦不是逐句无遗漏的原书全文翻译。遇到一个术语有多种通行译法时，正文采用本书统一译名，并在首次出现处附英文。“**解读**”与“**帮助理解**”方框是本书加入的解释和推演；“**论证重构**”把作者的意思整理成步骤，不等于原书的逐字表述。未特别标明的正文主要沿着作者的叙述推进，争议处仍应回到原著核对。

建议阅读方法

1. 首读先看“文献定位”和每节末小结，掌握问题与结论。
2. 二读逐项检查案例：主体相信什么、命题是否为真、理由是否充分、是否存在运气。
3. 三读遮住正文，仅根据论证图和公式复述前提、结论与反驳点。
4. 写作或讨论时回到英文原文核对措辞与页码；本书只作导读和学习辅助。

二、全学期板块与必读篇目

下表依据课程大纲组织。知识论、心灵哲学与形而上学的指定原文已收入本版；伦理学与语言哲学先建立固定位置，待相应原文加入后沿用同一体例扩充。

板块	课程主题	必读材料	本版状态
知识论	知识的分析	Richard Feldman, <i>Epistemology</i> , Chapters 2–3	已收入
知识论	认识辩护	Alvin I. Goldman 与 Matthew McGrath, <i>Epistemology: A Contemporary Introduction</i> , Chapters 1–2	已收入
知识论	怀疑论	Richard Feldman, <i>Epistemology</i> , Chapter 6	已收入
心灵哲学	心的形而上学、心灵因果性、意识官能	John Heil, <i>Philosophy of Mind</i> , Chapters 2–3; Tim Bayne, <i>Philosophy of Mind</i> , Chapters 3, 4, 5, 8, 10	已收入
形而上学	持存与变化、可能世界、自由意志	Alyssa Ney, <i>Metaphysics: An Introduction</i> , Chapters 6, 7, 9	已收入
伦理学	规范伦理学、元伦理学、应用伦理学	Russ Shafer-Landau, <i>Living Ethics</i> , Chapters 5, 6, 10; Earl Conee 与 Theodore Sider, <i>Riddles of Existence</i> , Chapter 10; <i>International Encyclopedia of Ethics</i> 相关条目	待加入原文
语言哲学	意义与表征	Thomas Nagel, “The Meaning of Words,” Chapter 5 of <i>What Does It All Mean?</i>	待加入原文

按授课周次使用

周次	日期	主题	对应材料
第 2 周	9 月 15 日	知识的分析	Feldman 第 2、3 章
第 3 周	9 月 22 日	认识辩护	Goldman 第 1、2 章
第 4 周	9 月 29 日	怀疑论	Feldman 第 6 章
第 6–8 周	10 月 13–27 日	心的形而上学、心灵因果性、意识官能	Heil 第 2–3 章；Bayne 第 3、4、5、8、10 章
第 9–11 周	11 月 3–17 日	持存与变化、可能世界、自由意志	Ney 第 6、7、9 章
第 12–14 周	11 月 24 日–12 月 8 日	规范伦理学、元伦理学、应用伦理学	Shafer-Landau、Conee 与 Sider、伦理学百科条目
第 15–16 周	12 月 15–22 日	意义与表征、人工智能与语言理解	Nagel 第 5 章；本板块不纳入期中或期末考核

第一板块

知识论

知识论板块导读

知识论 (*epistemology*) 研究知识、证成、合理信念以及怀疑论。²本板块的基本方法是分析：提出一个定义，分别检查它给出的条件是否必要、合取起来是否充分，再用反例定位失败之处。Feldman 与 Goldman 的文本分别展示了这种方法在“知识分析”和“证成理论”中的运作。

五篇文献之间的关系如下：

文献	中心问题	主要理论
Feldman 第二章	命题知识需要哪些条件？	真信念、传统知识分析
Feldman 第三章	得到证成的真信念为何仍可能不是知识？	盖梯尔反例、第四条件方案
Goldman 第一章	证成链条如何开始、延续或闭合？	无穷主义、融贯论、基础主义
Goldman 第二章	什么因素使信念得到证成？	证据主义、过程可靠主义、内在主义、外在主义
Feldman 第六章	可错的证据是否足以产生知识？	怀疑论、摩尔式回应、可错论

知识论板块主线

1. 认识论首先区分真理与证成：命题可以为真却无人有理由相信，也可以为假却得到高度合理的支持。

2. 推论证成产生理由倒退。无穷主义允许无尽理由链，融贯论诉诸系统内相互支持，基础主义则设置非推论性起点。

²课程课件将 *epistemic justification* 译为“认识辩护”，将 *justifier* 译为“辩护者”；本书正文沿用“证成”与“证成因素”。此外，“无限主义”与“无穷主义”、“融贯主义”与“融贯论”、“认识后退”与“认识倒退”在相关篇目中分别对应同一英文术语。

3. 温和基础主义允许经验提供可错的初步证成，并允许其他信念增强或击败这种证成。
4. Goldman 第二章进一步指出：当前证据并不总能解释一个实际信念为何合理，信念的形成与保存历史也具有认识论意义。
5. 证据主义与可靠主义的冲突，最终扩展为内在主义与外在主义关于证成因素性质的争论。

Feldman 第 2 章：知识的传统分析

Richard Feldman, *Epistemology*, Chapter 2, “The Traditional Analysis of Knowledge, pp. 8–24.

知道某件事，与碰巧猜对某件事，有什么不同？要回答这个问题，首先需要确定“知道”指哪一种认识关系，然后再逐项考察知识是否要求相信、真理和证成。

F2.I 知识的种类

A. 几种主要的知识

“知道”在日常语言中承担不同工作。Feldman 先列出六种常见句式：

类型	形式	例子
认识某人	S knows x	教授认识 J. D. 塞林格
知道是谁	S knows who x is	学生知道 J. D. 塞林格是谁
知道是否	S knows whether p	图书管理员知道馆内是否有塞林格的书
知道何时	S knows when A happens	编辑知道塞林格的书何时出版
知道如何	S knows how to A	塞林格知道如何写作
知道事实	S knows that p	学生知道塞林格写了《麦田里的守望者》

最后一种是**命题知识** (*propositional knowledge*): 主体知道某个事实，或知道某个命题为真。本书集中研究这种知识。但在确定研究对象之前，还要问：其他所有知识能否都还原为命题知识？

B. 所有知识都是命题知识吗

有些“知道”确实可以用“知道某命题”为真来解释。

D1：知道是否

$$S\text{知道是否}p =_{df} S\text{知道}p \text{ 或 } S\text{知道}\neg p.$$

如果图书管理员知道馆内是否有塞林格的书，那么真实情况无论是哪一种，她都知道相应的真命题：若有，她知道“有”；若没有，她知道“没有”。这里的“或”连接两种知识状态：她要么知道有书，要么知道没有书。它不同于只知道“有书或者没有书”这个排中式。一个完全不了解馆藏的人，也知道馆内要么有书、要么没有书，却未必知道究竟是哪一种情况。

怎样检验一个定义

$=_{df}$ 读作“定义为”。定义必须在**所有代人实例**中使左右两边同真同假；只要发现一种可能情形，使左边真而右边假（或相反），就得到**反例**（*counterexample*）。这是本书接下来检验“知识=得到证成的真信念”的基本方法：并非争论多数情形中是否符合，而是寻找条件合取后仍失败的个案。

必要条件与充分条件可以借一个普通定义区分。假定把“母亲”定义为“生育过孩子的女性”，养母就提出了疑问：她是母亲，却没有满足定义右侧的条件。这说明右侧可能不是必要条件。反过来，如果某个对象满足右侧条件，却不属于被定义的概念，那么右侧就不是充分条件。³

面对反例，可以限制左侧的分析对象，例如改问“生母”是什么；也可以修改右侧条件，例如加入收养关系；还可以拒绝把养母算作母亲。不过，第三条路线必须有独立理由，不能只因某个案例使定义失败，就宣布它不算数。知识分析中的争论具有相同结构：究竟应修改定义、缩小所分析的知识范围，还是重新判断案例，需要结合具体理由决定。

“知道何时”也可以借助命题来解释。编辑知道塞林格的书何时出版，意味着存在一个特定时间 t ，她知道“这本书在 t 出版”：

D2：知道何时

$$S\text{知道事件}A\text{何时发生} =_{df} \exists t (S\text{知道“}A\text{在}t\text{时发生”}).$$

“知道谁”“知道哪一个”“知道为什么”也可能通过类似方式分析，不过后两者会涉及更复杂的识别标准和解释要求；Feldman 在这里不急于给它们完整定义。

但是，**认识一个人**不等于仅仅知道关于此人的某些命题。一个读者可能知道大量塞林格的生平事实，却从未与他相识；因此，下列定义 D3 太宽：

³参见赵斌 (Bin Zhao)，北京大学《哲学导论》2026 年秋季课件，*The Analysis of Knowledge*，第 1 页。

S 认识 x $=_{df}$ S 知道关于 x 的某些事实.

这种熟悉关系称为**亲知 / 熟识知识** (*knowledge by acquaintance / familiarity*)。

给 D3 加上“知道很多事实”或“知道特别重要的事实”也救不了它：传记作者可以掌握大量重要事实，仍从未见过作者本人。认识某人更接近曾经与其接触、相遇并保持某种熟悉关系。这一反例不是要求微调 D3 中的数量，而是揭示它选错了**关系种类**。

“知道如何”则有歧义。曾经会滑雪、如今因伤不能再滑的人，仍可说自己知道滑雪的步骤；一个没有理论训练却能熟练滑行的儿童，也可说知道如何滑雪。于是应区分：

两种“知道如何”

D4a: S 知道如何做 $A =_{df}$ S 知道完成 A 的重要步骤是什么；

D4b: S 知道如何做 $A =_{df}$ S 具有完成 A 的能力.

原书的伤后滑雪教练可以准确讲解动作要领，故满足 D4a，却因受伤不能亲自滑行，不满足 D4b。一个会熟练滑行、却不会解释动作步骤的儿童恰好相反：她能滑，故满足 D4b，却不掌握教练那样的明确步骤知识。两例表明“知道如何”至少有两种用法，而不是一场简单的“知识究竟都是命题还是都是能力”的二选一。D4b 意义上的知道如何，是**能力知识 / 程序性知识** (*ability knowledge / procedural knowledge*)，不必等同于命题知识。

C. 小结

Feldman 由此区分至少三种基本知识：

1. **命题知识** (*propositional knowledge*)：知道某个命题为真；
2. **亲知知识** (*acquaintance knowledge*)：熟悉或认识某人、某物；
3. **能力知识** (*ability knowledge*)：能够完成某项活动。

三类知识的条件未必相同。关于“知道命题”的定义即使正确，也未必能直接用于判断一个人是否认识塞林格或是否会滑雪。以下以 S 表示认识主体，以 p 表示命题；“ S 知道 p ”所指的始终是命题知识。

F2.II 知识与真信念

A. 知识的两个条件

第一个条件是**真理**。一个人可以非常确信错误命题，甚至声称自己“知道”，但若 p 为假，他便不可能知道 p 。例如，Thomas Jefferson 并非美国第一任总统；无论主体多么确信，都不能知道“Jefferson 是第一任总统”。

例 2.1：侦探小说

读者在小说最后一章以前一直确信管家是凶手，结尾却揭示会计才是凶手。读者可能说：“我一直知道是管家干的，后来才发现不是。”这句话的交际作用是强调结局出人意料；按字面却不可能为真。读者此前只是非常确信，并没有知识。

第二个条件是**相信**。这里“相信”取广义：只要主体把 p 当真，哪怕态度有所保留，也属于相信。面对一个已经考虑的命题，主体通常有三种态度：相信 p 、相信 $\neg p$ 、悬置判断。

例 2.2：知道自己的名字

朋友 John 被问“你相信自己叫 John 吗”，回答：“我不只是相信；我知道。”这并不表示知识排斥信念。正如“我不是累，我是筋疲力尽”并非真的否认疲劳，John 是在用较强表达纠正较弱表达所产生的暗示：他当然接受“我叫 John”为真，只是不愿把自己的把握说得像猜测一样弱。

用 K_{Sp} 表示“ S 知道 p ”，用 B_{Sp} 表示“ S 相信 p ”，这两个必要条件便可写为：

$$K_{Sp} \implies p, \quad K_{Sp} \implies B_{Sp}.$$

B. 知识等于真信念吗

把两个必要条件合起来，得到一个简洁但错误的定义：

TB：真信念分析

真信念 (*True Belief*, TB) 分析将知识等同于真与相信两项条件的结合。

$$S \text{知道} p \quad =_{df} \quad (i) S \text{相信} p; \quad (ii) p \text{为真}.$$

必要条件合取后未必成为充分条件。下面两个例子分别展示“纯猜中”和“理由太弱却猜中”。

例 2.3：正确的比赛预测

纽约队与丹佛队实力相当，专家意见也五五开。主体毫无根据地预感丹佛队会赢，结果丹佛果然获胜。主体的预测为真，他也确实相信它，但这一正确结果并不能把无根据的预测变成知识。

例 2.4：悲观的野餐计划者

天气预报只说周六降雨概率略高于 50%。悲观者据此非常坚定地相信野餐会遇雨，而当天确实下雨。天气预报提供了某种支持，却不足以支持他如此坚定的判断；后来实际下雨，也不会使原来的微弱证据事后增强。

比赛案例还有一个细节：如果认为“尚未发生的事情原则上不可能被知道”，便可能把非知识归咎于“未来性”，而不是幸运猜测。Feldman 让主体不看比赛，散场后继续相信“丹佛已经赢了”；命题这时关乎**过去**，证据却丝毫没有改善，所以仍然不是知识。反例因此针对 TB 本身，而不依赖争议性的未来知识理论。

野餐案例又比纯猜测多了一层：降雨概率略超过一半，已经为相信“会下雨”提供**某种**支持；但一点点支持并不足以跨越知识门槛。判断信念能否构成知识，要问理由够不够强，不能只问“有”或“没有”理由。

缺失的第三项通常称为**证成** (*justification*)：主体必须以足够合理的方式相信 p 。至于“足够”如何界定，是后续证成理论要解决的问题。

F2.III 知识的传统分析

Feldman 把**知识的传统分析** (*Traditional Analysis of Knowledge, TAK*) 写成如下形式。它把知识分析为**得到证成的真信念** (*Justified True Belief, JTB*)；TAK 是这套分析的名称，JTB 指分析中的三个条件：

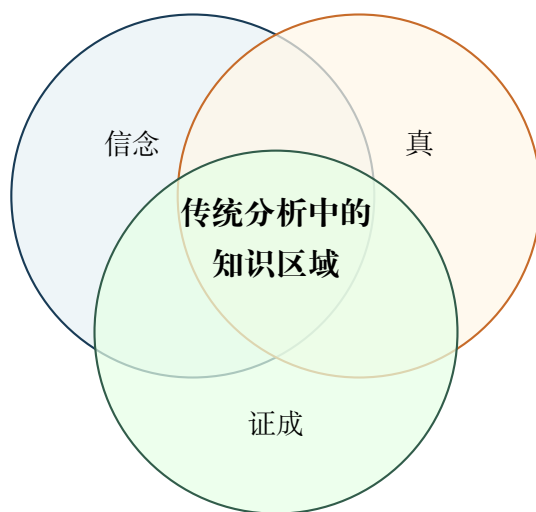
TAK：得到证成的真信念

$$S \text{ 知道 } p \quad =_{df} \quad \begin{cases} (i) & S \text{ 相信 } p; \\ (ii) & p \text{ 为真}; \\ (iii) & S \text{ 相信 } p \text{ 得到充分证成}. \end{cases}$$

柏拉图的《美诺篇》(*Meno*) 中已有相近的思想。苏格拉底说，真意见固然有用，却容易从心中逃走；给出理由，便像把它们系住，使其成为知识。一种可能的解释是：真意见要成为知识，还需要理由或证成。这里暂不讨论知识是否比普通信念更不容易被遗忘，而只取其中关于理由的要求。

罗德里克·奇泽姆 (Roderick Chisholm) 也曾提出：一个人知道某命题，当且仅当

他相信该命题、命题为真，并且该命题对他而言是“明证的” (*evident*)。最后一个条件关乎他相信这一命题有多合理。传统分析承接的正是这一思路。不过，“相信”“真”“得到证成”各自还需要解释；三个词并列起来，并不意味着问题已经解决。



A. 信念 (Belief)

相信一个命题，就是接受它为真。这种接受可以有所保留，也可以十分确信，因此“相信”并不专指证据不足时的猜测。面对一个已经考虑的命题，主体可以采取三种基本态度：相信它，否定它，或悬置判断。

否定 (*disbelief*) 在这里指相信该命题的否定。例如，否定“华盛顿是美国第一任总统”，就是相信“华盛顿不是美国第一任总统”。**悬置判断** (*suspension of judgment*) 则既不接受原命题，也不接受其否定。因而，日常所说的“我不相信”有时指否定，有时只是拿不准；分析知识时必须把二者分开。一个从未想过某命题的人，也不同于考虑后决定暂不判断的人。

相信是主体对**命题**采取的态度，不是主体与某个句子之间的关系。法国儿童 Pierre 用法语相信“George Washington 是美国第一任总统”，美国儿童 Peter 用英语相信同一事实；两人不接受同一个句子，却相信同一个命题。因此应区分：

句子：表达命题的语言工具 $\neq$ 命题：不同语言可以共同表达的真假内容。

B. 真理 (Truth)

作者采用**真理符合论** (*Correspondence Theory of Truth, CT*) 的朴素核心：

CT：符合论

一个命题为真，当且仅当世界正如该命题所说；若世界并非如此，命题为假。

意见不同，并不使同一命题有不同的真值

首先，命题是真是假，不由相信它的人数决定。即使许多人误以为杰斐逊是美国第一任总统，也不会改变华盛顿曾担任这一职务的事实。其次，同一个命题不能仅因甲接受、乙反对，就既“对甲是真的”又“对乙是假的”。两个人的态度确实不同，但他们争论的若是同一个历史事实，真假仍由那个事实决定。

这并不为独断辩护。认为自己的意见正确、对方的意见错误，是发生分歧时很普通的态度；由此却不能推出自己永不出错，更不能推出无需倾听对方。新证据可能要求我们改正信念；有时，发现一个认真思考的人持不同意见，本身就构成重新检查根据的理由。客观真理与开放的讨论并不冲突。

时间和地点改变时，句子可能表达不同命题

“华盛顿现在是美国总统”在 1789 年可以表达真命题，在 2003 年则表达假命题。这并不意味着同一事实同时既真又假。一种处理方式是把时间明确写进内容：“华盛顿在 1789 年是总统”与“华盛顿在 2003 年是总统”是两个命题。世界发生变化，关于不同时间的命题便可以具有不同真值。

地点也是如此。缅因州的人打电话给佛罗里达州的人，说“这里在下雪”；后者回答“这里没有下雪”。两句话可以同时为真，因为前一句报告缅因州的天气，后一句报告佛罗里达州的天气。相同或相近的句式，不保证表达同一个命题。只有先确定“现在”“这里”等表达的所指，才能判断双方是否真正矛盾。

趣味判断首先涉及语句的含义

“酸奶味道好”似乎特别容易支持“真理因人而异”的说法。不过，要判断它的真假，须先弄清这句话在说什么。Feldman 提出三种可能解释。

第一，它可能表示“我喜欢酸奶的味道”。这样，不同说话者谈的是不同人的喜好。甲确实喜欢、乙确实不喜欢，并没有使同一个命题对甲真、对乙假；他们表达的是关于各自口味的不同命题。

第二，它可能表示“大多数人喜欢酸奶的味道”。这时，不同说话者表达的是同一个关于多数人喜好的命题。它是否为真，要看多数人实际上喜不喜欢，而不取决于说话者本人爱不爱吃。

第三，它可能诉诸某种独立于个人喜好的味觉评价标准。按照这种解释，即使许多人不喜欢酸奶，它仍可能符合该标准。不过，是否存在这样的标准、标准是什么，都是需要另行讨论的问题。

三种解释各有困难，本节不必在其中作出最后选择。它们说明的是：有关个人喜好的事实也可以成为命题所符合的事实。命题谈论主观感受，并不意味着它的真假可以任

意改变。

真理独立于信念，不等于真理不可认识

有人担心：如果真理由独立于我们的世界决定，我们岂不是永远只能知道自己的看法，而接触不到真正的事实？这一结论并没有从符合论中推出。“事实不由我的相信决定”与“我无法认识事实”是两项不同主张。要从前者走到后者，还需要说明知觉、记忆或推理为什么不能使我们认识独立事实。

这正是怀疑论需要提出论证的地方。符合论本身既没有证明我们拥有知识，也没有证明知识不可能。Feldman 暂时沿用我们确实知道许多事情的标准观点，以考察知识的条件；怀疑论的挑战留待后面的章节专门讨论。

模糊性与边界情形

一个真正棘手的问题是**模糊性** (*vagueness*)。设想我们去机场接 Michael，只知道他是成年男性，别人告诉我们“Michael 很高”。若他身高 6 英尺 4 英寸 (约 1.93 米)，这句话通常为真；若他只有 4 英尺 10 英寸 (约 1.47 米)，通常为假；若他约 5 英尺 10 英寸 (约 1.78 米)，按照案例所设的普通语境，就可能处在“高”的边界上。

在最后一种情况下，困难未必来自信息不足。即使我们精确知道他的身高、成年男性的平均身高等事实，“高”这个词也可能没有一条精确界线。“健康”“富有”“明智”等词同样存在边界情形。

这时还要区分句子的模糊与命题的模糊。如果句子虽然含糊，却在具体语境中表达一个真值确定的命题，简单的符合论未必需要改变；如果命题本身也有不确定真值，那么把所有命题都分成真、假两类就可能过于简单。也许需要承认“不确定”的第三种真值，或者容许真有程度。Feldman 没有在这里裁决这些复杂问题，只指出符合论可能需要相应修订。

C. 证成 (Justification)

证成有程度。野餐案例中的天气预报提供一些证成，却未达到知识所需的强度；若主体亲眼看见窗外正在下雨，证成便显著增强。因而 TAK 第 (iii) 项应理解为**充分或很强的证成**，不是任何微弱理由。

此外，“主体有理由相信 p ”不蕴涵“主体实际相信 p ”。

例 2.5：缺乏自信的考生

Mr. Insecure 认真复习，模拟考试表现良好；正式考试题型熟悉，教师看过答案后也说答得不错。尽管他拥有充分理由相信自己通过考试，他因长期缺乏自信而不相信这一点。于是，他对“我通过了”具有命题证成，却没有相应信念。

证成还具有主体相对性与时间相对性。同一私生活命题，对本人可能证据充分，对陌生人则毫无根据；同一场野餐的降雨命题，在预报发布、乌云出现、雨真正落下的不同时间，证成程度也会改变。

拥有证成还不同于**能够当场提出一份辩护**。儿童可能有充分知觉理由相信玩具就在眼前，却还无法向哲学家准确表述自己的根据。将“能解释为什么”设成知识的普遍必要条件，会无端排除儿童与许多普通人的知识。

F2.IV 真正的知识与表面上的知识

一个社会、机构或权威可以把某些主张**算作**知识，但“被算作知识”不等于真正满足知识条件。掌权者可能让缺乏良好证成的官方观点获得“知识”名义，也可能压制证据更好的异议观点。社会机制如何分配“知识”标签，是重要的社会认识论问题；本章的分析目标却是**真正的知识** (*real knowledge*)，即主体实际上是否满足知识条件，而不是**表面上的知识** (*apparent knowledge*)。

例如古人可能有理由相信地球是平的、处于宇宙中心，甚至有理由以为自己知道这些命题；但命题为假意味着他们并不真正知道。一个时代公认某项内容是知识，与它是否实际上为真、是否满足知识条件，是两种不同的判断。

这一区分也防止一种混淆：TAK 不是关于社会如何使用“专家”“权威”称号的经验概括，而是要说明一个人实际上知道某件事需要满足什么条件。这种概念分析不同于描述社会怎样使用“知识”一词；其中的证成条件又包含对相信方式的规范评价。

F2.V 本章结论

本章从“知道”的多种用法中抽取命题知识作为研究对象，并逐步得到：

知识 $\stackrel{TAK}{=}$ 得到充分证成的真信念 (Justified True Belief, JTB).

但这只完成了候选分析的建立。若要证明定义正确，还需检查反方向：凡是得到充分证成的真信念，是否一定是知识？Feldman 第三章将说明答案是否定的。

Feldman 第二章总结

1. 本书研究命题知识，而非亲知或纯粹能力。
2. 真与信念是知识的必要条件，却不是充分条件；幸运猜中仍不构成知识。
3. 传统分析加入充分证成，得到“知识 = 得到证成的真信念”。
4. 真理、信念与证成三者不能互相还原：真理由世界决定，信念是主体态度，证成评价主体的认识根据。
5. 下一章的核心任务，是检验 TAK 的三个条件合取后是否充分。

Feldman 第 3 章：修正知识的传统分析

Richard Feldman, *Epistemology*, Chapter 3, “Modifying the Traditional Analysis of Knowledge, pp. 25–38.

真理排除了错误判断，信念排除了主体根本不接受命题的情况，证成又排除了无根据的猜测。这三个条件合起来，是否已经足以构成知识？盖梯尔问题表明，即使三项条件全部满足，认识上的偶然性仍可能使知识落空。

F3.I 对传统分析的反驳

一个定义正确，要求左右两边在所有可能案例中同真同假。TAK 不只是说知识必然包含得到证成的真信念，还说任何得到证成的真信念都足以构成知识。Edmund Gettier 的反例攻击的正是后一方向。

A. 三个反例

例 3.1：十枚硬币

Smith 看见 Jones 倒出口袋中的硬币并仔细数出十枚；Jones 能力很强，Smith 又听见老板说 Jones 将得到职位。因此 Smith 有充分理由相信：

(1) Jones 会得到职位，并且口袋里有十枚硬币。

他由此推出：

(2) 将得到职位的人口袋里有十枚硬币。

然而老板说错了，真正获聘的是副总裁的侄子 Robinson；巧合的是，Robinson 口袋里也恰有十枚硬币。(1) 为假，(2) 为真。Smith 对 (2) 有得到证成的真信念，却不知道 (2)，因为真结论只是巧合。

例 3.2: Nogot / Havit

同事 Nogot 开 Ford、持有 Ford 所有权文件，平时也诚实。Smith 据此相信：

(3) *Nogot*拥有一辆 Ford。

当地 Ford 经销商举办抽奖，同办公室只要有 Ford 车主便可参加。Smith 于是推出：

(4) *Smith*的办公室里至少有一人拥有 Ford。

事实上 Nogot 在假装车主，(3) 为假；但不为 Smith 所知的 Havit 确实拥有 Ford，所以 (4) 为真。Smith 再次得到一个有证成的真信念，却没有知识。

例 3.3: 田野中的羊

Smith 远望田野，看见一个与羊完全相似的东西，因而相信“那个动物是一只羊”，并告诉车内的孩子“田里有一只羊”。可见物其实是牧羊犬或逼真的羊模型；田野另一处、在 Smith 视线之外，碰巧真有一只羊。于是“田里有羊”为真，但使 Smith 相信它的可见对象不是那只羊。

若认为案例中的证据还不够强，可以增加可信文件、可靠证人的保证，或者使可见对象更逼真。只要充分证成仍容许偶尔出错，就仍可构造相应的误导情形。因此，反例的要点不是把微弱猜测冒充充分证成，而是让主体拥有通常足够好的理由，同时使理由所指向的事实与结论实际成真的事实分离。

这种分离在第二、第三个故事中尤其清楚。Ford 案例不是说“Smith 完全没有 Ford 相关证据”，而是说证据首先指向 *Nogot*，使结论为真的却是 *Havit*；田野案例不是说“Smith 没有看见类似羊的东西”，而是说他看见的不是使“田野里有羊”成真的那只羊。两例都把信念的证成路径与命题的成真路径拆开。

十枚硬币与福特汽车的另一种构造

同一类认识运气可以用不同情节构造。课堂采用的十枚硬币版本，最后得到职位的是 Smith 本人，恰好有十枚硬币的也是 Smith；Feldman 这里采用的版本则把这两个角色交给 Robinson。两者的关键相同：Smith 的理由指向 Jones，结论实际成真却靠另一个获聘者。人物变换不改变反例所针对的充分条件。⁴

⁴本段及 Ford / Barcelona 案例参见 *The Analysis of Knowledge*，第 3-5 页。课件十枚硬币案例中的“*And (1) is true*”与后文设定冲突，应为结论 (2) 为真；前提 (1) 因 Jones 未获聘而假。

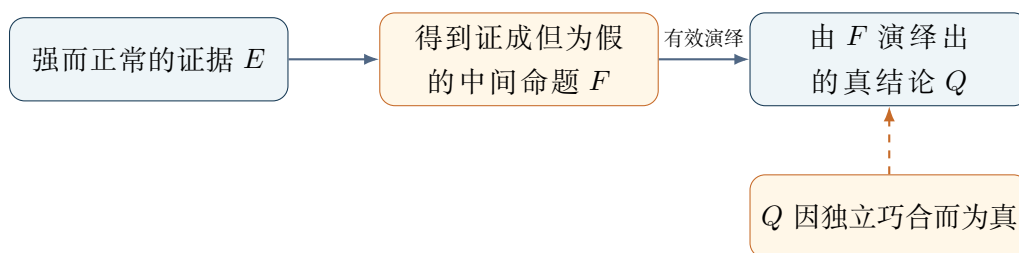
Ford 与 Barcelona

Smith 有很好的理由相信同事 Jones 拥有 Ford 汽车，却不知道朋友 Brown 在哪里。他由“Jones 拥有 Ford”推出：“Jones 拥有 Ford，或者 Brown 在 Barcelona。”事实上，Jones 开的车是租来的；但 Smith 全不知情的 Brown 恰好就在 Barcelona。

用 F 表示“Jones 拥有 Ford”，用 B 表示“Brown 在 Barcelona”，推理就是 $F \therefore F \vee B$ 。这是有效的析取引入：只要第一项为真，整个析取句便为真，不需要另外掌握第二项的证据。按照**证成演绎原则** (*Justified Deduction Principle, JD*)，Smith 对 F 的证成可以传给 $F \vee B$ 。但实际情形是 F 假、 B 真；使析取句为真的，恰好是他毫无了解的那一项。因此，他的真信念仍不是知识。

这些反例可以概括为两步：先让主体凭充分理由形成一个被误导的信念，再安排一个与那些理由无关的巧合，使最终结论为真。另一种维护 JTB 的办法，是干脆承认主体知道结论。然而，这会放弃一个很有吸引力的判断：知识不能仅仅因为与认识根据无关的运气而真。争论因此不只是例子讲得是否离奇，还涉及我们愿意保留哪些关于知识的基本条件。

B. 反例的共同结构



反例使用两条看似可信的原则：

JF 与 JD

得到证成的假命题原则 (*Justified Falsehood Principle, JF*): 主体可能有证成地相信一个假命题。

证成演绎原则 (*Justified Deduction Principle, JD*): 若 S 有证成地相信 p , p 蕴涵 q , 且 S 由 p 正确推出并因此相信 q , 那么 S 有证成地相信 q 。

JF 允许 Smith 合理地相信前面的假命题；JD 则使证成能够通过正确演绎传给后面的结论。这里不能把“推理有效”误解为“结论一定真”：有效演绎保证的是，如果前提为真，结论就不可能为假。前提实际为假时，有效推理的结论既可能假，也可能碰巧真。

十枚硬币案例的推理形式没有问题：若 Jones 获聘而且有十枚硬币，就确实能推出获聘者有十枚硬币。错误出在获聘者其实不是 Jones；结论又因为 Robinson 也有十枚硬

币而碰巧成立。Smith 有信念，结论为真，依据 JF 和 JD 也有证成，但他并不知道结论。这就构成了对 TAK 充分性的反例，并没有否认 TAK 中的任何一个条件可能仍是知识的必要条件。

F3.II 为传统分析辩护

A. 拒绝 JF

第一种辩护说：既然奇数编号命题为假，Smith 就不可能对它有证成。这样最终真结论也失去证成，案例便不再反驳 TAK。

Feldman 认为代价过高。设一个正常案例中主体拥有极强证据且命题为真；再构造一个主体拥有**完全相同证据**、却因罕见欺骗而命题为假的反常案例。若假命题绝不可能得到证成，反常案例中的信念无证成；但按照下列原则，正常案例也必须无证成：

SE：相同证据原则

相同证据原则 (*Same Evidence Principle*, SE): 若两个可能案例中，主体关于 p 的证据毫无差异，那么主体要么在两个案例中都被证成地相信 p ，要么在两个案例中都没有这种证成。

拒绝 JF 并接受 SE，将推出几乎没有日常信念得到证成，因为几乎任何正常证据都可能在某个欺骗案例中伴随假命题。这已放弃“我们拥有大量知识”的标准看法，不能作为在标准看法内部拯救 TAK 的轻微修补。

这一反对以 SE 为前提。它并不是单凭“有人可能受骗”就证明正常人没有证成，而是说：如果拒绝假信念的证成，又坚持相同证据必有相同证成，就很难保住普通信念。拒绝 JF 因此并非逻辑上不可行，却不能算作在保留日常知识的前提下对 TAK 所作的小修补。

B. 拒绝 JD

第二种辩护承认 Smith 对假前提有证成，却否认证成能由假前提传给其逻辑后果。问题是，这会要求 Smith 一方面合理地相信“Nogot 拥有 Ford”，另一方面对“办公室至少有一位 Ford 车主”悬置判断。既然后者由前者显然推出，这种态度组合并不合理。

而且，若把 Gettier 案例改成前提恰好为真的正常案例，主体的内部证据与推理完全相同。只因外部真假不同便把同一推理判为有证成或无证成，再次违背 SE。因而拒绝 JD 也不能令人满意地保存 TAK。

F3.III 修正传统分析

A. 无虚假根据理论

Michael Clark 提出的**无虚假根据理论** (*No False Grounds Theory, NFG*) 要求：知识不仅满足 TAK 三项，而且主体相信该命题的所有根据都必须为真。

NFG：无虚假根据

$$S\text{知道}p =_{df} \begin{cases} (i) & S\text{相信}p; \\ (ii) & p\text{为真}; \\ (iii) & S\text{相信}p\text{得到证成}; \\ (iv) & S\text{相信}p\text{的所有根据均为真}. \end{cases}$$

困难首先在于“根据”算哪些内容。

例 3.4：替代路线

Smith 观察到 Nogot 在办公室工作、开 Ford 并出示相应文件。他不把“Nogot 拥有 Ford”列作中间结论，而先得到：“办公室里有人开 Ford 并持有相应文件。”这句话是真的。他再据此相信“办公室里有人拥有 Ford”。最后一步是通常颇有根据、却不保证为真的非演绎推论：开车和持有文件并不逻辑蕴涵拥有该车。显式出现的前提和结论都真，但让结论成真的仍是 Smith 毫不知情的 Havit。因此，他仍没有知识。

若把根据狭义理解为**显式推理步骤** (G1)，例 3.4 中没有假根据，NFG 错误地判为知识。若广义理解为**所有发挥作用的背景信念和预设** (G2)，可以把“Nogot 拥有 Ford”算作隐含的假根据，挡住例 3.4；但理论又会太严格。

“根据”的划法	例 3.4：替代路线	例 3.5：额外理由
G1：仅显式步骤	步骤全真，错误地算成知识	若显式走过 Nogot 路线，也会排除知识
G2：还包括背景预设	可抓住隐含的“Nogot 拥有 Ford”	把可有可无的假背景也算进来，错误地取消知识

例 3.5：额外理由

Smith 既有关于冒牌车主 Nogot 的一组理由，也有关于真正车主 Havit 的一组独立且充分的理由。Havit 路线单独就足以使他知道“办公室有人拥有 Ford”。虽然 Nogot 路线包含假命题，Smith 仍然有知识，因为真实路线提供了独立支撑。

由此，G1 太窄，G2 太宽。知识不要求信念背景中没有任何错误；关键是错误是否对证成起不可替代的作用。

停摆钟与二十只钟

“根据”应当包括多少背景预设，也可以用钟表说明。⁵

停摆钟

Alice 看见钟面指向两点，于是相信现在是两点。实际上确实是两点，但这只钟在十二小时前就停了。她不知道钟已停摆，其判断只是碰巧赶上了钟面再次显示正确时间的时刻。

如果只把明确使用的前提算作根据，“钟面指向两点”是真的，目标信念也是真的，NFG 似乎无法排除这个偶然真信念。NFG 的支持者可以说，Alice 还预设了“这只钟正在正常运行”；这个隐含根据是假的，因此案例并未满足 NFG。

但把所有背景预设都纳入，又会重现“额外理由”案例的困难。假定 Alice 看见二十只钟都显示五点，其中十九只正常，只有一只停摆；她又预设这些钟全部正常。在其余十九只钟确实提供充分而适当的独立支持的设定下，一项多余的错误预设似乎不应取消她的知识。由此需要区分：错误是整个判断不可缺少的根据，还是在已有充分支持之外附带存在的错误。

B. 无击败者理论

另一方案关注尚未被主体掌握的真命题。若一个真命题 t 满足这样的反事实条件——主体一旦有证成地相信 t ，就不再有证成地相信 p ——那么按照这一方案， t 就是对 p 之证成的**击败者**（*defeater*）。

ND：无击败者

无击败者理论（*No Defeaters Theory*, ND）要求：在 TAK 三项之外，不存在真命题 t ，使得：如果 S 有证成地相信 t ， S 便不再有证成地相信 p 。

例如，Smith 若知道“Jones 不会获聘”或“眼前所见不是羊”，原来的证成就会被

⁵本例及其变式参见 *The Analysis of Knowledge*，第 6 页。

取消。ND 因此能处理最初的 Gettier 案例，却也排除了许多真正的知识。

例 3.6：收音机

Smith 坐在书房里，收音机关闭着，而且他知道它关着。此时 Classic Hits 101 正在播放 Neil Diamond 的歌曲 “Girl, You’ ll Be a Woman Soon”。设定 Smith 没有节目单等其他消息渠道；在这个案例中，他若有证成地相信电台正在播放这首歌，就是因为打开收音机、调到该台并听到了歌曲。可是，那样收音机便已开着，他也不再理由相信它关着。于是，“电台正在播放这首歌” 这个真命题竟满足 ND 的击败条件，使 ND 错误地排除 Smith 原本拥有的知识。

这个反例针对反事实条件的构造方式，而不是指出 Smith 的观察有任何缺陷。“若 Smith 知道电台正播歌” 这一条件在案例中必须通过**打开收音机**来实现：反事实世界里的 Smith 没有证成地相信“收音机关着”，正是因为反事实世界里的收音机开着。从“在另一个改变了事实的世界不能知道 p ” 跳回“在当前世界也不知道 p ”，是 ND 的结构性错误。

例 3.7：Grabit 偷录音带

Black 亲眼看见学生 Tom Grabit 把录音带塞进口袋并离开图书馆，因此知道 Tom 偷走录音带。Tom 住在精神病院的母亲却声称是 Tom 的双胞胎 Tim 所为；事实上 Tom 没有双胞胎，这只是她的妄想。真命题“Tom 的母亲说 Tim 拿了录音带” 若被 Black 孤立地得知，会削弱 Black 的证成，却是**误导性击败者** (*misleading defeater*)。Black 幸运地没有听到它，但这种幸运不使她原来的观察失去知识地位。

因此，“存在任何会削弱证成的真命题” 条件过强。必须进一步区分真正揭示证据缺陷的击败者，与本身会造成误导的表面击败者。

Grabit 案例又说明另一种失败：真命题“母亲说了 Tim 偷录音带” 本身是真的，但该话语内容是由妄想造成的假信息。只获得这半截事实的观察者可能动摇；完整了解母亲的精神状态、Tom 并无双胞胎及现场情况之后，原信念反而依然强。理论若只寻找“一句真而会使人动摇的话”，会把**误导性击败者**误判为真正揭示认知缺陷的反证。

D. 一个审慎的提议

Feldman 承认尚无公认的 Gettier 问题解法，但认为各案例仍呈现共同点：主体的证成本质上**依赖某个假命题**。在田野案例中，“我看见的东西是羊” 不可替代；在额外理由案例中，即使 Nogot 路线错误，Havit 路线仍可独立支撑结论，所以总体证成不本质依赖错误。

EDF：无本质性虚假依赖

原书用 EDF 标记以下分析,其第四项要求为: *S's justification for p does not essentially depend on any falsehood*, 即主体对 p 的证成不本质上依赖任何假命题。^a

$$S \text{知道} p =_{df} \begin{cases} (i) & p \text{为真;} \\ (ii) & S \text{相信} p; \\ (iii) & S \text{相信} p \text{得到证成;} \\ (iv) & S \text{对} p \text{的证成不本质上依赖任何假命题.} \end{cases}$$

^aEDF 为原书工作定义的标记; 相关表达为 *essential dependence on falsehood* (对假命题的本质依赖)。

“本质依赖”尚未得到严格分析,因此 EDF 是工作定义而非最终解答。它的优点是保留 TAK 核心,又直接指向 Gettier 反例的病灶: 错误在证成链中是否不可缺少。

F3.IV 本章结论**Feldman 第三章总结**

1. Gettier 案例证明: 得到证成的真信念仍可能因认识运气而不是知识。
2. 反例依赖 JF 与 JD。否认其中任一原则都会损害普通证成判断,无法轻易挽救 TAK。
3. “所有根据都为真”在窄解释下挡不住替代路线,在宽解释下又被额外理由案例反驳。
4. “不存在任何潜在真击败者”无法处理收音机和 Grabit 案例,因为有些真信息只是改变情境或提供误导。
5. 较稳妥的工作答案是: 知识为得到证成的真信念,且其证成不本质依赖错误。但“本质依赖”仍需进一步说明。

从知识分析到因果联系

盖梯尔问题也可以引出一种不同的修正方向: 不再只问主体有什么理由,而是追问信念与使它为真的事实之间具有什么联系。课堂在这里进一步讨论因果理论、可靠主义及知识优先认识论。⁶

⁶以下讨论依据 *The Analysis of Knowledge*, 第 6-8 页; 它们延伸了前面的知识分析问题,并非 Feldman 第三章新增的原书小节。

知识的因果理论

知识的因果理论 (*causal theory of knowledge*) 的初步主张是： S 知道 p ，当且仅当 S 相信 p ，而这个信念与 p 这一事实具有适当的因果联系。“事实”一词已经包含真理条件；如果 p 根本不真，就不存在使该命题为真的事实来建立所要求的联系。

它与 JTB 都要求真信念，区别在于第三个条件。JTB 诉诸理由与证成这样的评价性条件；因果理论试图用事实与信念之间的因果关系解释知识。在最简单的知觉案例中，面前的树通过光线和视觉系统使我相信那里有树，信念正是由它所表征的事实产生的。

阿姆斯特朗 (David Armstrong) 的温度计类比说明了这种想法：温度为 n ，导致正常温度计显示 n ；读数之所以正确，是因为它对实际温度作出反应。类比不必把温度计当作有意识的认知者，而是强调认知与环境之间有规律的因果依赖。

这可以解释为什么猜中不等于知道。猜测即使为真，成真的事实也未必参与信念形成。在 Ford 案例中，Brown 身处 Barcelona 这一事实没有使 Smith 形成相应析取信念；在羊的案例中，藏着的真羊也没有进入他的视觉因果过程。证言和解释推理则需要更长的因果链：例如，事实先被证人观察，再通过言语影响听者，而非直接刺激听者的感官。

有因果联系为何仍可能不够

头部受到撞击

Gerald 从楼梯上跌下并撞到头，脑部受到扰乱，形成若干毫无根据的信念：吃生菜会导致肥胖、芝加哥小熊队将赢得世界大赛，以及自己刚从楼梯上跌下。他对摔落没有任何感觉记忆；第三个信念与前两个一样，只是撞击造成的异常结果。

Gerald 确实跌倒了，这件事也确实导致了“我刚跌倒”的信念，但这样的直接因果联系并不使他知道自己跌倒。因果理论因而必须加入“以适当方式导致”。这项修正排除了某些异常联系，却也提出新的说明任务：什么使某种因果过程适当？若只是把能够产生知识的过程称为适当，解释就会循环。

另两类知识也难以套入简单因果模型。对于“所有人都会死”这样的概括，我们观察的是有限个具体事例，而非一个直接作用于感官的普遍事实。借助因果链的适当重构是否足够，还需要说明归纳推理的地位。数学和逻辑的困难更明显： $2 + 2 = 4$ 这样的事实怎样对一个人的心灵产生因果作用？至少，知觉知识的模型不能原封不动地覆盖所有知识。

假谷仓：环境为何影响知识

假谷仓地区

Henry 驾车经过乡间，看见一座真正的谷仓，相信自己看见了谷仓。他不知道附近布满了从公路上看与真谷仓一模一样的假立面。如果看向其中任何一个，他都会误认为那是真谷仓；这一次只是碰巧看向了真的。

与隐藏真羊的案例不同，这次使信念为真的对象确实被看见了：真谷仓正常地导致 Henry 形成真信念。但课堂采用的判断仍是，他似乎没有知识，因为在当时环境中，他的方法极易把假立面认成谷仓。问题从“实际因果链是否接上事实”，进一步转为“这项认识能力在相关环境中能否分辨真与假”。实际的一次成功，并未消除周围的错误风险。

可靠过程与认识成功

过程可靠主义 (*process reliabilism*) 可以用于分析知识：知识是真信念，并且由倾向于产生真信念的可靠过程形成。这个表述与 Goldman 第二章关于证成的可靠主义需要区分。后者首先说明信念为何得到证成，并容许可靠过程偶尔输出假信念；把它用于知识时，还要加上实际为真，并检验它能否排除妨碍知识的偶然性。

Temp 与被调节的室温

Temp 的工作是记录房间温度，他每次都查看墙上的温度计。温度计其实已经坏了，读数在一定范围内随机波动。但房间里藏着一个操纵恒温器的人，每当 Temp 查看时，便把室温调到与读数一致。因此，Temp 通过这种方式形成的温度信念总是真的。

案例把“总能得到真信念”与通常的测量关系分开了。正常温度计因实际温度而显示相应读数；这里却是隐藏者因读数而改变温度。Temp 本人既不掌握这一机制，也没有凭正常测量能力追踪温度。课堂据此提出反对：即使整个安排确保真输出，仍未必有知识。

这个结论针对的是简单的可靠过程分析；它并不立即推翻所有可靠主义版本。理论可以继续追问，评价的过程是否应包括隐藏者的操作、什么环境算相关环境，以及可靠性是否还须与主体的辨别能力建立适当联系。不过，这些条件需要解释，不能仅从“输出一直为真”中读出。⁷

⁷*Epistemic Justification* 第 6 页另列 True Temp 为未知可靠性的难题，但未展开情节；此处 Temp 的详细设定取自 *The Analysis of Knowledge* 第 7-8 页。

知识优先认识论

前面的方案都尝试用更基本的条件拼出知识：真、信念、证成，再加某种反偶然性或因果条件。**知识优先认识论** (*knowledge-first epistemology*) 改变了解释方向。它包含两项主张：消极主张是，知识不能被分析为独立于知识而规定的一组更基本条件；积极主张是，知识本身可以作为理论起点，用来解释信念、证据、证成与合理性等概念。

承认知识有必要条件，并不等于承认它必能被还原分析。红色必定是有颜色的；但这不保证我们能把红色定义为“有颜色，加上某个完全不提红色的独立条件”。同样，知道 p 要求 p 真，并要求主体相信 p ，却未必保证还能找到一个不借助知识概念的条件 X ，使“真信念加 X ”恰好等于知识。

当然，把 X 写成“使真信念成为知识的东西”总是做得到，但这只是重述问题，没有给出独立分析。知识优先论不是否认可以研究知识，也不是声称知识神秘莫测；它否认的是某一种还原方式，并要求我们检验以知识解释其他认识概念能否取得更好的理论效果。

知识分析的进一步问题

无虚假根据与无击败者方案试图修补 JTB；因果理论要求信念恰当地联系事实；可靠主义关注形成过程及其真输出倾向。头部撞击、假谷仓与 Temp 案例依次表明：实际因果联系、一次正常成功，乃至某种安排下始终为真，都需要继续接受知识条件的检验。知识优先论则追问，是否应当放弃把知识完全还原为其他概念的目标。

Goldman 第一章：证成的结构

Alvin I. Goldman 与 Matthew McGrath, *Epistemology: A Contemporary Introduction*, Chapter 1, “The Structure of Justification,” pp. 3–24.

许多信念之所以合理，是因为另一些信念支持它们。但支持结论的前提本身，也需要根据。继续追问下去，理由会无限延伸、形成相互支持的网络，还是终止于不再依赖其他信念的起点？这就是**证成倒退问题**（*regress problem*）。

这些可能的回答，分别说明证成关系如何组织。不过，只有结构还不够：即使画出一条完整的支持链，仍须说明链中的信念凭什么具有支持力。尤其是基础主义，既要解释非推论的起点如何得到证成，也要说明这些起点能否支持丰富的日常知识。

1.1 认识论的概念与问题

1.1.1 认识论研究什么

认识论（*epistemology*）研究知识以及与知识相关的思想、推理和理解活动。它并不只是描述人们事实上怎样思考，而是评价哪些思考、推理与意见形成方式较好，哪些较差。道德理论评价行动领域中的正确与错误；认识论则评价智识领域中的合理与不合理、得到证成与未得到证成。

智识评价之所以重要，首先具有实践理由。行动取决于对后果的判断；若关于行动后果的信念为真，主体就更可能作出促进自己、家庭或社会福祉的选择。医生、财务顾问和汽车技师的价值，也不仅在于善意，而在于能否形成真判断并将知识正确应用于新情况。

作者举了一个实际选择：学生因为预计某行业几年后就业机会大增而选择受训。若预测为真，这可能是好投资；若预测为假，认真训练也可能没有带来预想收益。例子并非要求学生每次预测都绝对正确，而是说明行动能否达到目的，常取决于其所凭借的信念是不是真的。选择医生也是同一逻辑：我们需要的不只是愿意帮忙的人，还需要能正确判断病情、预估治疗效果及副作用的人。

不过，真信念并非只有工具价值。人类也会出于好奇追问恐龙灭绝的原因，即使答

案没有即时实践用途。由此可见，追求真理同时具有实践价值和内在的智识价值。

真理值得追求，却不能仅靠希望结论正确就得到。在尚不知道一个命题是否为真时，我们仍要决定是否相信它，这就需要评价自己的证据与形成信念的方法。**证成** (*justification*) 与**保证** (*warrant*) 正是用于这种评价的概念。一个信念的根据好不好，与它最终是否为真有关，却不等于预先检查其真假；好证据和好方法也可能偶尔误导我们。

1.1.2 信念与信念态度

命题 (*proposition*) 是意图表达某项事实的内容。**信念** (*belief*) 则是主体对命题采取赞同态度，即认为命题所说的情况确实如此。

主体还可以对同一命题采取其他态度：

术语	英文	含义
相信	belief	赞同 P ，认为 P 为真
否定	disbelief	拒绝 P ，相信 $\neg P$
悬置判断	suspension of judgment	对 P 既不相信也不否定
信念态度	doxastic attitude	相信、否定、悬置及不同确信度的总称

“doxastic” 来自希腊语 *doxa*，意为意见。除三分法外，还可以用 0 到 1 之间的**信念度** (*degree of belief / credence*) 表示主体对命题的确信强度。

同一个命题，态度可以不同

“明天会下雨”是命题内容；相信它、相信其否定以及暂不判断，是主体对这同一个内容的不同态度。这个区分让认识规范的评价更精确：若现有证据很弱，适当态度可能是悬置判断，而非一定要在两个信念间选一个。作者本书通常采用相信、否定、悬置的三分法；信念度模型会在后续概率认识论中发挥更大作用。

1.1.3 证成是一种认识规范评价

证成 (*justification*)、**保证** (*warrant*) 和**合理性** (*rationality*) 都是认识评价术语。称某个信念“得到证成”，是在智识维度上给予正面评价；称其“未得到证成”，则是在批评其依据或形成方式。

这些词表达相近的正面评价，却并非在所有理论中同义。“合理性”有时用于评价不同强度的信念度；“证成”在这里主要用于评价相信、否定和悬置判断这三种态度。它们的精确关系取决于具体理论，不能预先视为完全相同。

例如，仅凭外貌认定某人品行恶劣，是一种较差的判断基础。它首先是智识缺陷，不必同时构成道德过错。因此，道德规范性与认识规范性都使用“应当 / 不应当”的结构，却属于不同类型的规范评价。

1.1.4 真理不等于证成

真理 (*truth*) 取决于世界实际如何；证成则取决于主体与命题之间的认识关系。二者不能等同。例如，某片海滩在某一时刻的沙粒数恰好为 N ，相应命题为真，却可能无人知道准确的 N ，也无人有充分理由相信这个具体数目。真理不需要等待主体取得证据。

反过来，一个人可能根据正常而有力的观察相信某事，却因为罕见的误导而相信了假命题。他的理由可以很好，世界却恰好不是理由使他以为的那样。证成评价的是他所处的认识情境或所采用的认识方式，不能直接用结论的真假代替。

因此：

$$\text{真} \neq \text{得到证成}, \quad \text{得到证成} \neq \text{真}.$$

1.1.5 知识与其他概念的关系

大多数认识论理论同意：

1. 知识蕴涵真理：若主体知道 P ，则 P 必须为真；
2. 知识要求主体相信 P ，或至少对 P 具有很高信念度；
3. 多数理论认为知识还要求证成，单纯“碰巧相信真命题”不构成知识。

把第三项理解为多数理论接受的条件，可暂用以下关系式表示：

$$K(P) \implies P, \quad K(P) \implies B(P), \quad K(P) \implies J(P).$$

其中 K 表示知道， B 表示相信， J 表示得到证成。本章不急于给知识下最终定义，而先研究证成的结构。

“知识蕴涵真理”又称知识的**事实性** (*factivity*)：我自称确信 P ，不能单凭确信使错误的 P 成为知识。原书说第三项“知识要求证成”是多数理论认可的联系，而非在本节已得到无争议证明。暂先研究 J ，是因为若连普通信念凭什么合理都说不清，知识分析便没有稳固起点。

1.2 认识论倒退问题

1.2.1 推论证成的基本结构

许多信念通过其他信念获得证成。Tony 相信纽约队明年会夺冠，并提出：

P_1 ：纽约队得到了多年来最有天赋的新秀；

P_2 ：纽约队还引进了一名联盟顶级内线；

C ：所以纽约队明年将夺冠。

朋友 Henry 提出异议：纽约队上一年表现很糟，Tony 凭什么认为它下一年会夺冠？Tony 的两项理由正是对这一质疑的回答。新秀和优秀内线可以改善球队实力，却不能排除伤病、对手进步等因素。因此， P_1 与 P_2 即使为真，也不保证 C 为真；这里提供的是非演绎支持。

同时，推论证成不以公开争辩为必要条件。即使 Henry 从未提问，只要 Tony 私下确实根据这些已得到证成的消息作出判断，他的信念也可能获得同样的支持。认识论要评价的是信念是否恰当地建立在理由上，而不只是一个人是否能在别人面前作出辩护。

推论证成可画成树：较低节点的前提信念向较高节点的结论信念传递证成。但如果前提本身没有证成，就没有证成可供传递。

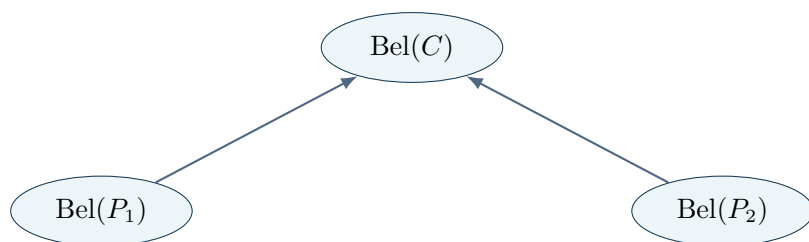


图 1.1 推论证成树。两项已证成的前提信念共同支持结论信念。箭头表示证成支持的方向，而非命题真值的蕴涵方向。

图中的 $\text{Bel}(P_1)$ 与 $\text{Bel}(P_2)$ 代表前提信念，它们也需要证成。Tony 可能援引新秀以往的得分记录、内线球员的篮板表现等材料，但这些材料又会引出新的问题：记录可信吗？观察准确吗？如果某个前提只是无根据的传闻，它并不能因为处在树的下方就获得证成。树形结构显示支持的方向，却没有自行解决支持从何而来的问题。

1.2.2 理由倒退

若继续追问“为什么相信 P_1 和 P_2 ”，Tony 可能给出更低层理由；这些理由又需要进一步支持。于是形成**认识论倒退问题**（*epistemic regress problem*）：理由链究竟应当无限延伸、在某处停止，还是形成循环？

四种主要回应如下：

立场	英文	对倒退的回答
无穷主义	infinetism	理由链无重复、无终点地无限延伸
基础主义	foundationalism	理由链最终停在非推论地得到证成的基础信念
融贯论	coherentism	信念形成相互支持的网络，部分支持路径回到先前节点
倒退怀疑论	regress skepticism	三种积极方案均失败，因此推论信念不可能得到证成

前三种想说明推论证成究竟如何可能：链无限往下走、链在非推论节点结束、或支持不再是单向的链。第四种倒退怀疑论则说三者都行不通，从而否认大量推论信念有证成。这个怀疑结论的力量取决于前三种结构是否穷尽积极方案，以及它们是否真的全部失败。作者此处只提出选择空间，尚未证明其中任何一项。

1.2.3 共时理论与历时理论

无穷主义、传统基础主义和融贯论通常是**共时理论** (*synchronic theory*)：它们只考察与目标信念同时存在的理由。相反，**历时理论** (*diachronic theory*) 允许过去的信念获取与保存事件影响当前证成地位。

评价的时间必须固定。如果 Tony 在作出预测时没有任何根据，后来偶然取得好证据，并不会使先前那个时刻的判断变得合理。不过，这不同于否认过去会影响现在：若他早已通过可靠来源获得消息，后来忘记来源，却一直保存着相应信念，先前的学习过程仍可能与当前证成有关。Goldman 第二章将借记忆案例进一步考察这种历时说明。

1.3 无穷主义

无穷主义 (*infinetism*) 认为，一个得到推论证成的信念背后存在无穷而不重复的理由链。它立即面临两个问题。

1.3.1 证成从何处产生

如果每个信念的证成都完全依赖支持链中更下一层的信念，而且链条没有起点，那么似乎没有任何地方首次产生证成。无穷主义者可以类比无限时间：一个序列不一定需

要第一项，证成关系也未必必须具有第一项。但批评者认为，这只说明无起点结构在逻辑上可以设想，并未说明依赖性证成如何在没有独立来源时真实存在。

时间类比和证成论证不能混为一谈：“没有第一天”并不蕴涵“每一天的存在都靠更早一天把某种资源交过来”。现在的疑问更精确：如果每个信念得到支持都依靠更低的、同样待支持的信念，那么哪一环能够拥有可以传递的支持？作者承认“必须有源头”在此仍是强直觉而不是严格证明，因而没有单靠这一点宣布无穷主义逻辑不可能。

无穷主义者还可以质疑“传递”这个比喻所包含的预设。若一开始就认定推论只能传递已经存在的证成，不能在推论关系中产生证成，那么寻找独立源头似乎势在必行。但这是否已经预先排除了无穷主义想说明的可能性？课堂提出的答辩是：为什么不能认为，证成恰恰产生于适当的推论结构之中？⁸

这一答辩并未给出完整的正面理论。无穷主义者仍须解释怎样的无限关系产生支持，以及为什么它优于另一条同样无限但不合理的链。不过，它揭示了争论的真正分歧：双方不只争论链有没有第一项，还争论证成是否一定要先在某个节点独立出现，才能通过推论传下去。

1.3.2 有限主体不可能拥有无限信念

人类不可能同时持有无限多信念。无穷主义者据此区分：

- **信念证成** (*doxastic justification*)：主体实际持有的信念得到证成；
- **命题证成** (*propositional justification*)：即使主体没有相信 P ，其认识情境仍使相信 P 成为合适态度。

无穷主义可以主张，存在无限多主体有理由相信的命题；主体不必实际相信全部命题，只须能够执行其中任意有限段推理。

这一回应不要求有限的心灵实际容纳无限多个信念，而是区分主体已经相信什么与有理由相信什么。例如，一个人正感到冷，即使没有想到“我感到冷”这个命题，他仍可能有理由相信它。

但这还没有说明未被相信的命题怎样完成推论证成。假定 P_1 由 P_2 支持， P_2 又由 P_3 支持，如此无穷。如果主体只相信 P_1 ，没有相信 P_2 ，那么他的实际信念如何已经基于 P_2 获得推论支持？若回答说他也相信 P_2 ，而 P_2 仍只能由 P_3 得到证成，同样的要求便会继续下去。

问题不在于命题证成是否存在，而在于它能否替代实际信念，在一条完全由推论组成的无限链中承担传递工作。

“我感到寒冷”即使未被注意，也可能因寒冷体验而具有命题证成；但这种证成来自

⁸本节关于无穷主义答辩和信念倾向的讨论，参见赵斌，*Epistemic Justification*，第2页。

非信念性经验，反而属于基础主义所说的基础证成，而不是无穷主义所需要的纯粹推论链。

1.3.3 信念的现实活动与倾向

有限心灵反驳还涉及“拥有信念”的含义。一个人并不会在每个时刻都想到自己相信的所有事情。正在想“明天有课”，是一项**当前发生的信念** (*occurrent belief*)；此刻没有想起自己的住址，却能在被问及时直接回答，则可以算作已经保存在心中的**第一阶倾向性信念** (*first-order dispositional belief*)。

课堂介绍的无穷主义答辩进一步诉诸**第二阶倾向性信念** (*second-order dispositional belief*)。这里所说的“第二阶”不是“相信自己相信”的反思层级，而是更间接的可获得性：命题目前既没有被考虑，也没有作为可直接唤起的信念存放在心中，但主体在适当条件下能够通过有与现有认识资源相联系的活动取得它。

例如，一个人并不记得蒙大拿州首府是 Helena，也不能从已有信念直接推出答案；但在需要时，他有查阅《世界年鉴》州府名录的倾向和能力。按照这项较宽的提议，相关命题对他仍然是可获得的。这里并不是说任何互联网上的信息都自动属于他的信念；答辩要求有实际可用、与主体认识活动恰当相连的取得途径。

这样，无穷主义就不必要求主体在有限时间内逐项想完无限理由。不过，新的问题随之出现：能够获得一个理由，是否等于现在已经拥有该理由？取得某个事实所用的年鉴、检索能力和判断又需要什么支持？这与命题证成和实际信念证成的区别相联系，却不完全相同；前者讨论支持的地位，倾向性方案则试图扩大主体可以拥有或调用的信念范围。

1.4 融贯论

1.4.1 从循环推理到整体网络

融贯论 (*coherentism*) 不应只被理解为允许“ P 支持 Q 、 Q 再支持 P ”的简单循环。它把得到证成的信念体系理解为一个整体系统，其中各部分通过逻辑、概率和解释关系相互支持。

基础主义承认某些从信念系统外部获得证成、再把证成传给其他信念的基础信念；融贯论则否认这种特权起点，认为证成产生于信念之间的相互关系，而且支持方向可以是双向或多向的。

纸牌屋是常用比喻：没有某张纸牌完全独立地支撑全部结构，各张牌通过彼此倚靠而共同站立。融贯论由此挑战倒退问题预设的线性、单向证成模型。

在这种整体观中，一项信念可以从多个方向得到支持，也可以反过来支持其他信念。因此，融贯论不只是把一条论证首尾相接；它要用整套信念之间的相互约束来说明证成。然而，纸牌屋只是结构比喻，还没有回答：究竟哪些关系使信念之间的支持具有认识价值？

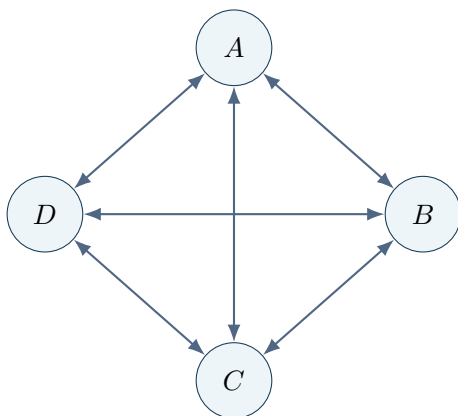


图 1.2 融贯网络的示意。支持不只沿一条直线运行；一个成员可由多个成员联合支持，也可反过来提高其他成员的可信度。

1.4.2 一致性不足以构成融贯

逻辑一致性是融贯的最低要求，但不是充分条件。以下命题彼此不矛盾：

1. 有猫；
2. 有香蕉；
3. 有时会下雨；
4. 有太阳；
5. 有月亮。

这五个命题可以同时为真，但在案例的设定中，它们没有形成显著的演绎或概率支持关系。由“有猫”推不出“有月亮”，把其他四句当真，也没有因此获得相信“有香蕉”的新理由。因此，逻辑一致性只排除矛盾，尚未构成融贯论所要求的相互依赖。

另一组命题的联系就明显得多：

1. Edna 喜欢卡布奇诺；
2. 星期二下午，Edna 想喝一杯卡布奇诺；
3. 当时她认为，最近能买到好卡布奇诺的地方是 Cafe Nero；
4. 星期二下午，她去了 Cafe Nero 买卡布奇诺。

她平常喜欢卡布奇诺，使她那天下午想喝一杯更有可能；“她想喝”与“她认为这家店最近”联合起来，使她前往那家店更有可能；她确实去了，又可以反过来支持关于当时欲望的判断。这些关系并非逐句的演绎蕴涵：喜欢一种饮料不保证任何时候都想喝，去了咖啡馆也可能另有原因。它们是能够双向发生的概率支持关系，因而构成较高程度的系统融贯。

1.4.3 高度融贯是否产生证成

作者的核心反对是：系统融贯程度与成员信念的证成质量并不必然相关。零散信念可能分别通过仔细观察形成；高度融贯的故事则可能来自愿望式思考。主体也可能分别形成若干信念，完全没有意识到它们之间的融贯关系。若融贯没有参与信念形成，就很难解释它为何能产生信念证成。

融贯、真理与证成

1. **系统性质与真假：**咖啡馆故事比猫与香蕉的列表更融贯，却没有理由因此断言前者包含更多真命题。
2. **系统性质与实际信念证成：**相信零散五句的人也许逐句仔细观察；相信咖啡馆故事的人也许只是希望故事成真。若两人形成信念的过程相反，光凭系统分数不能判决谁的实际信念更合理。
3. **未被主体把握的关系：**若四句确实互相支持，但主体分别相信它们、从未意识到它们彼此有关，这种关系至少不能自动解释他为什么以这种方式相信。若关系还复杂到主体无法把握，它连命题证成是否成立也可能有争议。

因此，有相互支持的关系，与这种关系实际为主体的信念提供证成，仍是两回事。

Fiona 的侦探幻想

Fiona 编织出情节严密、细节丰富、内部高度一致的侦探故事，并因幻想过于生动而相信故事中的每个事件真的发生。故事的内部融贯性很高，但她显然没有因此获得关于现实的证成。

侦探小说可以做到人物动机和时间线严丝合缝；从“如果故事中 A 发生，则 B 很合理”却不能跳到“现实中 A 发生过”。Fiona 的问题不在推理能力不足，而在故事整体与她认识现实所用的经验、证言等来源脱节。内部关系再严密，也不能单独填补这一缺口。

该案例说明：

内部高度融贯 $\nrightarrow$ 与现实建立适当联系。

1.4.4 隔离反驳

隔离反驳 (*isolation objection*) 指出，若证成只取决于其他信念，那么知觉、疼痛、冷热感觉等非信念性心理状态便无法直接影响证成。但“我感到冷”的实际感觉显然与“我现在感觉冷”是否合理有关；摩托车声也应当支持“街上有摩托车”。

注意“感到冷”与“相信我感到冷”在种类上不同：前者是非信念性经验，后者才是可真可假的信念。严格融贯论若只容许信念支持信念，即使眼下冷得发抖或清楚听到摩托车冲过街道，这些经验本身也无证成作用。一个人完全不理睬声音、只因旧信念一贯认定此处无车而否认摩托车，这种“仍与信念系统一致”并不使他的否认合理。即使主体对经验形成了信念，仍需说明：为什么实际发生的经验能够影响这些信念的合理性？只计算信念与信念的关系，尚未回答这一问题。

一个与鲜明经验冲突、却与旧信念系统相协调的信念通常仍不合理。严格融贯论使信念系统与世界通过经验提供的输入相隔离。

1.4.5 同样融贯的竞争系统

隔离反驳问的是经验怎样进入证成；**替代融贯系统论证** (*alternative coherent systems argument*) 则问：如果不诉诸经验，只看内部结构，怎样在彼此竞争的系统之间作出选择？⁹

设想两套关于同一事件的详细叙述。第一套把某人的行为解释为出于善意，第二套解释为蓄意欺骗；各自又配有动机、背景和后续行为，使内部联系十分严密。两套叙述不能同时完全正确，却可能具有相近的融贯程度。进一步说，我们可以构造许多彼此不相容而同样融贯的系统，现实却只能符合其中一部分。

这并非从“有两个解释”就推出任何一个都不能得到证成。它所反对的是，仅凭内部融贯程度便足以说明哪一套系统更值得相信。若系统间没有独立的比较依据，融贯与真理之间的联系就仍未得到解释。这个问题也说明，故事内部每一步都解释得通，不等于故事整体与事实相符。

1.4.6 序言悖论：合理信念必须共同一致吗

此前的反驳主要针对融贯的充分性：高度融贯也可能没有证成。**序言悖论** (*preface paradox*) 则质疑另一方向：逻辑不一致是否必然使相关信念失去证成？

设想一位作者认真研究并核查了一本书中的每项具体主张。他分别有充分理由相信 $p_1, p_2, \dots, p_n$ 。但他也知道，即使认真编写的长篇著作，通常难免有某个错误，因此在序言中承认“本书至少有一处错误”。后一信念也可能有不错的经验根据。

⁹ 此处的竞争系统与序言悖论，参见 *Epistemic Justification*，第 3 页。

若把这些信念合起来，便得到：

$$p_1, p_2, \dots, p_n, \quad \neg(p_1 \wedge p_2 \wedge \dots \wedge p_n).$$

它们不能共同为真。可是，关于每项具体主张的证据，与关于整本书可能有误的证据，似乎都不应轻易被忽略。

这没有证明矛盾无关紧要，也没有给出悖论的唯一答案。作者或许应降低对个别主张的确信度，或区分逐项接受与接受全部主张的合取。困难在于：若融贯论把全系统一致性直接设为每一成员获得证成的必要条件，就必须说明为什么这种普遍而谨慎的认识处境不合理。逐项有根据，与有根据地相信它们全都正确，不能未经说明便当作同一件事。

1.4.7 同一系统成员地位问题

若信念得到证成当且仅当它属于高度融贯的总体系统，那么同一系统中的成员似乎都拥有相同证成地位。但正常人的信念体系同时包含可靠常识、弱推测和明显偏见。

假如证成只由“整个系统是否高度融贯”决定，那么系统中的每个成员都会得到相同判定。然而，一个人刚看清钟表形成的时间信念，与他对陌生人动机的无根据揣测，可以同时存在且具有不同的证成地位。理论因而需要解释成员之间的差异，不能只给总体系统一次评价。如何在保留融贯论基本立场的同时作出这种修正，是本章第五道课后问题所提出的任务。

1.5 基础主义与基础信念

基础主义 (*foundationalism*) 认为，每一条成功的推论证成链都在有限步骤后终止于**基础信念** (*basic belief*)。基础信念具有两个核心性质：

1. 它不是从其他信念推论出来的；
2. 它仍然得到证成。

因此，基础信念不是毫无根据的武断信念；“非推论”与“得到证成”必须同时满足。

这里的树根有严格的结构含义：沿着一个推论所得信念往下追查支持它的前提，最终能够抵达某些并非由进一步的前提信念支持、却本身已有证成的信念。如果根节点也必须再由别的信念支持，基础主义便没有回应倒退难题。反过来，只有少数基础信念而不能推出足够多的日常信念，理论也无法抵御全面怀疑论。基础主义必须同时完成“找到地基”和“建起楼层”这两项工作。

基础信念拥有**直接证成** (*immediate justification*)，非基础信念则通过其他信念获得**间接证成** (*mediate justification*)。这里的“直接”是“不以其他信念为中介”，不是“发生得很快”。

一种完整基础主义必须回答四个问题：

1. 一个信念怎样才能非推论地获得证成？
2. 哪些内容的信念可以直接得到证成？
3. 直接证成必须有多强，是否必须达到确定性或不可错性？
4. 基础信念能否支持足够多的普通知识，从而避免全面怀疑论？

传统强基础主义倾向于把基础信念限制为关于当前心理状态且不可错的命题；温和基础主义则允许普通知觉信念获得可被击败的直接证成。

四个问题之间的关系

第一个问题追问**来源**：经验、感觉或其他因素如何直接证成一个信念？第二个问题追问**范围**：只有“我现在似乎看见红色”可以这样得到证成，还是“桌上有一只梨”也可以？第三个问题追问**强度**：这种证成是否必然确定无误？第四个问题追问**延伸**：即使前三题已有答案，它们是否足够支持关于过去、未来和外部世界的信念？这四项要求相互独立：找到了直接来源，不表示其内容范围和强度已经确定；即使这些问题都有答案，也还不能保证基础足够支持日常知识。

1.6 基础信念是自我证成的吗

基础主义内部，对基础证成的来源也有分歧。课堂以**信念的基础主义** (*doxastic foundationalism*) 指称诉诸自我证成基础信念的路线：即使基础信念报告知觉情况，相应知觉状态也不另充当其理由。**非信念的基础主义** (*non-doxastic foundationalism*) 则认为，经验等非信念状态能够为基础信念提供证成。¹⁰

“非信念的”修饰的是证成来源，不是说被证成的对象不再是信念。两条路线都想避免继续诉诸前提信念，但分别需要回答不同问题：信念怎样证成自身？经验又怎样支持一个可真可假的判断？

¹⁰本区分及随意性反驳、塞拉斯二难，参见 *Epistemic Justification*，第 4 页。这里沿用课件对两种基础主义的命名。

1.6.1 第一种理解：相信本身保证内容为真

一种观点把**自我证成** (*self-justification*) 理解为：信念的发生逻辑蕴涵其内容为真。例如，只要主体相信“我至少拥有一个信念”，他就确实拥有一个信念。

这种理解的吸引力在于：推论证成要求另一信念充当前提；如果一个信念仅凭被相信这一事实便保证了其内容，那么似乎不再需要另外寻找前提。不过，**保证为真**和**提供相信它的理由**是两个不同条件。

但多数关于心理状态的信念并不明显满足这种蕴涵。“我相信自己悲伤”是否必然推出“我确实悲伤”仍有争议，因为主体可能误判、误用概念或自我欺骗。

更重要的是，即使这种真理保证确实存在，它仍未必提供证成。设 L 是一条极其复杂的逻辑真理。由于 L 必然为真，“主体相信 L ”当然蕴涵“ L 为真”。但主体可能仅仅理解了命题内容，却完全不知道它为何成立，甚至只是猜中。

在这个例子中，蕴涵关系来自 L 本身的必然性，并不是主体掌握了好的理由。若把“只要相信便保证为真”当作证成的充分条件，就会把这种无根据的逻辑猜测也算作得到证成。因此：

信念内容不可能为假 $\nRightarrow$ 主体有理由相信。

真理保证不能替代认识根据。

1.6.2 第二种理解：事实本身证成信念

另一种理解是：若主体相信自己处于心理状态 M ，而主体确实处于 M ，那么 M 这一事实直接证成相应信念。Roderick Chisholm 据此主张，对于某些第一人称心理命题，主体可以通过重申事实说明证成。

此时变化的是证成来源：第一种理论强调这个信念被相信会保证内容为真；第二种理论强调相应事实确实存在会证成这个信念。Chisholm 用“ a 是 F ，因此我相信 a 是 F 有根据”这一形式表达想法，同时承认这种形式不能任意套用。对“月球上不可能有生命”，仅重说该命题显然无法说明为什么有根据；他认为某些关于自己当前信念、思想的陈述则可以这样做。

问题在于，为什么重申“我现在疼痛”能够说明证成，而重申“月球上不可能有生命”或“Alice 现在疼痛”却不能？Chisholm 把前一类状态称为**自我呈现状态** (*self-presenting state*)，并称它们“通过自身被把握”。但若没有说明这种把握如何产生认识权利，“自我呈现”只是标签，不是解释。

要填上空缺，至少还得说明：主体与自己的当前心理状态之间有什么特殊关系，使相应事实能够在无需另一个已证成信念的情况下成为证成来源；为什么这套说明不扩张

到任意真实陈述。只有说明这种特殊认识关系，“自我呈现”才不只是为待解释现象起一个名称。

1.7 驳斥若干对直接证成的驳斥

1.7.1 随意性反驳

随意性反驳 (*arbitrariness objection*) 认为，无论把理由链停在一个自我证成的信念，还是停在某个经验，似乎都只是任意宣布“到此为止”。如果其他信念要有支持，为什么这一项可以例外？

强基础主义的一种回答是，起点拥有认识上的确定性，因而不是随便挑出的信念。但这一回答随即需要说明哪些起点确定，以及确定性为什么产生证成；前一节关于逻辑真理的例子已经表明，内容不可能为假与主体有理由相信并不相同。

温和基础主义也可以从另一方向回答：停止追加前提信念，不等于取消一切根据。如果经验本身具有非推论的支持作用，链就不是停在任意信念上。争议因此应落实为非推论支持是否可能，而不能把“没有进一步的信念理由”直接等同于“没有任何理由”。

1.7.2 塞拉斯二难

塞拉斯二难 (*Sellarsian dilemma*) 把非信念基础主义的困难进一步明确化。经验要么具有断言式的表征内容，要么没有。所谓断言式内容，是经验把世界呈现为某种样子：它仿佛说“事情就是如此”，但仍可能呈现错误。

若经验具有这样的内容，批评者会问：为什么相信它呈现得正确？似乎还要另一项理由，说明这次经验可信。若经验完全没有这种内容，它又如何支持某个命题为真？一个纯粹发生的感受，怎样在理由关系中为“桌上有梨”提供支持？

二难的第一支试图把经验重新放回理由倒退，第二支则试图剥夺经验作为理由的资格。不过，它的力量取决于两项要求是否成立：经验内容只要可错，就一定需要额外证成吗？没有与信念相同的断言形式，就一定不能提供支持吗？基础主义者不能仅仅重说“经验就是基础”，批评者也不能不经论证便把这两项要求视为当然。后面的层级上升争论，尤其针对第一支中对额外理由的要求。

1.7.3 君主身份的类比

欧洲君主通常通过血统继承王位，但新王朝的君主也可能通过征服首次取得王位。类似地，证成既可以从已有信念传递，也可能在没有其他信念传递“证成力量”时首次产生。这个类比并未证明直接证成存在，却表明“所有证成都必须继承”不是概念真理。

类比的要害是区分取得一种地位的常见办法与取得这种地位的唯一办法。知道许多信念从其他信念继承证成，并不意味着任何证成都必须由已有证成传递而来。基础主义可以主张：如同王朝起点可以通过非继承方式取得王位，证成链也可以通过非推论来源获得起点。

William Alston 给出最低限度的定义：

间接证成与直接证成

- **间接证成** (*Mediate Justification*, MJ): 主体相信 P 得到证成, 是由于该信念与主体其他已证成信念之间的关系。
- **直接证成** (*Immediate Justification*, IJ): 主体相信 P 得到证成, 是由于不同于“与其他已证成信念之间关系”的某种因素。

IJ 只是一个否定性定义, 没有说明直接证成的正面来源。Alston 的目标也较有限: 不是完成基础主义, 而是反驳“直接证成原则上不可能”的论证。

接下来的反对意见涉及更高一层的要求: 主体仅仅拥有证成是否还不够, 必须另外有理由相信自己拥有证成?¹¹若后一要求本身缺乏论证, 就不能用它排除直接证成。

1.7.4 层级上升与门票类比

层级上升 (*level ascent*) 是把一阶条件升级为关于一阶条件的二阶证明要求。持有效门票通常足以进入演唱会; 若检票员还要求持票人提供“证明该门票有效的文件”, 就提高了标准。

同理, 应区分:

信念得到证成 与 主体有理由相信该信念得到证成。

后一项是更高阶状态, 不应在没有论证时被规定为前者的必要条件。

如果每次通过门都得先取得一张证明门票有效的证明, 而那份证明又需要证明自身有效, 问题只会在更高层次重演。认识论中的类似要求是: 要拥有关于 p 的根据, 先得拥有关于“我有关于 p 的根据”的根据。反对者需要说明为什么这种升级是必要条件, 不能仅凭主体有时能反思自己的理由, 推出总要先完成反思。

1.7.5 Sellars 的观察知识要求

Wilfrid Sellars 认为, 一个观察者在绿色物体前可靠地形成“它是绿色的”还不足以构成知识; 观察者还必须能把自己的判断表现理解为绿色物体存在的可靠标志。

¹¹Alston 原本的讨论以知识为对象; Goldman 在这里把相应反驳用于直接证成问题。

这是更具体的层级上升：从“观察绿色物体时形成绿色判断”前进到“认出自己的绿色判断可靠地指示外界有绿色物体”。后一命题关乎判断与世界之间的普遍联系，已经超出了当次观察的绿色本身。

Alston 认为，这要求主体拥有二阶信念：“我的绿色判断是对象为绿色的可靠迹象。”若任何一阶知识都要求这种二阶认识，二阶认识又可能要求三阶认识，从而重新制造倒退。儿童可以在不能解释视觉可靠性的情况下合理地相信面前有红球，正说明实际证成与反思性证明需要区分。

1.7.6 Bonjour 的高阶论证

Laurence Bonjour 设 Φ 为任何声称能够直接证成信念 B 的性质。他认为主体必须在认知上掌握：

B 具有性质 Φ ;
具有 Φ 的信念很可能为真;
 $\therefore B$ 很可能为真.

如果 B 的证成依赖主体对这两个前提的已证成信念， B 就不是直接得到证成。

例如候选性质 Φ 可以是“由某种直接经验引起”。Bonjour 并非简单否认经验可能与真理相关；他认为主体还需要有根据地把自己的这个信念归入那类可靠信念，并有根据地相信那一类倾向于为真。只有这样，主体才有资格由 Φ 走向 B 很可能为真。若每个直接来源都加上这两项前提，直接证成在结构上又变成间接证成。

Alston 的回应是：为什么必须要求主体知道这些高阶前提？Bonjour 把“证成必须被主体认知掌握”当作前提，却没有独立证明它。因而，他尚未证明直接证成不可能。

1.8 强度、内容、击败者与增强者

1.8.1 基础信念的内容

笛卡尔式基础主义 (*Cartesian foundationalism*) 把基础信念限制为关于第一人称当前心理状态的信念，例如“我似乎看见红色”。20 世纪的温和基础主义则允许“桌上有一个梨”这样的物理对象命题直接由知觉经验证成。

两种观点的分歧不在于主体是否使用视觉，而在于基础信念的内容：知觉是先直接证成关于经验的命题，再推导外部对象；还是直接证成关于外部对象的普通信念。

1.8.2 直接性与证成强度

笛卡尔把认识论理解为追求确定性，并将当前第一人称心理状态视为获得这种确定性的起点。但内容范围与证成强度之间没有明显的必然联系：一种信念可以因为内容特殊而适合直接得到证成，却未必因此不可能出错。现代基础主义也因此把来源与强度分开考察：

来源问题	证成是否通过其他信念获得，即直接还是间接
强度问题	证成是微弱、充分、很强，还是达到确定性

一个模糊的视觉经验可以直接提供较弱证成；一个数学结论则可能通过长篇证明间接获得极强证成。因此，直接性不蕴涵不可错性。

1.8.3 初步证成、击败者与增强者

当某个来源为命题注入正面支持时，命题对主体具有**初步证成** (*prima facie justification*)。综合全部相关因素后的地位称为**全盘考虑后的证成** (*ultima facie / all-things-considered justification*)。

削弱或取消初步证成的因素称为**击败者** (*defeater*)；增加证成支持的因素称为**增强者** (*booster*)。信念性理由和非信念性经验都可能发挥这两种作用。

初步有根据，不保证全盘考虑后仍有根据。一个清楚的视觉印象可能支持某判断，而关于照明异常或知觉受干扰的有力信息则可能削弱这种支持；负面因素足够强时，初步证成便被击败。反过来，单一来源的支持可能还不充分，另一来源的支持加入后，却足以使相信该命题成为合理态度。

因此，“初步”与“全盘考虑后”的区别，关乎是否已经计入其他相关因素；“直接”与“间接”的区别，则关乎支持是否通过其他信念而来。一个直接来源既可能很强，也可能很弱，还可能在整体评价中被增强或击败。

1.8.4 829 型飞机案例

主体在黄昏和云层中看到一架轮廓类似 829 型的飞机，但远视不佳，视觉证据不足。此时他想起该新机型当天会在当地试飞。视觉经验提供非推论支持，背景信念提供推论支持；二者单独都未达到相信门槛，结合后却足以证成“那是 829 型飞机”。

关键不是飞机恰好确为 829 型，而是主体拥有哪些相互补足的支持。云层、光线和视力分别解释视觉线索何以不够；已读到的试飞消息则解释为什么这架特殊机型此时出现在这一地区变得较可信。如果把整个信念一次性贴成“基础”或“非基础”，这两种支持就会被迫择一，无法描述案例中的证成结构。

因此，直接证成和间接证成不宜被视为对整个信念的互斥分类。更合理的做法是把它们用于**证成来源**：同一个信念可以同时具有直接来源和间接来源。

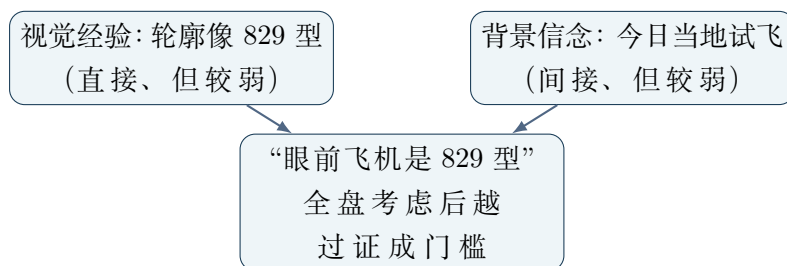


图 1.3 双重证成来源。“基础”与“推论”描述支持来源，不必把整个信念永久归入互斥类别。

融贯论者会指出，任何知觉信念都可能被背景信念增强或击败。基础主义可以接受这一点，却不必承认“只有信念间关系才能产生证成”。经验仍可提供原始支持，其他信念则参与最后的综合评价。

承认其他信念能影响直接证成，确实接受了融贯论的一部分见解：即使一个信念有直接的经验性来源，它最后能否成立仍可能取决于与其他信念的关系。这会使“基础信念”一词产生一点张力：若它被严格定义为全部证成都非推论，829 型飞机的信念便不是基础信念；若讨论的是基础性的证成成分，同一信念又确实含有直接来源。把“直接”与“间接”用于证成成分，便可以描述这种混合情形，而无须否认经验本身提供支持。

1.9 从地基走向更高楼层

1.9.1 从心理经验到外部世界

对于把基础限于第一人称当前心理状态的笛卡尔式基础主义，还有一项特殊困难：即使关于当前经验的信念直接得到证成，它仍须说明如何走向“桌上有梨”“外面有树”等外部世界命题。由“我似乎看见一个梨”不能演绎推出“桌上确实有梨”，因为梦境、幻觉和欺骗情形在逻辑上仍然可能。

许多日常和科学推理并非演绎，而是**最佳解释推理**（*Inference to the Best Explanation, IBE*），也称**溯因推理**（*abductive inference*）：在能够解释证据的竞争假设中，选择总体上最好的假设。

1.9.2 鹿迹与模拟器

主体在雪地中看到类似鹿的足迹，又知道鹿常在附近活动，于是推断有鹿经过。邻居提出竞争假设：足迹可能由一种不会留下轮迹的鹿迹模拟器制造。

邻居提到，名为“Toys for Tricksters”的公司已经推出这种机器，而且机器在恶作剧者中颇为流行。他声称眼前的足迹就是机器造成的；退一步，也至少坚持这是完全可能的。于是他断言：主体没有理由相信鹿经过这里。

两种假设都能说明足迹的形状，争议却尚未解决。“可以解释”不等于“解释得同样好”；仅仅存在一种替代可能，也不自动取消原来的根据。不过，邻居关于机器已上市且广受欢迎的消息，又确实构成必须考虑的背景信息。判断鹿假设是否仍值得相信，需要一项比较解释的原则。

1.9.3 解释推理证成原则

候选原则**解释推理证成原则** (*Justification by Explanatory Inference, JEI*) 为：

JEI

若假设 H 解释主体 S 的证据 E ，且不存在与 H 不相容、但能同样好或更好地解释 E 的假设 H' ，那么 S 有理由以 E 为基础相信 H 。

原书还给出一个较强的备选读法：把上述条件理解为“当且仅当”——既要求其足以让 S 有理由相信 H ，也要求只有满足此条件 S 才有理由相信 H 。两种读法不能混同：前一种是充分条件，后一种同时宣称它是必要条件。若采用较强的“当且仅当”版本，而又有不相容的假设同样好或更好，那么仅据这些证据肯定 H 便不符合原则，应当考虑悬置判断。若仅采用充分条件版本，条件未满足只说明该原则没有给出许可，不能仅凭这一点反推所有相信 H 的根据都不存在。哪一种版本正确，还需进一步论证。

解释优劣常由以下标准衡量：

- **解释简洁性** (*explanatory simplicity*)：在其他条件相同的情况下，引入较少基本设定；
- **与背景知识融贯** (*coherence with background knowledge*)：更符合已广泛接受的知识；
- **解释深度** (*explanatory depth*)：对现有资料给出更深入、更有启发性的说明；
- **避免特设成分** (*avoidance of ad hoc elements*)：较少使用仅为应付反例而添加、缺乏独立支持的补丁。

这些标准都以“其他条件相同”为前提。简单理论不必然为真，复杂现实也可能需要复杂解释。

1.9.4 Berkeley 的竞争解释

多数哲学家认为，独立于心灵的物质对象是我们持续、稳定经验的最佳解释。但 George Berkeley 提出：经验序列可以由上帝这一无限心灵有规律地产生，无须额外假设不可直接经验的物质实体。

Berkeley 甚至声称自己的解释更简单、更符合背景知识：我们已经知道心灵能够引起经验，却尚未独立证明“物质实体”能够因果地产生经验。这一比较本身可以争论，但它揭示了最佳解释推理的困难：不同理论对何为基本实体、背景知识和简洁性可能有根本分歧。

第一章总结：证成的来源与传递

第一章的论证可以整理为以下六步：

1. 认识论评价信念是否合理，证成不同于真理，也不同于道德正确性。
2. 大量信念通过其他信念获得支持；继续追问支持前提便形成证成倒退。
3. 无穷主义允许无限理由链，却难以解释证成来源及有限主体如何拥有所需信念结构。
4. 融贯论把线性链条改造为相互支持的整体网络，却面临虚构系统、经验隔离和成员地位差异问题。
5. 基础主义以非推论地得到证成的基础信念终止倒退；但“自我证成”不足以解释这些信念，批评者的高阶要求也未能证明直接证成不可能。
6. 温和基础主义因此采用更灵活模型：经验提供可错的初步支持，背景信念可以增强或击败；再通过最佳解释推理尝试从经验走向外部世界。

第一章结论

第一章没有宣布某一理论最终获胜。其较稳健的结论是：无穷主义与严格融贯论存在严重困难；基础主义仍有希望，但必须解释直接证成的正面来源，并说明基础经验如何合理支撑有关外部世界、过去与未来的普通信念。

第一章课后问题参考思路

以下为原书章末五题的讨论与参考解答。

问题一：道德规范性与认识规范性有何共同点和差别

题目所问

道德上有“应当 / 不应当”，认识活动中也有“合理 / 不合理”“得到证成 / 没有得到证成”。既然二者都是关于正确与错误的评价，为什么又说它们是两种不同的规范性？

二者的共同点在于，它们都不只是描述主体事实上怎样行动或怎样相信，而是在评价主体**应当**怎样。说“Alice 相信明天会下雨”，只是心理事实；说“Alice 根据当前证据应当相信明天会下雨”，则加入了认识规范。同样，“Bob 拿走了钱”是事实描述，“Bob 不应当拿走钱”是道德评价。

但是，两类评价追求的直接目标不同：

规范领域	直接评价对象	典型目标或标准
道德规范性	行动、意图、品格及其对他人的影响	权利、福利、公正、义务、德性
认识规范性	信念、悬置判断、推理与求证方式	真理、避免错误、证据契合、理性一致

因此，可以把二者看作同一大类中的不同种属：共同点是都涉及可评价的“应当”，差别是评价对象与成功标准不同。例如，医生关于病情的判断可以证据充分，但他泄露病人隐私的行为仍可能在道德上错误。这里被认识论评价的是诊断信念，被道德评价的是披露行为；一种评价的正面结果并不决定另一种评价的结果。类似地，善意的帮助也可能建立在未经核实的谣言之上。

同属规范评价，意味着两者都超出事实描述、提出应当如何的标准；属于不同种类，则意味着它们所依据的理由和直接评价对象有所不同。二者有共同形式，并不要求具体标准完全相同。

问题二：基础信念必须比非基础信念更强吗

基础主义首先是一种关于**证成来源**的理论，而不必是一种关于**证成强度排序**的理论。基础信念的定义性特点是，它的证成不依赖其他前提信念；非基础信念则至少部分通过推论获得证成。由此并不能直接推出：

$$\text{基础信念的证成强度} > \text{任何非基础信念的证成强度}$$

几条彼此独立的证据可以共同支持一个非基础信念，使其总体证成强于任一单独的基础信念。例如，模糊视觉只给“远处有人”提供较弱的直接支持；声音、定位数据与可信证言共同加入后，关于那个人身份的推论信念可能得到极强证成。

笛卡尔的“我思”之所以特殊，是因为在主体思考它时，试图怀疑它反而实例化了它：若我正在怀疑、判断或受骗，我至少正在思考，因而作为思考者存在。这常被描述为**自证性**（*self-verification*）或在实际思考中的不可错性。但应注意两点：

1. 这种性质说明“我正在思考”具有极强地位，却不等于所有基础信念都必须具有同等确定性；
2. 普通知觉信念可以非推论地得到初步证成，同时仍可能被幻觉、光线异常或新证据击败。

明智的基础主义者因此应采取**温和基础主义**：以“非推论来源”界定基础性，而不以“绝对不可错”界定基础性；允许基础信念证成有强弱、可以被击败，也允许非基础信念因多条证据汇合而比某些基础信念更强。

问题三：法庭中较融贯的故事是否证明融贯论正确

法庭上，控辩双方都试图提供能够把证言、物证、时间线和动机串联起来的故事。若控方故事解释更多证据、内部冲突更少，陪审员通常确实有理由更倾向于它。这说明**融贯性具有认识价值**，但尚不能推出严格融贯论。

严格融贯论的核心主张并不是“融贯性有时增强证成”，而是“证成完全来自信念系统内部关系”。法庭案例中，故事之所以值得相信，还取决于系统外部的输入：

- 指纹、录像和 DNA 检验是否真实、程序是否可靠；
- 证人是否看清、是否诚实、是否受到诱导；
- 证据链是否完整，是否存在被忽略的反证；
- 两个故事中哪一个更能解释这些独立取得的材料。

一个侦探小说可以高度融贯，却不因此成为真实案件记录。基础主义者可以说：知觉、记录、证言等提供初步或基础证成；故事之间的融贯性随后充当**增强者**（*booster*），通过最佳解释推理提高某一假设的总体可信度。这样既承认融贯性的作用，又否认它是唯一来源。

问题四：简单、较少特设成分的故事为什么更值得相信

简洁性与避免特设成分通常是**可被击败的解释美德**，不是独立保证真理的魔法规则。它们具有认识价值，至少有三种可能理由：

1. 在其余条件相同、额外设定又缺乏支持时，少依赖一项设定，便少承担一项尚未说明的正确性要求；不过，不能单凭假设的数量计算整个理论为真的概率；
2. 特设补丁只为保护理论免受某个反例攻击，却没有独立证据，因此可能是在“救理论”而非解释世界；
3. 较简单理论通常更容易产生清楚预测，因而更容易接受新证据检验。

然而，“较简单”不能自动压倒证据。真实犯罪可能本来就由多人、巧合和复杂动机共同造成；过分简单的控方故事也可能只是删除不利细节。在法庭案例中，即使控方故事相对更简单，也还需要另外判断其证据是否达到判罪所要求的强度。相对优于另一故事，与自身得到充分证明，并不是同一个条件。

简洁性的适用条件

简洁性最好理解为“其他条件相同时”的比较标准。它能在解释力、证据覆盖和背景知识大致相等时打破平局，却不能替代物证、可靠证言和对替代假设的认真检验。

问题五：怎样修改融贯论，使同一系统中的信念具有不同证成地位

原书批评一种粗糙的系统级融贯论：

$$B \text{ 得到证成} \iff B \text{ 属于一个高度融贯的总体系统.}$$

这一公式只评价“整个系统有多融贯”，于是同一系统中的所有成员似乎只能一起得到证成或一起失去证成。较有希望的修正，是把证成从纯粹系统级性质改成**信念相对于系统的局部融贯度**：

局部融贯修正版

信念 B 的证成强度取决于： B 与系统中多少独立信念相互支持；这些支持信念本身有多稳定； B 在解释系统其他部分时是否具有作用；是否存在未解决的冲突或击败者；删除 B 会使系统整体解释力下降多少。

这样，一个被多条记忆、观察和背景知识支持的中心信念可以高度证成；一个孤立、与多项证据冲突、只能靠临时补丁保留的边缘信念则可能不受证成，尽管二者属于同一总体系统。

不过，修正版仍面对两项困难。第一，如果所谓“独立支持”最终仍只是其他信念，隔离反驳并未消失；一个与现实完全脱节的幻想系统仍可能内部组织良好。第二，若理论允许知觉经验直接决定某些节点的权重，它就已经吸收基础主义成分，成为某种“基础—融贯混合论”。因此，这个修正可以解决“成员地位一刀切”的问题，却未必保住纯粹融贯论。

Goldman 第二章：证据主义、可靠主义与内外在主义

Alvin I. Goldman 与 Matthew McGrath, *Epistemology: A Contemporary Introduction*, Chapter 2, “Two Debates About Justification,” pp. 25–50.

一个信念是否合理，取决于主体此刻拥有的证据，还是也取决于这个信念实际上怎样形成、怎样保存至今？这一问题把证据主义与过程可靠主义区分开来。争论随后进一步涉及证成因素的范围：决定证成的条件，是否必须全部属于主体的心理状态，或全部能由主体直接把握？

2.1 证成与证据

2.1.1 传统认识论的共同预设

基础主义与融贯论虽然在证成结构上相互对立，却共享若干传统预设。基础主义允许经验和其他信念为目标信念提供证成；融贯论则把证成限制在信念或其他信念性状态之间的关系中。但两者通常都认为，真正能够影响主体证成地位的因素，是主体自己的心理状态。

证成因素

证成因素 (*justifier / J-factor*)，是任何能够正面或负面影响目标信念之证成地位的因素。它既可以增强一个信念，也可以削弱或击败一个信念。

传统理论主要承认两类证成因素：

1. **其他信念**。主体可以从已经相信的前提出发，推导目标信念。当主体回答“我之所以相信 P ，是因为 X, Y, Z ”时，所援引的理由通常就是其他命题性信念。
2. **经验状态**。例如视觉经验、听觉经验、疼痛经验，以及一种有意识的“似乎回忆起”的记忆经验。

融贯论只承认第一类；基础主义则允许第二类直接为基础信念提供证成。

这里“记忆经验”特指一次有意识的似乎回想起来，例如突然仿佛再度听到姐姐当年的话。它与后文的“记忆保持过程”不同：人可以把一个信念保存多年，而当下并没有发生任何鲜明的回想经验。若把这两种记忆混为一谈，就看不出 Jane、Nora 和 Mary 的案例为何对“只算当前经验”的理论构成挑战。

2.1.2 当下时间切片假设

传统认识论还常常接受一个更强的限制：主体在时刻 t 对 P 的信念是否得到证成，只由主体在同一时刻 t 所处的心理状态决定。过去发生的认知事件本身，不进入当前证成地位的计算。

Jane 的童年记忆

25 岁的 Jane 相信，自己 7 岁时曾因姐姐一句刻薄的话受到伤害。她现在没有关于当时感受或声音的鲜明回忆，只是仍然相信那件事发生过。传统理论在评估她当前信念时，要求只查看她当前的心理状态，而不把她 7 岁时真实经历过的心理事件直接算作当前证成因素。

这种立场可称为**当下时间切片假设** (*current-time-slice assumption*)，也被形容为“瞬间的唯我论”。它继承了笛卡尔式认识论的基本规划：从主体此刻能够把握的思想和经验出发，重建有关外部世界、过去与未来的信念体系。

以 Jane 为例，“七岁时确实受伤”关系到现在记忆信念是否为真；传统理论要求把这一过去事件排除出现在是否有理由相信的直接决定因素。于是理论还须给出三条推理道路：从现有感觉推到外部物体，从现在推到未来，从现在推到过去。作者在此描述的是传统理论承担的任务，而非事先断言 Jane 的信念缺乏证成。

理论困难

如果证成只能依赖主体此刻拥有的心理材料，那么记忆信念会形成困难：一个人可能曾经通过可靠来源学到某事实，随后忘记来源，却仍然稳定地保留该信念。若过去的获取过程完全不重要，就难以说明这个信念为何仍然可以是合理的。

2.1.3 信念证成与命题证成

评价实际信念，与评价主体所处的认识情境，需要分开：

概念	评价对象	基本问题
信念证成 (dox- astic justification)	主体实际持有的某个具体信念	这个信念是否以适当方式形成或维持?
命题证成 (propositional justification)	命题、主体及其认识情境	即使主体尚未相信 P ，在当前情境下相信 P 是否合适?

主体可能对 P 拥有命题证成，却没有信念证成。例如，Jack 面前的数据事实上支持候选人获胜，因此他处于适合相信该命题的证据位置；但如果他事实上仅凭愿望式思考形成该信念，那么这个具体信念并未得到适当形成。

两种评价可以针对同一时间、同一命题、同一个人得出不同结果。命题证成说明主体具备相信的根据；信念证成还要说明这个实际信念与根据之间的关系。因此，一个主要解释命题证成的理论，未必不加修正便能解释信念证成。

2.1.4 心灵主义证据主义

Conee 与 Feldman 所代表的心灵主义证据主义 (*mentalist evidentialism*) 主张：主体 S 在时刻 t 的全部正面与负面证成因素，都由 S 在 t 时刻拥有的证据性心理状态构成。这些状态包括经验和已证成的信念。

其核心思想可写成：

S 对 P 具有命题证成 $\iff$ 相信 P 是与 S 的总体证据相契合的态度。

这里的关键概念是**契合** (*fittingness*)：证据与所采取的态度之间必须具有适当关系。

“总体证据”同时包括有利和不利材料。例如，一项初步观察似乎支持 P ，后来一条清楚的反证却要求悬置判断；只选取前一项便不能确定“相信 P ”是契合的态度。契合也不表示“只要有任何支持就应当相信”：支持不足时，恰当态度可能是暂不判断。

推论性契合

当证据由命题性信念构成时，证据内容必须充分支持目标命题。支持可以是：

- 演绎支持：前提逻辑蕴涵结论；
- 归纳支持：过去的规律与当前观察使结论具有较高概率；

- 解释性支持：目标假设构成对证据的最佳解释。

例如：

E_1 ：过去在某地点出现烟时，那里通常有火；

E_2 ：现在地点 L 出现了烟；

H ：地点 L 现在有火。

E_1 与 E_2 并不严格演绎出 H ，却对 H 形成较强的归纳支持。

即使通常“见烟知火”，该处的烟也可能另有来源。因此，推论性契合不限于有效演绎，还包括强度可以变化、可能受新资料击败的归纳支持。第一章的最佳解释推理则是第三种候选支持关系。

演绎与溯因的差别还可以用课堂的两个例子说明。由“院子里有一只狐狸”，在狐狸是哺乳动物这一分类前提下，可以推出“院子里有一只哺乳动物”；在前提真实的条件下，结论不可能假。¹²

如果在 Smith 的汽车后备箱中发现一只带血手套，推测 Smith 是凶手，则属于解释性推理：凶手假设试图说明为什么手套出现在那里。但这个证据不逻辑蕴涵他有罪；手套也可能由他人放入，或者与另一事件有关。要使该推测获得充分证成，还必须比较替代解释，并结合其他证据。课堂的简式突出的是支持关系的类型，不能把它理解成一条单凭手套就足以定罪规则。

非推论性契合

经验如何直接支持信念则更加困难。理论至少需要回答：

- 经验是否具有可以由陈述句表达的命题内容？
- 如果经验没有命题内容，它如何与具有命题内容的信念发生“契合”？
- 视觉与触觉相互冲突时，应给哪一种经验更高权重？
- 清晰度和鲜明度如何改变经验的证成力量？
- 预感也是一种经验，它是否也为其内容提供证据？

此外，不能把任何信念都直接算作证据。一个未经证成的前提，即使碰巧与结论构成有效推理，也不能把证成传递给结论。因此，更准确的说法是：只有已证成信念才能充当推论证据。

¹²狐狸与带血手套的例子，参见 *Epistemic Justification*，第 5 页。

这样一来，证据与契合的说明又需要进一步的条件：判定某个前提能否作为证据之前，理论已经需要知道它是否有证成。至于视听冲突时该相信谁、疼痛经验与命题如何契合、预感是否算数，则表明非推论性契合仍有大量工作未完成。

本节结论

证据主义以“证据”和“与证据契合”解释证成，形式简洁，但“契合”本身并不简单。特别是经验与信念之间的非推论性契合，仍需一套正面理论。

2.2 证据主义面临的问题

2.2.1 证据共存是必要条件或充分条件吗

充分性：拥有证据，还需要怎样相信

如果证据主义只要求目标信念与适当证据同时出现在主体心中，那么这种共存条件不足以产生信念证成。

Jack 与选举结果

Jack 支持国会候选人 Cindy。开票当晚，他看到第一批选区的一组数据；若仔细计算，这组数据确实足以使人对 Cindy 胜选有较高信心。然而 Jack 情绪激动，粗心算出一个令人失望但错误的结果。他的注意力却更多放在另一批选区：那里的已出数据很少，他仍凭愿望认定那里会有大量人支持 Cindy。最终，他相信 Cindy 将获胜，真正起作用的是对后一批选区的愿望式猜测。

Jack 拥有支持结论的证据，也最终相信了这个受证据支持的结论。但是，他的信念实际上由愿望式思考和错误计算产生。因此：

证据存在且支持 $P \not\Rightarrow$ 实际的 P 信念得到证成。

Jack 对“Cindy 会获胜”可能拥有命题证成，却没有信念证成。问题在于，适当证据并未通过适当思考实际产生他的信念。

这个案例的关键，是把两批材料的作用分开。第一批数据提供了好证据，却因计算错误而没有成为 Jack 相信胜选的适当根据；第二批数据本身尚不足以支持乐观判断，却因愿望式思考而实际推动了信念。结论内容与证据相契合，并不能修复形成这个信念的方式。

证据主义者可以加入**基于关系** (*basing relation*)，要求信念不仅与证据相契合，而且确实基于适当证据。也可以区分“有理由相信”与**有根据地相信** (*well-founded belief*)，仅对后者提出这种要求。但“基于”需要进一步说明：如果它要求证据通过合适的思考

过程产生信念，那么因果形成史便已进入认识论解释。信念如何形成，就不能全部归入与证成无关的心理学问题。

证据共存是否必要

当前证据与信念共存也未必是信念证成的必要条件。

Nora 与地球周长

Nora 相信地球周长约为 25,000 英里。她现在想不起自己从哪里读到或听到这一事实，也没有当下的相关经验；但假设她最初确实从权威来源学到该事实，并通过记忆稳定保留至今。直觉上，她当前的信念仍可能是得到证成的。

Nora 案例说明：

当前没有相伴的原始证据 $\nRightarrow$ 当前信念不受证成。

这个判断有一项前提：承认 Nora 在忘记来源后仍可保有证成。如果认为证据遗忘已经使她失去证成，就还需要另外争论。但接受这一通常判断后，当前证据共存就不再是必要条件。这与 Jack 案例形成对照：Jack 当前有证据，实际信念却无证成；Nora 当前没有相伴的原始证据，信念却仍可能合理。

证据主义者可能援引 Nora 关于自己记忆可靠性的信念。不过，Nora 未必形成过“我的记忆通常可靠”这样的高阶判断；即使追问时能够形成，它也未必在此前实际维持她的周长信念。案例中真正有解释力的，似乎是她过去可靠地学到了这个事实，并通过记忆稳定保存至今。这里的困难不只是 Nora 说不出理由，而是当前没有相应的支持经验或前提信念，却仍可能有一段良好的获取与保存历史。

内省是否总伴随经验性证据

人们通常认为，“我现在头痛”或“我认为自己的社会保障号码是某串数字”等关于当前心理状态的信念，可以非推论地得到证成。若所有非推论证据都是经验，那么内省似乎也应伴随某种经验性证据。

然而，内省没有视觉、听觉那样独特的感觉性质。若把目标心理状态本身视为证据，又会出现“状态为自身提供证据”的疑问：

- “使我有理由相信自己悲伤的，是我确实悲伤”，似乎具有一定直觉吸引力；
- 但“使我有理由相信梭状回正在激活的，是梭状回确实正在激活”，显然不能成立。

证据主义必须说明，为什么主体的悲伤可以直接支持关于悲伤的信念，而某个真实脑状态却不能直接支持关于该脑状态的信念。单纯重复事实不能解释这种认识论差异。

还要区分两个对象：头痛可能具有显著的感觉特征，但“我现在相信自己的社会保障号码是某串数字”所指的判断并无相应的特殊感觉质。如果把一切内省依据都说成感觉经验，就很难说明后一类第一人称信念的证成从何而来。作者是在提出解释负担，而不是已经证明任何内省必定没有经验。

预感问题

预感是一种心理经验，但一般不被认为能为其内容提供正当证据。一个人突然预感灾难将发生，并不因此就有理由相信灾难将发生。

最自然的解释是：预感并不是其内容为真的可靠指标。但一旦使用“可靠指标”解释哪些经验具有证据作用，证据主义就引入了并非纯粹心理状态的外在因素。

2.2.2 “证据”指什么

证据主义依赖“证据”解释“证成”，却需要给证据一个非循环定义。

一种定义是：

X 是 Y 的证据，当且仅当，拥有 X 能够提高主体相信 Y 的证成。

若证成又被定义为“符合证据”，便形成循环：

证成 $\rightarrow$ 证据 $\rightarrow$ 证成.

这类似于把“巫师”定义为制造魔法事件的人，再把“魔法事件”定义为巫师制造的事件。两个定义互相指向，却无法帮助判断现实中是否存在巫师或魔法。

另一种定义是：

证据是对真理的可靠指示，即通常能够正确指向事实的标志。

该定义可以解释为什么闪电是雷声即将出现的证据，也可以解释为什么预感通常不是证据。但“对真理的可靠指示”并非主体内部的心理状态。若可靠指示性影响证成，它本身就是一种外在证成因素，从而削弱心灵主义证据主义的核心主张。

这不是说视觉经验本身不再是心理状态，而是说经验之是否可靠地指示外部事实取决于世界与经验的实际联系，不能仅从主体当下的心理状态读出。证据主义现在面临精确的两难：若把证据解释为“能增加证成的东西”，解释会循环；若把证据解释为“可靠的真理迹象”，它又放弃“全部证成因素均为心理状态”的限制。

证据主义的三重困难

1. **不充分**：主体可能拥有支持 P 的证据，却因愿望式思考形成 P 。
2. **非必要**：主体可能忘记原始证据，却因可靠学习和记忆保存继续合理地相信 P 。
3. **定义困境**：用证成定义证据会循环；用可靠性定义证据又会引入外在因素。

2.3 证成与过程可靠性

2.3.1 为什么仅仅“基于证据”仍不充分

Jack 案例表明，信念必须以适当方式从证据产生。但即使信念确实以证据为输入，也未必得到证成。

标准的析取三段论 (*disjunctive syllogism*) 是：

$$\frac{P \vee Q, \neg Q}{\therefore P}.$$

如果主体对两个前提拥有已证成信念，并正确应用该规则，那么结论可以继承证成。

Chad 的过度推广规则

Chad 在看到“ $P \star Q$ ”和“非 Q ”时，总是推出 P ，其中 $\star$ 可以是任意二元真值联结词，包括析取、合取或条件句。该规则在析取情况下偶然与正确规则一致，却在大量其他情况下无效。假设 Chad 正是通过这套荒谬的一般规则，从“ $P \vee Q$ ”与“非 Q ”推出 P 。

虽然 Chad 当前的前提恰好逻辑蕴涵结论，而且结论确实“基于”这些前提，他的信念仍不合理。原因是，他所使用的认知过程一般会频繁产生错误。

要判断这里的失误，必须观察他实际采用的规则，而不只观察这一次碰巧成立的逻辑关系。严格的析取三段论只在“ P 或 Q ”和“非 Q ”下推出 P ；Chad 的习惯却把中间的 $\star$ 替换成任何二元联结词。例如，若前提是“如果 P 则 Q ”和“非 Q ”，正确推理应当得到“非 P ”，而 Chad 却仍推出 P 。令 P 、 Q 都为假，就能看到两个前提为真、他的结论却为假。眼下“或”的实例正确，并未将他所用的糟糕方法变成好方法。

因此：

符合证据 + 基于证据 $\not\Rightarrow$ 信念证成。

理论还必须评价把输入转化为输出的信念形成过程。

2.3.2 过程可靠主义的核心思想

愿望式思考和 Chad 的无效推理具有共同缺陷：它们产生真信念的比例很低。正常知觉、内省和有效推理则通常产生较高比例的真信念。

过程可靠主义

过程可靠主义 (*process reliabilism*) 认为：一个实际信念是否得到证成，取决于产生或保持该信念的认知过程是否具有较高的真理产出倾向。

这一思路评价的是过程产生真信念的倾向，而非某一次输出碰巧是真是假。愿望式思考有时也能猜中，正常知觉有时也会出错；单次结果不足以判定方法的认识价值。

2.4 过程可靠主义的表述

2.4.1 从初步原则到两种可靠性

最初步的表述是：

$$B \text{ 得到证成} \iff B \text{ 由可靠的信念形成过程产生。}$$

“足够可靠”的精确百分比可能是模糊的，因为“证成”本身也存在边界模糊性。理论不必承诺一个固定的 80% 或 90% 阈值。

但这一表述还没有区分不同类型的过程。对于直接产生信念的知觉过程，可以考察它通常有多少真输出；对于以已有信念为前提的推理过程，则不能不问输入是否为真。再好的推理也不能保证从错误前提得到真实结论。

知觉等过程不以已有信念为输入，而推理和记忆通常以已有信念为输入。二者应采用不同的可靠性标准。

两种可靠性

- **无条件可靠性** (*Unconditional Reliability*, UR)：不以其他信念为输入的过程，在一般情况下倾向于直接产生真信念。
- **有条件可靠性** (*Conditional Reliability*, CR)：以其他信念为输入的过程，在输入信念为真时，倾向于产生真输出。

第二项的“有条件”限定的是过程类型的真率测量：统计那些输入信念均为真的情

形，看这种规则通常会不会输出真信念；它不是说某次输入实际上必须全真，才能让当次输出获得证成。后一个问题由下面的 CR 回答：当次输入应当得到证成。

若推理输入本身为假，即使过程完全有效，也不能保证输出为真。评价有条件过程，因而需要把方法本身与这一次输入的地位分开。可靠主义提出：

(UR) B 若由无条件可靠过程产生，则得到证成；

(CR) B 若由有条件可靠过程产生，且全部信念输入都得到证成，
则 B 得到证成。

这里要求输入信念得到证成，而不要求它们事实上全真。认识主体能够被合理要求做的，是采用良好程序；即使良好程序偶尔出错，也不自动取消证成。

2.4.2 历史性原则

将无条件过程、有条件过程和记忆保存连接起来，可以得到历史性的可靠主义原则：

可靠主义原则 R

时刻 t 的信念 B 得到证成，当且仅当， B 是一个信念形成或信念保持过程序列的输出；其中每一步使用的过程都是无条件可靠或有条件可靠的，而后者所接受的每个信念输入来自此前符合这一要求的步骤。这样，从最初形成到后来保存的整条认识历史都进入评价。

Mary 与娱乐新闻

Mary 从娱乐杂志 *Gossip* 读到 Kenneth 与 Gwyneth 申请离婚。她先通过正常视觉形成“杂志如此报道”的信念，又有一个此前已得到证成的背景信念：“这本杂志报道此类消息通常可信。”她据此推断离婚消息为真，随后通过记忆把该信念保存一周。一周后，她可能已经忘记最初读过这本杂志，却仍然保留该信念。

Mary 的信念经过三个相互衔接的阶段。视觉读取消息是无条件过程；从杂志的报道和已确立的可信度推到离婚消息，是以前信念为输入的有条件过程；跨越一周的记忆保存也是有条件过程。这里没有要求 Mary 在一周后重新经历一次“我似乎记得 *Gossip* 这样写”的意识事件，也没有要求她现在能够说出最初的来源。假设期间没有新的反证，过去获取与保持的可靠链条便能说明现在的证成。

如果知觉、推理和记忆保持过程都可靠，那么 Mary 一周后的信念仍可得到证成。当前证据不再出现并不重要，因为其获取与保存历史仍然良好。

可靠主义由此引入了传统理论没有充分承认的一类证成因素：**信念形成和保持过程**。这些过程并不是主体在辩护时必须援引的证据或理由，却能解释其信念为何具有特定证

成地位。

2.4.3 可靠主义与命题证成

过程可靠主义首先解释实际信念的信念证成。若主体尚未形成 P 信念， P 没有实际因果历史，原则 R 不能直接应用。

一种扩展方案是：若主体当前拥有某些心理状态，把这些状态输入可靠过程便会产生 P 信念，那么主体对 P 具有命题证成。

例如，主体正在观看山景，视觉经验足以使正常观察者通过可靠辨识形成“远处有三座山峰”的信念；即使主体尚未注意山峰数量，也可以说他有理由相信该命题。

证据主义与可靠主义的分界

证据主义关注主体当前**拥有什么理由**；过程可靠主义关注信念事实上**如何形成与保存**。前者以当下证据为中心，后者以跨时间的因果历史与真理产出能力为中心。

2.5 可靠主义的吸引力与困难

2.5.1 吸引力一：降低怀疑论门槛

怀疑论者常要求主体排除恶魔、缸中之脑等竞争假设，才能合理相信外部世界。可靠主义提出更温和的要求：实际使用的信念形成过程只需确实倾向于产生真信念。主体不必先证明自己不是缸中之脑，也不必拥有关于自身可靠性的完整理论。

这使普通知觉信念较容易获得证成。不过，批评者会说：可靠主义可能只是绕开怀疑论，而非真正回应怀疑论提出的反思要求。

2.5.2 吸引力二：自然连接证成与真理

得到证成不保证真理，但通常应提高信念为真的可能性。融贯论很难仅凭系统内部协调解释这种真理联系；可靠主义则直接把高真率写入证成条件：

可靠过程 $\implies$ 其输出具有较高真值比例 $\implies$ 证成与真理形成系统联系。

2.5.3 元证成要求

批评者可能要求：主体不仅要通过可靠过程形成 P ，还必须**有理由相信自己的 P 信念是可靠地产生的**。可靠主义者反驳：如果一阶可靠性不足以产生证成，为什么再增加

一个同样未经证明的二阶可靠信念就足够？若继续要求更高层级，就会出现无穷倒退。

较合理的修正是加入较弱的**反击败条件**：主体不能已经相信，或可靠地相信，自己的信念形成方式在当前情形中不可靠。主体无须证明过程可靠，但若拥有明确反证，也不能将其忽略。

2.5.4 Wright 的进一步怀疑论挑战

关于反怀疑论，原书还讨论另一项批评。可靠主义给出“如果普通认知过程实际上可靠，普通信念便可能得到证成”的条件；怀疑论者 Crispin Wright 追问：这仅说明人可能得到证成，何以说明我们实际上已经得到证成？反思中的主体甚至还想有理由相信“自己的普通信念有证成”。可见，回应怀疑论至少有三个不同目标：

1. 给出一种普通信念可能得到证成的条件；
2. 证明现实中的这些信念实际满足条件；
3. 说明主体如何有根据地反思并相信自己满足条件。

简单可靠主义主要解决第一项，不能仅靠定义完成第二项。作者也提醒：若第三项又按可靠主义理解为“主体通过可靠方法相信自己一阶过程可靠”，怀疑论者尚未证明这种二阶证成不可能；如果怀疑论者拒绝这种理解，仍需准确说明希望增加的条件。这里双方的争论尚未因提出 Wright 的要求就结束。

2.5.5 一般性问题

一般性问题 (*generality problem*) 源于同一个具体认知过程可以归入许多不同类型，而不同类型的可靠性不同。

阅读《纽约时报》得知中奖

George 通过《纽约时报》形成“自己中奖”的信念。这一具体过程既可以分类为“通过阅读《纽约时报》形成信念”，也可以分类为“通过阅读形成信念”，甚至可以进一步描述为“通过阅读印刷文字形成信念”。前一类型可能高度可靠，后一宽泛类型则可能包含大量不可靠材料。

可靠主义必须说明，应以哪一种过程类型计算可靠性。分类太窄，可能只包含当前一次过程，使可靠率失去解释力；分类太宽，又会把不同来源混在一起。

可靠主义者的一项回应是：任何要求信念“基于证据”的理论，都要说明基于关系由何种因果过程实现，因此其他理论也可能面对过程类型划分问题。一般性问题或许不是可靠主义独有，但这并不等于问题已经解决。

需要注意，这里问的并非“同一报道是否真实”，而是同一具体读报行为应该算哪种过程的实例。若只取“这次读报并得出中奖”作为类型，只有一个样本，可靠比例也没有独立解释力；若泛化成“读任何东西”，类型包含了大量质量参差的文本。靠扩大或缩小类型任意调节真率，便无法稳定判断这一个信念是否得到证成。

2.5.6 新恶魔问题

新恶魔世界

设想一个主体拥有与正常人完全相同的知觉经验，也同样谨慎地根据经验形成信念，却被恶魔系统性欺骗。因此，其知觉过程在该世界中几乎总是产生假信念。

简单可靠主义似乎必须判定：恶魔受害者的知觉信念不可靠，所以没有得到证成。但许多人直觉上认为，该主体与现实世界的正常主体在内部状态、理性谨慎程度和可用证据上完全相同，因此其信念也应得到证成。

推理所针对的是在恶魔世界实际运行的视觉过程。如果该过程在当地持续输出假信念，原始可靠主义就会给出不受证成的判决；反对者则用“与正常人的内部状态完全相同”施压。它与 Norman 案相反：恶魔世界中内部情况好而实际真率差；Norman 内部缺少通常可援引的根据而实际真率高。

这形成外在主义与内在主义的尖锐冲突：

可靠主义判断	过程在实际环境中系统性出错，故信念不受证成。
内在主义直觉	主体已经根据自己能够把握的全部证据尽责地推理，故信念应受证成。

2.5.7 认可清单可靠主义

Goldman 提出一种以**证成归因者**为中心的解释。人们在现实世界中观察到视觉、听觉等过程通常可靠，于是在心中形成“认可过程清单”；猜测、愿望式思考等则进入“不认可清单”。面对恶魔世界这类奇异思想实验时，人们可能继续使用现实世界形成的清单，于是仍把知觉信念判断为得到证成。

这属于方法上的转折。我们在读思想实验时首先直接掌握的是自己或他人给出的证成判断，而不是通过某种测量仪器读取案例人物“实际有多少证成”。作者于是提出两阶段的解释假说：归因者先在现实环境中给过程类型贴上好、坏标签；再在现实或假想案例中凭已储存的评价给人物的信念贴上“证成 / 不证成”标签。比如现实中眼见通常可靠，所以面对恶魔受害者，归因者可能继续称正常视觉所生的信念合理。

这种**认可清单可靠主义** (*approved-list reliabilism*) 不是直接说恶魔世界中的知觉过程在那里可靠，而是解释人们为什么仍倾向于作出“已证成”的归因。

因此，“归因者在现实中认可视觉”与“恶魔世界里的视觉实际上可靠”是不同主张。前者是对我们的归因心理的推测性解释，后者在该案例设定中为假。哪怕前者解释了直觉，还须继续论证：这些直觉为何是正确的规范判断，或归因模型为何足以构成证成理论。

2.5.8 Norman 的可靠千里眼

无证据的可靠千里眼

Norman 实际上拥有极可靠的千里眼能力，该能力使他突然相信总统正在纽约。他不知道自己具有这种能力，也没有任何普通知觉证据或理由支持该信念。

原始可靠主义会把 Norman 的信念判为得到证成，因为产生它的过程事实上高度可靠；但多数人认为这一结论反直觉。认可清单理论可以解释这种直觉：千里眼、超感知等神秘过程通常不在正常归因者的认可清单上，所以人们拒绝赋予 Norman 证成。

不过，这也产生新的问题：理论是在分析信念真正的证成地位，还是只是在解释人们如何使用“证成”一词？若公众清单存在偏见或知识不足，归因习惯未必等于规范真理。

2.5.9 Olsson 的归因比较思路

原书接着提出 Erik Olsson 建议的经验检验：设两组受试者分别观看约二十段普通人形成信念的生活片段。第一组逐段判断片中人物的信念是否得到证成；第二组独立判断人物获得该信念的途径是否可靠。如果两组对相同片段的判断高度相关，这会为可靠主义解释日常证成归因提供初步支持。

两组要分开，目的在于不先用“可靠”一词暗示第一组的回答。即使发现高度相关，也只表示人们的证成归因与可靠性归因一起变化；它既没有直接测得人物实际的可靠率，也没有自动证明人们对规范地位的判断正确。更重要的是，作者在这里介绍的是可以如何检验的研究构想，而非报告一项已经完成并证实理论的实验。

2.5.10 历史性与“下游后果”

一项批评把可靠主义类比为结果主义：若某信念后来导致大量真信念，是否就会因好结果而得到证成？

反对者所援引的**认识结果主义** (*epistemic consequentialism*) 把获得真信念、避免假

信念视为值得追求的结果，再借伦理功利主义的类比提出质疑：倘若以后带来了许多真信念，最初信念会不会“因此”变得正当？需要区分一种形成方法通常导向真理与某个已形成信念后来带来有益后果。

为了研究经费而相信上帝

一位长期无神论者为了获得宗教机构的研究经费，向欲望屈服并改变信念，开始相信上帝。后来他获得经费，研究成果又产生大量科学真理。

可靠主义不必承认最初的宗教信念得到证成。它是一种**历史性、上游导向**的理论：决定目标信念证成地位的是产生和维持该信念的上游过程，而不是该信念随后造成的下游后果。欲望屈服是不可靠的形成过程；后续研究产生再多真理也不能反向证成原信念。

可靠主义的得失

可靠主义以信念的真实形成史解释 Jack、Nora、Mary 和 Chad 案例，并自然连接证成与真理；代价是必须处理过程类型、恶魔世界、可靠千里眼以及归因者清单是否具有规范权威等问题。

2.6 内在主义与外在主义

2.6.1 争论框架

证据主义属于**内在主义** (*internalism*) 家族，过程可靠主义属于**外在主义** (*externalism*) 家族。二者争论的是证成因素的性质：

两种立场

- **内在主义**：全部证成因素都是“内在的”。
- **外在主义**：至少某些证成因素是“外在的”。

“内在”有两种主要解释：

其中，把内在性解释为主体可经反思直接接近，称为**可达及主义** (*accessibilism*)；把它解释为完全由主体心理状态构成，则称为**心理主义** (*mentalism*)，也常译为心灵主义。这里的“心理主义”专指证成因素的限制，不是把认识规范简单化为“自己觉得对就对”。

1. **心灵状态解释**：内在因素就是主体自己的非事实性心理状态，如经验、信念、似乎如此的感觉。
2. **可接近性解释**：内在因素是主体在作出信念决定时可以直接接近的事项，包括内省以及某些先验反思。

可靠性与真理通常属于外在因素，因为主体可能不知道自己的过程是否可靠，也可能无法直接判定目标命题是否为真。

两种“内在”标准也不可混同。主体的许多心理状态确实在其心中发生，但未必都能被当场准确辨识；反过来，可接近性论者也可能允许某些通过先验反思直接认识的逻辑关系，虽不是心理状态，仍算“内在”。因此说某理论“内在主义”时，须问它要求的是只取心理因素，还是只取主体可直接接近的因素。

2.6.2 规范规则模型

可以把认识论规范类比为道德规范。行动符合正确道德规则时是正当的；同样，信念态度符合正确认识规则许可时便得到证成。正确规则的前件会列出主体的认识条件，后件则规定在这些条件下允许相信、否定或悬置判断。

因此，进入正确规则前件、并影响许可结论的条件，就是证成因素。

这种写法使抽象争论可检验：假定一条认识规则说“若主体有 E ，便可相信 P ”，就问 E 的含义是否完全由当下心理状态确定；若还需 E 与 P 间的客观蕴涵、过程的实际可靠性或规则本身的正确性，强内在主义便承担新的说明负担。

2.6.3 新手与观鸟专家

啄木鸟识别

新手与专家同时清楚地看到近处树上的一只鸟。专家立即知道它是啄木鸟；新手即使也猜“啄木鸟”，却没有同等证成。

内在主义解释是：专家拥有关于啄木鸟外观的合理背景信念和储存表征，而新手没有。若新手获得同样的内部心理条件，也会获得同样证成。

外在主义解释是：专家拥有一种可靠分类能力，能够把当前视觉表征与典型啄木鸟表征进行有效匹配；新手缺乏这种真实的辨别能力。二者差异最终涉及谁更能作出真分类，因此解释具有外在主义性质。

这场讨论并不否认专家可能拥有更多内部知识。争议在于：是“脑中有啄木鸟外观的合理信念”已经足以解释差异，还是仍须追问这些表征在现实鸟类识别中是否能可靠分辨。因此同一案例能够推动两种理论，不能仅凭“专家有不同心理状态”便判定所有证成因素都是内在的。

这两个解释不必完全排斥：专家的内部知识结构可能正是其可靠能力的心理实现。不过，两方对真正构成证成的规范因素有不同理解。

2.6.4 可实施配方与无误指导观

内在主义的一项动机是：正确认识规则应像写给新手的食谱，使用主体能够理解并执行的条件。若规则写成“只采用事实上可靠的过程”，主体可能根本无法直接知道的过程是否可靠，因而无法保证遵守规则。

按照**无误指导观** (*infallible-guidance conception*)，正确规则不必保证主体得到真理，但应保证谨慎主体能够无误地判断自己是否遵守了规则。这样，认真尽责的主体至少能够确保自己的态度得到证成，即使仍可能相信假命题。

外在主义的反问是：一个规则难以被主体自觉执行，是否足以说明它不是正确的证成规则？规范地位未必完全取决于主体能否从第一人称视角确认自己合规。

2.7 内在主义与推论规则

如果证成完全取决于内在因素，那么正确推论规则的条件，也应当能够完全用内在因素表述。演绎推理提供了一个具体检验：主体相信前提、前提实际蕴涵结论，以及主体认为这种蕴涵成立，分别发挥什么作用？

2.7.1 INF：仅要求主体相信前提

初步规则以 INF 标记，表示这里讨论的推论规则 (*inference rule*)。INF*、INF**、INF*** 是它的逐次修正版，不是三个独立理论的通行简称：

INF

如果主体 S 在 t 时刻相信 K, L, M, N ，且这些命题的合取逻辑蕴涵 P ，那么 S 在 t 时刻被允许相信 P 。

该规则包含两类条件：

1. 主体拥有若干前提信念；
2. 前提内容与结论之间客观存在逻辑蕴涵关系。

第二项明显不是主体的心理状态。复杂逻辑关系也并非所有主体都能直接、无误地把握。因此，如果 INF 正确，逻辑蕴涵本身就是外在证成因素。

但 INF 并不正确，因为未经证成的前提不能仅凭逻辑关系把证成传给结论。即使迷信前提演绎出某结论，结论也不会因此获得证成。

2.7.2 INF*: 要求前提信念得到证成

修正规则为：

INF*

如果主体 S 在 t 时刻对 K, L, M, N 拥有已证成信念，且这些命题的合取逻辑蕴涵 P ，那么 S 在 t 时刻被允许相信 P 。

“拥有已证成信念”并不明显是纯心理条件。根据 Nora 案例，当前信念的证成可能依赖过去的学习事件与记忆保持过程，而这些事件此刻既未被回忆，也不可直接接近。因此，INF* 中的前件仍包含外在成分。

2.7.3 INF**: 主体看出蕴涵关系

内在主义者可能把客观蕴涵改为主体“看出”蕴涵：

INF**

如果 S 对 K, L, M, N 拥有已证成信念，并且 S 看出这些命题的合取逻辑蕴涵 P ，那么 S 被允许相信 P 。

但“看出 X ”是事实性状态：若主体真的看出 X ，则 X 必须为真。它不仅描述主体怎么想，还包含客观蕴涵确实成立这一事实。因此，INF** 仍然悄悄保留外在因素。

2.7.4 INF***: 主体认为存在蕴涵

若进一步改为非事实性的“主体认为”：

INF***

如果 S 对 K, L, M, N 拥有已证成信念，并且 S 认为这些命题的合取逻辑蕴涵 P ，那么 S 被允许相信 P 。

关于蕴涵关系的那一项条件终于换成纯粹心理性的“主体认为”；但规则中“前提信念已得到证成”这一项仍可能涉及过去的历史，整个 INF*** 并未因此完全成为纯心理规则。而且单看新的蕴涵条件，它过于宽松：一个极不擅长逻辑的主体可能错误地认为任意前提蕴涵任意结论。若只要他真诚地“认为蕴涵成立”就获得证成，逻辑无能反而不会受到认识论惩罚。

这表明：

纯粹主观地认为推理正确 $\not\Rightarrow$ 推理得到证成。

正确推理似乎必须以某种方式联系客观逻辑关系、真实能力或可靠判断，而这些都是具有外在性质。

内在主义者可以退一步说：“大多数”而非“所有”证成因素都属内部。原书认为，这比绝对的全称主张更有希望，但也没有仅凭让步就得到证明。客观蕴涵关系与前提信念的实际证成史，已经构成两类可能的外在因素。不过，决定证成的还可能包括另一类事实：所采用的规则是否正确，以及它凭什么正确。

2.7.5 规则正确性与更高层级的证成因素

证成因素不一定是对信念产生因果影响的东西。更一般地说，证成因素是能够解释一个信念**凭什么**拥有其证成地位的事实。

如果 INF* 是正确规则，而且主体按照该规则形成信念，那么规则本身的正确性也是解释信念为何得到证成的一部分。这类似法律判断：一个人构成犯罪，不仅取决于其行为事实，也取决于法律规则如何规定。法律本身不必因果地产生行为，却是解释法律地位的规范因素。

认识规则的正确性既不是主体的心理状态，也通常不能由主体直接接近，因此属于外在证成因素。

更高一层还存在**规则正确性的标准**。一个有吸引力的标准是：正确认识规则应促进获得真信念并避免错误信念。真理获取、错误避免和可靠性都涉及心灵之外的事实，因此也属于外在因素。这里是候选标准：如果正确认识规则最终确以导真和避错为根据，最高层级的说明也就不能只诉诸主体内部状态；作者并未在这一节独立证明一切认识规则只能由这一标准判定。

强内在主义的困难

作者并不主张所有证成因素都必须都是外在的；信念和经验显然是重要的内部因素。但逻辑蕴涵、过去的形成历史、过程可靠性、规则正确性以及以真理为导向的规范标准，似乎都不能还原为主体当前的心理状态。因此，强内在主义所主张的“全部证成因素均为内在因素”难以成立。

第二章总结：证据、过程与证成因素

第二章的基本推进方式，是从证据主义的简洁公式出发，通过反例逐步加入形成过程、历史和真理联系，最后把具体理论争论提升为内在主义与外在主义之争。

四种立场的关系

理论	典型阵营	证成核心	主要困难
心灵主义证据主义	内在主义	相信 P 应与主体当前的总体心理证据相契合	共存不充分、非必要；证据定义循环；预感与内省问题
过程可靠主义	外在主义	P 信念由高真率的形成或保持过程产生	一般性问题、新恶魔问题、Norman 千里眼
基础主义	可内在也可外在	存在非推论地获得证成的基础信念	基础证成来源及向上层信念的传递
融贯论	通常内在主义	信念通过与总体系统的相互支持获得证成	与经验隔离、虚构系统、真理联系不足

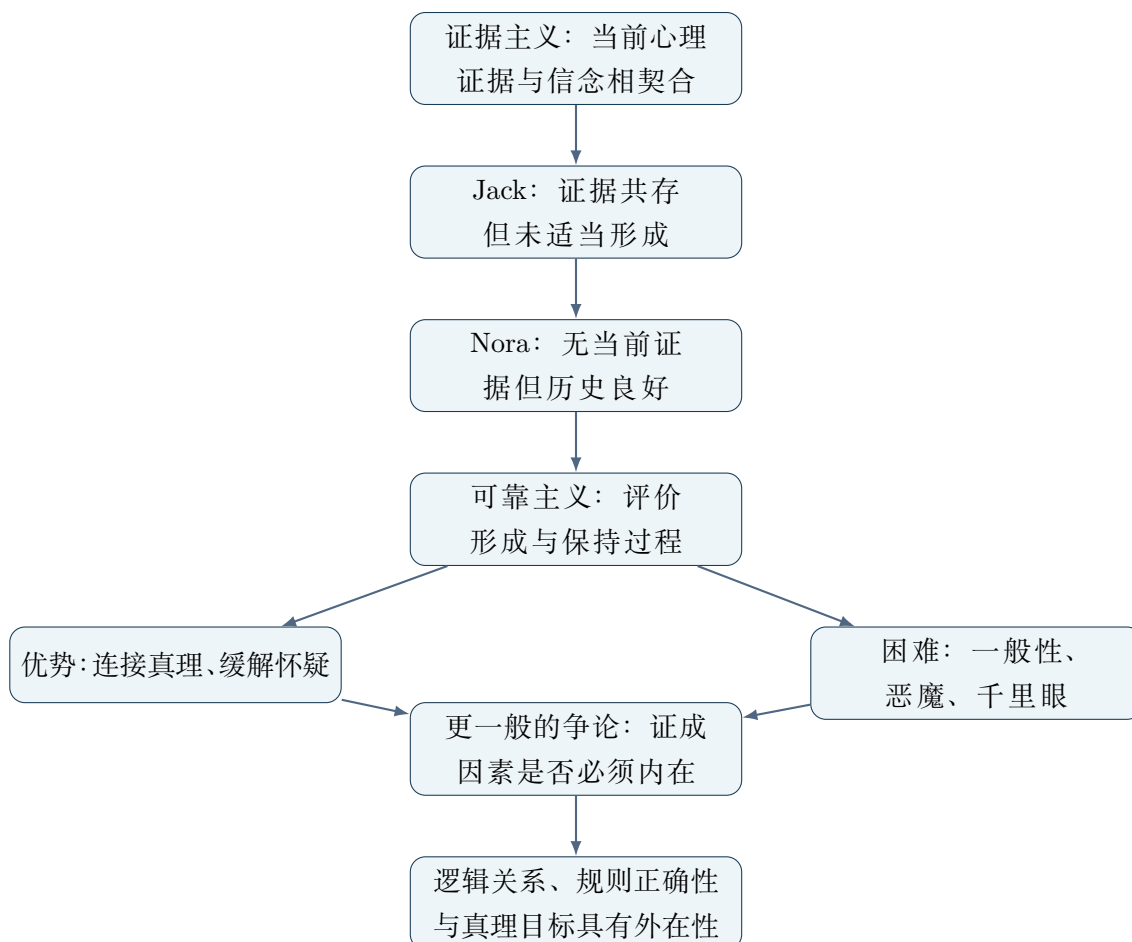
六个核心案例

案例	直觉判断	所反驳或支持的主张	核心概念
Jack 选举	虽有支持结论的数据，实际信念仍不合理	证据与信念共存不足以产生信念证成	基于关系、因果形成
Nora 周长	忘记原始来源后仍能合理相信	当前证据共存不是证成必要条件	历史性、记忆保持
Chad 推理	当前推理实例恰好有效，信念仍不合理	符合并基于证据仍可能不充分	过程类型、可靠性
恶魔世界	内部状态完全正常，但过程在该世界中不可靠	简单可靠主义与内在主义直觉冲突	新恶魔问题
Norman 千里眼	过程事实上可靠，但主体毫无理由	裸可靠性可能过于宽松	可靠性、认可清单
观鸟专家	专家与新手同样看见鸟，证成地位不同	内外在主义可对同一差异给出竞争解释	内部知识与外部能力

几项关键区分

1. **命题证成不等于信念证成**。有足够证据可以相信 P ，不表示实际形成的 P 信念使用了这些证据。
2. **证据存在不等于信念基于证据**。Jack 拥有支持材料，却没有通过适当思考使用它。
3. **信念基于证据不等于形成过程可靠**。Chad 确实从前提推出结论，却使用一条普遍不可靠的规则。
4. **可靠不等于主体知道可靠**。外在主义不要求主体具备关于自身可靠性的高阶信念。
5. **内在不只等于心理**。可接近性内在主义允许某些可由反思直接把握的非心理事实；但复杂逻辑蕴涵未必满足这一条件。
6. **外在主义不等于否认理由**。它只是认为，理由之外还有过程、真理联系或客观规则等证成因素。

从证据关系到证成因素



第二章结论

证据主义准确把握了理由与主体第一人称可辩护性的重要性，但只考察当前心理证据，难以说明信念的因果形成、记忆保存与真理联系。过程可靠主义利用信念的实际历史弥补这些不足，却面临过程类型、新恶魔与可靠千里眼等问题。作者最终倾向于一种外在主义结论：证成不能完全还原为当前心理状态；逻辑关系、过程可靠性、规则正确性和导真目标至少构成部分外在证成因素。

第二章课后问题参考思路

以下为原书章末八题的讨论与参考解答。

问题一：头痛的哪些性质能够成为证据

头痛具有许多真实性质：强度、位置、持续时间、神经实现方式等。主体并不会仅因处于头痛状态，就对每个描述该状态的真命题都获得证成。尤其是“该状态伴随梭状回激活”这类神经命题，通常不由主体直接经验到。

证据主义可增加一个**现象呈现原则**：

现象呈现原则

经验状态 E 只为关于其**向主体显现的性质**之命题提供初步证成； E 实际具有、但未进入主体现象意识或注意范围的隐藏物理性质，不仅凭 E 的发生就成为主体的证据。

若疼痛向主体呈现为尖锐、持续并位于左侧，主体可对这些性质获得初步证成；神经激活不属于现象内容，因而被排除。但“呈现”不宜被限制为主体已经明确注意并形成概念判断，否则婴儿、动物或注意力未集中者会被过度剥夺证成。较合理的说法是：该性质必须在经验中可供主体直接把握，而不是只有神经科学家从第三人称检测才能发现。

困难在于，“向主体呈现”仍需进一步分析。若把它理解为经验可靠指示某性质，又会重新引入外在主义；若只诉诸主体感觉，则可能无法解释错觉与预感为何权重不同。

问题二：预感是否是证据

不能把所有经验赋予相同证据权重。内在主义者可以提出三种路线：

1. 否认预感具有断言式现象内容：它只是焦虑或倾向，并不把世界呈现为“灾难将发生”；
2. 承认预感提供极弱的初步证成，但其力量通常不足以越过相信门槛；

3. 要求证据经验具有可反思把握的来源特征，例如正常知觉具有稳定、结构化的感官呈现，而无来源预感缺少这种结构。

第一条路线会碰到某些预感确实仿佛带有明确内容的反例；第二条路线与“预感完全没有证据作用”的直觉有张力；第三条路线若把稳定性解释为实际可靠率，又会向外在主义靠近。因此，本题揭示的不是一个容易修补的小漏洞，而是纯粹内在主义怎样区分不同经验认识价值的核心难题。

问题三：可靠主义怎样解释命题证成

过程可靠主义首先解释已经形成的信念为何具有**信念证成**。若主体尚未相信 P ，则不存在产生该信念的实际过程，因而需要反事实式扩展：

主体 S 对 P 具有命题证成，当且仅当， S 当前拥有一组相关输入状态，并且存在一种 S 实际具备、在当前情形中可正常运行的可靠认知过程；若 S 把该过程用于这些输入，它会形成 P 信念。

潜在问题包括：

- 同一输入可以通过许多过程产生不同结论，应选择哪一种过程？
- 如果主体没有能力运行某个高度可靠但极复杂的算法，仅仅存在该算法是否足够？
- “可正常运行”是否取决于主体有时间、有注意力并掌握相关概念？
- 反事实过程的类型划分会重新引入一般性问题。

较好的修正是要求可靠过程属于主体已经拥有或容易调用的认知能力，而非任何逻辑上可能的过程。

问题四：不可靠形成的信念能否后来获得证成

历史主义不应推出“一次形成不可靠，永远不能得到证成”。假设主体最初因谣言相信 P ，后来亲自查阅权威档案、核对反证并重新确认 P ；直觉上，他当前的信念可以得到证成。

可靠主义可把评价对象定义为**当前信念状态的有效历史**，而不是从最早起源开始、永远不可修复的整条历史：

更新后的历史原则

时刻 t 的 P 信念之证成，取决于对该信念当前维持具有决定性作用的最近形成、重新确认、更新和保持过程，以及此后是否出现未被适当处理的击败者。

这里仍有“同一个旧信念被修复，还是形成了一个新的信念个例”的本体论问题，但两种描述都允许认识地位随可靠复核而改变。

问题五：认可清单理论最有希望和最薄弱的应用

认可清单理论最有希望的用途，是解释**新恶魔案例**：我们之所以仍把受骗者的正常知觉信念判断为合理，是因为归因者沿用现实世界中形成的过程评价，继续认可视觉、记忆和正常推理。

它较薄弱的用途，是处理 Norman 的可靠千里眼。若 Norman 的能力确实稳定可靠，公众只是因为不了解这种能力而没有把它列入认可清单，那么公众无知不应自动决定 Norman 真正的规范地位。该理论因此更擅长解释人们**怎样作出证成归因**，却未必单独解释**什么使信念真正得到证成**。

还可以提出折中：认可清单必须是信息充分、反思平衡的归因者在掌握过程真实表现后会接受的清单。但一旦加入“真实表现”，理论又回到了客观可靠性。

问题六：一般性问题是否属于所有理论

只要理论需要说明信念“基于”何种理由、经验或能力，就可能面对某种过程分类问题。例如，证据主义也要区分：主体是通过正常演绎、粗心联想，还是愿望式思考从证据到达结论；德性认识论也要说明应评价“视力”“鸟类辨识能力”还是“当前光线下的远距离辨识能力”。

但一般性问题对可靠主义尤其尖锐，因为过程类型的统计可靠率就是其核心条件。若类型不同，证成结论可能直接相反。其他理论即使面对类似分类，也未必把全部证成判断建立在过程频率上。因此，问题具有普遍形式，却对可靠主义施加更重压力。

问题七：新手与专家，哪种解释更强

内在主义者会说，专家与新手并非真正内部相同：专家拥有关于羽色、喙形、叫声和栖息地的背景信念，其经验也可能因训练而具有更精细的可识别结构。专家的判断与这些内部材料更好地契合。

外在主义者则说，决定性差异是专家获得了一种可靠辨识能力；即使她无法完整说出全部规则，其训练历史和实际成功率仍能使判断获得较高证成。

若只问主体第一人称能够援引什么理由，内在主义解释更自然；若问这些理由或表征为什么具有认识价值，外在主义提供更直接的导真说明。较稳妥的综合观点是：内部表征构成可靠能力的心理实现，二者分别说明主观可辩护性与客观真理联系。

问题八：INF*** 为何过于宽松，以及如何修正

INF*** 只要求主体认为前提蕴涵结论，因此会使逻辑能力极差但十分自信的人轻易获得证成。设主体已证成地相信 P ，却荒谬地认为 P 蕴涵任意 Q ；按 INF***，他便被允许相信 Q 。主观确信由此取代了推理正确性。

内在主义可以尝试加入：

- 主体在充分反思后仍无可责备地判断蕴涵成立；
- 推理符合主体可直接把握的基本逻辑规则；
- 主体具有在相似推理中辨别有效与无效形式的能力。

但第一项仍可能奖励无可责备的逻辑无能；第二项必须说明何谓“可直接把握”；第三项若诉诸能力的实际成功率，就向外在主义靠近。若干脆加入“蕴涵事实上成立”，则直接承认外在逻辑事实。正因如此，INF*** 反例不是表面漏洞，而是对纯粹内在主义的重要压力。

Feldman 第 6 章：怀疑论 (I)

Richard Feldman, *Epistemology*, Chapter 6, “Skepticism (I), pp. 108–129.

我们的日常理由不能保证绝不出错。这是否意味着，我们从来没有真正知道那些自以为知道的事情？怀疑论的困难，正在于它似乎能够从一些很容易接受的前提出发，得出与日常认识相冲突的结论。

日常知识与怀疑论的挑战

按照 Feldman 在第一章介绍的**标准观点** (*the Standard View*)，我们知道许多普通事情：自己正在读一本书，最近吃过晚饭，昨晚看过一场电影，窗外有一棵枫树。这些判断通常不需要复杂论证。我们看见、记得或者通过其他熟悉的方式得知它们，并据此生活。

怀疑论的挑战却是：这些熟悉的方法有没有真正使我们获得知识？在梦中，我们也可能觉得自己正在读书；一个人的记忆可能出错；欺骗者也可能使不存在的东西看起来存在。如果这些错误情形不能被彻底排除，我们还可以说自己“知道”吗？

这里有两种不同的不安。有些人开始怀疑自己原来是否真的知道；另一些人仍然相信自己知道，却发现很难说明怀疑论的推理究竟在哪一步出了问题。Feldman 尤其重视后一种困难。一个结论听起来极不可信，并不等于我们已经找到了反驳它的理由。**即使我们确信结论不对，也仍须解释论证究竟错在何处。**

知识的分析为这场争论提供了几个必要区分。一个人相信某命题、该命题为真、他有充分理由相信它，是三件不同的事。“证成” *justification* 在这里指信念所得到的合理支持，不是指已经给出绝不可能失败的证明。盖梯尔问题又提醒我们：即使理由很好、结论恰好为真，真与理由之间仍可能存在使知识落空的意外。本章并不撤销这些区分，而是集中追问其中一个条件：理由必须好到什么程度，才有资格成为知识的理由？

这个问题对于**证据主义** (*evidentialism*) 尤其直接，因为证据主义把证成与主体拥有的证据、理由联系起来。因果论、追踪论、可靠主义和适当功能论等理论则可能强调另一类条件：信念与事实之间有怎样的因果关系，认知过程是否可靠，认知机制是否正常运作。采用这些理论的人，可能不要求主体能拿出同样强的理由，就承认他拥有知识。

Feldman 并未因此把怀疑论问题留给证据主义者独自处理。即使一种理论容许人在理由不够好的时候拥有知识，我们仍可以关心：我们平常的理由到底好不好？若答案是“很不好”，这本身就值得担忧。因而，避开某一种知识定义上的难题，未必已经回答了关于普通理由的挑战。

事实、证成与知识

以“门已经锁好”为例：**门事实上锁好了**，是在说世界怎样；**我有充分理由相信门锁好了**，是在评价我的证据；**我知道门锁好了**，则至少要求前两方面恰当地结合。门可能锁好了而我毫不知情；我也可能记得自己锁过门，却因记忆错误而没有知识。怀疑论主要质疑第二方面能否达到知识要求，并不需要先证明第一方面总是失败。

F6.I 怀疑论的种类

在日常语言中，“怀疑”可能只是暂时拿不准，或者不轻易相信传闻。这里的**怀疑论** (*skepticism*) 则是一类哲学主张：我们没有，或者不能具有，某一范围内的知识。评价这样的主张，首先要问清两个问题：它否认哪些知识？它把这种否认说得多强？

A. 全面怀疑论与局部怀疑论

全面怀疑论 (*global skepticism*) 认为没有人知道任何事情。**局部怀疑论** (*limited skepticism*) 只否认某一领域的知识。例如，一个人可以认为我们知道眼前有什么，却不知道未来；也可以承认数学知识，却否认有关宗教的知识。两种立场都没有否定一切知识，所以属于局部怀疑论。

为什么先作这一区分？因为并非每一种“我们不知道”的判断都在挑战常识。原书让我们考虑关于大沙漠沙粒总数的一组命题：撒哈拉沙漠恰有 n 粒沙，莫哈韦沙漠恰有 x 粒沙，等等。这里按思想实验的通常设定，暂且把计数时刻和沙漠范围固定。即使存在某个准确数字，我们几乎都不认为现实中有人知道这些精确答案。因此，几乎人人都可以是某种“沙漠怀疑论者”。

这种承认不令人不安，因为标准观点本来就没有主张我们无所不知。真正产生冲突的是另一类局部怀疑论：没有人能通过知觉获得知识，或者没有人能通过记忆获得知识。如果这些主张成立，受影响的不只是沙粒数这种特殊问题，而是“眼前有一本书”“我昨晚看过电影”等大量普通判断。

所以，“局部”只是在说明否认的范围，不是在评价其后果轻重。否认一切关于外部世界的知识，范围仍小于全面怀疑论，却已经足以对标准观点构成严重挑战。后文的梦、恶魔与缸中之脑，主要用于提出这种挑战。

B. 怀疑主张的强度

否认同一个领域的知识，还可以采取不同强度。原书以关于未来的知识为例，区分以下三句：

SF1、SF2 与 SF3	
SF1	任何人知道任何关于未来的事情，都是不可能的。
SF2	没有人具有知道任何关于未来的事情的能力。
SF3	事实上，没有人知道任何关于未来的事情。

第一句把障碍放在所要求的事情本身。设想要做出一个有九个面的立方体盒子：问题不在木匠不够熟练，而在于它不能既是立方体，又有九个面。“已婚的单身汉”也一样；再努力或者再聪明，都不能使这个描述成为可能。SF1 试图对未来知识提出类似判断：不是现有人类太笨，而是这种知识本身不可能。原书注释另提醒，若把“任何人”严格限于人类，也有人可能把不可能性解释成人类本性的限制；正文先采用最强的读法说明区别。

第二句则承认目标未必自相矛盾，只说它超出现有人类的能力。原书的类比是两分钟跑完一英里：不能做到，不是因为“跑完一英里”与“两分钟”在概念上冲突，而是人类没有这种能力。一个能力不同的存在者也许可以做到。相应地，SF2 不是说未来知识毫无意义，而是说我们达不到它。

第三句最弱。它只报告知识事实上没有出现，并不判断人是否有能力获得。**还没有做到，不等于做不到；做不到，也不等于所要求的事情本身不可能。**例如，对于某个目前无人解出的难题，我们还要另外研究：是尚未有人发现答案，还是人的能力确实不足。

主张	障碍在哪里	仍需区分什么
SF1	所要求的知识本身不可能	不是靠提高认知能力就能解决
SF2	超出现有人类的认知能力	未必排除其他种类的认知者
SF3	知识事实上尚未出现	未必排除人类后来获得它

Feldman 认为，许多有实质内容的怀疑论论证，最好理解为针对能力的论证，即类似 SF2 的主张。它们并非仅说“我们目前偶然少知道一些事情”，而是说现有的知觉、记忆或推理能力不足以产生相应知识。外部世界怀疑论所质疑的，正是我们的认识能力能否达到知识要求。

怀疑的是知识、证成，还是证据

还可以从另一维度说明怀疑主张的力度。SF1-SF3 区分的是不可能、无能力与事实上没有；课堂则另用**力度** (*force*) 表示怀疑论否认的是哪一种认识地位，用**范围** (*scope*) 表示它针对哪一类信念。¹³

“我们不知道外部世界是什么样”可以承认普通信念仍有相当好的理由；“我们甚至没有理由相信外部世界是这样”则进一步否认证成；“我们没有任何相关证据”又提出关于证据的主张。它们不能仅因都被称为怀疑论而互相替换。知识还要求真理及其他条件，所以不能从知识缺失直接推出证成或证据缺失。

同样，否定某个对象的存在，与怀疑我们能否知道它存在或不存在，也应分开。一个断言上帝不存在的人，是在对相关事实作否定判断；他不必要是关于这一问题的认识论怀疑者。

F6.II 怀疑论者究竟主张什么

范围和强度之外，还需要澄清两种误解。第一种把“不能知道某领域的事实”听成“该领域没有事实”；第二种把“我们的信念可能错”听成“我们的信念实际上都错了”。两种误解都会使我们反驳一个并非怀疑论者真正提出的主张。

A. 存在但不为我们所知的真理

一个人说“我怀疑其他星球上有生命”，通常是在怀疑那种生命是否存在。他未必怀疑我们的认识能力；他甚至可能认为，如果那里真的有生命，我们就能够发现。这首先是关于宇宙中有什么的判断。

认识论怀疑论的重点不同。它可以承认某个问题有真实答案，只认为人类无法知道答案。原书以伦理学说明这种区别：

伦理怀疑的两种含义

ES1: 存在伦理真理，但我们不能知道这些真理是什么。

ES2: 不存在伦理真理，因此我们也不能知道相应的伦理事实。

接受 ES1 的人承认行动究竟对不对有事实可言，却认为我们无法达到这些事实。例如，他可能主张，判断对错需要知道上帝的心意，或者需要知道行动延伸到无穷未来的全部后果，而人类缺少这类能力。问题在于事实与认识能力之间的距离。

接受 ES2 的人可能对道德语言采取另一种解释。原书举出两句话：“她把旧家具捐

¹³ 参见赵斌，*Skepticism*，第 1 页。课件的 *force* 与本节 SF1-SF3 所区分的模态强度不是同一维度。

给慈善机构，这是好的”；“他从慈善机构偷走家具，这是坏的”。按照书中用来说明 ES2 的看法，前一句近似于“好啊，她捐了家具”，后一句近似于对偷窃喝倒彩：它们主要表达赞同或反对，而不是报告额外的道德事实。这类似棒球观众为球队欢呼。比赛本身有胜负等事实，但欢呼这个举动是在表达态度，不是在描述另一个事实。

这个例子的作用是区分问题，而非在本章证明某一种元伦理学理论正确。若根本没有某一类事实，就当然没有关于那些事实的知识；但这与“有事实，只是我们的认识方式达不到”不是同一个论题。Feldman 在这里主要研究 ES1 式的结构：过去、未来和周围世界存在事实，而怀疑论者质疑我们能否知道它们。

否定知识与否定事实

“我不知道门锁没锁”不等于“我知道门没锁”。同样，“我们不知道外部世界是什么样”不等于“我们知道外部世界不存在”。后一句自己也作出了强烈的知识主张，不能由前一句直接推出。

B. 怀疑论、真理与证成

知识要求真理，因此一个实际为假的信念不是知识。但这一点能否证明：一个可能为假的信念也不是知识？原书用两个关于地球形状的论证，把问题具体化。这里沿用书中“古人”的示例说法，其论证只需要设想某些人确实持有地平信念，并不需要把它当成对全部古代思想史的概括。

论证 6.1：古人关于地球形状的判断

- (1-1) 古人相信地球是平的，这一信念为假。
- (1-2) 所以，古人不知道地球是平的。

这一步推理依赖一个没有写出来的前提：如果一个信念为假，它就不是知识。这个前提来自知识的真理条件。即使持有地平信念的人十分确信，也不能把假信念变成知识。

论证 6.2：现代人关于地球形状的判断

- (2-1) 我们相信地球大致是球形的，这一信念可能为假。
- (2-2) 所以，我们不知道地球大致是球形的。

表面看，第二个论证只是把古人换成了我们。但它还换掉了一个关键表达：“为假”变成了“可能为假”。要从新的前提推出结论，就需要一个新的原则：只要一个信念可能为假，它便不是知识。这个原则比“假信念不是知识”强得多。

论证 6.3：古人与我们

(3-1) 如果论证 6.1 是健全的，那么论证 6.2 也是健全的。

(3-2) 论证 6.1 是健全的。

(3-3) 所以，论证 6.2 也是健全的。

健全的论证 (*sound argument*) 是推理有效、前提又都为真的论证；**有效** (*valid*) 只保证，如果前提都真，结论就不可能假。原书在 6.3 中比较的是前两个论证是否同样健全。Feldman 拒绝 (3-1)：前两个论证使用不同的隐含原则，不能因为第一个好，就认定第二个同样好。

可以用更直观的方式理解。若已经查明某证人说错了，我们当然不能把他的那句话算作知识。若只指出“证人是可能说错话的人”，却还没有查明这一次说错了，我们就尚未得到同一个结论。要把后一种情形也排除出知识，必须另外论证：知识要求连错误的可能性都不存在。

这一澄清没有证明论证 6.2 一定错误。怀疑论者仍然可以为它的强原则辩护。它只是告诉我们，争论已从“知识必须为真”转向“知识需要怎样的理由”。怀疑论者通常不打算断言地球其实不是球形；他想说的是，既然我们的理由仍与错误相容，它们便不够好，不足以使我们知道。下一节将把这个想法发展成四个论证。

F6.III 四个怀疑论论证

梦、恶魔和缸中之脑为什么值得讨论？因为怀疑论者并不需要先证明我们正遭受这些欺骗。他要借这些设想指出，某些经验在正常情形和错误情形中都能出现。**经验存在这件事，并不自动保证经验所指向的外部事实存在。**

但到这里还没有推出“没有知识”。从“证据不保证绝不出错”走到“证据不足以产生知识”，中间仍需要一项关于知识标准的前提。四个论证的区别，正是在于它们怎样说明这一步。它们虽然共享欺骗情形，却依赖不同的知识条件；相应的反驳也必须指向各自的前提。

A. 错误可能性论证

原书引用笛卡尔《第一沉思》的一段反思。笛卡尔似乎正在清醒地看着面前的纸，移动头部，有意识地伸出手。他起初觉得眼前的一切比梦清晰得多；但随后想到，自己过去也曾在梦里被相似印象欺骗。如果梦中也会有“我现在很清醒”的感受，那么仅凭这个感受，就没有获得一种绝不会失效的梦醒判据。

恶魔设想把欺骗扩大：一个强大的欺骗者使我们产生眼下全部感官经验，而世界实

际上不是经验呈现的样子。当代的**缸中之脑** (*brain in a vat*, 简称 *BIV*) 以技术故事表达类似结构：一颗大脑泡在液体中，连接计算机；计算机发出的电脉冲使它仿佛拥有身体，正在房间里看书和活动。它的经验可以与普通人的经验相同，但它并没有经验中那样的身体与周围环境。

三个故事的细节不同，共同用途却是显示：对于大量外部世界信念，主体即使有眼下这样的经验，也仍可能弄错。

论证 6.4：错误可能性论证

(4-1) 对任何人关于外部世界的几乎每一个信念，该信念都有可能出错。

(4-2) 如果一个信念有可能出错，它就不是知识。

(4-3) 所以，任何人关于外部世界的几乎每一个信念都不是知识。

(4-1) 与 (4-2) 承担不同工作。第一项列出可能的错误；第二项规定，一旦有这种可能，知识资格便被取消。若两项都成立，结论确实跟着成立。因此，单纯说“缸中之脑听起来很荒唐”，并没有正面回答第二项。

“几乎”这个限定也有用途。正在受骗的人至少必须存在，否则无从受骗。因此，“我存在”被笛卡尔当作著名例外。有些哲学家还把简单数学信念或关于当前心灵内容的某些信念列为例外。例外范围可以争论，但论证无须否定一切知识才具有威胁：只要否定大量日常外部世界知识，就已经冲击了标准观点。

逻辑可能性与概率

怀疑论者可以承认某个骗局极不常见，仍坚持它在逻辑上没有当前证据排除。若 (4-2) 真的要求错误完全不可能，再低的错误可能性也不够。可错论稍后要质疑的，正是为什么知识必须满足这种要求，而不只是争论骗局发生率有多低。

B. 不可分辨性论证

第二个论证不先说“有错误可能就没有知识”，而从主体能否分辨两种情形入手。它的吸引力是：如果从我所能把握的理由看，两种信念完全处于同等地位，为什么其中一个算知识，另一个却不算？

原书例 6.1：Jones 警探的调查

琼斯警探 (Detective Jones) 正在调查一宗犯罪。她拥有非常有说服力、彼此完全可比的证据，表明布莱克 (Black) 与怀特 (White) 都无罪。这些证据足以让标准观点的支持者通常承认知识，却并不是逻辑上排除一切错误的证据。

实际情况是：Black 无罪；White 有罪，却收买证人并伪造了其他证明自己无罪的材料。Jones 并未因为偏见或粗心而偏向谁；就她拥有的理由而言，两人的无罪主张同样可信。还可以设想，她有充分理由相信第三个人才是罪犯。

先分别考察两个信念。关于 White，结论其实是假，所以 Jones 不知道 White 无罪。这只使用了知识的真理条件，没有争议。关于 Black，结论是真的，理由也极好。为什么不能说 Jones 知道 Black 无罪？怀疑论者要在这里加入一项原则：知识案例与非知识案例，不应当从主体内部看来不可分辨。

如果接受这个原则，两个案件就必须一起判断：要么都知道，要么都不知道。既然 White 一案不可能是知识，Black 一案也被排除。注意，Black 一案被排除不是因为查出了 Black 有罪，而是因为 Jones 无法仅凭内部视角发现两案知识地位的不同。

比较项目	Black 无罪	White 无罪
Jones 所见的证据	极有说服力	同样有说服力
Jones 的主观偏见	设定为没有	设定为没有
命题实际真假	真	假
知识地位	正是争论所在	不是知识，因为不为真

这里需要两个术语。**可错证据** (*fallible evidence*) 是这样的证据：即使它存在，被支持的命题仍可能为假。它不等于糟糕证据。大量可靠证词、文件与现场材料可以提供极强支持，却仍不能在逻辑上排除精心伪造。

内省上不可分辨 (*introspectively indistinguishable*) 则指，从主体自身的视角看，两种情形没有可供他分辨的相关差别。原书给出两种用法。第一种是在两个可能情形之间比较：Jones 都凭完全相同的理由相信 Black 无罪，但 Black 在一个情形中确实无罪，在另一个情形中有罪。第二种就是原案中的 Black 与 White：两个信念对象不同，材料也不必逐字相同，但主体内部没有理由认为一案比另一案更可信。第二种“不可分辨”是证据地位上的对等，不是 Jones 连两个人的名字都分不清。

有了这些区分，原书把推理概括如下：

论证 6.5：内省不可分辨性论证

- (5-1) 如果人能够凭可错证据获得知识，那么就可能有知识案例与非知识案例在内省上不可分辨。
- (5-2) 但是，知识案例与非知识案例不可能在内省上不可分辨。
- (5-3) 所以，人不能凭可错证据获得知识。(由 5-1、5-2)

(5-4) 我们关于外部世界命题的全部证据都是可错的。

(5-5) 所以，我们不能获得关于外部世界的知识。(由 5-3、5-4)

为什么 (5-3) 从前两项推出？先暂时假设可错证据能够产生知识。那么按照 (5-1)，应当允许一个内部相同、信念却为假的对应案例；但 (5-2) 恰好禁止这一结果。为了保住 (5-2)，就必须撤回最初假设，得到 (5-3)。最后，既然外部世界证据都可错，而可错证据一概不能产生知识，外部世界知识便被排除。

Feldman 把压力集中在 (5-2)。它要求知识地位的差异必须能够在内部体现出来。这个要求有一定吸引力，却并非“可错证据”的定义本身。第四节将讨论：知识既然要求真理，它是否本来就不完全由主体内部发生什么来决定？

C. 确定性论证

第三个论证最简洁。它把前面隐约出现的高标准直接说出来：知识要求绝对确定，而我们对外部世界达不到绝对确定。

论证 6.6：确定性论证

(6-1) 如果主体 S 知道命题 p ，那么 S 对 p 绝对确定。

(6-2) 没有人对任何外部世界命题达到绝对确定。

(6-3) 所以，没有人知道任何外部世界命题。

这里 S 只是任意一个人的代号， p 只是任意一个命题的代号。例如， S 可以是正在读书的人， p 可以是“我面前有一本书”。论证不是靠符号产生力量：它实际说的是，若知道面前有书必须拥有最高强度的理由，而我们没有这样的理由，就不能说知道。

“确定”有两种不同含义。**心理确定性** (*psychological certainty*) 涉及主体有多确信、心里是否还感到疑惑。**认识确定性** (*epistemic certainty*) 涉及理由本身有多强。一个人可以毫无根据地十分自信；另一个人也可能已经有极强证据，却因焦虑而仍感到不安。感受强弱不等于证据强弱。

如果 (6-2) 只是说“没有人对外部世界感到完全确信”，它很容易被固执而自信的人反驳。怀疑论者真正想说的是：这种确信没有得到最高强度证据的支持。梦与恶魔等设想表明，经验留下了错误的余地。因此，这里的绝对确定应按认识上的含义理解。

论证于是把问题收拢到 (6-1)：**知识需要很好、足够好的理由，还是必须需要再也不可能更强的理由？**两者若不同，第二个要求就不能仅凭“知识应当有充分证成”得到。

D. 传递性论证

第四个论证引入一个很有吸引力的原则：如果我已经知道一件事，又知道另一件事必然从它推出，那么我可以通过相应推理知道后者。例如，我知道盒子里有三支铅笔，又知道有三支铅笔就意味着至少有一支铅笔；经过这个推理，我可以知道盒子里至少有一支铅笔。

原书称之为**知识传递原则** (*transmissibility principle*)，也提到它常被称为**闭合原则** (*closure principle*)。本章使用的是有限而自然的想法：主体理解了有关逻辑关系，可以借推理获得结论。它不是说，只要知道任何命题，就已经自动知道它的一切复杂逻辑后果。

现在回到布赖恩 (Brian)。设定两个命题：

BA ： Brian 有手臂；

BIV ： Brian 只是一颗连接人工生活模拟器的缸中之脑。

“只是一颗”不能省略。它明确规定 Brian 没有完整身体，因而没有手臂。所以，若 BA 为真， BIV 就必须为假。

假如 Brian 知道自己有手臂，也懂得上述关系，那么他就可以推出并知道自己不只是缸中之脑。怀疑论者却认为，他不能知道自己不是缸中之脑。因此，怀疑论者倒过来推断：Brian 原先以为拥有的手臂知识，必定也不是知识。

传递原则在论证中的作用

顺向：知道自己有手臂，加上知道“有手臂就不只是缸中大脑”，应当能得到“知道自己不只是缸中大脑”。

怀疑论的倒推：如果最后一种知识根本不可能得到，而传递关系又没有问题，那么最初的“知道自己有手臂”就不能成立。

这里推翻的是知道自己有手臂，并不是直接证明自己没有手臂。

原书把这个具体故事推广。令 O 表示一个普通外部世界命题， SK 表示与 O 不能同时为真的怀疑假设。符号 $\neg$ 表示“并非”， $\rightarrow$ 表示“如果……那么……”。因此， $O \rightarrow \neg SK$ 只是说：如果普通命题为真，相应的怀疑假设便为假。

论证 6.7：传递性论证

(7-1) S 不能知道 SK 为假。

(7-2) O 蕴涵 SK 为假，而且 S 知道这一关系。

(7-3) 如果 S 知道 O 为真，又知道 O 蕴涵 SK 为假，那么 S 可以知道 SK 为假。

(7-4) 所以， S 不知道 O 。

(7-2) 由案例设定保证。即使 Brian 真是缸中之脑，也未必因此丧失逻辑理解能力；模拟器欺骗他的感官，不必同时让他不理解“如果有手臂，就不只是大脑”。Feldman 在本章也不打算否定 (7-3) 的基本思想。最值得追问的是 (7-1)：我们为什么不能知道怀疑假设为假？

这一点标志着第四个论证与前三个的结构差异。前三个论证提出知识的一般要求，再说明普通经验达不到要求；第四个则直接把一项知识缺失放进前提。它能把“不知道怀疑假设为假”传递成“不知道普通命题”，但这一步本身并未说明前一种不知道为何成立。

论证	从什么现象出发	需要额外检查的前提
6.4 错误可能性	普通信念仍可能出错	有错误可能就不能是知识 (4-2)
6.5 不可分辨性	内部看来相同，真假却不同	知识与非知识必须内部可辨 (5-2)
6.6 确定性	普通理由达不到绝对确定	知识要求绝对确定 (6-1)
6.7 传递性	知识可经已知蕴涵传递	不能知道怀疑假设为假 (7-1)

这张表只定位争点，尚未完成反驳。要回应怀疑论，需要说明为什么拒绝这些前提是合理的，而不只是为了保护日常结论而临时降低标准。

F6.IV 回应怀疑论

四个论证都把我们带向一个与标准观点冲突的结论。现在不能只选一个最容易反驳的故事，再把其他论证一起放下；只要其中一个论证健全，相应的怀疑结论就成立。原书先考察两种常见回应，说明它们为什么尚未完成任务，然后详细展开可错论。

A. 怀疑论会自我反驳吗？

最自然的反击是反问：“你说我们什么都不知道，那么你知道自己的这句话吗？”原书把类似思路写成一段对话。以下保留其推理顺序作译述。

原书对话：怀疑论者知道自己的前提吗？

反怀疑论者：你知道怀疑论论证的前提是真的吗？例如，你知道我们的信念几乎都可能错，或者我们的证据几乎都可错吗？

怀疑论者：是的，我知道这些前提为真。

反怀疑论者：这些是关于人类信念和证据的事实。你一面说我们没有这类知识，一面又承认自己知道这些事实，立场便不一致了。

怀疑论者：那我承认，我不知道这些前提为真。

反怀疑论者：那么，你也不能凭这些自己不知道为真的前提，建立你知道怀疑论结论为真的证明。

这个反击确实揭示了一种困难。“我知道没有人知道任何事情”不可能为真：如果我知道这句话，我就至少知道了一件事；这样，它声称我知道的内容——无人知道任何事——反而是假。对于局部怀疑论，反击还需要检查被声称知道的前提是否真的落在被否认的领域内，不能只凭“怀疑论”三个字便断定自相矛盾。

日常行为也会造成类似压力。怀疑论者与别人交谈，似乎承认有人在场，也知道别人说了什么；他避开驶来的卡车，似乎知道站在车前有危险。如果他真的一无所知，为什么还这样行动？

Feldman 首先提醒，行为不一定包含那么强的承诺。怀疑论者可以相信卡车危险，所以躲开，同时否认这个信念达到知识。他也可以谨慎地不声称自己知道论证前提。因为相信与知道不同，他未必仅凭行动就与自己的结论产生逻辑矛盾。

但即使退一步，承认怀疑论者实际上难以始终保持一致，最重要的问题仍未回答：**四个论证究竟哪一项前提为假？**一个人说话前后不一致，不会自动使他刚才列出的每个论证都失效。若论证形式有效、前提又真，其结论仍然真。批评提出论证的人，并不能替代检查论证。

因此，原书区分**辩论目标** (*dialectical goal*) 和**哲学目标** (*philosophical goal*)。前者是让怀疑论者陷入困境、无法继续坚持原话；后者是找出论证中不合理的前提或错误。本章要完成后者。

这一选择还有实际理由。若把成功标准规定为“使一个拒绝一切前提的人心服口服”，任何证明都会陷入无休止的追问。每一次解释都需要某个前提，对方总能继续问“你怎么知道？”就像不停追问为什么的孩子。我们需要评价理由，而不是把一个顽固对手愿不愿让步作为理由是否充分的标准。

B. 摩尔式回应

乔治·爱德华·摩尔 (G. E. Moore) 提出另一条路线。他比较的不是怀疑论者是否自治，而是两边主张各自有多可信。原书引述他的大意是：与其相信那些复杂的怀疑论前提全都为真，不如相信“我知道这是一支铅笔”“我知道你具有意识”。后两项在他看来更确定、更有理由被接受。

这个回应并不是从“人人都这么说”推出“它必定正确”。它是在面对一个有效论证时作判断：我们必须接受结论，或者拒绝至少一个前提；若某个前提反而比被否定的常识知识更不可信，拒绝该前提并不是不讲逻辑。

以确定性论证为例，可以把分歧简化为以下形式。怀疑论者说：“如果知识需要绝

对确定，那么我不知道这里有铅笔；知识需要绝对确定；所以我不知道。”摩尔式回应则说：“我知道这里有铅笔，比‘知识需要绝对确定’更可信；因此应当放弃那个要求。”两边争论的不是有效推理是否必须服从，而是冲突出现时应当保留什么前提。

摩尔还有一个著名的外部世界证明：举起一只手，说“这里有一只手”；再举起另一只，说“这里还有一只手”；由此推出外部世界存在。若前提为真，又承认手是外部对象，结论确实能推出。这个证明很难仅凭其形式被驳倒。

然而，Feldman 认为，这仍未给出充分的诊断。即使同意日常知识比怀疑前提更可信，我们还想知道：究竟哪个前提出了问题？为什么那个前提如此有吸引力？它具体混淆了什么？摩尔式回应告诉我们某处有错，却没有系统解释错误所在。

还应区分“给出一个健全论证”和“给出双方都接受的证明”。怀疑论者恰好质疑“这里有手”能否被知道，因此摩尔的前提不具备争论中立性。原书把这种区别作为进一步说明，但其主要不满仍然是解释不足，而不只是对方没有被说服。

C. 可错论

可错论 (*fallibilism*) 提供了更具体的回应。按照 Feldman 在这里辩护的版本，知识需要真信念、很好的理由，以及避免盖梯尔式意外所需的条件；它不要求理由在逻辑上彻底排除一切错误。

这句话容易被听成“知识有时也是错的”，必须立即排除这种误读。可错论允许的是**证据可错**，不是**知识内容为假**。同样的好证据，在正常情形中支持一个真信念，在罕见骗局中可能支持一个假信念；前者可以是知识，后者不是知识。它没有把后者也收入知识。

这也不是把知识标准降到“只要本人相信就行”。没有好理由的猜测、实际为假的信念，以及真得过于偶然的盖梯尔信念，仍然需要被排除。分歧只在于：排除这些情形以后，是否还必须要求主体的理由达到不可错的最高程度。

确定性论证直接表达了双方在知识标准上的分歧。从它出发，可以进一步说明：为什么错误可能性不必取消知识，为什么相同内部理由未必带来相同知识地位，以及为什么知识传递不必导向怀疑论。

C1. 知识与绝对确定：足够好的理由为何不必是完美理由

可错论首先拒绝 (6-1)：“如果 S 知道 p ，那么 S 对 p 绝对确定。”它并不急于否定 (6-2)；它可以承认我们对许多普通事情没有绝对确定，却否认这足以取消知识。

面前的桌子

面前确实有一张桌子，光线良好，视觉系统正常。主体根据视觉印象和背景信息相信“面前有桌子”。虽然从主体角度仍可设想罕见的幻觉或骗局，他没有具体理由认为眼下正在出错，反而有极好的理由相信自己看见了桌子。

按可错论，只要这里没有盖梯尔式意外，这可以是知识。若面前实际上没有桌子，则不是知识；良好的主观理由不会把不存在的桌子变成存在。

问题随即出现：为什么允许“很好但不完美”的理由？原书首先诉诸我们如何使用“知道”这个词。两个人驾车离家，一人问另一人是否知道刚才锁了门，对方答“知道”。接着第一人追问：“你绝对确定吗，确定到再也不可能更确定吗？”对方可以合理地回答“不”。他未必是在撤回第一句；第二个问题可能确实增加了第一问没有提出的要求。

“我知道”与“我完全确定地知道”听起来也不完全重复。后一种说法似乎表达了更强要求。这些语言现象支持把知识与绝对确定分开。不过，Feldman 并未把它们当成决定性证明：怀疑论者仍可能认为，日常用法只是比较宽松。

原书于是再提出一个理论上的考虑。假如我们同意把“知识”专门留给绝对确定，那么大量日常案例都要改称为非知识。但我们仍需要区分几种明显不同的情形。记得自己刚锁好门，与猜明年世界大赛的冠军、预测下个月选举结果，并不是同样的认识状态。我们还需要把正常的强证据真信念，与盖梯尔式的偶然真信念区分开。

不妨给第一类正常案例另取名字，叫作“几乎知道”。接下来，我们又会问：几乎知道需要什么理由？为什么猜测不算？为什么盖梯尔案例不算？最后可能发展出一门“几乎认识论”，研究的正是原先认识论研究的许多差别。可错论者认为，没有必要先把“知道”让给一个极端标准，再用新词把日常知识讨论重新建立一遍。

这不是用改名解决所有哲学问题。它支持的较有限判断是：把“知道”理解为允许很好而非完美的理由，与普通语言和理论任务都相当协调。至于多好才足够，还需要进一步说明。

C2. 知识与错误可能：实际没错，不等于根本不可能错

对于论证 6.4，可错论拒绝 (4-2)¹⁴：“如果一个信念有可能出错，它就不是知识。”要理解这一拒绝，先看支持 (4-2) 的一种自然推理：若可能错，就达不到绝对确定；若达不到绝对确定，就不是知识；所以可能错就不是知识。第二步恰好又使用了 C1 拒绝的标准。因此，前一个回应会直接影响这个论证。

另一个使 (4-2) 显得可信的因素，是日常“如果……那么……”和“不能”表达中的歧义。原书用单身汉作类比。以下两句看起来相近，意思却不同：

1. 如果一个人是单身汉，那么他没有结婚。

¹⁴英文原文后文对论证 6.4、6.5 的个别前提回指有误，此处依各论证首次出现时的编号。

2. 如果一个人是单身汉，那么他不可能结婚。

第一句当然正确。“单身汉”这个词本来就包含未婚条件。但第二句若说的是这个人的能力或可能人生，就不正确。一个选择不结婚的三十岁成年人，可以属于有能力结婚的人，只是实际尚未结婚。原书把他与一个三周大的婴儿比较：婴儿在当前年龄不具有参与婚姻所需的能力，成年单身汉的情况却不同。

这里有一个常见追问：“他一旦结婚，就不再是单身汉，怎么还能说单身汉可能结婚？”答案是，可能发生变化的是这个人；第一句只规定，在他仍是单身汉的情形中，他没有结婚。它没有规定这个人必须永远保有单身汉身份。

知识也可以作相似区分：

两句话之间的差别

A: 如果主体知道 p ，那么他对 p 事实上没有弄错。

B: 如果主体知道 p ，那么他属于根本不可能对 p 弄错的情形。

可错论接受 A，拒绝仅凭 A 推出 B。在一个对应的欺骗情形中，主体可能保有相同理由而弄错；此时取消的是该情形中的知识资格，不是允许一个错误信念仍叫知识。

比较项目	正常看见桌子	设想的精心骗局
主体相信什么	面前有桌子	面前有桌子
主体可把握的理由	清楚经验与良好背景理由	设想保持同样的内部理由
实际有无桌子	有	没有
是否可以成为知识	可以，仍须排除盖梯尔式意外	不可以，因为信念为假

两种情形的比较说明，同一套理由并不逻辑保证相同真值。可错论允许第一个情形有知识，同时承认第二个情形是可能的。

必然性的作用范围

令 Kp 表示“主体知道 p ”， $\Box$ 表示“必然”。知识的真理条件可以表达为：

$$\Box(Kp \rightarrow p).$$

意思是，在任何情形中，只要在那里确实知道 p ， p 就为真。它不等于：

$$Kp \rightarrow \Box p.$$

后一句会把一个被知道的普通事实，变成在所有情形中都成立的必然事实。例如，知道门锁了，并不使门在所有可能情形中都锁着。

对于本章更直接的争点，令 E 表示主体的证据。“ E 可错”是说， E 存在与 p 为假能够相容；“ E 演绎保证 p ”才是说这种组合不可能。知道 p 要求 p 实际为真，不等

于要求 E 具有后一种保证。这一区分表达了可错论的基本立场：保留知识的真理条件，并不要求证据必然保证真理。

原书也承认，某些特殊信念确实不容易以这种方式出错。“存在信念”“我存在”“我有一个信念”是其例子：若一个主体正在相信“我有一个信念”，相信这个动作本身就使该命题为真。但这不足以把特殊保障提升成一切知识的必要条件。

但这种特殊保障仍有范围限制：某人不能在相信自己存在时把它相信错，不等于他不能错误地否认自己存在。一个人可能受怀疑论影响，错误地相信“我不存在”。正面相信无法为假，与对于整个问题不可能作任何错误判断，不是一回事。

最后，可错论并不让具体反证失效。原书排除的是把实际错误、积极的出错理由或盖梯尔式意外也无条件算作知识。仅仅能够设想视觉被骗的故事，与已经收到自己视觉受干扰的具体证据，认识地位不同。前者提出逻辑可能，后者可能改变眼下的证成。可错论不能仅凭“人总可能错”便忽略后者。

C3. 内省不可分辨性：理由相同，为什么知识可以不同？

回到 Jones 警探。可错论者可以承认她对 Black 和 White 同样认真、同样有理由，却不必承认她对两人拥有同样的知识。原因在于，知识不只是对主体努力程度的奖励。它还要求信念为真，而真不真部分取决于主体之外的事实。

对于 White，信念实际为假，所以没有知识。对于 Black，如果理由足够好，信念为真，也没有盖梯尔式意外，就可以有知识。这个不对称并不等于说 Jones 应该在行动或主观信念上毫无根据地偏向 Black；它只是在从完整案例的角度评价两项信念是否满足知识条件。

因此，可错论拒绝的是 (5-2)：知识与非知识可以从主体内部看来不可分辨。把主体的信念和理由都保持不变，再改变外部真假，正可能改变知识地位。若知识需要真理，就不能要求知识的每一个条件都表现为内心的某种特殊感觉。

Feldman 把这一点称为主体与世界的某种“合作”。主体以恰当方式相信，有很好的理由，就完成了自己这一方面的工作；世界还必须配合，使这次不是一个有证成的假信念，也不是一个盖梯尔案例。只有双方条件适当地结合，才产生知识。“合作”只是比喻，并不是说世界有意帮助主体。

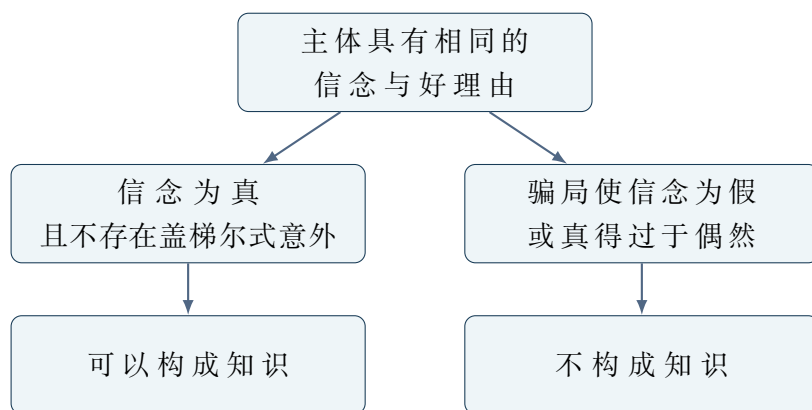


图 F6-1 内部理由与知识地位

原书还有一个生动说法：真信念不会比假信念发出更亮的光，也不会在心眼前闪烁一个表示“真”的字母 T。无论是否接受可错论，我们都可以理解，两个从内部看来相同的信念可能一真一假。既然知识要求真理，知识与非知识之间也就未必具有普遍可见的内部标志。

这一回应并没有否认我们通常能够判断自己的理由好不好。Feldman 甚至保留一种可能看法：有证成与无证成也许比有知识与无知识更容易从内部区分。不过，这里没有证明所有证成都必须内部可辨；原书注释明确说，非证据主义者可能反对这类主张。“**知识不完全由内部状态决定**”，也不等于“**对证成的一切外在主义理论都已成立**”。

进一步的反对：如果连两种树都分不清，凭什么说知道？ 怀疑论者仍可提出一个看来更实际的疑问。假如我分不清柏树（cypress）与雪松（cedar），看见它们时的经验对我没有区别，我怎么能知道眼前是哪一种？如果这里不能知道，那么看见真树与梦见树内部相同，为什么还能知道自己真的看见树？

原书的回答是，分不清外观，不必等于缺少一切相关证据。植物学专家可以告诉我树种；我也可能有独立而有说服力的理由，知道当地根本不生长雪松。这些信息并没有使两种树的外观突然变得不同，却可以使我知道眼前是哪一种。

原书树木案例的论证作用

假设我自己无法凭视觉分清柏树与雪松。

若专家认定眼前是雪松，并且我有充分理由信赖这一认定，证言可以补充单凭外观得不到的支持。若我另有充分理由知道当地没有雪松，相关背景证据也可能支持眼前是柏树。

所以，“某两种视觉经验对我不可分辨”最多先说明单凭这项视觉差异无法区分，不能未经论证就推出“我用任何证据都不可能知道”。

但这里也不能跳得太快。树种例子引入了新的独立信息；梦、恶魔与 BIV 的强版本却可以继续追问：专家证言和背景经验是否也在欺骗中？Feldman 并不是用一次专家指

认就演绎消灭了全部怀疑。他要说明的是，没有绝对可靠的内部判据，仍可能拥有很好的总体理由；若要说明所有这些理由都不够，需要重新论证知识必须绝对确定。下一小节 C4 及原书第七章还将继续处理这个压力。

进一步的反对：即使偶尔真的知道，我能知道自己知道吗？ 读到这里，可能产生另一种不安：“世界配合时，我有知识；世界不配合时，我没有。既然内部感受可以一样，我是不是永远无法判断自己属于哪一种？”原书把这个疑问表述为：

论证 6.8：知道自己知道的论证

- (8-1) 如果可错论为真，那么人永远不能判断自己是否拥有知识，也就是永远不知道自己拥有知识。
- (8-2) 可错论为真。
- (8-3) 所以，人永远不知道自己拥有知识。

Feldman 指出，结论并非自相矛盾，却很奇怪。可错论既然允许我们知道各种普通事实，也允许我们知道别人知道某事，为什么唯独“我自己知道”必须永远不可知？要看出 (8-1) 哪里有问题，应把不同层次拆开。

以“灯亮着”为例。一阶知识 (*first-order knowledge*) 是我知道灯亮着。高阶知识 (*higher-order knowledge*) 在这里则是我知道“我知道灯亮着”。后一句的对象不只是灯，还包括我与这个事实的认识关系。

在正常情形中，我观察到灯光，有充分理由相信灯亮着；灯也确实亮着，且没有盖梯尔式意外。这样可以得到一阶知识。随后，我还可以想到：自己是在认真观察，而非任意猜测；在此类判断上通常正确；自己拥有正常而有力的理由。按照 Feldman，这些又可以构成很好的理由，支持“我知道灯亮着”。若这个高阶信念为真并满足相应知识条件，就可以得到高阶知识。

第二层理由仍然不是绝对可靠。一个被巧妙欺骗的人，可能错误地以为自己知道。但这并不比第一层理由可错更令人惊讶。若第一层允许很好而非完美的理由，第二层却突然要求一个永不出错的鉴别器，就在不加论证地提高标准。

高阶知识的可能性与成立条件

Feldman 在此拒绝的是“可错论使高阶知识永远不可能”。他并没有因此主张，任何主体只要知道 p ，就已经自动知道自己知道 p 。高阶知识仍需要主体形成相关信念并满足相应条件。

同样，“可能判断错”也不等于“永远不能判断对并获得知识”。可错论在第一层与第二层采用的是同一原则。

C4. 传递性论证：为什么不能知道自己不是缸中之脑？

可错论对最后一个论证的回应，可以直接回到原来的前提编号。它不必放弃 (7-3) 的知识传递，而可以拒绝 (7-1)：“主体不能知道怀疑假设为假。”

为什么有权拒绝？Feldman 认为，使 Brian 有很好理由相信“我有手臂”的普通经验，也为拒绝“我仅仅是缸中大脑”提供很好理由。故事里没有给出任何实际证据表明存在这种模拟装置，更没有给出 Brian 恰好身处其中的证据。我们能够描述一种欺骗，并不等于这个描述已经成为击败普通证据的理由。

这里不能把论证缩成“没有证据说我是 BIV，所以我知道不是”。单纯没有找到某事的证据，并不普遍保证知道它为假。Feldman 诉诸的还包括支持普通命题的正面理由：这些理由本来就支持“我有手臂”，而有手臂与“仅仅是大脑”的设定不相容。

怀疑论者可能继续说：“这些经验仍可能由模拟器制造。”可错论者同意这不是逻辑上完美的排除，却追问：为什么这就不足以产生知识？如果回答只是“只要尚有错误可能，就不能知道”，论证便回到了 (4-2) 或 (6-1)，而这正是前面已经受到挑战的高标准。

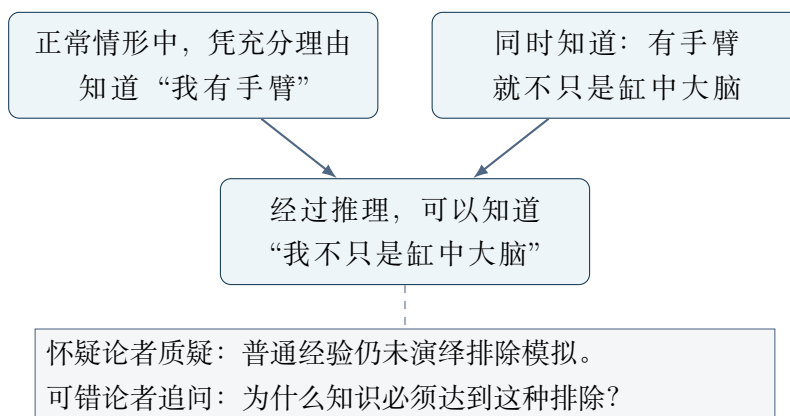


图 F6-2 可错论与知识传递

这个回应不等于宣布所有怀疑论者都会接受它。原书明确保留更深的反对，留待第七章讨论。就本章而言，重要的是看清论证 6.7 没有自行证明 (7-1)。它表明，如果不能知道怀疑假设为假，普通知识也受到威胁；它并未单凭这一传递关系，证明前一种知识确实不可能。

因此，准确的阶段性结论是：可错论者有理由要求为 (7-1) 提供独立辩护。若其唯一辩护仍是“经验不绝对可靠”，争论便再次落在知识是否需要绝对确定上。不能仅因为怀疑论者提出了该前提，就把它当作双方已经接受的事实。

知识闭合与实际推理

传递性论证也常以**认识闭合原则** (*epistemic closure principle*) 表述：如果 S 知道 p ，保留这项知识，并从 p 有能力地推出 q 、由此相信 q ，那么 S 获得关于 q 的知识。¹⁵

这些限定有实际作用。原则不说一个人自动知道自己知识的所有逻辑后果，也不说他只要真诚地认为推理有效便获得知识。它要求适当完成推理、因此相信结论，并且在过程中没有失去原来的知识。否则，一个连结论都没有想过的人，会被错误地算作已经知道它。

令 O 为“我有两只手”， SK 为课件设定的“我是没有手的缸中之脑”。在主体把握 O 蕴涵 $\neg SK$ 并适当推理的条件下，闭合原则支持：

$$K_S O \longrightarrow K_S \neg SK.$$

怀疑论者接受这一联系，却否认后件，因而否认前件；摩尔式回应则肯定前件，并由此肯定后件。双方并非分别选择了正确和错误的推理形式，而是在有冲突的主张之间作出了不同取舍。

F6.V 阶段性结论

本章处理的是一类**高标准怀疑论** (*high-standards skepticism*)：它借助知识需要确定性、不能容许错误可能等要求，否定普通外部世界知识。在错误可能性论证和确定性论证中，这种要求直接写进前提；在不可分辨性和传递性论证中，Feldman 认为，它为关键前提提供了重要而较隐蔽的支持。

注意，这是一项经过分析才提出的诊断，不是说四个论证从字面上完全相同。特别是传递性论证，它可以在形式上不出现“确定性”这个词；Feldman 质疑的是，为什么相信“不能知道怀疑假设为假”，以及这种相信是否又依赖了高标准。

三种回应的分工也已经清楚。自我反驳指控考察怀疑论者的整体立场是否自洽；摩尔式回应比较常识知识与怀疑前提的可信程度；可错论进一步说明，哪些前提可以拒绝，以及为什么“知道”不必等同于“不可错”。前两种回应不是毫无价值，而是不足以代替后一种逐前提解释。

¹⁵本节采用 *Skepticism* 第 1、5 页对闭合原则的表述。

论证	回应位置	可错论的具体回答
6.4 错误可能性	拒绝 (4-2)	知识要求信念实际为真，不要求证据把所有错误可能都演绎排除。
6.5 不可分辨性	拒绝 (5-2)	内部理由相同，外部真值与盖梯尔因素仍可不同，所以知识地位未必相同。
6.6 确定性	拒绝 (6-1)	极好而非最高强度的理由，可以满足知识所需的证成。
6.7 传递性	拒绝 (7-1)	支持普通命题的好理由，也可以支持与之不相容的怀疑假设为假。
6.8 高阶知识	拒绝 (8-1)	知道自己知道也可以是可错的；不必要求比一阶知识更高的绝对标准。

Feldman 认为支持可错论的理由相当有力：普通语言区分知道与绝对确定；它保留日常认识差别而不必另造“几乎认识论”；单身汉类比揭示了一种关于“必须”和“不能”的混淆；主体与世界的合作说明了内部相同为何不保证知识相同；高阶知识也不因允许可错性便被彻底排除。

不过，有力不等于决定性。可错论尚未给出一个在所有案例中都能机械使用的证据门槛；它也没有在本章消除所有关于普通证据与怀疑假设的争论。Feldman 因此称结论为“阶段性的”。他把这一路线与本书第四章的温和基础主义联系起来：经验可提供非推论的初步支持，但这种支持不必不可错。Goldman 第一章关于直接证成、击败者与增强者的讨论，也呈现出类似的区别：经验可以提供支持，而这种支持是否充分，要结合其他相关因素判断。

Feldman 第六章：整章思路

本章先问清怀疑论否认的范围与强度，再把“没有事实”“事实为假”和“我们不知道”分开。完成这些澄清后，梦、恶魔和缸中之脑才能获得明确用途：它们显示普通经验可以与错误相容。

随后出现真正的争论：**与错误相容的证据，是否一概不能产生知识？**四个怀疑论论证分别通过错误可能、内部不可分辨、绝对确定和知识传递，把这个压力加到日常信念上。

可错论并不否认我们可能被骗，也不允许假信念成为知识。它主张，真实且非盖梯尔式的信念，可以凭很好而非完美的理由成为知识。这样便能逐一质疑四个论证的关键前提，同时说明高阶知识为何也仍然可能。

留下的问题是：什么样的好理由确实足够？哪些错误可能只是可描述的情形，哪些会真正削弱当前证成？这些问题使怀疑论讨论继续展开，而非在一句“人当然会错”之后结束。

思考题与讨论

1. Jones 对两个人都有同样好的理由，为什么可错论容许知识地位不同？先区分证成与真理，再说明知识既取决于理由，也取决于信念是否为真及是否存在盖梯尔式意外。不能用“她其实更相信 Black”解释，原案恰好排除了这一差别。
2. “知识不可能为假”为什么不能直接推出“知识必须有不可错证据”？前一句说，假信念不能保有知识资格；后一句说，证据必须逻辑保证真理。可错论允许相同证据在另一个错误情形中出现，只是那个情形里的信念不再是知识。
3. 传递性论证为何不能单凭自身证明没有普通知识？因为它需要先接受“不能知道怀疑假设为假”。传递原则说明前后两种知识的联系，却没有独立证明后一种知识不可获得。作答还应说明，可错论拒绝的是这个知识缺失前提，而非本章采用的传递原则。

怀疑论的其他回应

可错论通过区分充分理由与绝对保证来保留日常知识。但还可以从不同方向质疑怀疑论：缸中之脑能否用我们的词语表达同一个假设？知道普通事实是否要求排除所有可设想的替代可能？“知道”的使用标准是否随谈话语境改变？这三组问题分别引向语义外在论、相关替代项理论与语境主义。¹⁶

普特南：词语的意义是否完全在心中

希拉里·普特南 (Hilary Putnam) 的反怀疑论论证以**语义外在论** (*semantic externalism*) 为出发点：词语的意义或指称，至少部分由说话者之外的因素决定。它不同于 Goldman 讨论的证成外在主义。前者问词语和思想指什么，后者问信念凭什么得到证成；接受其中一个，并不自动等于接受另一个。

含义 (*meaning*) 关乎表达式所表达的内容，**指称** (*reference*) 关乎表达式指向何种对象或事物。在接下来的思想实验中，重点是：即使两人的内部状态相同、发音相同，所处环境及与对象的联系不同，词语也可能指向不同物质。

孪生地球

设想孪生地球与地球十分相似，只是河流、湖泊和海洋里的液体不是 H_2O ，而是另一种外观和日常性质相似的物质 XYZ。那里的人说一种结构上与英语相同的语言，也使用“water”这个声音。但按照语义外在论，地球人的“water”指 H_2O ，孪生地球

¹⁶以下讨论依据课程课件 *Skepticism*，第 3-7 页；它们延伸怀疑论讨论，不属于 Feldman 第六章原有的分节。

人的 “water” 则指 XYZ。

两人都可以望着各自海洋说 “There is water here”，并各自说真话。没有矛盾，因为他们表达的命题不同。地球人说的是真水存在，孪生地球人说的则是 XYZ 存在；不能先把后者的 “water” 直接译成地球意义的 “水”，再指责他面前没有水。

语言环境	“water” 的指称	“There is water here” 为真的条件
地球	H ₂ O	当地有 H ₂ O
孪生地球	XYZ	当地有 XYZ

由此转向一种始终生活在模拟环境中的缸中之脑。它没有通过正常历史接触实际的树、脑和缸；它使用的 “树” 可能指向图像中的树，或者产生树经验的电子刺激，而不指向我们说的树。因而，当它发出 “我是缸中之脑” 这一串声音时，它未必在表达我们用同样字词表达的命题。

这一步依赖特殊的语言史设定。刚刚被放进缸里、此前一直生活在普通环境中的人，可能仍然用 “树” “脑” “缸” 指真实对象。不能把关于长期模拟生活的语义论证直接推广到所有可能的缸中之脑。

普特南论证的两种重构

第一种重构依赖自己所表达的内容：

- 1. 我说或想 “我不是缸中之脑” 时，表达的是：我不是实际的脑放在实际的缸里。
- 2. 一个始终处于相应模拟环境的缸中之脑说或想同样字词时，表达的是另一内容，例如它不是图像中的脑放在图像中的缸里。
- 3. 因而，我不是这种缸中之脑。

如果前两项都已得到恰当说明，内容的差异便能用于区分双方。然而，一个正在怀疑自己是否处于缸中的人，为什么能够先肯定第一项？语义外在论说明环境会影响指称，却没有仅凭这一一般原则就确定我实际处于哪种环境、我的词语实际指什么。如果第一项已经要求掌握这种区别，它就不能轻易用来解除最初的怀疑。

这一困难与摩尔证明相似。由 “我有手” 和 “缸中之脑没有手” 可以推出 “我不是缸中之脑”；但已经怀疑自己是否有手的人，不会仅凭这项推理接受第一个前提。论证有效、甚至事实上健全，与它能否让怀疑者获得先前没有的根据，是不同问题。

第二种重构试图先不选定自己处于哪种环境，而按两种可能分别推理：

- 1. 我或者是缸中之脑，或者不是。

2. 如果我是那种长期处于模拟环境中的缸中之脑，“我不是缸中之脑”就表达：我不是图像中的脑处于图像中的缸。
3. 在这一设定下，这句话表达的内容为真：实际的模拟主体并不是图像内的那个对象。
4. 如果我不是缸中之脑，这句话按普通含义表达：我不是实际的脑处于实际的缸。
5. 在后一情形下，这句话表达的内容也为真。
6. 所以，无论哪种情形，我说“我不是缸中之脑”时所表达的内容都为真。

这里的第三步还依赖第二步对图像指称的具体解释，并非单凭“意义有外在因素”就成立。但即使接受整个分情况论证，结论也需谨慎理解：它证明的是同一串字词在两种环境中各自表达的命题为真，而这两个命题并不相同。能否由此进一步得到普通含义下“我不是实际缸中之脑”的知识，正是争议所在。

“刚刚入缸”的可能又带来独立限制。如果一个人原有正常的语言联系，后来才被接入模拟器，他也许仍在表达关于实际脑和缸的命题。此时第二步关于指称变化的设定不再适用。因此，这条语义路线即使能够排除某种极端假设，也未必排除了所有知觉欺骗。

相关替代项理论

相关替代项理论 (*relevant alternatives theory*) 主张，知道 p 需要排除与 p 竞争的相关替代可能，而不必排除所有逻辑上可描述的可能。

回到柏树与雪松的例子。若当地确有两类外观相近的树，而主体完全无法区别，那么“眼前是另一树种”就是需要认真处理的替代项。相比之下，在普通认树情境中，一个没有任何具体依据的全局模拟故事，未必属于同一组相关替代项。问题由“能否彻底排除一切错误”转为“哪些错误可能在当前判断中必须排除”。

但“相关”不能只解释为“知道时必须排除的”，否则理论只是用另一个词重述知识条件。它还需要说明替代可能为何进入或离开比较范围。课堂以德雷茨克 (Fred Dretske) 的思路强调这种选择性，并借此考察闭合原则。

三个不能同时接受的主张

1. 主体不知道怀疑假设为假： $\neg K_S \neg SK$ 。
2. 普通知识在适当推理下传递到对怀疑假设的否定： $K_S O \rightarrow K_S \neg SK$ 。
3. 主体拥有普通知识： $K_S O$ 。

摩尔式回应拒绝第一项；怀疑论拒绝第三项；课堂讨论的相关替代项路线则拒绝第二项。它认为，判断“我有手”与判断“我不是缸中之脑”时，所需排除的替代项可能不同。前者可以在普通比较范围内得到知识，而后者在新的比较范围下没有得到知识；因

此，不能仅由逻辑蕴涵就保证知识继续传递。

这一回答的代价是放弃相应的闭合原则。我们很自然地认为，从已知前提作适当演绎可以扩展知识；相关替代项理论必须说明为什么某些推理会越出原有的知识范围。并非凡是强调“相关性”的理论都必须作同一种取舍；这里针对的是课件所讨论的拒绝闭合的版本。

语境主义：谁在什么语境中说“知道”

语境主义 (*contextualism*) 把注意力转向知识归属句：“S 知道 p” 是否说真话，部分取决于作出这项归属的谈话语境。这里变动的是归属者语境 (*context of ascription*)，并不必然是被评价主体的证据、信念或外部处境。

“高” 提供了一个类比。Henry 九岁，身高接近五英尺。在谈论平均约四英尺六英寸的同龄同学时，说 “Henry 很高” 可以为真；在谈论 NBA 球员的身高时，说 “Henry 不高” 也可以为真。他的身高没有在两次谈话之间改变，变化的是比较标准。这与 Feldman 第二章关于 “Michael 很高” 的边界问题相关，却并不相同：那里突出标准不精确时的模糊性，这里突出不同语境采用不同标准。

谈论双手的两种语境

在酒吧里，朋友谈起一位因滑雪事故失去双手的人，并说：“至少你还有双手，应该感到庆幸。” 主体回答：“当然，我知道自己有双手。”

在认识论课堂上，大家详细讨论没有手的缸中之脑。有人追问：“如果你也是这样的脑，一切经验仍会完全相同。你怎样知道自己有双手？” 同一个主体回答：“这样说来，我不知道自己有双手。”

语境主义允许两次回答都真。在普通语境中，知识标准较低，正常经验足以满足要求；在怀疑语境中，先前被忽略的错误可能成为谈话焦点，标准提高，原有证据不再满足此时“知道”的使用条件。

归属语境	“我知道自己有双手”	“我不知道自己有双手”
普通语境	真	假
怀疑语境	假	真

这不表示“有手”这个事实因谈话改变，也不表示主体在谈话中突然失去了手或原来的视觉证据。变化的是知识归属句的评价标准。因此，严格说来，怀疑假设被提出后，不是有某种知识实体从主体那里消失了，而是在新语境下，原先的肯定归属不再为真。

这与拒绝闭合的相关替代项路线也不同。语境主义可以要求在同一语境、同一标准下保留闭合，而把冲突解释为论证中知识标准发生了转移。在高标准下，主体既不满足

“知道自己有手”，也不满足“知道自己不是缸中之脑”；在普通标准下，则可以作相应的肯定归属。

语境主义留下的困难

第一项困难是**认识上的下降** (*epistemic descent*)。既然怀疑可能一旦被提出便提高标准，之后怎样回到日常标准？不能只是因为我们希望重新说“知道”，就任意宣布那些可能不再相关。理论需要解释谈话语境怎样转换，以及哪些变化足以使高标准不再支配当前归属。

第二项困难是**语义盲点** (*semantic blindness*)。若怀疑论长期困扰哲学家的原因，在于他们没有注意一个普通词语的语境变化，那么为什么这种变化如此难以识别？“高”通常能够容易地补出比较范围，“知道”是否真有完全类似的语义机制，需要额外理由。

第三项困难是**不可说出的知识** (*unspeakable knowledge*)。按照课件所讨论的语境主义版本，有些命题可以在普通标准下被知道；但某些关于它们的知识宣称，一旦被明确提出，就会使相关错误可能凸显、抬高标准，从而不再能够真确地说出。难点不在于主体没有语言表达能力，而在于表达行为似乎改变了原本使归属为真的条件。并非每一次说“我知道”都必然提高标准，但理论必须说明这种效应何时发生，为什么不使日常知识宣称普遍失效。

几种反怀疑论路线的分歧

可错论质疑不可错的知识标准；摩尔式回应优先保留常识知识；普特南借语言与环境的联系限制某些怀疑假设的表达；相关替代项理论限制必须排除的错误范围，某些版本为此放弃闭合；语境主义则把知识归属的真值与谈话标准联系起来。它们并不是同一种回答的不同名称，理论代价也各不相同。

知识论板块总结：从知识分析到怀疑论

五篇文献与课堂讨论围绕知识、证成与怀疑论形成相互联系的讨论：

- 1. **Feldman 第二章建立候选定义。**知识至少要求信念、真理和充分证成。
- 2. **Feldman 第三章揭示认识运气。**三项条件全满足时，结论仍可能因与理由无关的巧合而真，因此还需排除 Gettier 式偏差。
- 3. **Goldman 第一章追问证成的结构。**任何推论性证成都引出理由倒退；基础主义以经验提供非推论起点，融贯论强调系统联系，无穷主义接受无尽理由链。
- 4. **Goldman 第二章追问证成的来源。**当前证据与信念共存并不够；形成和保持信念的历史、过程可靠性、逻辑关系与导真规范都可能影响证成。
- 5. **Feldman 第六章检验证成标准。**若证成必须绝对排错，怀疑论似乎成立；可错论则主张很强而非完美的证据已经足够。

问题层次	核心提问	本板块的阶段性答案
知识分析	哪些条件构成知识？	真信念、充分证成、并排除 Gettier 式认识运气
证成结构	理由链怎样组织？	温和基础主义仍有优势，但须吸收融贯关系与击败者
证成来源	为什么一个实际信念合理？	既看主体证据，也需考察基于关系、形成历史与真理联系
怀疑论标准	知识必须绝对确定吗？	可错论否认；知识可由很强但可错的证据产生

知识分析的后续路线进一步表明，排除运气未必只靠增加一项关于理由的条件。因果理论转向事实与信念的联系，可靠主义转向过程的真理产出倾向，知识优先论则质疑还原分析本身。头部撞击、假谷仓与 Temp 分别检验这些联系是否足够，而不是重复同一个“碰巧猜对”的故事。

关于证成结构，有限心灵与序言悖论把问题推向更细的层次：哪些可获得的信息算作主体拥有的信念？逐项有理由相信，是否等于有理由相信全体合取？塞拉斯二难则继续追问经验为什么能够成为理由。这些问题说明，基础主义、融贯论和无穷主义都需要对支持关系作正面解释。

怀疑论的进一步回应又把认识论与语言哲学联系起来。语义外在论关乎思想内容如何由环境部分决定；相关替代项理论关乎必须排除哪些错误；语境主义关乎何时能够真正地归属知识。比较它们时，应分别检查所保留的知识、闭合原则和语言条件，不能把“有一种反怀疑论回答”当作争论已经结束。

综合思考题

1. 为什么真信念不是知识？为什么得到证成的真信念仍可能不是知识？
2. JF、JD 与 SE 分别在 Gettier 论证中承担什么作用？
3. 基础主义、融贯论、无穷主义如何回答倒退问题，各自的主要困难是什么？
4. 可错论怎样分别回应四个怀疑论论证，它拒绝的是哪项关键前提？

第二板块

心灵哲学

心灵哲学板块导读

本板块沿两条互相交叉的路线推进。第一条是**心身关系路线**：心灵是否是一种非物质实体？如果不是，心理状态应当被还原为行为、脑状态还是功能角色？第二条是**心智能力路线**：知觉如何让世界呈现给我们，思想以什么形式运作，意识为什么似乎无法由物理事实完全说明？两条路线最终在**心灵因果性**（*mental causation*）问题上汇合：如果心理状态真实存在，它们怎样在物理世界中产生作用？

原书章节	核心任务	在板块中的作用
Heil 2–3	建立笛卡尔二元论，并追踪其历史后继者	说明“非物质心灵”既符合某些直觉，又制造心身因果难题
Bayne 3	比较行为主义、同一论、功能主义与取消主义	展示物理主义不只有一种版本
Bayne 4–5	分析知觉与思想	从具体心智能力检验表征论、功能主义与人工智能
Bayne 8	讨论意识的解释鸿沟	集中考察物理主义最困难的对象
Bayne 10	讨论心灵因果性	检验心理解释能否在因果封闭的物理世界中保有位置

板块总路线

笛卡尔二元论 → 物理主义的多种实现
→ 知觉、思想与意识的个案检验
→ 心理状态能否真正成为原因

Heil 第 2 章：笛卡尔二元论 (Cartesian Dualism)

文献定位

本章不是把笛卡尔二元论当作一个早已淘汰的历史怪论，而是先展示它为何自然：心理经验似乎非空间、具有独特质感，并且只向主体本人直接开放。随后，Heil 把这些直觉组织成实体二元论，再说明真正的压力来自**心身互动**与**物理因果封闭**的冲突。

2.1 科学与哲学

我们通常认为自己和他人都有心灵，但“有心灵”究竟意味着什么？可能的回答差别很大：心灵可以被看成暂居身体的灵魂，也可以被看成身体以某种方式组织起来时具有的状态，还可以直接被认作大脑。这里首先要避免一种误解：科学给出答案，哲学只提出没有答案的问题。

Heil 的要点是，经验材料本身往往不能唯一决定其形而上学解释。神经科学可以告诉我们某种脑活动与疼痛报告稳定相关，却未必单凭实验就能决定“疼痛就是脑活动”“疼痛由脑活动实现”或“疼痛是脑活动造成的非物质状态”哪一种说法正确。科学提供事实和模型；如何理解这些事实所承诺的世界图景，仍需概念分析与整体理论比较。

原书开篇还请读者留意一种常见的分工误解：若把“科学发现事实、哲学在事实之外空谈”当作先决条件，就已经替某种心灵理论作出了判断。哲学家不仅要服从成功的经验研究，也要厘清研究中的“状态”“解释”“同一”“原因”究竟是什么意思。反过来，单靠自己觉得心灵神秘，也不能反驳任何神经科学发现。以下考察二元论，正是让这两类考量在同一个问题中碰面。

经验问题与形而上学问题

经验问题 (*empirical question*) 原则上可由观察或实验裁决，例如某脑区是否在疼痛时激活。**形而上学问题** (*metaphysical question*) 追问这些事实中的对象和性质究竟是什么、如何依存，例如疼痛是否与某种神经状态同一。两者相关，但不能简单互相

替代。

本节作用

本节为整本心灵哲学定下方法：哲学不能无视科学，却也不是等待科学替它回答所有问题。竞争理论可能容纳同一批实验结果，差异在于它们如何解释结果。

2.2 笛卡尔的二元论

笛卡尔实体二元论 (*Cartesian substance dualism*) 主张，世界中有两种根本不同的实体：物质实体与心灵实体；人是二者结合而成的复合存在。Heil 先列出三组表面上支持二元论的差异。

2.2.1 空间性与非空间性

物体占据空间，有形状、大小和位置；思想、欲望、感受似乎没有同样意义上的空间尺度。“想吃汉堡的欲望有几厘米长？”听上去就是范畴错误。身体感觉似乎是反例：脚趾痛好像位于脚趾。但**幻肢痛** (*phantom pain*) 表明，脚趾已经不存在，主体仍会体验到“脚趾处”的疼痛。由此最多能说疼痛被体验为位于那里，不能直接推出疼痛事件本身占据那个空间。

2.2.2 物理性质与经验性质

神经科学家可以观察放电频率、神经递质和脑区连接，却不会在脑组织中看见疼痛的刺痛感。经验的质感与物体的质量、形状、运动等性质似乎不是同一类性质。二元论因此说：意识经验具有**现象性质** (*phenomenal qualities*)，它们不能由物质的性质穷尽。

2.2.3 公共可见性与第一人称私密性

物体原则上可以由不同观察者从适当位置观察；我当前的疼痛、图像和思想却以特殊方式直接向我呈现。笛卡尔把这种优势表达得很强：

- **透明性** (*transparency*)：只要主体处于某心理状态，就知道自己处于该状态；
- **不可纠正性** (*incorrigibility*)：只要主体判断自己处于某心理状态，该判断就不会错。

现代心理学关于无意识、动机误认与认知偏差的研究削弱了这两项强主张，但未必消除**特权通达** (*privileged access*)：我对自己当前经验的了解是直接的，他人即便借助最精密的扫描设备，也需从行为或脑活动作推断。

透明性和不可纠正性不是一回事

若我实际上在嫉妒却毫未察觉，这是对透明性的反例：有状态，却不知道。若我把疲惫误判为悲伤，这是对不可纠正性的反例：以为有状态，实际上没有相应状态。某种心理状态可能允许一种错误而不允许另一种错误，因此不能把“第一人称有优势”偷换成“第一人称无所不知也永不出错”。

物质身体 (material bodies)	心灵 (minds)
空间性的、有形状与大小	表面上非空间性的
具有物理性质	具有独特的心理或现象性质
公共可观察	第一人称直接、第三人称间接

图 2.1 Heil 对身体与心灵的三重对照。

不要把差异直接当作结论

上述三组差异首先是二元论的**动机**，还不是决定性证明。物理主义者可以回应：心理状态或许有空间实现却不按日常方式描述；经验性质或许就是我们从第一人称把握的物理性质；通达方式不同也不必推出对象种类不同。

2.3 实体、属性与样态

要准确理解笛卡尔，必须先掌握三个术语。**实体** (*substance*) 不是“水、油漆”一类材料，而是能够作为个体存在、承载性质的具体对象。桌子、树、鸟以及组成它们的部分，都可以在传统意义上叫实体。实体不同于事件、集合和数，也不同于它所具有的性质。

笛卡尔进一步区分：

- **属性** (*attribute*)：使一种实体成为其根本种类的规定；
- **样态** (*mode*)：实体以其根本属性存在的具体方式。

物质实体的根本属性是**广延** (*extension*)，即占据空间；球形、体积和表面纹理是广延的不同样态。心灵实体的根本属性是**思维** (*thought*)；感觉、想象、情绪、信念和欲望都是思维的样态。笛卡尔的“思维”比日常用法宽，几乎包括所有意识状态；但它又比现代认知科学窄，因为笛卡尔基本不承认无意识思想。

这套区分的论证作用比术语本身更重要。说身体有温度或形状，不是在身体之外再加一个“温度实体”；说心灵在疼痛，也不是心灵之外另放一个独立的“疼痛东西”。但

笛卡尔进一步认为，若一物的根本属性是广延，它的样态就都要以广延为基础；若一物的根本属性是思维，其样态就以思维为基础。正是由这个根本属性不可混同的判断，他才得出实体分立的结论。反对者可以接受“性质不同”，却质疑它是否足以推出“承载性质的实体不同”。

一个容易混淆的层级

实体：这一个心灵或身体 $\supset$ 属性：思维或广延 $\supset$ 样态：疼痛、信念；形状、运动
属性不是额外漂浮的对象；样态也不是脱离实体自行存在的东西。

2.4 笛卡尔二元论的形而上学结构

由“每种根本实体只有一个定义性属性”与“思维和广延互相排斥”，笛卡尔得到：有广延者不思维，思维者无广延，因此心灵不等于身体。不过，**不同不等于无关**。我的心灵与我的身体具有特别紧密的关系：我的手碰火时是我痛，我决定起身时通常是我的身体移动。

笛卡尔的完整图景因此包含双向因果：外界和身体产生感觉，思想、欲望和决定引起身体动作。原书“坐到图钉上”的例子可重构如下。

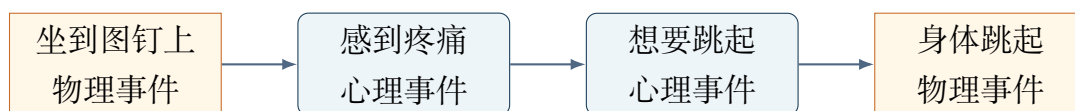


图 2.2 笛卡尔式心身互动：因果链跨越物理域与心理域。

这种图景有两项吸引力。第一，它贴近日常的自我观：我有一个身体，但似乎可以想象自己换身体、离开身体或在身体死亡后继续存在。第二，它解释了科学描述与经验世界之间的差异：物理学描述无声、无色的微观结构，心灵却拥有鲜明的颜色、声音与痛感；二元论把后者安置在心灵中，并以稳定共变说明经验为何能指示外界。

Heil 用卡夫卡小说中醒来变成巨型昆虫的格里高尔、科幻小说中的换身体与出体经验，说明为何“我不是这具身体”容易被设想。但这个思路是不对称的：我们也会说“我有一个心灵”，然而“把我的心灵换掉，我仍是我”就很难说清；日常的“改变主意”只是改信念，不是换掉一个心灵实体。这里的想象首先解释二元论为何诱人；从一个故事容易想象，不能直接推出它在形而上学上可能。

另一条原书线索是经验与科学图景如何对接。物理学描述表面微结构、光线与声波，经验却有红球的鲜红和管弦乐的音色；二元论给这些经验品质留了位置。它仍须说明为何红色表面可靠地引发相应视觉经验，进而让我们凭经验认识世界。故二元论不只是“两种东西”的清单，还要有一套跨两域的稳定对应关系。

2.5 心身互动难题

二元论的优势也是它的核心负担。若心灵没有空间位置、身体却只能通过空间中的接触和运动发生作用，两者如何互动？波希米亚的伊丽莎白公主向笛卡尔提出的质疑是：与其承认无广延者能推动物体，似乎还不如承认灵魂本身具有某种广延。

难题可以分成三层。

1. **接触问题**：若物理因果依靠接触，非空间心灵如何接触身体？
2. **传递问题**：若原因把运动、能量或某种量传给结果，而心灵与物体没有共有性质，究竟传递了什么？
3. **封闭问题**：现代科学假定物理事件若有充分原因，就有充分的物理原因；额外的非物质原因似乎会破坏完整解释。

二元论者可说心身因果是 *sui generis*，即独一类的基本关系，不应套用物理接触模型。但这只是允许它存在，尚未说明它怎样可能。更严重的是，现代科学的工作方式通常预设**物理因果封闭** (*causal closure of the physical*)：

物理因果封闭原则

对于每一个具有原因的物理事件，都存在一个充分的物理原因；列出全部物理原因，便得到完整的物理因果解释。

笛卡尔曾把松果腺视为心身联系点，并设想心灵不增加运动总量，只改变粒子方向。但后来的动量与能量守恒理论使这种方案更困难。量子概率也不是明显出口：若基本概率本身就是定律的一部分，心灵系统性地“挑选”结果仍会改变真实概率分布，而不是在定律允许的空隙中无痕工作。

“没有说明机制”与“原则上不可能”须区分

伊丽莎白公主的问题首先要笛卡尔解释：两个被规定得如此不同的实体怎样作用。接触或运动转移只是在当时看来最自然的物理模型；若原因不一定需要接触，单凭“没有接触”未必构成逻辑反驳。然而若物理学能够为每个动作给出完整物理原因，心灵另外造成同一个动作便涉及重复解释。后一压力来自因果封闭，比单纯想象不到连接机制更强。

2.6 一个因果封闭的宇宙

Heil 最后强调，因果封闭不是通过某一个实验直接证明的逻辑真理，而是一条极有力量的方法论承诺。科学家遇到暂时没有物理解释的事件时，通常继续寻找隐藏的物理

原因，或承认它可能是无因的随机事件，而不会立即引入非物质心灵。二元论若要保住互动，必须说明为什么这一成熟的解释策略在某些脑事件上失效。

互动二元论面对的三岔路

1. 否认物理因果封闭，允许心灵补入物理因果链；
2. 接受封闭但把心理与物理原因看成系统性过度决定；
3. 修改因果概念，使心灵不以“额外能量输入”的方式发挥作用。

Heil 第三章与 Bayne 第十章将分别考察这些路线的变体。

Heil 第二章整章总结

1. 心灵与身体在空间性、性质类型和认识通达上呈现三重差异，这构成二元论的直觉基础。
2. 笛卡尔把身体界定为以广延为属性的实体，把心灵界定为以思维为属性的实体；思维与广延互斥，故心身不同一。
3. 二元论能解释第一人称经验与科学描述的差别，也贴近日常“我有一个身体”的自我理解。
4. 但笛卡尔还必须承认双向因果：身体产生感觉，心灵引起动作。
5. 非空间实体怎样作用于空间实体并遵守守恒定律，是互动难题；物理因果封闭使问题更尖锐。
6. 本章没有直接判决二元论为假，而是确定其说明责任：它必须给出可理解的心身因果模型。

Heil 第 3 章:笛卡尔的遗产(Descartes's Legacy)

文献定位

本章考察笛卡尔之后的几种“修补方案”。它们共同承认心理与物理之间存在稳定对应，却以不同方式放弃或重写互动：平行论否认跨域因果；偶因论让上帝协调；唯心论取消物质实体；副现象论只保留物理到心理的单向作用；Lowe 则尝试一种不把自我说成非物质的实体二元论。

3.1 无互动的二元论：平行论

心物平行论 (*psychophysical parallelism*) 保留两种实体，却否认心理与物理之间真正因果互动。坐到图钉、感到疼痛、想要跳起和身体跳起仍依次出现，但心理序列与物理序列只是精准同步。

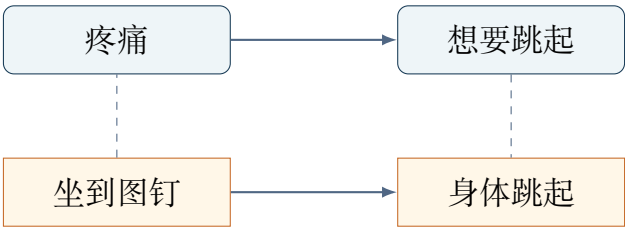


图 3.1 平行论：实线为域内因果，虚线仅表示同步，不表示因果。

问题是：如此细密而普遍的同步为何存在？若回答“这是原始事实”，理论显得特设，因为互动本来提供了更直接的解释。若诉诸上帝预先设定两条序列，如钟表匠调好两只永远同步的钟，又必须给出独立的上帝论证，否则只是把“不可解释的同步”改写成“上帝保证同步”。

试把图中的虚线认真地读一遍。若疼痛不是坐到图钉造成的，则我们说“坐到图钉所以疼”的日常解释只是在指出两条序列中恰好相邻的节点；同样，“想跳起所以跳起”也不是因果解释。平行论不否认这些时序事实，争议在于它能否说明为什么恰好是这样

的一一对应。既然两条链本来可以不同步，前定和谐或更深的共同根据就承担了实质论证任务。

3.2 偶因论

偶因论 (*occasionalism*) 给上帝更积极的角色。身体坐到图钉并不造成疼痛，只是上帝在这个“场合”产生相应疼痛；决定跳起也不推动身体，而是上帝使相应动作发生。

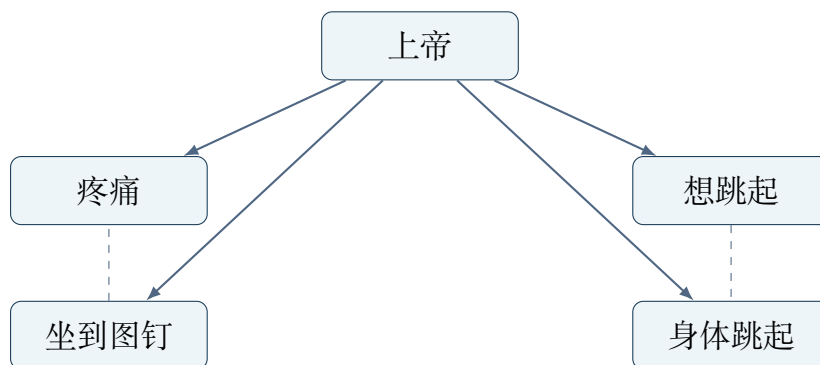


图 3.2 偶因论：表面的横向联系由上帝的纵向作用保证。

偶因论并不必然把人当木偶：主体仍有自己的思想与意志，只是上帝确保知觉对应环境、意志对应身体动作。然而它仍需回答，第三种实体“上帝”如何作用于心灵和物质；若没有进一步因果理论，问题只是被推移。

“偶因”的字面意思值得留心：图钉不是疼痛的真正原因，却是上帝在那个场合令疼痛发生的契机。这既不同于平行论的一次性预设，也不同于笛卡尔二元论里心灵与身体相互作用。平行论问“为何两钟始终同步”，偶因论回答“因为每个时刻都有维持秩序的行动”；接下来我们自然追问，这个维持行动是否比原先的互动更易理解。

3.3 因果关系与偶因论

偶因论较深的动机来自**因果纽带问题** (*problem of the causal nexus*)：当说事件 A 造成事件 B 时，A 与 B 之间究竟多了什么？笛卡尔用运动传递理解因果；休谟式理论则认为，我们实际观察到的只是相似类型事件的稳定接续，所谓“必然联系”是由习惯投射出来的。

Heil 在这里扩大问题：若连两个物理事件之间的“使之发生”都不是我们直接看到的东西，那么不该只因为心—物之间看不见一根传力的绳索，就认定心身因果特别可疑。这可以是对笛卡尔的一种宽容辩护；偶因论者却得出相反结论，认为无论心—物还是物—物，有限事物都没有真正的因果能力。注意这一步从“没有看见内在纽带”到“除上帝外没有真正原因”并非演绎必然，须另有关于因果的形而上学论证。

若因果本来就只是规律性，那么心理—物理因果并不比物理—物理因果更神秘：两者都没有可观察的内在纽带。偶因论者进一步问，为何世界具有如此有序的规律？可把时间想成一系列互相独立的宇宙切片；前一切片本身不能生产后一切片，上帝在每一刻重新创造世界并维持秩序。

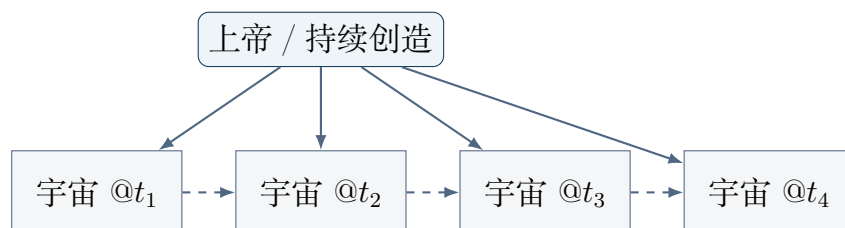


图 3.3 横向因果被改写为每个时间切片与上帝之间的纵向依赖。

这使偶因论成为一整套关于因果与时间的形而上学，而不只是临时修补心身问题。但代价很高：必须接受时间切片的形而上学独立性、所有横向因果只是表象，并承认上帝的持续创造。

3.4 唯心论

唯心论 (*idealism*) 采取更激进的路线：若心灵与物质无法协调，就取消独立的物质实体，只承认心灵及其经验。坐到图钉、身体跳起和实验仪器，在本体论上都被分析为有序的经验集合。贝克莱式版本再用上帝说明为什么不同主体的经验稳定、有规律并彼此协调。

唯心论能“保存现象”：如果它为真，我们仍会有坚硬石头、红番茄和实验室仪器的经验。踢石头或观察云室轨迹都不能简单反驳它，因为观察结果本身仍是经验。它也很节俭，只承认心灵及其内容。

Heil 接连把约翰逊博士踢石头和科学家使用云室作对照。前者使人切身感到阻力，后者提供精密仪器留下的粒子轨迹；两者的认识价值相差很大，但从唯心论者的立场看，二者仍都以观察为终点，尚未在逻辑上把我们带到“独立于一切心灵的物质”。这不是劝人放弃实验，而是要区分“实验数据是什么”与“哪种本体论最好地解释数据”。这一难题也呼应前面 Goldman 基础主义由经验推断外部世界的任务。

然而这种优势也造成不满：唯心论通过把实在认作现象来取消“表象如何对应独立实在”的问题。许多人希望保留表象与实在的区别，而不是把整个物质世界塞进心灵。

3.5 心灵与意义

贝克莱不仅说物质世界不存在，更说“完全独立于经验的物质对象”不可思议。以番茄为例：我们想到的是红、圆、酸，以及在相应条件下会有的触觉、嗅觉和味觉；这

些全是现实或可能经验。若再减去一切经验内容，还剩下什么可理解的“心外番茄”？
论证可重构为：

贝克莱式可设想性论证

1. 我们关于普通对象的可理解内容由实际或可能感觉经验构成。
2. 经验只能存在于心灵中，经验只与经验相似。
3. “心外物质对象”要求保留对象内容，同时排除一切经验内容。
4. 排除全部经验内容后没有剩余的可理解内容。
5. 因而“完全独立于心灵的物质对象”是空洞表述。

反对者会说，可以以因果角色、结构或最佳解释来思想不可直接经验的对象；我们不必先形成一幅相似图像才能有意义地指称它。Heil 在此不裁决争论，而是让读者看到：唯心论的力量来自一套意义理论，而非单纯怀疑外界。

不要把贝克莱的说法缩成一句口号

“你无法想象一棵没有人看见的树”，乍听便有反驳：我现在不就在想一棵无人看见的树吗？贝克莱更有力的追问是，想象中的颜色、形状、质地是不是仍以可能经验刻画对象；若全去掉，这个独立基底还有什么积极内容？对手则不必否认经验材料的重要性，可以说“产生这些经验的共同外在原因”也是内容。争论点是一个不可像经验那样呈现的因果角色能否使物质概念有意义。

3.6 副现象论

副现象论 (*epiphenomenalism*) 接受物理因果封闭，同时允许物理事件产生心理副产品；心理事件本身没有任何物理或心理结果。它类似松动的风扇皮带同时造成发动机过热与哨声：哨声伴随过热，却不造成过热。

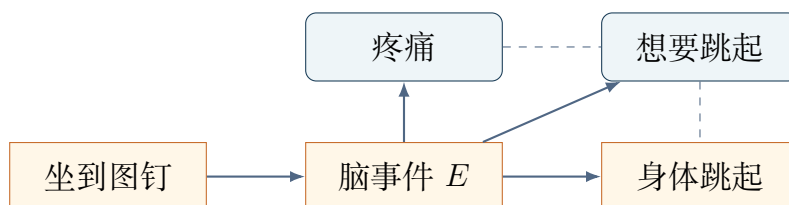


图 3.4 副现象论：脑事件产生经验与动作，经验不反过来产生作用。

这能解释为什么行动总与决定相符：造成决定的脑事件也造成动作。反事实“若我没有决定，就不会伸手”仍可能为真，因为去掉决定也意味着去掉共同的脑原因。可是副现象论承认一种特殊的“悬垂因果关系”：物理事件能产生心理事件，心理事件却永远

不能进入后续因果链。这种不对称本身仍需说明，并受到**奥卡姆剃刀** (*Ockham's Razor*) 的压力：若较简单理论能解释同样事实，就不应增加这种特殊关系。

3.7 尾声：非笛卡尔式二元论

E. J. Lowe 提出**非笛卡尔实体二元论** (*non-Cartesian substance dualism*)。自我与身体是两个实体，类似铜像与构成它的铜块：它们暂时重合并互相依赖，却有不同同一性和持存条件。自我是没有实体性部分的简单实体，但并非完全非物质；它可以非派生地具有身高、质量等物理性质，也具有思想与感受。两只手则只是自我通过身体派生地具有的性质。

原书的铜像例子逐项检验同一性：铜块熔成铜锭后铜块还在，雅典娜铜像不在；换掉雕像一只铜臂，雕像或许还在，原来那块铜却不完整了。若两个东西有不同的保存条件，二者便不能严格同一。Lowe 认为身体与自我也如此：身体作为生物复合体可保留其物理部分而失去自我，一个自我也不等于脑这一身体部分；“我有脑”是因为我有一具含脑的身体，不等于“脑是我作为简单自我的一部分”。

“简单”不能误读成“没有性质”。感觉、记忆和想象能力属于自我，却不是像心脏那样能够作为独立实体存在的组成部分；思想和感受也可归属于这个简单自我。同时，自我的身高与质量无需靠两个相互分离的部件才可归属于它，而“两只手”则需要一个有部分的主体。这正是它与笛卡尔灵魂的分歧：实体二元并不逻辑地要求其中一种实体毫无物理性质。

Lowe 也重写心灵因果。传统图景把决定 M_1 看成在 t_0 插入神经链，最终产生 t_1 的腿部运动 B_1 。但真实神经因果史不断向前分叉、向后交织，无法找到一条只通向 B_1 而不通向眨眼 B_2 或摆臂 B_3 的孤立链。

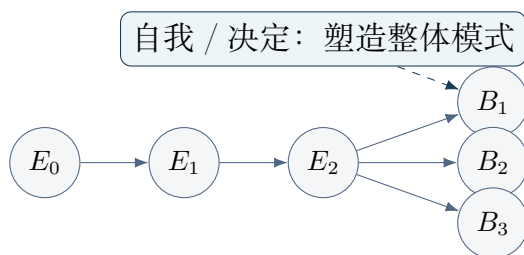


图 3.5 Lowe: 自我不是凭空开启一条链，而是使物理因果网络具有某种行动模式。

他以蜘蛛与蛛网作类比：网不必在每一步推动蜘蛛，却为某种运动提供和塑造条件。类似地，自我可使宇宙中的因果链呈现某种行动结构，而不是额外注入能量。这个方案只是研究起点，却说明“二元论”不等于“非物质灵魂撞击松果腺”。

沿图往回追踪右腿移动的原因，脑信号会再连到先前的身体与环境事件，而且与摆臂、眨眼的原因交织，不容易找出一个在决定发生前就只通向右腿的孤立通道。在这样

一张网络中寻找“心灵从这里插入”的唯一断口，可能从开始便误认了行动的因果结构。Lowe 的主张较含蓄：决定使整个世界的因果网络呈现会产生某类行动的模式；蜘蛛网之于蜘蛛的行动正是“提供条件、规定路径”而非“额外推一下”。类比能给出一种可能思路，却还没有建立自我到底以何种机制选择模式；Heil 也明说这里只是展示非笛卡尔选项，不是解决全部互动疑难。

Heil 第三章整章总结

1. 平行论以心理、物理两条同步链取代互动，但必须解释同步为何不是巧合。
2. 偶因论让上帝保证每个心理与物理事件的对应；其说服力取决于更一般的因果与持续创造理论。
3. 唯心论只承认心灵及经验，因而消除心物互动，却付出取消心外物质世界的代价。
4. 贝克莱的核心不是“我没看到物质”，而是“脱离可能经验的物质概念没有内容”。
5. 副现象论服从因果封闭，允许物理到心理而不允许心理到物理；它保住意识存在，却牺牲意识的因果效力。
6. Lowe 把自我说成依赖身体但不等于身体的实体，并用“塑造因果模式”替代“插入因果链”。
7. 全章的共同教训是：心身理论不能脱离因果、实体、时间、解释与简约性等更一般的形而上学问题。

Bayne 第 3 章：如何成为物理主义者 (How to Be a Physicalist)

文献定位

物理主义说心理现象“不超出”物理现象，但这句口号有多种实现方式。本章比较四条路线：把心理状态分析成行为倾向的**行为主义**，把它们认作脑状态的**同一论**，把它们认作因果角色的**功能主义**，以及否认某些日常心理种类存在的**取消主义**。作者的首要目标是把这些理论分清，而非立即选出胜者。

接续上一章，Bayne 把“不超出”分成两个要求。**模态要求** (*modal constraint*) 说：不可能两个世界在一切物理事实上完全一样，心理事实却不同；这通常用**形而上学随附性** (*metaphysical supervenience*) 表述。**可理解性要求** (*intelligibility constraint*) 进一步说：这种随附不能只是毫不理解的原始巧合，理论应说明物理事实为何足以固定心理事实。“每当神经状态出现，上帝额外造出疼痛”也许给出稳定对应，却没让疼痛在物理意义上不再是额外东西。四种路线要逐一接受这双重检验。

3.1 行为主义

心理学中的方法论行为主义主张，科学应集中研究可观察行为及环境、学习史对行为的改变，避免用不可观察的内部状态解释行为。哲学上的**分析行为主义** (*analytical behaviourism*) 更强：它认为一个心理状态就是一组行为倾向，而且这种分析包含在心理词语的意义中。

例如，相信外面下雨，就是倾向于出门前穿雨衣、在被问“外面下雨吗？”时点头等。这里的行为必须以非心理方式描述；若“穿雨衣”被分析为“有意穿上自己相信是雨衣的东西”，心理词就被偷偷放回定义，分析并未完成。

行为主义有两项重要动机。其一是认识论动机：我们根据微笑、喊叫、行动和言语合理地归属快乐、愤怒和信念；若心理状态就是行为倾向，这种他心知识便不神秘。其二是反笛卡尔动机：心理语言的意义离不开公开使用标准，心灵不是只有主体本人能进

入的“机器中的幽灵”。

这里需要辨认理论的强度。研究时优先观察行为是心理学方法选择；“疼痛”一词的意思就是若干行为倾向则是哲学分析。只有第二种强意义才声称解释疼痛完全无须先解释独立的内部心理物。原书也提醒：赖尔与维特根斯坦批评完全私有的心灵图景，并不因此自动赞成所有心理状态都是行为倾向。

但理论面临三类反例：

1. **私密性**：行为倾向原则上能由第三人称观察，不能解释主体对自身当前思想与经验的特殊通达。
2. **超斯巴达人和演员**：超斯巴达人 (*super-spartans*) 有痛却完全不哭喊、退避，甚至没有这种倾向；“演员”没有痛，却具有全部疼痛行为倾向。两种情形在概念上似乎可能，说明疼痛与行为的联系不是分析必然的。
3. **循环性**：相信下雨是否导致带伞，还取决于主体是否想保持干燥、是否相信雨马上会停、是否想隐藏信念等。分析一个心理状态总会引用其他心理状态，无法把整体心理词汇完全还原为非心理行为。

局部成功不等于一般理论成功

“害羞”“好奇”等性格特征很可能主要是行为倾向，但不能由此推断疼痛、图像和信念都只是行为倾向。行为主义提醒我们行为对心灵概念至关重要，却没有证明行为穷尽心灵。

三种批评针对不同环节

“超斯巴达人”检验必要性：有疼痛是否一定有相应倾向？“演员”检验充分性：有全部倾向是否就一定痛？“相信下雨还会不会带伞”检验可完成性：能否只用身体动作把信念定义完，而不在条件句里循环诉诸别的信念与欲望？即使行为主义者质疑一项思想实验，另外两项负担仍在。

3.2 心脑同一论

心脑同一论 (*mind-brain identity theory*) 主张心理状态类型与脑状态类型同一，例如疼痛也许就是“疼痛矩阵”中某种活动。它不是由词义分析得到的先验真理，而是类似“水是 H_2O ”“闪电是放电”的后验、综合同一，由科学研究发现。

必须区分两种同一：

- **类型同一论** (*type identity theory*)：每一种心理性质等同于一种神经性质；
- **个例同一论** (*token identity theory*)：每一个具体心理事件等同于某个具体物理事

件，但同类心理事件未必属于同类神经事件。

戴维森的**异常一元论** (*anomalous monism*) 是个例同一论：不存在严格的心物桥接定律，但每个心理事件都是某个物理事件。“异常”指心理描述不受严格定律统摄，“一元”指具体事件最终只有一套。

这里“每一个”与“每一种”的差别可以通过两人头痛来看：张三的这次疼痛等于张三的某次脑事件，李四的这次疼痛等于李四的另一次脑事件，却未说二人的脑事件必须属于同一个精细神经类型。个例同一较易接纳不同生物都能痛，但也失去直接回答“疼痛这种性质是什么”的统一类型公式。异常一元论更须解释：同一个事件在物理描述下服从严格定律，在心理描述下为什么不服从同类定律。

同一论避开行为主义的循环，却仍面对私密性与**多重实现** (*multiple realization*)。如果疼痛等于人类某神经类型，那么神经结构完全不同的章鱼、外星生物或人工系统就不可能疼；但心理状态似乎能够由不同材料实现。这里还要区分：疼痛事实上是否已在不同生物中多重实现，以及它是否可能多重实现。即使经验事实尚有争论，后一个模态直觉仍给类型同一论压力。

反对者提醒，多重实现有时来自分类粒度不对称：把心理状态描述得很粗，却把神经状态描述得极细，自然会得到“一对多”。公平比较需使用相近粒度。然而脑损伤后功能由其他区域接管等现象，仍表明同一种能力可能在同一物种内有不同实现。

3.3 功能主义

功能主义 (*functionalism*) 不问一个状态由什么材料构成，而问它在系统中做什么。凸轮轴可以由不同材料制成，只要它把旋转运动转为往复运动；同样，心理状态由它在认知系统中的**功能角色** (*functional role*) 确定。

最常见的功能角色是因果轮廓：什么输入会造成该状态，该状态与哪些其他心理状态相互作用，又造成什么行为和内部输出。疼痛可能由组织损伤或创伤记忆引起，导致心率上升、退避行为和买止痛药的意图。因而功能主义不是简单行为主义：它明确把其他内部状态纳入角色，呈现整体性心灵。

功能主义常以**实现** (*realization*) 解释心物关系。疼痛是“具有某个承担疼痛角色的状态”这一二阶性质；在人类中由某神经状态实现，在章鱼或机器人中可由其他状态实现。因此它与多重实现天然相容，但严格说来不自动推出物理主义，因为功能角色原则上也可能由非物理状态实现。

“实现”不是“等于某一种材料”

设角色 R 是“受到损伤时出现、促成退避与求助，并改变对受伤部位的判断”。某脑过程 N 实际承担 R ，于是主体拥有“有某状态承担 R ”这一心理性质。 N 是承担者，

R 是说明为何把不同承担者归为同一心理类的角色。若外星生物以硅基状态 S 承担同一个角色，它也可能有相应心理状态；但只列输入与输出仍未回答两者疼起来是否一样。

两组内部区分不可忽略：

1. **长臂 / 短臂功能主义**：长臂观点把外界对象和历史纳入角色，例如由榆树而非橡树引起；短臂观点只考察主体内部的输入输出结构。
2. **先验 / 后验功能主义**：先验观点从民间心理学的常识网络分析角色；后验观点认为角色只能由心理学、认知科学发现，又称**心理功能主义** (*psychofunctionalism*)。

功能主义特别适合信念、欲望和意图，却在现象状态上受挑战。疼痛“是什么感觉”似乎是内在质感，而功能角色是关系性质。若两个系统因果结构相同却经验质感不同，功能主义就无法区分；这一问题将在 Bayne 第八章集中出现。

怎样建造人工心灵？

行为主义者只要求复制正确的行为倾向；同一论者要求制造合适的“人工大脑”；功能主义者要求复制内部因果组织。这个思想实验清楚显示三者标准不同：外显行为、物质构成和因果结构并不是同一件事。

3.4 取消主义

取消主义 (*eliminativism*) 认为某些民间心理学种类根本有所指，正如燃素和以太被成熟科学淘汰。要把它与两种立场分开：**还原论** (*reductionism*) 说 X 存在但就是 Y ；**修正主义** (*revisionism*) 说 X 存在，不过日常关于 X 的许多信念错了；取消主义则说没有真正的 X ，或 X 不是科学上有用的种类。

保罗·丘奇兰以信念、欲望等**命题态度** (*propositional attitudes*) 为对象。他认为民间心理学长期停滞，且其种类无法整合进进化论、生理学和神经科学，因此可能被更成熟理论取代。批评者回应，这要求高层科学种类必须整齐映射到低层种类，标准过强；生物学、心理学与经济学本来就可能形成“斑驳的”跨层联系，而不是逐项同一。

丹尼特以疼痛为例：日常概念似乎同时要求疼痛本身令人厌恶，并要求主体对是否疼痛不会错。吗啡使用者可能说“仍有痛感，但不再难受”，迫使我们放弃至少一个条件。可是更温和的结论是，疼痛包含感觉、情感和动机等可分离成分；这属于修正概念，不等于取消疼痛。

若取消主义扩展到**意向性** (*intentionality*) 和意识本身，它将极其激进。取消信念的人不能只被反问“你是否相信没有信念”，因为他会拒绝用信念描述自己的态度；真正难

题是，若替代状态完全没有内容，它怎样能指向命题“取消主义为真”？彻底取消意向性似乎会连真理、理由和理论主张本身一起动摇。

3.5 科学的角色

四种理论不仅结论不同，方法也不同。分析行为主义与先验功能主义相信扶手椅概念分析能揭示心理本质；早期同一论和后验功能主义把决定权更多交给科学；Lewis-Armstrong 式同一论先由哲学给出功能分析，再由科学找出实现者；取消主义看似“科学优先”，却仍依赖关于心理词怎样指称的语言哲学假设。

原书的比较还有一层：Lewis-Armstrong 式同一论与先验功能主义都先用常识确定“疼痛角色”，却在终点分开。前者说疼痛就是实际担当该角色的那种状态，寻找担当者仍需科学；后者说疼痛就是有某状态担当该角色这一功能性质，其本质在原则上已由角色分析给出。后验功能主义则连究竟什么角色是关键，都留待心理学查明。因而“科学参与多少”不是所有物理主义者都给出同一答案。

理论	心理状态是什么	主要优势	核心困难
行为主义	行为倾向	他心知识、公开标准	私密性、超斯巴达人、循环
类型同一论	脑状态类型	清晰还原、符合科学	多重实现、第一人称经验
功能主义	因果 / 功能角色	材料中立、解释多重实现	现象质感是否超出角色
取消主义	某些日常种类并不存在	容许科学彻底改写常识	指称标准、自我瓦解风险

Bayne 第三章整章总结

行为主义、同一论、功能主义和取消主义都是“物理主义友好”的方案，但它们绝非同义。前两者直接把心理现象认作广义物理现象；功能主义通过物理实现功能角色来说明依赖；取消主义则通过否认某些心理种类来解除还原任务。不同版本在先验分析与经验科学的分工上差异很大，而且可以针对不同心理现象混合使用。复习时应能分别陈述每种理论的本体论主张、核心动机和至少一个针对性反例。

Bayne 第 4 章：知觉 (Perception)

文献定位

本章依次回答三个问题：知觉经验与对象是什么关系？知觉的**现象特征**是否完全由其**表征内容**决定？知觉能直接呈现哪些性质，且知觉是否必然有意识？狐狸案例贯穿全章，使幻觉、错觉、临场感和高阶内容落在同一框架中。

4.1 三种知觉理论

设想你透过窗户看到一个红色、蓬尾的小动物，随后认出它是狐狸。知觉不仅有对象，还把颜色、形状、距离和运动等性质归给对象，因此具有**意向内容** (*intentional content*)：经验把世界呈现为某种样子，并据此具有准确条件。

必须区分三种情形：

- **真知觉**：狐狸存在，并具有经验所呈现的性质；
- **幻觉** (*hallucination*)：根本没有狐狸，却有仿佛看见狐狸的经验；
- **错觉** (*illusion*)：狐狸存在，但因红灯照明等原因被呈现成它并不具有的样子。

内容论 (*content view*) 或**表征论**把三者看成同一基本种类的经验：都有内容，只是准确程度不同。**感觉材料论** (*sense-data theory*) 认为即使幻觉也必有直接对象，故直接对象是私人心理对象“感觉材料”；正常知觉也先直接把握感觉材料，再间接把握狐狸。其困难是感觉材料本身神秘：有几维、能否在无人感知时存在、能否被误知？更重要的是，经验中的狐狸似乎位于公共空间，而不像私人图像。

感觉材料论的推理值得慢读：在幻觉中“似乎看见红狐狸”；幻觉里没有公共世界的红狐狸；若每一次“似乎看见”都必须有一个所见之物，只能引入私有对象。接着把真知觉也分析为首先见到同一种私有对象。这最后一步并非从幻觉事实直接演绎出来：内容论会拒绝“似乎见到 X 必有某个 X 被见到”，朴素实在论则拒绝从坏情形推断好情形的本质。这是三派真正分手的岔口。

内容论回应幻觉时不必创造特殊对象。我们能想到不存在的黄金国，不代表有一个

对象“黄金国”存在；同理，“似乎看见狐狸”不推出有某物被看见。表示不存在对象是思想、欲望和知觉共同面对的问题。

虚拟现实如何检验临场感

Bayne 提出：一段真正沉浸的虚拟现实即使没有被观看的普通狐狸，也能让物体仿佛就在眼前。可改变画面精细度、人在移动时画面随之变化的规律，以及视听触等感官是否一致，再看临场感如何改变。如果低细节画面仍有临场感、而破坏运动恒常性便使其骤减，内容论便有线索：临场感可能来自经验所表征的行动与视角依赖，不必总由真实外在对象构成。这是研究方案，不是一个实验已彻底判定朴素实在论。

第三种理论是**朴素实在论** (*naïve realism*) 或关系论。正常知觉不是与狐狸内容发生关系，而是狐狸本身成为经验的构成部分。这解释了**知觉临场感** (*perceptual presence*)：狐狸不只是被想象或判断存在，而是世界本身向主体显现。

原书这里并非否认真知觉有内容，而是问“仅有内容够不够”。我相信窗外有狐狸、想象一只狐狸、真正看见狐狸，或许都涉及“窗外有狐狸”，却并不像同一种遭遇。朴素实在论以现实狐狸本身进入经验关系来解释这个差别；内容论则可诉诸独特的知觉态度和随视点变化而稳定的知觉恒常性。两派对临场感从何而来的说明不同，而非只有一方承认狐狸存在。

问题在于错觉和幻觉中没有合适对象构成经验。朴素实在论因此通常接受**析取论** (*disjunctivism*)：真知觉与不可分辨的幻觉不是同一种内部状态；“要么真看见，要么幻觉”不能被一个共同心理种类穷尽。内容论则可用特殊的感官态度和**知觉恒常性** (*perceptual constancy*) 解释临场感：我们不仅表征硬币当前投影，还表征移动视角会怎样规律改变外观。虚拟现实也表明，临场感可在普通物体不构成经验时产生。

4.2 意向论与现象特征

知觉意向论 (*perceptual intentionalism*) 主张，知觉的现象特征等同于或至少随附于其意向内容：内容完全相同，感觉方式就不可能不同。反对者认为经验还含有属于经验本身的感觉性质 (*sensational properties*)。

比较“看狐狸”和“听圆号”，可以看出区分的必要：一方面经验告诉我们世界是什么样，因而有可真可假的准确条件；另一方面这次经验如何感觉又不能靠说“它是真的”说明。意向论既不是说经验没有质感，也不是说所有经验都准确；它给出的是一个依赖主张：只要正确说明经验表征了什么、以什么方式表征，就没有剩余的独立质感差异。

Block 区分两类感觉性质。**心理颜料** (*mental paint*) 是经验自身的性质，但被系统用来表征外界；**心理乳胶** (*mental latex*) 则完全没有表征功能。意向论否认两者，感觉论者至少接受其一。

疼痛显示争论不限于视觉。疼痛可被看作表征身体损伤的陈述式状态：幻肢痛便是不准确表征；也可被看作向主体发出“不要给脚踝负重”的**命令式内容** (*imperative content*)。问题不仅是疼痛有无内容，还包括它的内容是描述世界还是指导行动。

4.3 支持和反对意向论

最著名支持是**透明性论证** (*transparency argument*)。凝视一棵树并转向经验本身时，我们仍只注意到树的绿、形状和位置，而找不到经验自身的“内在颜色”。论证为：

透明性论证

1. 内省知觉时，我们只觉察对象所呈现的性质，未觉察经验自身的感觉性质。
2. 若感觉性质存在，内省理应把它们揭示为经验的内在性质。
3. 因而没有理由承认感觉性质，现象特征可由内容解释。

模糊视觉不是轻易的反例：摘下眼镜时，物体不被表征为自身变模糊，而是被以较低精度表征。反对者也可拒绝第二前提：心理颜料恰恰通过“透明地”指示外界发挥作用，所以存在却不会以内在性质的身份出现。

把摘眼镜前后比作高、低分辨率照片容易把握要点：我们没有说树突然长出“模糊”这一物理性质，而是经验所呈现的边缘、细节与空间关系不再明确。意向论仍可将精度差异纳入内容。感觉论者则可接受我们始终看到“树的颜色”而非“脑内一块颜色”，同时主张经验里使用了有表征作用的感觉载体。透明性一方面确实反对毫无表征作用的“心理乳胶”，另一方面未必排除有作用的“心理颜料”。

两个反对意向论的案例更棘手。第一是**共同可感性质** (*common sensibles*)：同一枚圆硬币可以被看成圆，也可以被摸成圆，内容似乎相同但感觉不同。回应是，两种经验整体内容不同：视觉把形状与颜色结合，触觉与压力结合；精度不同；并各自表征眼、头、手指的运动。

第二是**颠倒光谱** (*inverted spectrum*)。Bonnie 看红物的体验像 Clyde 看绿物的体验，但二人的分类与言语完全相同。若内容相同而质感不同，意向论失败。意向论者可把颜色解释为“在正常条件下使某类观察者产生某种体验”的倾向性质，使二人表征不同颜色；但“某种体验”似乎又必须用非意向的现象性质解释，可能循环。争论因此触及第一性质与第二性质：颜色、气味、味道究竟在对象中，还是依赖感知者？

4.4 知觉允许什么内容

保守主义 (*conservatism*) 认为视觉直接呈现的只有颜色、形状、纹理和运动等低阶性质；**自由主义** (*liberalism*) 允许更丰富内容，例如“是一棵松树”“正在打碎杯子”“脸上有喜悦”“苹果可食用”甚至“情境不公正”。这就是**可容许内容争论** (*admissible contents debate*)。

不要从日常“看起来像松树”直接推出知觉含有松树性质，因为“看起来”可能只是比较用法。Siegel 的**现象对比法** (*phenomenal contrast method*) 更有力量：Kylie 成为树艺师前后看同一松树，经验显然改变；最佳解释也许是她现在直接看到“松树性”。保守派有两条回应：

- 高路线：变化来自新的判断倾向，而不是经验内容；
- 低路线：训练改变了对针叶形状、树皮纹理等低阶特征的注意与敏感性。

高路线难解释判断能力为何改变；低路线较可信，却未必穷尽专家“看出它是松树”的现象变化。争论关系到知觉是否为高阶信念直接提供证据，也关系到概念能否渗透知觉。

速度表：判断很快不等于直接看见车速

若指针指向每小时 100 公里，熟悉仪表的驾驶者可以立即说“我看到正在开到 100 公里”。但视觉也许只直接呈现了指针、刻度与数字，车速是结合“仪表正常工作”的背景信念推知的；即使推理自动、毫不费力，仍未因此成为视觉内容。若驾驶者知道表坏了，他仍会看见指针指向 100，却不再据此相信车真以 100 行驶。这给保守派提供一个清晰的“经验与判断”分离测试。

对比法的推论距离

对比法先建立“学会识别松树前后，主观上确有不同”，再问不同位于哪一层。从有现象差异到“松树性成为视觉内容”，还须排除同时发生的注意改变、低阶感官分辨改变、判断与联想改变。反之，保守派也不能简单说“只有判断改变”，而须具体解释为何学习带来的变化看起来像一次不同的观看。

4.5 无意识知觉？

盲视 (*blindsight*) 患者声称在盲区看不见刺激，却能在强制选择中高于随机地辨别线条方向。**连续闪烁抑制** (*continuous flash suppression, CFS*) 则向一只眼呈现快速变化图案、另一只眼呈现目标，目标不进入意识却能影响注意。它们是否证明无意识知觉？必须同时满足：主体确实无意识，以及主体层面确实在知觉。

盲视的“无意识条件”有争议：使用“完全没看到—微弱一瞥—几乎清楚—清楚图像”的四级量表后，报告与辨别能力的分离会缩小，患者可能有低清晰度经验。CFS 较可能满足无意识条件，却未必满足知觉条件：视觉系统加工信息，不等于主体看见。若主体层面状态应能为判断和行动提供理由，那么仅在不自知情况下改变注意的表征也许仍属次人格加工。

原书的 CFS 例子更具体：一只眼看到快速变化的高对比度图案，另一只眼被呈现目标图像；参与者说只看到前者，对目标作二选一猜测也可能接近随机，但目标类别仍在不知不觉中影响注意的方向。于是证据并非“他暗中看见了，却拒绝承认”，而是“信息在系统内有效，但是否成为他看到的内容尚有争论”。若把一切视觉加工都叫知觉，结论很容易成立，却换掉了原本关于主体知觉的问题。

Bayne 第四章整章总结

1. 内容论以准确条件统一真知觉、错觉与幻觉；感觉材料论引入私人对象；朴素实在论让外界对象构成正常经验。
2. 析取论拒绝正常知觉与幻觉共享基本心理种类；临场感是三种理论竞争的关键测试。
3. 意向论说现象特征由内容决定；透明性支持它，共同可感性质与颠倒光谱挑战它。
4. 可容许内容争论考察知觉能否直接呈现种类、因果、心理、行动和道德性质。
5. 盲视与 CFS 迫使我们区分视觉系统加工、主体层面知觉与意识知觉。

Bayne 第 5 章：思想 (Thought)

文献定位

本章先刻画思想的核心特征，再解释“思想语言假说”如何用内部符号说明思想的开放性。随后比较联结主义与自然语言支架，审查中文房间论证，最后以图灵测试讨论人工系统何时算真正思想者。

5.1 思想的种类与特征

Bayne 不追求思想的严格充要条件，而提出三个中心特征。思想是**刺激独立的** (*stimulus-independent*)：要看狐狸需与它保持因果接触，要想狐狸则不必；思想是**抽象的** (*abstract*) 和非模态的，能涉及进化史、数字、虚构对象及遥远时空；思想是**开放式的** (*open-ended*)，能整合多个感官、背景信念和目标，形成新结论与计划。

思想与知觉的边界未必锐利。玻璃观景台上的人相信平台安全，却仍恐惧退缩。Gendler 称这种状态为 **alief** (拟信念 / 联想式接受)：它表征危险，引发焦虑和退避，却不受反思信念直接控制，也不能自由进入推理。无论是否接受这一新种类，案例都说明行为引导不只来自理性信念。

非人动物也可能思想。狒狒能追踪家族和个体的社会等级：这种表征超出当前刺激，涉及不可直接感知的“高低地位”，并能处理许多关系，因而满足三项标准。人类思想的独特处不是有无思想，而是系统性和开放程度。

依照三项特征检查狒狒：它不必让竞争者此刻出现在视线内，便可以记住对方所属群体（脱离当下刺激）；“谁在社会等级中更高”不能直接像毛色一样被看到（抽象关系）；它可据多个个体和群体之间的关系选择结盟或退避（整合与开放）。这不证明狒狒具有人类的所有概念，而是解释 Bayne 为何拒绝简单地用“会不会说话”当思想有无的界线。

思想序列可由联想连接，也可由逻辑与证据连接。系统 1 式推理快速、自动、直觉；系统 2 式推理缓慢、可控、反思。球拍和球共 1.10 1 时，直觉答 10 美分，但正确答案是 5 美分。即使知道答案，错误冲动仍存在，类似知觉错觉，显示自动与反思加工可分离。

5.2 思想语言假说

思想语言假说 (*language of thought hypothesis, LOT*) 认为, 思想是内部语言“心语”(Mentalese) 的句子, 思考是受句法规则控制的内部句子转换。符号既有物理系统可识别的**句法性质** (*syntactic properties*), 也有意义或指称的**语义性质** (*semantic properties*)。

思想“狐狸吃了鸡”由分别表示狐狸、吃、鸡的原子成分组合而成; 推理“所有狐狸会死; Martha 是狐狸; 所以 Martha 会死”可由只对句法结构敏感的机制完成。机器无需先理解内容再操作符号; 若转换规则保真, 就能“照管句法, 让语义自行得到照管”。LOT 描述的是思想机制, 不是思想在内省中像什么。

“按句法处理”不是说心灵没有意义, 而是把机制层与解释层接上: 推理器可根据“所有 F 是 G , a 是 F ”的形式输出“ a 是 G ”; 由于这种形式对应有效推理, 它不必在每一步再询问句子的含义。可是符号最初为什么表示狐狸而不表示另一物, 以及脑中符号是否真有合适的组合结构, 仍是思想语言假说应继续解释的问题。

LOT 可被理解为一切思想的必要本质, 也可仅是关于人类认知架构的经验假说。Fodor 主要诉诸两项证据:

- **生产性** (*productivity*): 我们能形成从未想过的无限多新思想, 如“袋鼠身上不长草”;
- **系统性** (*systematicity*): 能想“Mary 爱 John”的主体通常也能想“John 爱 Mary”, 能从 $P \wedge Q$ 推出 P 也能推出 Q 。

语言通过有限词汇和组合规则解释同类能力, 因而内部表征很可能也有组合结构。不过组合表征不必是狭义句子, 也可以是地图或图结构; “思想语言”有宽、窄两种用法。

5.3 思想语言的替代理论

联结主义 (*connectionism*) 使用多层节点和带权连接; 学习改变权重, 处理是分布式、并行的。网络整体可表示“猫坐在垫子上”, 却未必有某个离散部件单独表示“猫”。这与经典 LOT 的局部、串行、可分解符号不同。

Fodor 的系统性两难说: 若联结网络解释系统性, 它必在更高层实现 LOT; 若不实现 LOT, 就无法保证系统性。回应是, 任何思想架构也许只需容许系统性, 而不必由本质保证它。动物和婴儿有思想却未必系统; 因此 LOT 也得借助学习和发展解释成熟人类为何系统。

这场争论不能仅靠指认“网络里没有一句明确的‘猫坐在垫子上’”结束。联结主义者须证明, 分布于节点及连接权重的表示怎样支持角色互换、组合迁移与新情形推理; 经典主义者则须说明, 拥有同一组词和句法能力为何实际就会把所有结构相近的思想都

学会。若网络在较抽象层面实现有组合结构的表征，两派争议可能部分是对同一个系统采用不同描述层次。

自然语言可能是关键支架。人不一定逐句用汉语或英语思考，但学习“所有”“如果……那么”“因为”等表达会塑造推理能力。外部符号还能把高阶问题转成可操作的一阶问题：学会用不同标签表示“相同 / 不同”的黑猩猩，随后能判断两组关系本身相同还是不同。纸笔、算盘、电脑、教师和对话也构成**认知支架** (*cognitive scaffolding*)。成熟思想可能分布在脑、语言、工具与社会互动之间。

这不是在“有内部心语”与“只用日常语言”之间硬选一个。一个儿童可能早就会对物体形成简单想法，学到公开的“相同”“不同”后，却能把关系本身当作可再比较的对象。前一种思想不靠成句语言，后一种新能力又确实有赖语言教学。外部纸笔也是类似情形：它不必作为独立的思考者，却能改变一个人能够稳定保持、重新排列和检验的思想复杂度。

5.4 中文房间论证

Searle 设想不懂中文的人被关在房间里，按英文规则匹配输入、输出汉字符号，外部表现与中文使用者一样。房中人只操作形式，不懂意义。因此：

中文房间论证

1. 中文房间确实进行形式符号操作。
2. 中文房间没有中文理解或思想。
3. 所以，符号操作不足以产生理解或思想。

系统回应 (*systems reply*) 拒绝第二前提：房中人不懂中文，不推出“人+规则+记忆+处理系统”整体不懂。会思想的人也由不会思想的神经元构成，从部分没有性质推到整体没有性质是合成谬误。即使房中人把规则全背下，个人作为英语思考者和他所实现的中文系统仍可在功能层级上区分。

按原书步骤，关键不是“房中人诚实地报告自己不懂中文”这件事是否可信，而是这一报告指向谁。若能力归属对象是整套规则、记忆与输入输出系统，个人没有直接听懂问句，并不能推出系统没有理解；若把全套规则都记在脑中，仍可问这个脑内是否实现一个不同于作为英语使用者的中文处理系统。系统回应不自动证明它懂中文，只是切断从“部分不懂”到“整体不懂”的推理。

机器人回应 (*robot reply*) 质疑第一前提是否涉及真正符号。像汉字的墨迹不自动具有意义；若把房间装进能感知、行动的机器人，使内部状态由电话引起并指导找电话，它们才与世界建立语义联系。Searle 说这已超出纯符号操作，但 LOT 从未必须否认符号

的内容依赖外界因果关系。中文房间直觉很强，却没有决定性证明整个具身系统无理解。

5.5 图灵测试与思想边界

图灵测试 (*Turing test*) 把“机器能否思想”替换为“机器能否在开放问答中使审问者无法稳定区分它与人”。它重视广泛的会话能力，却未规定测试时长、审问者知识等全部参数。

测试有吸引力，是因为我们通常也从复杂语言、适应情境、解释理由和持续对话判断别人有思想；若机器能在足够开放的条件下做到这些，断然拒绝它会显得标准不一致。但把人类会话作为唯一门槛也可疑：会思考的幼儿、不会说话的人以及社群性的动物可能过不了测试。Bayne 因此将其作为检验思想归属的一个指标，而不是把“思想”重新定义为“模仿成人聊天”。

Block 的 **Blockhead** (用) 穷尽式查找表储存每个问题的答案，理论上能过测试却似乎不思想。若图灵测试是思想的定义，这个逻辑可能性足以反驳；若它只是思想的经验指标，则关键是这种表实际能否建成。20 词句子的可能数量极其庞大，实时查表在现实中不可行。

“Lovelace 反对”说机器只能做程序员让它做的事，不能原创。但自我学习系统会发展设计者未预见的策略；AlphaZero 的围棋风格正说明“由人制造”不等于“每步由人预写”。另一个质疑是机器可能无意识地通过测试；目前我们既不知道思想是否必然意识，也不知道实际通过开放测试是否需要意识。

图灵测试最大的限制是单向性：通过也许是有思想的强指标，未通过却不能证明无思想。要求异质人工系统像人类会话，好比要求所有能飞者都像海鸥。更一般的问题是：一个系统只要能以刺激独立、抽象、开放方式表征环境就算思想者，还是还必须有意意识、内在语言或主体性？

Bayne 第五章整章总结

思想以刺激独立、抽象和开放为核心；LOT 用组合符号解释生产性与系统性，但联结网络、自然语言和外部支架提供替代或补充。中文房间显示句法与语义的张力，却未排除整个具身系统理解。图灵测试是有用指标而非无误定义，且失败不能推出无思想。人工思想问题最终迫使我们说明：思想的关键在行为、内部结构、世界联系、意识，还是这些因素的组合。

Bayne 第 8 章：意识的形而上学 (The Metaphysics of Consciousness)

文献定位

物理主义最困难的考验是意识。蝙蝠、黑白房间里的神经科学家 Mary 和哲学僵尸分别从视角、知识与可设想性制造压力。Bayne 把三者统一成**先验可推知性论证**，再比较 A 型物理主义、B 型物理主义、幻觉论、二元论和罗素式一元论。

8.1 蝙蝠、神经科学家与僵尸

8.1.1 蝙蝠论证

Nagel 问“成为一只蝙蝠是什么样？”我们可以想象人类用回声定位，却未必把握蝙蝠自身的经验视角。关于神经生理、行为和声波处理的客观知识似乎不会自动给出主观质感。论证的关键并非我们今天资料不足，而是现象事实与特定观点本质相关，而物理理论力求脱离观点。

蝙蝠论证

1. 一种生物的经验现象只能从该类生物的观点把握。
2. 若物理主义为真，完整客观物理说明原则上应穷尽这些现象事实。
3. 但完整客观说明仍不提供蝙蝠观点。
4. 所以物理主义至少是当前不可理解的，甚至可能为假。

Nagel 本人较谨慎：他没有简单宣布物理主义为假，而说我们尚无概念理解它如何为真。

从“我无法想象”到“物理主义为假”

先要区分三种欠缺：一个普通人不懂蝙蝠神经科学；即使学完神经科学，人类仍不能亲历蝙蝠回声定位；即使掌握一切物理事实，原则上也不能推出蝙蝠体验。只有第三种欠缺直接威胁物理主义。Nagel 强调的是我们形成经验概念受自身视角限制，因此原书既认真对待挑战，也保留他“尚不知如何理解物理主义”的谨慎语气。

8.1.2 Mary 与知识论证

未来神经科学家 Mary 知道颜色视觉的一切物理事实，却终生生活在黑白房。第一次看见红色时，她似乎学到“看红是什么感觉”。若她此前已知全部物理事实而仍学到新事实，则现象事实不是物理事实。

知识论证 (knowledge argument)

1. Mary 第一次看红时获得新知识。
2. 若物理主义为真，她在房中掌握全部物理事实后不应有新事实可学。
3. 因而物理主义为假。

8.1.3 哲学僵尸

哲学僵尸 (*philosophical zombie*) 与我们在物理、功能和行为上完全相同，却完全没有经验。物理主义认为物理事实形而上学地必然决定意识事实，故僵尸不可能。反对者却说僵尸可设想，并由“可设想 $\rightarrow$ 形而上学可能”推出其可能，从而反驳物理主义。

这里最脆弱的是从可设想到可能的桥梁。水不是 H_2O 看似可设想，但若水本质上就是 H_2O ，我们真正设想的只是“外观像水的非水物质”。物理主义者会问：僵尸是否也只是由于概念无知才显得可设想？

8.1.4 解释鸿沟与困难问题

三个论证共享**解释鸿沟** (*explanatory gap*) 直觉：从物理 / 功能状态到意识质感没有深层可理解的解释连接。还应区分两种鸿沟：

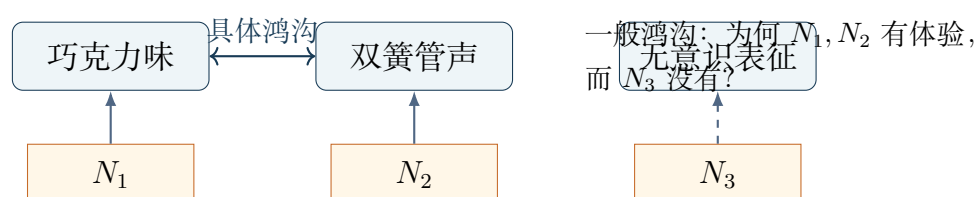


图 8.1 具体解释鸿沟与一般解释鸿沟。

具体鸿沟问为何某物理状态对应巧克力味而不是双簧管声；一般鸿沟问为何某些物理状态有任何体验而其他状态没有。Chalmers 把后者称为意识的**困难问题** (*hard problem*)；识别、整合信息和报告内部状态虽技术上很难，却至少已有研究方法，故被称为“容易问题”。

可用**先验可推知性** (*a priori scrutability*) 统一挑战：若全知且算力无限的拉普拉斯妖仅凭全部物理真理仍不能推出现象真理，物理主义为何说后者不超出前者？

不可推知性论证

1. 若物理主义为真，现象真理可从完整物理真理先验推出。(桥接前提)
2. 现象真理不能从完整物理真理先验推出。(不可推知性前提)
3. 因而物理主义为假。

A 型物理主义 (*Type-A physicalism*) 否认第二前提；**B 型物理主义** (*Type-B physicalism*) 否认第一前提。

把论证写成这两条前提有助于避免“物理主义者都否认 Mary 学到了东西”的误读。A 型认为足够完整的物理说明最终可使现象真理可推知，或认为所谓所得不是新事实；B 型可以承认 Mary 凭物理知识推不出亲历体验，却坚持本体上的同一不要求先验可推出。两派真正的分歧是物理事实确定现象事实是否必须伴随用理性从前者看出后者。

8.2 A 型物理主义与不可推知性前提

第一种策略是**能力假说** (*ability hypothesis*)：Mary 获得的不是新的“知道某事实”，而是识别、想象、记忆红色经验的“知道如何”。这样她有所得，却没有发现非物理事实。困难是现象知识能支持条件句与新问题，如“草莓是看起来像这样还是那样？”Marianna 先只看抽象色块，已经知道红的感觉，却仍不知道草莓是什么颜色；这暗示新能力之外仍有命题知识。

Marianna 案例须分两次学习：第一次见抽象红色色块时，她获取“红看起来如此”的体验；第二次见真正草莓时，才将这体验与“草莓”对象联系起来。若第一次之后她已经能够辨别自己尚不知草莓应归哪种色块，便似乎拥有可以进入疑问与条件推理的内容。能力假说要说明这种新提问能力为何仍只是纯粹技能；其反对者也得说明新内容为什么一定是一个额外的非物理事实。

第二种策略质疑思想实验中的想象可靠性。我们并不知道完整神经科学是什么，更不是拉普拉斯妖，凭什么确信 Mary 无法在房中推知红色体验？科学曾把“生命如何从无生命物质产生”的神秘化解，意识也可能如此。这是**乐观归纳** (*optimistic induction*)。

反驳是生命概念大体由结构与功能刻画，意识概念却恰恰包含第一人称质感，两种解释任务未必同类。

神秘主义 (*mysterianism*) 反向使用认知局限：意识是完全自然的现象，但人类大脑也许像狐狸无法掌握量子力学那样，缺少解决心身问题的能力。它不是超自然论，而是关于我们的认识边界。

第三种策略是**意向论** (*intentionalism*)：不要只看脑组织，而要看神经活动如何表征世界。若蝙蝠经验的质感由其回声定位系统能探测什么、如何引导行动来决定，完整表征理论也许能关闭具体鸿沟。但即使说明“为何感觉是这个而不是那个”，仍需解释“为何这些表征会有任何感觉”，所以一般鸿沟可能保留。

8.3 B 型物理主义与桥接前提

B 型物理主义承认物理概念与现象概念之间存在不可消除的认识鸿沟，却否认这推出两个本体领域。第一种路线指出，高层科学事实一般也未必能从基本物理事实扶手椅式推出；若“水是 H_2O ”等同一需要经验发现，意识—脑同一也可以是**后验必然真理** (*a posteriori necessity*)。

Kripke 的反驳是，水非 H_2O 的表面可设想性可解释为“水样物不是水”；疼痛非 C 纤维放电则难作同样解释，因为疼痛似乎直接由其感觉本身指称。意识的后验同一因此比普通科学同一更难消除偶然性外观。

请把两个思想实验并排：在另一个世界，透明液体拥有我们这里水的外观与用途，却由别的分子组成；我们会说那里看似有水，实际上没有我们这里的水。而若一个世界有完全一样的痛感却没有相应脑态，似乎很难再说“只是看似痛，实际没有痛”。B 型回应不能只重复“水是 H_2O 也要经验发现”，还须针对这一不对称性解释为什么意识的外观与实在仍可区分。

第二种路线是**现象概念策略** (*phenomenal concept strategy, PCS*)。同一事实可通过不同呈现方式知道：知道 Michael Caine 主演某片，不代表知道 Maurice Micklewhite Jr. 主演，虽然二者是同一个人。房中的 Mary 用物理 / 功能概念知道红色经验；出房后获得只能由亲历形成的现象概念。她获得“新知识”，但对象仍是“旧事实”。

PCS 面临两问。其一，若第一人称现象概念永不误用，它还是受正确性标准约束的概念吗？其二，Mary 至少学到“这个现象概念与那个物理概念共指”这一新事实；若确有新事实，策略便没有完全满足物理事实穷尽论。

8.4 幻觉论

幻觉论 (*illusionism*) 不试图填平鸿沟，而试图解释为何我们错误地觉得有不可化约的现象性质。魔术师仿佛从空帽取兔，解释任务不是说明兔子怎样从空无出现，而是说明它为何看起来出现；同样，大脑也许只制造“经验具有私密、内在、不可言说性质”的表象。

多数幻觉论者不否认我们有意识、看见颜色或闻到气味，而是否认经验具有被直接亲知的内在 *qualia*。Dennett 的咖啡品尝师 Chase 和 Sanborn 都承认体验发生变化：前者说味道没变而喜好变了，后者说喜好没变而味觉变了。若二人的一切辨别、情感和反应相同，是否还有事实决定究竟“质感”变了还是评价变了？案例削弱质感既内在又直接可知的假设。

这例子的力点不是证明喝咖啡“毫无感觉”，而是看在最有利于自我内省的场合，当事人能否凭一个私密、独立于一切辨别与反应的感觉核心裁定变化原因。如果不能，传统 *qualia* 图景中的“内在且无需推断即可完全掌握”便可疑；可疑的仍只是那种说明经验的方式，不是咖啡不带来味觉经验。

最强反驳是：舞台魔术可区分表象与实在，意识却似乎就是事情如何显现。“似乎有疼痛”本身难道不已经是一种经验？因此幻觉论不能只宣布 *qualia* 不存在，必须具体说明大脑怎样生成这些顽固直觉。

8.5 二元论与罗素式一元论

当代**性质二元论** (*property dualism*) 通常不承认非物质心灵实体，只说现象性质非物理，心物联系由基本心理—物理定律规定。它绕开物理还原，却把意识变成 Feigl 所谓“法则上的悬垂物”：其他高层现象都整合进物理世界，意识却由一组无进一步解释的基本相关关系挂在其上。

罗素式一元论 (*Russellian monism*) 提出，物理学只告诉我们基本实体做什么，即关系和倾向结构；意识或原现象性质则是物质的内在、范畴本性。现实只有一种底层性质，却有第一人称现象面与第三人称物理面，故又称双面一元论。

它避免“意识从完全非意识物质突然涌现”，也不需外加心物定律，但自然导向**泛心论** (*panpsychism*)：电子、夸克也有微小经验。较温和的**泛原心论** (*panprotopsychism*) 只赋予既非物理也非现象的原现象性质，却难以积极说明这种性质是什么，并可能只是把鸿沟搬到原现象—现象之间。

更严重的是**组合问题** (*combination problem*)：微观主体如何组成宏观主体？微小经验怎样必然产生统一的人类经验？目前缺乏可检验或概念上充分的组合模型。

三种“填补鸿沟”的成本不能混淆

性质二元论承认鸿沟反映世界里另外的性质，却须说明原始心物定律与因果效力；罗素式一元论将意识放进物质的内在本性，却须解决微观到宏观的组合；幻觉论拒绝原题中“不可化约 qualia”的存在，却须解释为何我们如此确信它存在。它们不是对同一个前提作三个同义回答，而是分别修改世界里有什么、物质的内在面是什么与我们对自己经验的表象是否可靠。

立场	对解释鸿沟的诊断	主要代价
A 型物理主义	鸿沟原则上可关闭，或源于我们无知	可能低估第一人称直觉
B 型物理主义	认识鸿沟不推出本体鸿沟	需解释为何意识鸿沟如此特殊
幻觉论	应解释鸿沟直觉，而非填平不存在的 qualia	“意识只是幻觉”疑似自我挫败
二元论	鸿沟反映真实的两类性质	原始心物定律、因果问题
罗素式一元论	意识是物质内在本性	泛心论直觉与组合问题

Bayne 第八章整章总结

蝙蝠、Mary 和僵尸把同一疑问从不同方向表达出来：完整第三人称物理说明似乎仍遗漏“是什么感觉”。A 型物理主义挑战遗漏直觉，B 型物理主义把遗漏放在概念而非世界中，幻觉论试图解释这种遗漏感；二元论与罗素式一元论则修改本体论。每条路线都能保住一部分直觉，却分别面对过度乐观、概念鸿沟、表象与实在难分、悬垂法则或组合问题。

Bayne 第 10 章：心灵因果性 (Mental Causation)

文献定位

我们以信念、欲望和决定解释行动，并自然认为这些状态造成行动。但如果每个身体运动已有充分物理原因，心理原因还做什么？本章先建立心灵因果的动机，再分析 Kim 的因果排斥、语义外在主义与 Libet 实验三种挑战。

10.1 为什么需要心灵因果

举起左手或右手时，你也许体验到考虑、决定和行动之间的联系。不过内省是否直接呈现“心理状态造成动作”仍有争议：也许经验只呈现动作实现了意图，或呈现“我”作为行动来源，而不是决定作为一个事件推动身体。

更稳固的动机来自解释。你去埃菲尔铁塔，是因为相信约定地点在那里并想见朋友；这里的“因为”似乎同时给出理由与原因。若心理状态完全无因果效力，行动者控制和心理解释都受威胁。

原书先把“我感到自己发动动作”的第一人称证据与“信念、欲望在解释中不可少”的第三人称证据分开。内省也许只报告时间先后或符合意图，单凭它未必看见因果纽带；但心理解释能预测并区分行动：想见朋友而相信朋友在巴黎，和想见朋友而相信朋友在罗马，通常引出不同路线。若“信念内容”在这个区别里从不产生作用，解释看似成功就需要另有交代。

所需的不是心理事件以某种物理性质造成结果，而是作为心理状态发挥作用。高音震碎玻璃，是由于音量而非歌词意义；同理，一个决定造成抬手，应当是由于它具有“决定抬手”这一心理内容。Bayne 称此要求为**作为心灵的因果作用** (*mental quausation*)。

还要区分两种因果观：**生产论** (*productive account*) 把因果理解为力、能量或能力的传递；**差异制造论** (*difference-making account*) 把因果理解为反事实依赖：若改变原因，结果会怎样改变。这一区分决定后面对排斥论证的回应是否成立。

10.2 因果排斥反对

令 M 为决定抬手的心理状态， P 为实现它的脑状态， E 为手臂运动。若 P 已充分造成 E ， M 似乎成为多余竞争者。

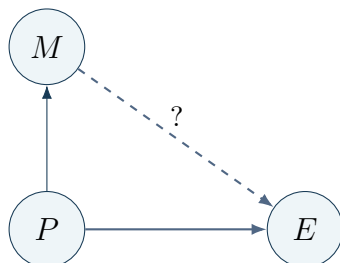


图 10.1 因果竞争：P 已足够时，M 还有什么工作？

Kim 式因果排斥论证

1. P 对 E 因果充分。(完整性 / 封闭性)
2. 若 P 充分造成 E ，不同于 P 的东西不能再造成 E ，除非 E 被过度决定。(排斥)
3. M 不同于 P 。(非还原物理主义)
4. E 不是由两个独立充分原因过度决定。(非过度决定)
5. 所以 M 不造成 E 。

互动二元论可拒绝第一前提，在物理链中留下心理缺口；但这与科学通常的因果封闭承诺紧张。也可说 M 与 P 像两个同时击中受害者的子弹一样共同充分，但日常行动若总是巧合式过度决定十分不可信。类型同一论拒绝第三前提： $M = P$ ，正如闪电与放电不是两个竞争原因。然而多重实现又给严格类型同一压力。

个例同一论说每个具体心理事件都是物理事件，能保住事件因果，却未说明它因心理性质而有效，也难解释“羞辱比损失更易造成抑郁”这类心理性质层面的因果概括。

排斥论证的对象是哪一种“两个原因”

若 M 只是同一个事件 P 的不同名称，就没有两颗子弹式的两个事件。但非还原论者主张的心理性质仍可能不同于神经性质；仅说“事件是一个”未回答动作为何恰好由决定抬手这一内容而非事件的某个偶然物理细节造成。反对意见因此可在事件层被化解，却在性质层继续出现。

排斥论证还会普遍化：若脑状态排斥心理状态，化学状态也可排斥生物状态，微观物理状态又排斥神经状态，最终所有高层因果都消失。既然特殊科学充满可靠因果解释，论证很可能在某个一般原则上过强。

10.3 非还原物理主义的回应

Jackson 与 Pettit 的**程序解释** (*program explanation*) 说，玻璃的脆性不直接打碎玻璃，却保证某种分子结构存在并在受击时造成破裂。类似地，决定也许不直接造成动作，却“编程”一个会造成动作的神经状态。这保住心理解释的因果相关性，却承认心理性质不真正有效，只是退守方案。

Yablo 的红旗反例更积极。公牛会向任何红色色调冲锋；眼前旗子恰为猩红。猩红足以保证红，却不是最合比例的原因，因为换成深红仍会冲锋；“红”才是改变后会改变结果的差异制造者。

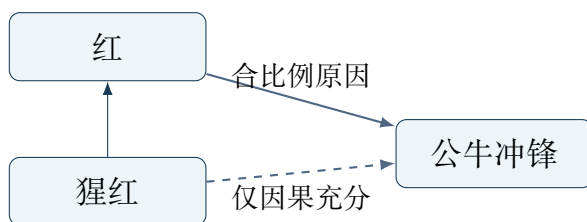


图 10.2 因果充分不等于最合比例的原因。

若因果是差异制造，神经实现 P 虽对动作充分，却可能太具体；决定 M 能由多种神经状态实现，干预决定比只干预某个实现者更稳定地改变动作。因此排斥前提并非普遍真理。这个回应只证明心理性质可以胜过物理性质，并未证明每个心理解释实际都如此。

公牛例须保留原书的条件性：如果它只对猩红冲锋，应选“猩红”作原因；如果其他红色也同样使它冲锋，应选“红”。这不是宣布高层原因永远更好，而是要求原因的细致程度恰好对应结果实际依赖的区别。将之移用于心理状态也必须做同类检验：抬手究竟依赖某一种精细神经实现，还是依赖可由不同神经实现承担的意图角色？

10.4 外在主义与心灵因果

内容外在主义 (*content externalism*) 说心理内容部分取决于环境、历史和社会语言，而非全在头脑中。问题是，原因的力量似乎只依赖对象的内在性质：真假硬币若质量形状相同，都能启动售货机。因此：

外在主义挑战

1. 心理内容不是纯内在性质。(外在主义)
2. 行为只由主体内在状态造成。(局部性)
3. 所以心理状态不能因其宽内容造成行为。

一种退守是区分宽、窄内容：日常心理归属使用宽内容，真正因果只靠内部窄内容。Dretske 则区分**触发原因** (*triggering cause*) 与**结构原因** (*structuring cause*)：移动鼠标触发光标移动，系统布线解释为何该动作有该效果。环境塑造神经系统，故宽内容可作为结构原因。不过决定通常被理解为当下动作的近因，而不是只解释能力如何形成，这个方案可能给得太少。

更直接的回应是否认局部性。晒伤包含“由紫外线造成”的关系历史，却显然会造成疼痛、呻吟和就医。批评者可说真正原因只是皮肤损伤，而非其晒伤历史；但在 Twin Earth 案例中，若宽状态比共享的内部状态更准确追踪行为差异，差异制造论仍可把关系性质当原因。争论再次取决于因果究竟是局部能量生产，还是反事实差异制造。

10.5 Libet 的挑战

Libet 让参与者在约 20 秒内随时抬手，同时看计时器并报告最初意识到意图 / 冲动的时刻 (*W* 判断)。脑电记录显示**准备电位** (*readiness potential, RP*) 平均在动作前约 500 毫秒开始，而 *W* 判断在动作前约 200 毫秒，*RP* 领先意识决定约 350 毫秒。

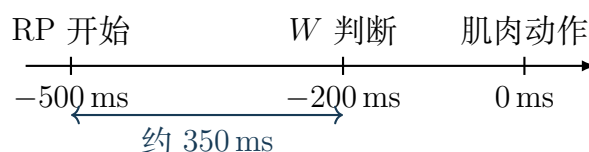


图 10.3 Libet 实验的典型时间关系。

由此产生**意志幻觉论** (*willusionism*)：RP 已造成动作，意识决定只是副现象。论证需要四步：RP 造成 Libet 动作；若如此，决定不造成动作；这类任意动作代表一般行动；所以决定通常无效。

仅仅“RP 较早”本身不能推出以上任何一个完整结论。要从较早的脑电波走到心理决定无因果性，至少还得证明 RP 不仅预示而且造成动作，并且排除早期脑过程与较晚的决定共同参与动作的可能。再从“实验室里何时随意抬手”走到“是否为了履行约定而举手”，还要证明这两种行为在与理由、计划和控制有关的方面足够相似。Bayne 因而把实验结果与对结果的哲学解读逐项拆开。

后两步都可质疑。Libet 动作被刻意剥离理由，而真实决定通常比较利益、目标和规范；涉及向机构捐款等理由选择时，神经轮廓不同且可能没有同类 RP。参与者也可能只在开始实验时作出有意识决定，随后把具体时机交给自动过程；或者体验的是“我造成抬手”，而不是“一个决定事件推动手”。

RP 早于报告也不证明 RP 已决定结果。原因本身可以有原因；即使 RP 导致决定，决定仍可能继续导致动作。更重要的是，单次 RP 被神经噪声淹没，研究通过以动作时

点对齐并反向平均许多试次识别它；若有 RP 但最终没有动作的试次，它们不会出现在这种平均中。因此 RP 是否对动作充分，数据本身并未确定。

Bayne 第十章整章总结

1. 心灵因果要求心理状态作为心理状态发挥作用，而不只与某个物理原因偶然同一。
2. 因果排斥把封闭性、排斥、心理—物理不同一与非过度决定结合，直接威胁非还原物理主义。
3. 类型同一论、程序解释和差异制造论分别提供还原、退守与正面回应；Yablo 指出“充分实现者”未必是最佳比例的原因。
4. 外在主义挑战宽内容的因果效力；结构原因、关系性质和差异制造提供有限回应。
5. Libet 实验的时间差真实而重要，却不能直接从任意抬手推广到有理由行动，也没有证明准备电位对动作充分。
6. 最终判断取决于我们需要的是严格生产性因果、差异制造、因果相关解释，还是行动者层面的控制。

心灵哲学板块总结：从心身区分到心理解释

这九章形成一条回环式路线。Heil 先从经验私密性、现象质感和非空间性说明二元论为何诱人，再让互动问题把我们推向其他方案。Bayne 第三章提出行为主义、同一论、功能主义和取消主义，但知觉、思想与意识的具体案例显示，没有一种简短还原可以自动覆盖全部心理现象。最后，心灵因果问题检验每一种本体论是否真能解释行动。

核心张力	一端	另一端
通达方式	第一人称直接经验	第三人称科学说明
本体论	独立心理实体 / 性质	心理不超出物理
心理个体化	内在材料与脑结构	功能、内容及环境关系
意识解释	缩小或解释鸿沟	接受真实本体鸿沟
因果说明	能量生产与局部实现	反事实差异与高层解释

板块复习路线

先能清楚区分“心灵是什么”“我们怎样知道他人有心”“心理状态怎样造成行动”三个问题。再用同一组案例交叉检验理论：超斯巴达人反驳行为还原，多重实现反驳类型同一，中文房间质疑纯句法，Mary 与僵尸挑战物理主义，因果排斥挑战非还原论。最后说明每个反例攻击的具体前提，而不要把它们笼统说成“证明心灵很神秘”。

第三板块

形而上学

形而上学板块导读

形而上学追问世界最一般的结构。本板块选择三类与日常经验直接相连的问题：同一物体怎样在变化中持续存在；“可能、必然、本质”在世界中对应什么；人的行动能否真正由自己决定。三章并非互不相干：持存理论使用时间和同一性，模态理论说明“本可以不同”的含义，自由意志则把时间、法则、可能性与因果共同用于行动者。

三章的概念接口

持存：跨时间还是同一个吗？ → 模态：它本可以是别样的吗？ → 自由：行动者本可以做别的吗？

Ney 第 6 章：持存 (Persistence)

文献定位

本章用忒修斯之船、雕像与泥块、暂时内在性质三个问题比较**持续论**、**延续论**与**阶段论**。争论焦点不是东西会不会存在一段时间，而是这种存在究竟由跨时数值同一、时间部分，还是时间对应物构成。

6.1 变化之谜

持存 (*persistence*) 表面上要求同一对象在不同时刻存在，但对象持存时又不断变得不同。巴门尼德据“是者不能变成不是者”否认变化；亚里士多德回应，“是”可表示存在，也可表示谓词归属。年轻者变成不年轻，不等于存在者变成不存在。

亚里士多德认为真正变化必须有某物承受改变，否则只是一个对象毁灭、另一个对象产生。他用**质料形构论** (*hylomorphism*) 说明：实体由质料与形式构成，变化中质料连续而形式改变。但忒修斯之船逐块换木板、人体细胞不断更新，表明对象甚至能在质料不连续时持续存在，因此还需更一般理论。

两个经典难题构成背景。其一，忒修斯之船 S_1 的木板逐渐替换，产生保持连续使用的 S_2 ；旧木板又被重装成 S_3 。 S_2 和 S_3 各有理由声称是原船，却不能都与 S_1 数值同一，否则由同一关系传递性得到 $S_2 = S_3$ 。其二，泥块 Lump 被塑成雕像 Goliath；二者位置和材料相同，却有不同生灭条件：压扁会毁掉雕像而泥块仍在。按莱布尼茨律，它们似乎不是同一个对象，迫使我们接受同地同时的两个物体。

同一性不能靠“差不多一样”代替

原船与逐板更换之船有功能与历史连续，与旧板重组之船有材料连续。哪一种更像原船并不是核心；若说 $S_1 = S_2$ 而且 $S_1 = S_3$ ，就必须接受 $S_2 = S_3$ ，但两船最后停在不同位置。若改说只是两船“在不同意义上延续原船”，便已经放弃把两个声称都理解为严格数值同一，向时间部分或对应关系迈出一步。

6.2 关于持存的三种观点

6.2.1 持续论 / 三维论

持续论 (*endurantism*) 认为对象通过跨时数值同一而持存：在 t_0, t_1, t_2, t_3 出现的是全部同一个 Lump。对象在每一存在时刻都**完全在场** (*wholly present*)，即其全部部分都在该时刻存在。持续论常与**三维论** (*three-dimensionalism*) 结合：物体只有空间部分，没有时间部分。

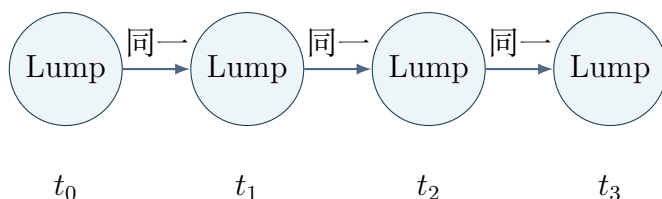


图 6.1 持续论：每一时刻完全在场的是数值同一对象。

6.2.2 延续论 / 虫体论

延续论 (*perdurantism*) 认为对象像在空间中展开一样在时间中展开，拥有不同的**时间部分** (*temporal parts*)。完整 Lump 是一条四维时空虫； t_0 处只有其早期阶段， t_3 处只有后期阶段，二者像左手与右手一样不数值同一。它是**四维论** (*four-dimensionalism*) 的一种。

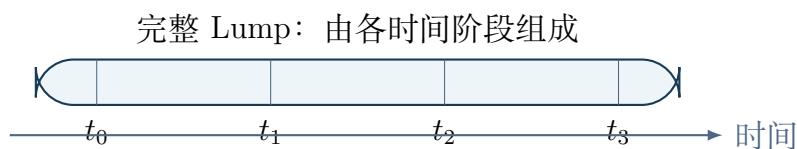


图 6.2 延续论：对象在时间中伸展，各时刻只呈现一个时间部分。

延续论者也可说“同一个 Lump 在两时刻存在”，意思是两个阶段属于同一四维整体。为与持续论真正区分，后者必须强调“完全在场”，不能只说“同一对象存在于多时刻”。

6.2.3 阶段论 / 远续论

阶段论 (*stage theory*) 或**远续论** (*exdurantism*) 也承认所有时间部分，却把日常对象“Lump”“Lewis”“金星”指向当前短暂阶段，而不是整条时空虫。跨时陈述由阶段之间的**时间对应物** (*temporal counterpart*) 关系获得真值。两种四维论在本体论上可承认完全相同的对象，区别在于普通名称指虫体还是阶段。

这里最易犯的错误是把延续论和阶段论画成“相信时间部分”与“不相信时间部分”。两者都承认时间部分以及由它们组成的较长整体；争执是“这一个日常对象”究竟指较长整体还是当下阶段。由此衍生看似语言学、实则形而上学的分歧：说“我明天仍会在”，是同一条虫拥有明天的部分，还是今天的我有一个恰当的明天对应物？

三种持存理论一览

- 持续论：一个完整对象在多个时刻数值同一地存在；
- 延续论：一个四维整体以不同时间部分占据不同时刻；
- 阶段论：普通对象是瞬时阶段，“将来仍存在”由未来对应物说明。

6.3 物质构成悖论的解决

延续论重写忒修斯之船的问题。 S_1, S_2, S_3 不是竞争“哪一个与原船同一”的三个完整对象，而是数值不同的阶段。若接受**部分整体普遍主义** (*mereological universalism*)，有一条四维船虫包含 $S_1 + S_2$ ，另一条包含 $S_1 + S_3$ ；二者都可被叫作忒修斯之船，严格说应命名为 A 船和 B 船。

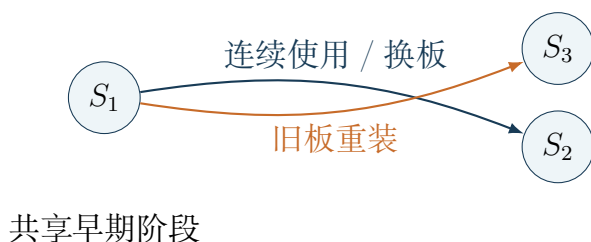


图 6.3 两条船虫时间重叠，共享 S_1 而在后来分叉。

这确实意味着早期“有两条船在场”，但不是两个完整物体占同一时空区域，而是两条虫共享同一阶段，类似 I-95 与 I-87 共享一段道路。时间重叠被类比为空间重叠。

道路类比恰好显示收获与代价。两条高速公路共享一段路，不会使那一段沥青突然变成两段；同理，两条四维船可以共享一段早期历史，而不让早期阶段有两份木头。但若问“那时港口到底停着几条船”，四维虫论须容忍“按完整对象数，是两条；按共同阶段数，是一段物体”的非日常回答。阶段论稍后正是针对这种计数困扰提出不同的普通名称解释。

雕像与泥块也成为两个部分重叠的四维对象。泥块在成像前已有阶段，雕像没有；二者在雕像存在期间共享阶段，却不共享完整历史。于是它们符合莱布尼茨律下的不同对象，又不要求每时刻有两个数值不同、完全重合的物体。

这个方案仍有极端难例：若泥块与雕像同一时刻产生、同一时刻毁灭并在全程共享所有阶段，延续论如何区分？它可能需要诉诸模态性质、种类或不同持存条件，而不能

只靠实际时空部分。

6.4 暂时内在性质问题

Lewis 称此为反对持续论的“主要而决定性”理由。形状似乎是**内在性质** (*intrinsic property*)：对象仅凭自身就是直或弯。Lewis 在 t_1 站直、在 t_2 坐下弯曲。持续论说 t_1 的 L_1 与 t_2 的 L_2 是数值同一的 Lewis；但同一个对象似乎就同时具有直与非直，违反莱布尼茨律。

持续论有两种回应。第一，与**现在主义** (*presentism*) 结合：在 t_1 只有“直”现实，“将弯”尚不现实；到 t_2 只有“弯”现实，“曾直”不再现实。Lewis 认为这以否定过去和未来存在来逃避持存问题，但现在主义者可说过去式与将来式陈述仍然为真。

第二，把性质时间化：对象不是无条件地直，而是“在 t_1 直”；“在 t_1 直”与“在 t_2 弯”是不同关系性质，无矛盾。Lewis 说这错误地把明显内在的形状变成关系。Haslanger 回应，涉及时间不必取消内在性：两个对象若在各时刻一个红、一个不红，就不是内在复制品；相对于时刻的性质仍可刻画对象本身。

延续论的解法较直接：直的是 t_1 时间部分，弯的是 t_2 时间部分，不同部分拥有相反性质并无矛盾，像左臂弯、右臂直。但阶段论者反问：严格说完整时空虫本身从未直或弯，只有部分如此；若普通名称指阶段，便可字面说“Lewis 现在是直的”。

把矛盾写清才容易看出分歧。持续论者说“同一个 Lewis 在 t_1 直、在 t_2 弯”；若省掉时间修饰，便误写成“同一个 Lewis 既直又非直”。Lewis 的批评不是不知道可以加时间，而是说若直与弯本来都是对象自身的内在形状，加上“与 t_1 相关”是否改变了我们原本要说明的性质。三维论者可以反驳：一种性质依时间实例化，不等于它在该时刻不由对象自身决定。

6.5 阶段论与时间对应物

阶段论还改善重叠计数：在忒修斯船分叉前只有一个阶段 S_1 ，故日常问题“这里有几条船？”可答一条，而虫体论答两条共享阶段的船虫。

代价是普通对象严格说只存在一瞬。阶段论以**语义上升** (*semantic ascent*) 解释日常跨时陈述：

“金星一百年后仍存在”

$\iff$ 一百年后有一个阶段，与当前金星保持时间对应物关系。

对应物关系由显著相似、因果连续、相近位置与构成等确定。类似地，“你小时候学

过钢琴”由过去某个对应阶段学琴为真。批评者会说这把“我曾做”改成“另一个与我相关者做”，没有真正保存主体同一；阶段论者则说日常跨时语言的真值已经获得说明，不必额外承认数值同一。

6.6 为三维论辩护

持续论的第一优势是常识：现在的我似乎就是昨天的我，而非只与昨天某阶段相关。Wiggins 进一步区分**持续体** (*continuant*) 与事件：要找手机或人的全部边界，只需在某时刻检查其空间边界；要找一场战争或演奏的全部部分，必须沿历史追踪开始与结束。事件有时间部分，不推出物体也有。

因此三维论的辩护不只是“我不喜欢四维虫”。它要求四维论说明为何物体和事件要以同一方式分割。一次演奏少了最后一小节就不是完整的那次演奏；一个人在某个时刻却能以全部空间部分站在那里，即使其生命史远未结束。四维论者会说完整的人也是跨时对象，日常只观察到一个截面；双方分歧在于“观察到全人”是字面真，还是对当时截面的便利说法。

van Inwagen 允许上帝创造只活一年的“笛卡尔阶段”，却不理解两个一年人怎样成为一个两年人的前后部分。在完整两年人的第一年结束时，并没有一个对象终结；第二年开始时也没有新对象产生。四维论者不能只重复空间类比，必须给出何以有“时间部分”的明确部分整体标准。

还要区分**延展** (*extension*) 与有部分：一个对象可占据非点状空间，却是没有空间部分的**延展简单物** (*extended simple*)；弦理论中的基本弦常被当作概念例子。因此“物体在时间中延展”本身不自动推出有时间部分。

三维论者最后指控，四维论把持存偷换成一串对象的产生和消失。延续论回应，整条虫始终与自身数值同一；阶段论以对应物语义保存持存句。争论于是转为：日常所需的“同一个”是严格数值同一，还是适当的部分 / 对应关系已足够？

理论	主要收益	主要代价
持续论	保住跨时数值同一 与常识主体	暂时内在性质、重合对象
延续论	优雅处理变化、分叉 和部分重叠	时间部分直觉、早期重叠计数
阶段论	现时谓词字面为真、 计数自然	普通对象瞬时存在、依赖对应物语义

Ney 第六章整章总结

持存问题不是简单问对象活多久，而是问跨时同一的形而上学结构。持续论用完整对象的数值同一，延续论用四维虫的时间部分，阶段论用瞬时阶段与时间对应物。忒修斯之船、雕像泥块和暂时内在性质分别测试分叉、重叠与变化。四维论解决许多悖论，却必须为时间部分和日常主体性付出解释；三维论保住常识，却要处理时间化性质与同地对象。

Ney 第 7 章：模态 (Modality)

文献定位

模态 (*modality*) 研究可能、必然、不可能与偶然。本章先区分逻辑可能与自然法则可能、命题模态与事物模态；再审查可能世界分析的三种本体论版本；最后讨论本质主义及“本质是否等于必要性质”。

7.1 作为真理方式的可能与必然

普通事实告诉我们现实怎样，模态事实告诉我们事情可以、必须或不可能怎样。命题若既非必然真又非必然假，就是**偶然的** (*contingent*)。例如“华盛顿是美国第一任总统”现实为真却本可为假； $2 + 2 = 4$ 通常被看作必然真； $2 + 2 = 5$ 必然假；“Abigail Adams 成为第一任总统”虽不现实，却可能。

Wittgenstein 式约定论曾把必然性理解为我们拒绝放弃某些句子的语言实践，而非世界中的事实。多数当代形而上学家仍希望把模态判断视为有真值的事实，并寻找非模态基础说明它们为何真。

请先分开两个任务：语义任务是告诉我们“可能有会说话的驴”在什么条件下为真；形而上学任务是告诉我们为何那些条件本身成立。“某可能世界有会说话的驴”在语义上很方便，但如果“可能世界”本来只是“可能发生的一种方式”的别名，分析就没有解释可能性从何而来。本章的具体世界、抽象替身、虚构主义、约定论之争，大多围绕这个第二任务。

7.2 可能性与必然性的种类

自然法则可能性 (*nomological possibility*) 指与实际自然法则相容。相对论若禁止超光速载具，它在此意义上不可能。**逻辑可能性** (*logical possibility*) 指不包含矛盾；超光速载具不等于圆方，故仍逻辑可能。通常：

自然法则可能的集合 $\subseteq$ 逻辑可能的集合.

不同可能性需保持相对化。独角兽可能逻辑上可有却不符合实际生物史；电子同时正负电若由“电子”“电荷”的理论本质排除，究竟是逻辑不可能、形而上学不可能还是自然法则不可能，需要进一步说明。

把限制条件设得越多，“可实现”的集合就越小。一个人的实际日程不允许他在同一时间同时参加两场会议，未必表示自然法则禁止分身；自然法则禁止的超光速旅行，也未必是纯逻辑矛盾。讨论“过去相同，我还可能做别的吗”之前，必须说清是固定哪些法则、哪些过去、哪些描述，否则争论双方可能在谈不同强度的“可能”。

另一组区分是：

- **言模态** (*de dicto modality*) 讨论命题整体的模态地位，如“必然地，所有三角形有三边”；
- **物模态** (*de re modality*) 讨论具体对象的性质是否本质，如“数字 3 必然是奇数”“苏格拉底是否可能不明智”。

后者涉及**本质性质** (*essential properties*)：对象若失去该性质便不再是那个对象或不能存在。

7.3 模态逻辑与可能世界分析

模态逻辑以菱形 $\Diamond$ 表示“可能”，方框 $\Box$ 表示“必然”。两者满足：

$$\Box A \equiv \neg \Diamond \neg A, \quad \Diamond A \equiv \neg \Box \neg A.$$

但符号只表示结构，不回答什么使 $\Diamond A$ 为真。最有影响力的回答是**可能世界分析** (*possible worlds analysis*)：

$\Diamond A$ 为真 $\iff$ 至少一个可能世界中 A 为真；

$\Box A$ 为真 $\iff$ 每一个可能世界中 A 为真。

“猪可能长翅膀”为真，因为有某个可能世界有翼猪；“三角形必有三边”为真，因为没有可能世界存在四边三角形。分析清楚，但立即引出更深问题：可能世界究竟是什么？若用“可能的世界”解释“可能”，是否循环？

7.4 Lewis 的模态实在论

模态实在论 (*modal realism*) 把可能世界理解为与我们一样具体真实、但在时空上完全断开的宇宙。一个世界是最大时空连通区域；“现实”像“这里”“我”一样是索引词，每个世界居民都把自己的世界叫现实。其他世界不是思想、描述或虚构。

它以巨大本体论换来透明还原：

$$\Diamond A \iff \exists w(A \text{ 在 } w \text{ 真}), \quad \Box A \iff \forall w(A \text{ 在 } w \text{ 真}).$$

Lewis 承认“难以置信的凝视”是无法完全消除的反应，却主张理论收益足够大。

第一项收益是分析其他模态概念。 A 偶然，当且仅当某世界 A 真、某世界 A 假。**反事实条件句** (*counterfactual*) “若袋鼠没有尾巴，它们会跌倒”可理解为：在与现实最相似且袋鼠无尾的世界中，它们跌倒。这里“最相似”要尽量保持对象、性质、规律与事实，只作满足前件所需改变。

第二项收益是分析性质。类别唯名论把性质认作实例集合，若现实中所有红物恰好都是圆物，就会错误地把红与圆认作同一性质。把所有现实和可能实例纳入集合后，其他世界的红而不圆物可区分两者。

第三项收益是分析思想内容。信念内容可看作使其为真的世界集合；学习意大利建国于 1861 年，就是把“可能是现实”的世界缩小到意大利于 1861 年建国的世界。信念真，当且仅当现实世界属于该集合。

此分析有粗粒度问题：所有必然数学真理在全部世界都真，因而拥有同一世界集合；相信 $2 + 2 = 4$ 与相信一个极复杂定理却显然不是同一内容。可能世界集合未必足以区分一切认知内容。

模态实在论不等于物理学多重宇宙

宇宙学多重宇宙若共享更大时空或同一基本物理结构，按 Lewis 定义仍可能只是一个世界的不同区域；更重要的是，物理多重宇宙只实现自然法则允许的情形，而 Lewis 需要容纳不同法则、甚至所有逻辑可能情形。

7.5 替代模态实在论

替代模态实在论 (*ersatz modal realism*) 保留可能世界分析，却把世界理解为现实存在的替身，如最大一致句子、命题集合或抽象结构。它与**现实主义** (*actualism*) 相容：凡存在者都现实存在；Lewis 的**可能主义** (*possibilism*) 则承认只可能而不现实的对象。

语言替代论 (*linguistic ersatzism*) 把世界看作最大一致的“造世界语言”句子。最大意味着对每个原子句 A ，世界句含 A 或 $\neg A$ ；一致意味着不同时含矛盾。于是可能就

是“至少一个最大一致句子说它为真”。

但造世界语言面临 Lewis 两难。

7.5.1 丰富语言之角

若语言直接含“有会说话的驴”“Obama 是世界级网球手”等丰富谓词，如何决定哪些组合一致？“Obama 从未拿球拍却是世界级网球手”没有 $A \wedge \neg A$ 形式，却似乎形而上学不可能。若用“真正可能的句子才一致”定义，就用模态解释模态。加入一套“形而上学真理”也只是把任务推迟到如何发现这套真理。

“一致”有两种读法必须分清：形式一致只排除形如 $A \wedge \neg A$ 的直接矛盾；形而上学可实现还得排除表面语法不矛盾、内容却不能共存的描述。丰富语言若只选前者，会把太多不可能情况算作世界；若偷偷选后者，又先知道了什么可能，才制造可能世界，未完成模态还原。

7.5.2 贫乏语言之角

若语言只用基本粒子、基本性质，或用时空点与“占据 / 未占据”0、1 编码，一致性容易非模态地规定。但句子不直接说“有独角兽”。替代论者必须说某种微观排列蕴涵独角兽存在；“蕴涵”又意味着该排列不可能不构成独角兽，重新引入模态。

贫乏语言看似可以任意排列 0 与 1，无须先问现实中什么可有；但从编码回到我们关心的“那算一头独角兽吗”“那算一个人吗”，还需要从底层到高层的规则。若规则本身说“只要如此排列，必然出现独角兽”，模态词已经躲进翻译步骤。这是 Lewis 所谓两角困境的对称性：说得丰富，筛选有模态；说得贫乏，解码有模态。

替代论的核心困难

具体世界本体论昂贵，但模态解释直接；抽象替身本体论节省，却容易把“可能”“蕴涵”“一致”暗藏在世界构造规则中。争论是解释成本与本体成本的交换。

7.6 拒绝可能世界分析

模态虚构主义 (*modal fictionalism*) 否认任何具体或替代世界存在，却容许可能世界话语作为有用虚构：

A 可能 $\iff$ 按照“存在可能世界”的虚构，有一世界使 A 真；

A 必然 $\iff$ 按照该虚构，每一世界使 A 真。

Lewis 的《论世界的复多性》本身可以充当这种虚构。问题是有限文本是否说够全部模态真理，以及模态事实是否荒谬地在文本写成后才出现。虚构主义必须区分“虚构存在”与“根据虚构为真”的基础。

约定论 (*conventionalism*) 完全绕过世界：必然真理由语言规则和词义蕴涵，如“单身汉未婚”；偶然真理还取决于世界，如“欧盟国家都在赤道以北”。圆方不可能，则源于“圆”“方”的使用规则冲突。

Kripke 与 Putnam 的两类案例挑战约定论：

- **后验必然** (*necessary a posteriori*)：“金的原子序数为 79”“水是 H_2O ”若为必然真，却只能由经验发现，不包含在普通词义中；
- **先验偶然** (*contingent a priori*)：“标准米尺在规定条件下长一米”由指称规则可先验知道，但那根物体本可被折断或未被选作标准，故对象具有该长度是偶然的。

这打破“必然 = 先验 = 词义真”“偶然 = 后验 = 世界事实”的简单对应。模态地位、认识方式和语言意义必须区分。

7.7 本质主义与反本质主义

本质主义 (*essentialism*) 主张对象独立于我们的分类方式，具有某些不得失去的性质。Quine 以既是数学家又是骑车者的 Jones 质疑：作为数学家，他似乎必然理性而不必两足；作为骑车者，似乎必然两足而不必理性。同一对象的“本质”随描述改变，是否还有脱离分类的事实？

形式上，范围差异揭示言模态与物模态：

言模态： $\Box \forall x (Mx \rightarrow Rx),$

物模态： $\forall x (Mx \rightarrow \Box Rx).$

前者说“必然地，所有数学家理性”：每个世界中只要谁是数学家就理性；后者说每个现实数学家在任何其存在的世界都理性，即把理性当个体本质。反本质主义者可以接受前者而拒绝后者。

跨世界说“Jones 仍是 Jones”时也须小心：他在那里不当数学家，是否还保有理性？仅知道“数学家依定义会理性”，不能回答“这个人不当数学家时会如何”。由类别的规则到个体的本质，需要额外原则；这正是 Quine 的疑问和 Kripke 后来以起源案例回应的焦点。

7.8 当代本质主义

Kripke 以**起源本质主义** (*origins essentialism*) 复兴物模态：伊丽莎白二世若由完全不同父母所生，那会是极相似、甚至成为女王的另一个人，而不是她；一张桌子若由完全不同木料制造，也不是这张桌子。对象的起源不是可随意替换的偶然描述。

Quine 式回应是，直觉仍依赖分类：把女王当生物个体时亲本重要，把她当“统治六十年的英格兰女王”时制度角色更显著。争论焦点不是能否构造相似对象，而是跨可能情形追踪“这一个对象”是否有独立标准。

较弱的**种类本质主义** (*sortal essentialism*) 说对象必然属于其基本种类：这个人不可能不是人，这张桌不可能不是某种桌。更弱的是自同一性。不过仍要区分：

$$\Box \forall x (x = x) \quad \text{与} \quad \forall x \Box (x = x).$$

前者只是必要的一般真理，后者给每个对象赋予必要自同一性质。传统本质主义更关心有些对象而非所有对象必具的区分性性质。

7.9 本质与必然性的关系

Kit Fine 反对直接把本质定义为“对象必然具有的全部性质”。苏格拉底必然具有“ $2 + 2 = 4$ ”这一性质，但它显然不说明苏格拉底是什么。本质应类似真实定义，刻画对象凭什么成为其所是；本质会产生必要性，但并非一切必要性质都进入本质。

理论	可能世界是什么	首要难题
Lewis 模态实在论	具体、时空隔绝且同样真实的宇宙	本体论极度膨胀
替代模态实在论	现实存在的句子、命题或结构	一致性 / 蕴涵中暗藏模态
模态虚构主义	不存在；仅在有用虚构中谈论	虚构何以覆盖并奠基全部模态真理
约定论	不需世界；模态来自语言约定	后验必然与先验偶然

Ney 第七章整章总结

1. “可能”至少有逻辑与自然法则两种强度；“必然地某命题真”不同于“某对象本质上有性质”。
2. 可能世界语义使模态推理清楚，却不自动解决世界的本体论。
3. Lewis 以具体世界换取非模态还原及对反事实、性质、内容的统一分析；替代论降低本体代价却容易循环。
4. 虚构主义和约定论拒绝世界本体，但分别面对虚构基础和 Kripke–Putnam 案例。
5. Quine 质疑脱离描述的物模态；Kripke 以起源本质主义回应；Fine 最终把本质理解为定义性根据而非必要性质清单。

Ney 第 9 章：自由意志 (Free Will)

文献定位

自由意志问题不是泛泛问“我想做什么就做什么”，而是问行动和欲望最终是否出自自我。决定论威胁“本可做别的”，不决定论又威胁控制。本章以结果论证为轴，比较相容论、自由意志主义、强决定论与强不相容论。

9.1 什么是自由意志

行动自由的最小含义是“这件事取决于行动者”。Kane 区分：

- **表层自由** (*surface freedom*)：能按自己的现有欲望行动，不受外部阻碍；
- **意志的终极自由** (*ultimate freedom of the will*)：不仅满足欲望，而且自己是这些欲望及行动的最终来源。

设想一整天都按愿望生活，看球、读书、骑车；这有表层自由。但若所有愿望由恶意神经科学家植入，行动虽满足欲望，仍似乎不真正属于你。现实中也没有完全隔绝的“纯自我”：家庭、教育、社会规范、阅读和偶然经历塑造我们的愿望。深层问题因此是，这些塑造是否仍能让行动者以足够方式拥有自己。

9.2 自由意志与决定论问题

决定论 (*determinism*) 是关于自然法则的主张：给定某时刻宇宙的完整状态和全部法则，未来任一时刻的完整状态原则上都可推出。若它为真，我们今天的选择早在出生前、甚至大爆炸时已由过去与法则固定。

van Inwagen 的**结果论证** (*Consequence Argument*) 把直觉形式化：

结果论证

1. 我们现在无力改变过去。
2. 我们现在无力改变自然法则。
3. 决定论若真，当前行动是过去与法则的必然后果。
4. 对我们无力改变的事实之必然后果，我们也无力改变。（无力的传递）
5. 因而我们无力使当前行动不发生，也就不能做别的。
6. 自由意志要求能够做别的。
7. 所以决定论若真，就没有自由意志。

关键争点有两处：无力是否按必然后果传递，以及自由是否要求**替代可能性** (*alternative possibilities*)。

结果论证的有效性和前提可信度

若否定的是第 4 步，相容论者得解释为何“我改不了过去与法则”仍可能与“我本可以做别的”共存；若否定第 6 步，他可以承认当时没有真实替代，却主张实际行动的来源与控制足以自由。两种回应不必联用，也不能只用“决定论太悲观”代替逐前提分析。

立场地图如下：

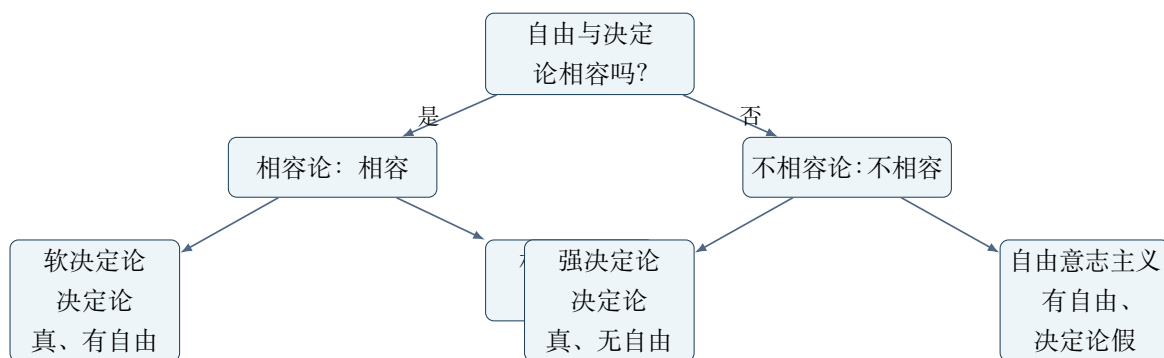


图 9.1 自由意志争论的主要位置。

不相容论 (*incompatibilism*) 否认自由与决定论相容；若接受决定论，成为强决定论；若坚持自由，成为**自由意志主义** (*libertarianism*)。**相容论** (*compatibilism*) 则认为决定论即使为真，仍有重要的行动者控制。

9.3 决定论有何根据

牛顿力学的预测成功使 Laplace 设想一个知道所有粒子位置、速度与力并有无限算力的理智，即**拉普拉斯妖** (*Laplace's Demon*)；对它，未来和过去一样确定。但可预测性是认识概念，决定论是法则结构；有限主体预测失败不证明不决定，完美预测也不必证明形而上学决定。

不决定论 (*indeterminism*) 只是决定论的否定，可能有三种来源：根本没有法则；法则不覆盖所有区域或情形；法则普遍却只给概率。量子力学的薛定谔方程本身是决定性的，是否还需随机“波函数坍缩”取决于解释；多世界解释便不加入这种随机定律。因此“量子力学已经证明不决定论”过于简单。

即使微观有概率，宏观系统中的概率效应通常相互抵消，未必为人类控制提供空间。极小随机偏差不自动成为自由；它也可能只是噪声。Kane 稍后会尝试让微观不决定性在特殊冲突行动中被放大。

9.4 相容论

第一类相容论回应攻击“无力传递”。David Lewis 说，我们确实不能在实际世界制造违反实际法则的例外，但“我本可以喝果昔而非吃麦片”可表示：有一个过去到选择前完全相似、我喝果昔且因此法则略有不同的可能世界。这不是行动者在实际世界施法破坏定律，而是若行动不同，最佳系统式法则描述也会不同。

这个回应依赖**休谟式法则观** (*Humeanism about laws*)：法则由整个世界的事件分布与最佳概括决定。反休谟论把法则视为超出事件、支配事件的基本事实，就更接近 van Inwagen 固定法则的前提。因此自由争论与“法则是什么”不可分。

Lewis 的“不同世界有不同法则”不是说我能在现实世界施法修改一本法律手册。他的比较是：早餐前历史保持相同，我喝果昔的另一个世界如果仍是决定论的，就不可能还遵守完全相同的实际法则；故这个世界的法则有细小差异。van Inwagen 会说我们原本要问的是在实际过去和实际法则已固定时还能不能换早餐，两人于是对“本可以”需要固定的背景有不同理解。

第二类相容论回应拒绝“自由要求能做别的”。Locke 的房间案例中，一个人自愿留在房内，却不知道门从外锁住；他无法离开，却似乎自愿留下。**Frankfurt 案例** (*Frankfurt case*) 更精确：Alice 自愿刺杀 Beatrice，Carlo 暗装装置，只有 Alice 打退堂鼓时才强制她执行。实际过程中装置从未启动，Alice 按自己的理由行动；她似乎自由且负责，却没有真正替代可能。

这说明未启动的背景限制未必破坏实际行动的自由。批评者会争论 Alice 是否仍有“开始反悔”之类微小替代，或装置能否在不预先决定她的情况下可靠介入；但案例至少

把焦点从“能否做别的”转到“实际因果序列是否属于行动者”。

洛克房间主要测试“愿意留下是否需要能离开”，Frankfurt 装置把这个想法推到行动过程：若 Alice 一直自行决定，背景监控者没有进入实际因果史；若她想不刺杀，装置便会介入，确保相同结果。因此批评者即使同意她实际上自愿，仍可问哪怕极小的反悔迹象能否算替代可能，或道德责任是否真的与“没有替代”兼容。这两类追问不能混成“装置曾控制她”。

Ayer 的简单相容论把自由理解为**无约束行动** (*unconstrained action*)：行动符合主体欲望，而非由枪口、支配或不可控制冲动迫使。强迫症窃贼虽想偷，却可能并未按自己希望拥有的意志结构行动，这推动更深版本：

- Frankfurt 的层级理论：自由行动来自主体认同的**二阶欲望** (*second-order desire*)，即“我希望自己想要什么”；
- Fischer 的**理由反应性** (*reasons-responsiveness*)：产生行动的机制能理解并在适当可能情形下回应理由。

决定性的法则甚至可能帮助控制：若意图与行为之间完全随机，主体无法可靠实现计划。相容论的真正任务不是否认因果，而是区分正常的行动者因果史与操纵、强迫、成瘾或失控的因果史。

9.5 自由意志主义

自由意志主义同意自由要求不被过去与法则完全决定，却立即面临**运气反对** (*luck objection*)：若在完全相同过去下行动 A 或 B 随机发生，结果为什么更属于主体，而不是更少受控？它必须同时避开决定与纯偶然。

9.5.1 行动者因果自由意志主义

Chisholm 的**行动者因果论** (*agent causation*) 说，自由行动有原因，但原因不是此前事件或状态，而是行动者这个实体本身。若欲望或意图已充分决定行动，行动者不是最终来源；若行动无因，又只是偶然。因此主体必须在事件因果链之外直接引起行动。

其难点是，这种行动者是什么，怎样在不受法则支配的同时产生物理改变？它与心身二元论相容，但继承非物质或特殊实体进入物理因果链的解释负担。

9.5.2 事件因果自由意志主义

事件因果自由意志主义 (*event-causal libertarianism*) 只承认事件与状态为因果项，却允许某些关键过程不决定。Kane 把焦点放在少数**自我形成行动** (*self-forming actions*)，

而非日常每次刷牙、吃早餐。

为何少数行动便可能足够？一次关键选择若在真实价值冲突中由行动者的努力形成品格，以后许多已经由品格和习惯决定的行为，仍可间接追溯到早先的自我形成。这样 Kane 不必声称每次日常行动都伴随量子不决定性；不过若第一批塑造品格的选择本身最终靠运气，追溯链仍未给出终极控制。这正是运气反对没有被“偶尔开放”自动化解的原因。

Sartre 的学生在二战时必须加入抵抗运动与照顾病母之间选择。两套价值和动机都真正属于他，并在脑中形成竞争过程；混沌动力学可能放大量子层面的不决定性，使任一过程胜出都未由过去充分固定。选择不是凭空出现，因为两个结果都由主体当下努力和理由网络支持；实际选择又塑造他日后成为什么人。

Kane 方案的双重条件

1. **开放性**：选择前存在至少两个真实可能，过去与法则不固定唯一结果；
2. **归属感**：每个可能结果都由行动者的理由、价值冲突和自我形成努力产生，而非外来随机插入。

随机性位于主体努力之中，不是替代主体。但批评者仍会问：若最终胜出由量子偶然决定，为什么是主体控制？

van Inwagen 因此认为自由也许是谜：与决定论相容很难理解，与不决定论相容同样难理解。

9.6 自由意志怀疑论

强决定论 (*hard determinism*) 接受决定论与不相容论，故否认自由。Pereboom 的**强不相容论** (*hard incompatibilism*) 更彻底：决定论下结果被过去固定，不决定论下结果含终极偶然；无论哪种，所需终极自由都不存在，因而无需先裁决物理法则究竟决定还是随机。

否认自由会牵动**道德责任** (*moral responsibility*)。幼儿、病患或被强迫者的行为不完全取决于他们，所以责备减弱；若所有人都不自由，报应性赞扬和谴责似乎全面失去基础。个人成就的骄傲、对罪犯的责难、友谊与爱的自愿意义也会被重写。

务必区分“行为可否评价”与“行动者是否终极应得奖惩”。自由意志怀疑者仍可说偷窃伤人、求助值得肯定，也可以出于保护和改变行为实施回应；他们质疑的是，人在独立于自身所有成因的意义上创造出自己，因此单凭应得便可以让他受苦。持自由论者反过来会问：若完全取消应得，承诺、道歉、感谢与责难还能否保持原来的道德含义？章节以这种实践后果收束，而没有让物理学替伦理学自动裁决。

Wegner 认为维持自由幻觉也许足以维持生活投入；Pereboom 则试图认真接受后果：可以出于隔离危险、康复、威慑和社会保护监禁罪犯，而不是因为他们终极“应得”痛苦；人际关系仍可因关怀、回应和共同历史有价值，而不必依赖终极自我创造。

立场	自由条件	主要难题
相容论	无强迫、行动来自认同欲望或理由反应机制	操纵案例、是否只有表层自由
行动者因果论	主体实体直接发起行动	特殊因果关系难理解
事件因果自由意志主义	理由支持但未决定的自我形成行动	运气与控制的冲突
强决定论	决定论真，故无自由	决定论证据与生活后果
强不相容论	决定与随机都排除	能否保留责任、关系与意义 终极自由

Ney 第九章整章总结

1. 表层自由是按欲望行动，终极自由要求主体还是欲望和品格的适当来源。
2. 结果论证由过去固定、法则固定、无力传递和替代可能性推出决定论下无自由。
3. 相容论可攻击无力传递或替代可能性，并以无约束、二阶认同、理由反应性重建自由。
4. 自由意志主义拒绝决定论，却必须说明不决定性为何不是随机；Chisholm 诉诸主体因果，Kane 诉诸自我形成行动。
5. 强不相容论认为决定与随机都不产生终极控制，并迫使我们重想道德责任、惩罚与人际关系。
6. 最深层问题不是“行动有没有原因”，而是何种因果来源、反事实能力与主体认同足以让行动真正算作“我的”。

形而上学板块总结：时间、可能性与行动者

Ney 三章逐步扩大“同一”与“可能”的应用范围。第六章问不同时刻的对象如何算同一个；第七章问现实对象与其他可能情形中的对象如何保持同一并具有本质；第九章问一个行动在其他可能情形中能否不发生，以及这种可能性是否构成自由。

整板块的共同方法

1. 先从常识判断出发：船持续、猪可长翅、行动者本可选择别的；
2. 用同一律、模态逻辑、自然法则和因果原则形式化；
3. 构造分叉船、具体可能世界、Frankfurt 装置等反例；
4. 比较理论保住哪些判断，以及把困难转移到何处。

板块复习时必须建立的三组联系

1. **时间与模态**：对象的持存条件通常也是它在反事实情形中被追踪的本质条件。
2. **模态与自由**：“能够做别的”必须说明是哪一种可能、固定哪些过去与法则。
3. **因果与责任**：决定论、不决定论、行动者因果和理由反应性分别给出不同的“行动来自我”的含义。

第四板块

伦理学

伦理学板块导读与扩充目录

本板块覆盖规范伦理学、元伦理学与应用伦理学。预定材料包括 Russ Shafer-Landau《*Living Ethics*》第 5、6、10 章, Earl Conee 与 Theodore Sider《*Riddles of Existence*》第 10 章, 以及《*International Encyclopedia of Ethics*》应用伦理学相关条目。编写时将特别区分三种层次: 什么行动是对的、道德判断在谈论何种事实、理论怎样进入动物与环境等实践问题。

第五板块

语言哲学

语言哲学板块导读与扩充目录

本板块不纳入课程期中或期末考核，主题为意义与表征，并延伸到人工智能与语言理解。必读篇目为 Thomas Nagel 《*What Does It All Mean?*》第 5 章 “The Meaning of Words”。后续译读将从“词语如何获得意义”进入心灵内容、指称、公共语言与理解条件。

附录

附录：中英术语表

术语后的括号列出正文所用简称。INF 及其带星号形式、R、D1–D4、G1–G2、SF1–SF3、ES1–ES2 等用于标记特定规则、定义或主张，不都具有独立的英文全称。

英文术语	中文译法	简要说明
epistemology	认识论	研究知识、证成、推理及相关认识规范的哲学分支
proposition	命题	能够表达事实并具有真假值的内容
belief	信念	主体对命题采取赞同态度, 认为其为真
propositional knowledge	命题知识	知道某个事实或知道命题 p 为真
acquaintance knowledge	亲知 / 熟识知识	认识或熟悉某人、某物, 不等同于只知道相关事实
ability knowledge	能力知识	实际具有完成某项活动的能力, 也称程序性知识
True Belief (TB)	真信念	真与相信的结合; TB 分析将知识等同于真信念
Justified True Belief (JTB)	得到证成的真信念	传统知识分析的三项条件: 证成、真理、信念
Traditional Analysis of Knowledge (TAK)	知识的传统分析	把知识分析为得到充分证成的真信念, 缩写为 TAK
correspondence theory of truth (CT)	真理符合论	命题为真当且仅当世界正如命题所说
Gettier problem	盖梯尔问题	得到证成的真信念仍可能因认识运气而不是知识的问题
Justified Falsehood Principle (JF)	得到证成的假命题原则	JF: 主体可能有证成地相信一个假命题
Justified Deduction Principle (JD)	证成演绎原则	JD: 证成可通过主体实际完成的有效演绎传给结论

英文术语	中文译法	简要说明
Same Evidence Principle (SE)	相同证据原则	SE: 证据完全相同的案例应具有相同证成地位
No False Grounds Theory (NFG)	无虚假根据理论	要求知识信念的所有根据均为真
No Defeaters Theory (ND)	无击败者理论	要求不存在会取消目标信念证成的真命题
EDF	无本质性虚假依赖方案	原书标记; 要求证成 does not essentially depend on any falsehood (不本质依赖任何假命题)
Justification by Explanatory Inference (JEI)	解释推理证成原则	用竞争假设之间的解释优劣说明推论证成
essential dependence	本质依赖	目标信念的证成是否不可替代地依赖某个假命题
doxastic attitude	信念态度	相信、否定、悬置判断及不同信念度的总称
justification	证成	对信念或信念态度所作的正面认识规范评价
warrant	保证、认识保证	与证成相近, 指使信念具有正当认识地位的因素
epistemic regress problem	认识论倒退问题	对理由继续索求理由时, 证成链应如何组织的问题
infinetism	无穷主义	主张推论证成依赖无终点、无重复的无限理由链
coherentism	融贯论	主张证成来自信念系统内部的相互支持与整体融贯
foundationalism	基础主义	主张推论证成最终依赖非推论地得到证成的基础信念
basic belief	基础信念	不从其他信念推论得出但仍获得证成的信念
immediate justification (IJ)	直接证成	不以主体其他已证成信念为中介的证成
mediate justification (MJ)	间接证成	通过与其他已证成信念的推论关系获得的证成
self-justification	自我证成	试图以信念自身、其内容或对应事实解释其直接证成

英文术语	中文译法	简要说明
self-presenting state	自我呈现状态	Chisholm 用于描述可通过自身被主体把握之状态的术语
level ascent	层级上升	把一阶证成要求提高为主体还须证明自己得到证成
prima facie justification	初步证成	只考虑某个支持来源时得到的暂定性证成
defeater	击败者	削弱或取消原有初步证成的因素
booster	增强者	为命题增加证成强度的因素
inference to the best explanation (IBE)	最佳解释推理	从竞争假设中选择对证据解释最优者的推理
abductive inference	溯因推理	最佳解释推理的另一名称
explanatory simplicity	解释简洁性	在其他条件相同时偏好较少基本设定的解释标准
ad hoc element	特设成分	为应付局部反例临时添加且缺少独立支持的理论补丁
justifier / J-factor	证成因素	能够解释或改变信念证成地位的因素
doxastic justification	信念证成	对实际持有的具体信念及其形成方式的评价
propositional justification	命题证成	主体是否处于适合相信某命题的认识情境
mentalistic evidentialism	心灵主义证据主义	只把主体当前心理证据视为证成因素
fittingness	契合、适切	相信某命题是否与总体证据相配
basing relation	基于关系	目标信念是否真正建立在相关理由之上
co-presence	共存	证据与目标信念同时出现在主体心中
process reliabilism	过程可靠主义	以信念形成或保持过程的真理产出倾向解释证成
unconditional reliability (UR)	无条件可靠性	不依赖信念输入时的高真率
conditional reliability (CR)	有条件可靠性	输入为真时，输出通常为真的性质
generality problem	一般性问题	同一具体过程可归入多种可靠率不同的类型

英文术语	中文译法	简要说明
new evil-demon problem	新恶魔问题	内部状态相同而环境使过程系统性失真的反例
approved-list reliabilism	认可清单可靠主义	以归因者储存的过程评价解释证成判断
internalism	内在主义	主张全部或主要证成因素在主体内部或可直接接近
externalism	外在主义	主张至少某些证成因素位于主体直接反思范围之外
factive state	事实性状态	处于该状态蕴涵其内容为真, 如“看出 P ”或“知道 P ”
epistemic rule	认识规则	规定在特定认识条件下允许采取何种信念态度的规范
truth-conduciveness	导真性	一种方法或规则促进真信念并避免假信念的性质
The Standard View	标准观点	认为我们拥有广泛日常知识、普通证据通常足以产生知识的出发立场
skepticism	怀疑论	否认主体拥有或能够拥有某类知识的立场
global skepticism	全面怀疑论	否认任何人拥有任何知识的立场
local skepticism	局部怀疑论	只否认某一领域知识的立场
brain in a vat (BIV)	缸中之脑	用模拟输入制造正常经验的怀疑论设想
fallible evidence	可错证据	在逻辑上与目标命题为假相容的证据
introspective indistinguishability	内省不可分辨性	两种情形从主体第一人称视角看来相同
indistinguishability principle	不可分辨原则	主张知识案例不能与非知识案例在主体内部看来完全相同
psychological certainty	心理确定性	主体感到完全确信的的心理强度
epistemic certainty	认识确定性	理由达到最大强度并完全排除错误
transmissibility principle	传递原则	知识可通过主体所知的逻辑蕴涵传给结论
closure principle	知识闭合原则	若主体知道前提并适当掌握其逻辑后果, 则知识对该后果保持闭合
Moorean response	摩尔式回应	以普通知识比怀疑论前提更确定为由拒绝论证

英文术语	中文译法	简要说明
fallibilism	可错论	主张知识要求强理由但不要求绝对确定或不可能出错
high-standards skepticism	高标准怀疑论	通过把知识门槛提高到确定性或不可错性来否认普通知识的立场
self-verification	自证性	一个判断的真诚形成或思考本身使其内容得到保证的性质
Cartesian substance dualism	笛卡尔实体二元论	把心灵与身体看作分别以思维和广延为属性的两类实体
privileged access	特权通达	主体对自身当前心理状态所具有的特殊第一人称认识途径
causal closure of the physical	物理因果封闭	任何有原因的物理事件都有充分物理原因的原则
psychophysical parallelism	心物平行论	心理与物理序列规律同步却不跨域因果互动的观点
occasionalism	偶因论	把表面因果事件当作上帝产生相应结果之场合的理论
idealism	唯心论	只承认心灵及其内容为根本实在、否认独立物质实体的观点
epiphenomenalism	副现象论	心理事件由物理事件产生却不再造成任何结果的观点
analytical behaviourism	分析行为主义	把心理状态分析为以非心理方式描述的行为倾向
type identity theory	类型同一论	主张心理性质类型与脑性质类型同一
token identity theory	个例同一论	主张每个心理事件与某个物理事件同一而不要求类型同一
multiple realization	多重实现	同一心理或功能性质可由不同物理结构实现
functionalism	功能主义	以状态在系统中的因果或功能角色确定心理种类
eliminativism	取消主义	否认某些民间心理种类真正存在或构成科学种类
intentional content	意向内容	心理状态把世界表征成某种样子的内容
sense datum	感觉材料	感觉材料论所说的私人、直接知觉对象
naïve realism	朴素实在论	主张正常知觉由主体与外界对象的直接关系构成

英文术语	中文译法	简要说明
disjunctivism	析取论	否认真知觉与主观不可分的幻觉属于同一基本心理种类
intentionalism	意向论	主张经验现象特征由其意向或表征性质决定
inverted spectrum	颠倒光谱	行为和分类相同而颜色体验系统颠倒的思想实验
blindsight	盲视	主体报告无视觉意识却能辨别盲区刺激的神经心理现象
language of thought	思想语言	把思想视为内部组合符号、思考视为句法转换的假说
systematicity	系统性	拥有某些思想能力时规律地伴随拥有相关思想能力
productivity	生产性	由有限成分形成开放乃至无限多新思想的能力
Chinese room argument	中文房间论证	以不理解符号的规则操作质疑句法足以产生语义理解
Turing test	图灵测试	以开放会话中能否与人类不可稳定区分来检验机器思想
explanatory gap	解释鸿沟	物理 / 功能说明与意识现象质感之间的可理解性缺口
knowledge argument	知识论证	以 Mary 首次见色获得新知识来反驳物理事实穷尽意识事实
philosophical zombie	哲学僵尸	与人类物理功能完全相同却没有意识经验的设想存在者
phenomenal concept strategy	现象概念策略	用同一事实的物理概念和现象概念两种呈现方式回应 Mary
illusionism	幻觉论	把不可化约现象性质视为大脑造成的系统性表象
Russellian monism	罗素式一元论	把意识或原现象性质视为物质内在本性的理论
combination problem	组合问题	微观主体或微观经验怎样组成统一宏观意识的问题
mental causation	心灵因果性	心理状态尤其因其心理性质而产生行动或其他结果的能力
causal exclusion	因果排斥	充分物理原因似乎排除不同的心理原因之论证

英文术语	中文译法	简要说明
difference-making	差异制造	以反事实变化而非能量传递刻画因果的进路
content externalism	内容外在主义	主张心理内容部分由主体环境、历史或社会关系决定
readiness potential	准备电位	Libet 类任务中在自发动作与意识意图报告前出现的脑电活动
persistence	持存	对象在多个时刻存在并经历变化的现象
endurantism	持续论	认为完整对象以跨时数值同一的方式持存
perdurantism	延续论 / 虫体论	认为四维对象通过在不同时刻具有不同时间部分而持存
exdurantism	远续论 / 阶段论	把日常对象认作瞬时阶段并以时间对应物解释跨时陈述
temporal part	时间部分	只占据一个时间区域、作为四维对象之部分的实体
temporary intrinsic	暂时内在性质	同一持存对象在不同时刻具有相反内在性质所造成的问题
modality	模态	关于可能、必然、不可能与偶然的领域
nomological possibility	自然法则可能性	与实际自然法则相容的可能性
de dicto / de re modality	言模态 / 物模态	分别关于命题整体及具体对象本质性质的模态
modal realism	模态实在论	Lewis 关于具体可能世界像现实世界一样真实的理论
ersatz modal realism	替代模态实在论	用句子、命题或抽象结构充当可能世界的理论
actualism / possibilism	现实主义 / 可能主义	分别否认或承认存在仅仅可能而不现实的实体
essentialism	本质主义	主张对象独立于分类而必然具有某些定义性性质
origins essentialism	起源本质主义	主张生物的亲本、物体的原始材料等起源对其同一性本质
determinism	决定论	给定完整过去状态与法则即可推出唯一未来的理论

英文术语	中文译法	简要说明
Consequence Argument	结果论证	从过去和法则不可改变推出决定论下不能做别的论证
compatibilism	相容论	主张自由意志与决定论可以同时为真
libertarianism	自由意志主义	主张自由与决定论不相容、且人至少有时自由
Frankfurt case	Frankfurt 案例	行动者不能做别的却似乎仍自愿且负责的思想实验
agent causation	行动者因果	把行动者实体自身而非此前事件看作自由行动原因
self-forming action	自我形成行动	Kane 所说塑造行动者品格、在价值冲突中作出的关键选择
hard incompatibilism	强不相容论	主张决定论与不决定论都不能容纳所需自由意志

本版结语：三大板块的共同问题

本版完成的知识论材料从一个看似简单的问题开始：“知道”比“碰巧相信真命题”多了什么？Feldman 第二章用证成排除裸猜测，第三章又用 Gettier 案例表明普通证成仍未完全排除认识运气。知识分析因此把问题转交给证成理论：什么是充分证成，证成怎样产生？

Goldman 第一章说明，推论信念的证成不能只靠重复“我还有理由”，必须解释理由链如何开始、延续或形成网络；第二章则指出，即使当前拥有看似充分的证据，也必须考察信念如何形成、如何保存，以及相应认知方法是否与真理稳定相连。Feldman 第六章最后检验知识门槛：若知识要求完全排错，怀疑论便占优势；若很强而可错的理由已经足够，普通知识仍可成立。

贯穿五篇文献的统一张力是：

知识既要从主体的理由看来合理，也不能只凭世界的偶然配合而为真。

证据主义、融贯论和内在主义较突出主体可以援引或反思把握的理由；可靠主义、外在主义与 Gettier 后的第四条件更突出真理联系和非偶然性；温和基础主义与可错论则试图让经验提供原始而可错的支持，并允许背景信念、击败者和形成历史共同决定最终地位。知识论的许多后续争论，正是在“第一人称的合理性”与“第三人称的导真性”之间寻找不牺牲任一方的解释。

心灵哲学把相同张力改写为第一人称与第三人称的关系：经验怎样向主体呈现，与科学怎样说明脑和行为，是否能由同一理论兼顾？二元论保存质感与私密性，却面临因果封闭；物理主义保存科学统一，却面对多重实现、意识鸿沟与因果排斥。

形而上学进一步揭示，这些争论依赖关于同一、可能性、法则和因果的背景理论。对象跨时间是否完整同一，决定了我们怎样理解变化；可能世界是否真实，决定了“本可以不同”的分析；自由意志能否存在，又取决于怎样理解法则固定、替代可能性与行动来源。

三大板块因此共享一种哲学方法：从强烈直觉出发，把它形式化为前提，再用反例检验前提是否真正必要。好的理论不是简单消除所有反例，而是明确说明自己保留哪些现象、修改哪些常识、承担哪些本体论与解释成本。

伦理学与语言哲学仍按课程大纲保留扩充位置。取得相应原文后，目录、文献定位、案例框、论证框、术语表与章末总结可沿用现有格式继续写入。