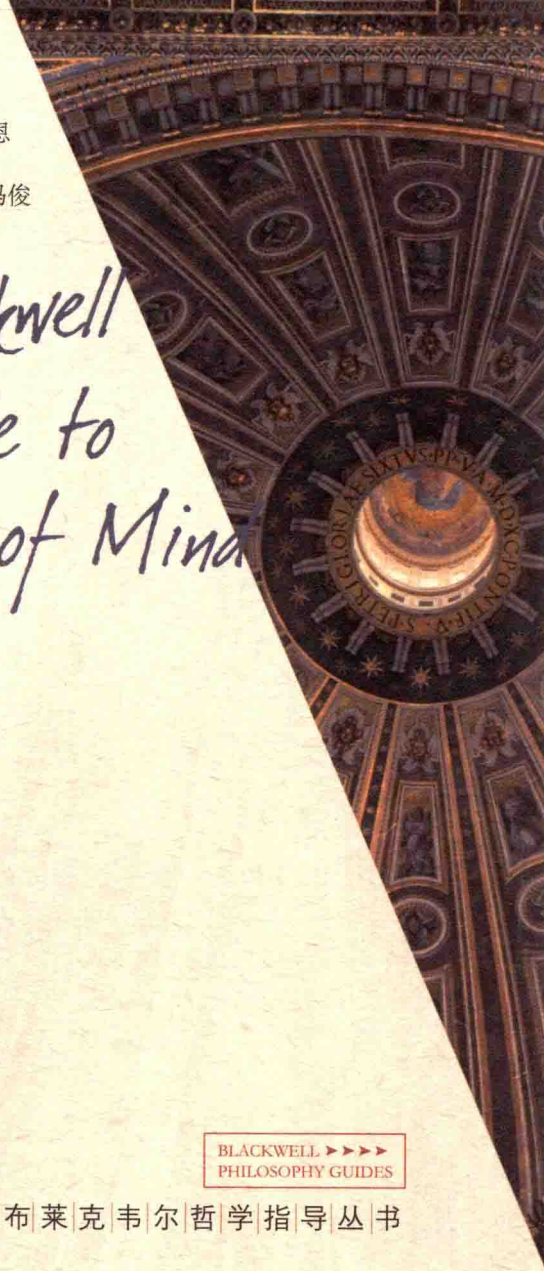
BLACKWELL PHILOSOPHY GUIDES

丛书总主编：[美]斯蒂文·M·卡恩
(Steven M.Cahn)
中文翻译总主编：冯俊

The Blackwell Guide to Philosophy of Mind



BLACKWELL PHILOSOPHY GUIDES

布莱克韦尔哲学指导丛书

斯蒂芬·P·斯蒂克 (Stephen P.Stich)
特德·A·沃菲尔德 (Ted A.Warfield) /主编
高新民 刘占峰 陈丽 等/译

心灵哲学

WILEY

中国人民大学出版社

著作权合同登记号
图字：01-2003-5266号

BLACKWELL PHILOSOPHY GUIDES

The Blackwell Guide to Philosophy of Mind

WILEY

www.wiley.com

布|莱|克|韦|尔|哲|学|指|导|丛|书

BLACKWELL ▸▸▸▸ PHILOSOPHY GUIDES

上架建议：西方哲学史

ISBN 978-7-300-18158-5

9 787300 181585 >

定价：78.00元

BLACKWEL

丛书总:

中文翻译总主编：冯俊

心灵哲学

斯蒂芬·P·斯蒂克（Stephen P.Stich）/主编
特德·A·沃菲尔德（Ted A.Warfield）
高新民 刘占峰 陈丽 等/译

The Blackwell Guide to Philosophy of Mind

BLACKWELL
PHILOSOPHY GUIDES

布莱克韦尔哲学指导丛书

中国人民大学出版社
·北京·

图书在版编目（CIP）数据

心灵哲学/斯蒂克，沃菲尔德主编；高新民，刘占峰，陈丽等译. —北京：中国人民大学出版社，2013.10
（布莱克韦尔哲学指导丛书）
ISBN 978-7-300-18158-5

Ⅰ.①心… Ⅱ.①斯… ②沃… ③高… ④刘… ⑤陈… Ⅲ.①心灵学 Ⅳ.①B846

中国版本图书馆 CIP 数据核字（2013）第 237887 号

布莱克韦尔哲学指导丛书
丛书总主编：[美] 斯蒂文·M·卡恩（Steven M. Cahn）
中文翻译总主编：冯俊

心灵哲学

斯蒂芬·P·斯蒂克（Stephen P. Stich）
特德·A·沃菲尔德（Ted A. Warfield） 主编
高新民 刘占峰 陈丽 等 译
Xinling Zhexue

出版发行	中国人民大学出版社		
社　　址	北京中关村大街 31 号	**邮政编码**	100080
电　　话	010－62511242（总编室）	010－62511770（质管部）	
	010－82501766（邮购部）	010－62514148（门市部）	
	010－62515195（发行公司）	010－62515275（盗版举报）	
网　　址	http://www.crup.com.cn		
	http://www.ttrnet.com（人大教研网）		
经　　销	新华书店		
印　　刷	涿州市星河印刷有限公司		
规　　格	165 mm×235 mm 16 开本	**版　　次**	2014 年 6 月第 1 版
印　　张	31 插页 3	**印　　次**	2014 年 6 月第 1 次印刷
字　　数	526 000	**定　　价**	78.00 元

版权所有　侵权必究　印装差错　负责调换

内容简介 ▼

本书内容广泛，收录了由一流学者撰写的许多具有特殊使命的篇章，它为心灵哲学中的核心主题给出了一种最新的审视。引领读者穿越范围广阔的主题，包括人工智能、意识、二元论、情绪、民间心理学、自由意志、个体主义、人格同一性和心身问题等。对于有兴趣更深入地理解这个领域的任何人都是一部重要的著作，对大学生是通俗易懂的导论，而对研究生和研究人员也提供了新颖、有趣的见解。

主编简介 ▼

斯蒂芬·P·斯蒂克（Stephen P.Stich）是罗格斯大学哲学教授，他当前的研究兴趣是心灵哲学和认知科学。他出版的著作有：《从民间心理学到认知科学》（*From Folk Psychology to Cognitive Science*, 1983）、《理性的断裂：一种实用主义认知评价理论的序言》（*The Fragmentation of Reason:Preface to a Pragmatic Theory of Cognitive Evaluation*, 1990）和《解构心灵》（*Deconstructing the Mind*, 1996）。他在心灵哲学、认知科学、语言哲学和认识论领域撰写有论文。

特德·A·沃菲尔德（Ted A.Warfield）是圣母玛利亚大学哲学副教授，他当前的研究兴趣是心灵哲学、认识论和形而上学。他出版的著作有：《心理表征文选》（*Mental Representation:A Reader*, 1994，与斯蒂芬·P·斯蒂克合编）和《怀疑主义》（*Skepticism*, 与K.德罗斯合编）。他还撰写有关于决定论、道德责任、人的自由和知识等方面的论文。

《布莱克韦尔哲学指导丛书》书目

哲学逻辑 (已出)

形而上学

认识论

科学哲学

➤ 心灵哲学

教育哲学

社会科学哲学

社会政治哲学 (已出)

伦理学理论 (已出)

工商伦理学

大陆哲学

古代哲学

近代哲学家

策划编辑：杨宗元
责任编辑：符爱霞
版式设计：赵星华
封面设计：王强工作室

《布莱克韦尔哲学指导丛书》
— 中文翻译编委会 —

主编：冯　俊

顾问：Alan Montefiore（英国牛津大学）
Nicholas Bunnin（英国牛津大学）

编委：（以姓氏笔画为序）
冯　俊（中国浦东干部管理学院）
江　怡（北京师范大学哲学与社会学学院）
朱志方（武汉大学哲学系）
张志伟（中国人民大学哲学院）
赵敦华（北京大学哲学系）
高新民（华中师范大学马克思主义学院）
童世骏（华东师范大学哲学系）
傅有德（山东大学哲学系）

教哲学和学哲学的方法指南

——《布莱克韦尔哲学指导丛书》中译本序

冯　俊

每一个民族和每一种文化都会有自己的哲学思想，而作为成熟的理论形态的哲学在世界的几大文明中都同样散发着智慧的光辉。哲学并非像黑格尔和德里达所言只是西方人的专利——只是起源于古代希腊，并用希腊文、拉丁文和德文等西方语言表达出来的一种独特的概念思维，哲学的形态、风格和语言表达都应该是多元的。19 世纪末和 20 世纪初中国人从西方引进“哲学”一词和西方哲学时，只是把它作为与中国传统哲学不同的另一种新奇的哲学，而从未把它看做是世界上唯一的哲学。一百多年来，中国人在不断地引进、移植和改造着西方哲学并将其本土化，在对西方哲学的理解中渗透了中国文化的独特视角，从中国社会的现实和独特发展中对于西方哲学产生了独特的理解和感悟。应该说，中国人在引进和研究西方哲学的过程中对于西方哲学也是有理论贡献的。

然而，在我们引进和研究西方哲学的过程中更多地是注意哲学的内容和学理，而很少去注意西方人是怎样学习和教授哲学的，实际上，哲学研究的方法和哲学教学的方法是两种不同的但是相辅相成的方法，学习哲学和教授哲学的方法能够帮助我们更好地了解和掌握哲学的精神实质和学理路径。早在 17 世纪，笛卡儿就注意到，研究者自己发现问题和解决问题的方法与他把这些发现表述出来告诉别人的方法是不同的，而霍布斯还专门研究了发现的方法和教导的方法之间的区别，这表明他们在研究哲学的同时还是十分注重哲学的教学方法，把自己的哲学表述清楚、传扬出去、教授给别人也是哲学家研究哲学的重要目的。因此，对于西方哲学的教学方法的研究应该是一个重要的课题。

哲学的教学方法一般说来包含三个方面，一是如何教，二是如何学，

三是学什么，即教学的组织形式、学习形式和教材。在过去的半个世纪中，我们教授哲学和学习哲学的方法日显僵化，主要是老师讲，学生听，老师像是在宣示真理，学生却是在死背教条，丰富鲜活的哲学变成了枯燥无味的说教。没有辩驳、讨论，更谈不上与哲学家对话。哲学教科书，一方面远离社会现实，不能对人类面临的重大问题作出回应，无法反映现时代的精神；另一方面也远离学术前沿，大多是已成定论老生常谈，知识结构陈旧。同时，“教科书文化”大行其道，每一个学校、每一个地区甚至全国都在编写同一门课的教材，编者更多关心的是市场占有份额。教科书千书一面，大同小异，大多是介绍性、转述性的二手资料，学生通过教科书根本就不能了解哲学家本人的原著和思想理路，哲学系的学生学习了几年的哲学，没有读过几本原著，更没有读过原文原著，甚至不知道某位哲学家的名字原文该如何书写、如何发音。这种现状是否也值得哲学教育者们警醒呢？在我们大力引进和研究西方哲学的同时，是否也需要研究西方的哲学教育方法和教材的编写方法呢？

中国古代的哲学也有自己的研究方法和教学方法，就教学方法而言，古时学哲学也不是靠一些导论性的教科书、二手的介绍材料，重视哲学家的原著，背诵、注解原著是一个突出的特点，学生在熟读和牢记哲学家原文原著的基础上，可以自然地体悟和阐释这些哲学的意蕴。而这种传统的方法也被我们丢掉了。今天我们是否需要借鉴西学或是反观传统来重新审视和检讨我们该如何教哲学、如何学哲学呢？

我在此想以在牛津大学的一些体验来简要地考察一下西方大学里成熟的哲学教学方法。

首先，哲学的教学方法和教学的组织形式是多种多样的，大致分为：(1) **个别辅导**（tutorial）。牛津大学的本科生是导师制，在学院里生活，学生每周和导师见面一次，导师布置书目，列出思考题，学生去图书馆借书、看书，一周之后学生和导师再次见面时，学生向导师汇报读书心得，导师答疑解惑，与学生讨论切磋，再布置新的书目和新的思考题。周而复始，几年下来，学生阅读了大量的书籍，思考了大量的问题，与导师也进行了充分的交流，导师在辅导和答疑中也在不断收集新的问题、不断思考和提高，做到了教学相长。(2) **讲座课程**（lecture）。除了在学院里有导师个别辅导外，学生还要听全校公开开设的大课。课堂小则几十人，多则几百人；课程提纲挈领，统观全局，给学生以该学科的概貌、前沿问题和争论热点的梗概，为学生继续学习提供指引。课程不照本宣科、不面面俱

到，浓缩凝练，一门课一般只有八讲。当然，也有几个老师开同一门课的现象，但是它们的内容和侧重点必定是有区别的，用不同的号码、代码区别开来以便学生选课。每一个老师，既需要在学院里做导师对学生进行个别辅导，也需要在学校里对全校开设大课。不管资格多老的教师或多有名的教授，都要给本科生上课，直到退休为止。(3) **讨论班**（seminar)。讨论班是高年级学生和研究生的主要教学组织形式，围绕着问题或原著，由学生们轮流主讲或宣读小论文，老师只是一个旁观者、参与者或引导者，在学生们讨论完之后稍做总结和归纳。讨论班中学生们都是主动地学习、研究型学习，经历了发现问题解决问题的全过程，收获非常大，表述自己的研究心得，与他人开展辩驳和争论，在争论中擦出思想的火花。参加一次讨论班，对学生的学习能力和研究能力是一次全面的锻炼。(4) **学术报告会**（colloquium)。学术报告会往往是提前一个学期或数月就安排好、隔周一次或定期性的学术活动，一般是以聘请校外名家为主就学术热点问题和前沿问题作学术报告，一般在报告人报告完后，还有评论人对学术报告进行评论，然后学生们再自由提问，报告人回答问题。往往一些好的学术报告可以使学生开阔眼界，为学生提供了与名人见面和现场讨教的机会。(5) **学术研讨会**（symposium)。研究生经常会和导师一起参加学术研讨会，宣读自己的学术论文，回答学者们提出的问题，和大家一起研讨，这是研究生走进学术圈，认识学术界朋友，同时也让学术界认识自己的一个很好的场所。(6) **学术大会**（conference)。各个学会、协会一年一度的学术大会，或全国性、国际性的专业大会，往往是围绕一个主题，多学科的学者多角度地进行研讨，规模宏大。一般说来，只有极少数学生或研究生在学习期间才有幸参加这样的大会。以上六种方式，经过几百年的实践，已经成为一种学术规范，被西方的大学和学术界广泛地采用。近几年来，我国学术界也已经部分地采用了这种规范，进行学术研讨或进行教改实验，但是，对于哲学教学特别是本科生的教学影响很小。

其次，就课程设置和学生的学习方式而言，与国内通行的教学相比，西方大学哲学系里开设的导论性、通史性的课程较少，专题课、原著研读课较多。例如，他们不会花大片的时间学习西方哲学通史，而是读柏拉图、亚里士多德、奥古斯丁、阿奎那、笛卡儿、洛克、休谟、康德、黑格尔、维特根斯坦和海德格尔等人的代表作，或开设专题讲座，他们更注重专深，而不太注重全面。由于名师名家与学生见面的机会多，学生参与学术活动的机会多，以及交流沟通工具和旅行的便利，在西方著名的大学里

学生有更多的机会与名师名家面对面地交流、讨教、对话，甚至向名家追问、质疑。学生们学习和研究的界限越来越模糊，从开始学习时起就能走入学术的前沿，参与学术热点问题的讨论和对话。而我们的学生则是经过漫长的打基础的过程，读二手书，听二手课，讨论二手问题，上学几年没有见过本校名牌教授、没见过国内名家，更难设想与世界级学者对话。

再次，西方大学里很少组织编写教材，既没有统一的教材编写大纲，也没有教育部推荐教材，他们更多地是把哲学家的原著作为教材、把那些经过时间检验或被学术界广泛认可的学术专著选作教材，而这些都是由老师来定的，一些导论性的教材也是老师个人多年的讲义，而不是由官方或校方要求或组织写的。他们不做重复建设，如果哪本书好，被大家选作教材广泛使用，就可以不断修订，一版再版，没有人再去编写地方版、地区版、高教版、成教版、职教版等，不再做花样翻新、遍地开花的工作。教科书不要求人手一本，不要求每年新生来都要买教科书，既节约了纸张能源，也节约了经费、减轻了学生的负担，学生主要是在图书馆借书看，或者新生向老生借书看。

当然，不能因此就说在西方没有成系列的大学教科书。但是这些教科书不是由官方或大学来组织编写的，而是由出版社来主导编写的，他们组织作者队伍，有计划、有目的地编写系列教材供大学选用或列为参考书。他们通过教科书体系的完备、品种的齐全、体例的新颖、作者的影响力、内容的丰富和较高的质量来影响大学的教学。

在这里我想以英国著名的 Blackwell 出版公司的哲学教科书系列为例，来考察一下英语世界的哲学教材编写的特点和风格。Blackwell 的哲学教材分为五个系列：（1）**一般导论**（general introduction）系列，主要是给大学本科生教授基本哲学概念和哲学线索的导论性教材，例如《哲学导论》，《哲学家的工具箱》，《以哲学的方式思维——批判反省和理性对话导论》，《从笛卡儿到德里达——欧洲哲学导论》，尽管是导论性的教材，但是在写法上专著的味道还是很浓。（2）**著作选读**（readers）系列，主要是哲学家的著作选集，以人为线，原著选读。（3）**专题文选**（anthology）系列，例如，*Blackwell Philosophy Anthologies*（《布莱克韦尔哲学专题文选》），该系列以不同的理论热点为线来编纂，其中包含有《从现代主义到后现代主义》、《心灵和认知》、《心灵、大脑和计算机——认知科学的基础》、《当代政治哲学》、《后殖民话语》、《形而上学》、《认识论》、《分析哲学》、《逻辑哲学》、《数学哲学》、《技术哲学》、《宗教哲学》、《生命伦理学》、《环境伦

理学》、《美国哲学》、《非洲哲学》等。（4）**指南**（guide）系列，例如，*Blackwell Philosophy Guides*（《布莱克韦尔哲学指导丛书》），这是论文选集，但基本上按照哲学的学科体系来分册，将该学科最具前沿性和代表性论文收集整理成一个体系，让它基本上能反映出该学科的前沿问题。（5）**伴读**（companion）系列，例如，*Blackwell Companions to Philosophy*（《布莱克韦尔哲学伴读丛书》），该系列有两种体例，一种是类似专题词典，对于该学科内的一些重要知识点，列出词条，按字母顺序排列，进行详尽阐释，每一个条目后还附有参考书目，全书后面附有索引，如同专业词典；一种是专题综述文章，文章涵盖了学科体系的方方面面，每一篇文章后面都附有参考书目，类似于专业辅导书，一书在手，就可以对于这个学科有一个全面的了解。以上五大系列，构成了一个完整的教材体系，蔚为大观，供各大学选用，供学生参考。当然，出版社只能靠教材的质量取胜，而不可能通过某种行政的力量或手段来推广。

我们在这里组织翻译了其中的一个系列，*Blackwell Philosophy Guides*，该系列一共十余本，包括《形而上学》、《认识论》、《科学哲学》、《社会科学哲学》、《哲学逻辑》、《心灵哲学》、《教育哲学》、《伦理学理论》、《社会政治哲学》、《古代哲学》、《近代哲学家》、《大陆哲学》、《工商伦理学》等。该丛书具有如下特点：首先，学术含量高。每一章都是由该领域的著名专家所写的一篇高质量的文章，编者按照文章所涉及的主题将这些文章编成体系章节，看上去是一本完整的书，但是它比一般的教科书有更高的学术含量和更密集的学术信息。其次，学科前沿性。每一本书都涉及了该学科的学术热点、前沿问题和学科发展方向，能够将读者一下子带到学术的前沿，让读者了解该学科的新进展、新动态。再次，专业性和可读性相兼顾。该丛书讨论的问题专业性较强，适合作为研究生包括硕士生和博士生的参考教材，但是，作者也十分注意丛书的可读性，力争让非哲学专业的读者也能读懂，让他们了解哲学领域内争论的问题和兴趣取向，从而扩大了该丛书的读者面和社会影响。最后，指导书和工具书相统一。该丛书由于前述的学术含量高、学科前沿性和专业性，因而它是读者学习哲学的指南、指导书，同时，它也给我们提供了非常全面的参考书目，给读者提供了一种向导，使他们能够进一步地学习和研讨，是难得的工具书。

引进此套丛书并不仅仅是为了介绍一种新的教学方法和教材编写方法，可能更有用处的是，这套丛书能够在很大程度上改变我们学生的知识

结构。几十年来我们哲学教科书的知识结构已经陈旧老化，对于当代哲学的前沿问题知之甚少，与西方哲学界流行的话语体系距离越来越远。我们希望该丛书能够给我们学哲学的学生带来一些新的东西，改变一下哲学在他们心中的面目，让他们知道当代哲学正在讨论什么问题、用什么方法研究问题。我们期望，通过这套丛书潜移默化的作用，能够为中国新一代的哲学学人进行哲学的自主创新打下良好的知识基础，提供一些有益的启示。如果能够达到此目的，那么该丛书的“指南”或“指导丛书”的名字也就名副其实了。当然，该丛书也有明显的缺陷，它主要表现的是英语世界的哲学特别是分析哲学思潮的主要内容，而对于大陆哲学特别是现象学思潮、结构主义和后结构主义思潮、西方马克思主义思潮等则很少涉及。因而，还必须指出，该丛书提供的哲学及其知识结构也是不全面的，对于哲学的热点问题和前沿问题的观照也是有偏颇的。可以设想，如果是德国哲学家或法国哲学家编辑出来的哲学指导丛书，必定是给读者不同的指引和不同的向导。如前所述，文化是多元的，哲学是多元的，那么哲学的指南或指导丛书也会是多元的，就不难理解了。

2002 年春我率代表团访问了剑桥大学出版社、牛津大学出版社和布莱克韦尔出版公司，在访问期间就产生了翻译该套丛书的想法，2002 年底正式提出出版计划，2003 年初签订了版权合同。我组织我的一些同事、同学、朋友和我自己的博士生参与了该套丛书的翻译工作，对整套丛书的体例和质量要求提出一些基本原则和规范，我所做的工作更多地是一种行政组织管理工作，他们都是不同领域的专家，他们自己会对他们所翻译的东西负责的，我希望他们的翻译不要给读者留下太多的遗憾。

作者简介

弗雷德·亚当斯（Fred Adams）：特拉华大学哲学教授。

肯尼斯·艾扎瓦（Kenneth Aizawa）：路易斯安那世纪学院查尔斯·T·比尔德（Charles T. Beaird）哲学教授。

约翰·比克尔（John Bickle）：辛辛那提大学哲学教授和研究生神经科学课程教授。

大卫·J·查默斯（David J. Chalmers）：亚利桑那大学哲学教授。

安迪·克拉克（Andy Clark）：印第安纳大学哲学教授和认知科学项目主任。

伦道夫·克拉克（Randolph Clarke）：佐治亚大学哲学副教授。

保罗·E·格里菲思（Paul E. Griffiths）：匹兹堡大学历史和科学哲学教授。

约翰·海尔（John Heil）：戴维森学院保罗·B·弗里兰（Paul B. Freeland）哲学教授。

斯蒂芬·劳伦斯（Stephen Laurence）：谢菲尔德大学哲学高级讲师。

柯克·路德维希（Kirk Ludwig）：佛罗里达大学哲学副教授。

威廉·G·莱肯（William G. Lycan）：北卡罗来纳大学小威廉·兰特·凯南（William Rand Kenan Jr.）哲学教授。

埃里克·马戈利斯（Eric Margolis）：赖斯大学哲学副教授。

安德鲁·梅尔尼克（Andrew Melnyk）：密苏里大学哲学副教授。

肖恩·尼科尔斯（Shaun Nichols）：查尔斯顿学院哲学副教授。

埃里克·T·奥尔森（Eric T. Olson）：剑桥大学哲学讲师兼丘吉尔学院研究员。

霍华德·鲁宾逊（Howard Robinson）：中欧大学哲学教授。

斯蒂芬·P·斯蒂克（Stephen P. Stich）：罗格斯大学哲学和认知科学

理事会教授。

特德·A·沃菲尔德（Ted A. Warfield）：圣母玛利亚大学哲学副教授。

罗伯特·A·威尔逊（Robert A. Wilson）：阿尔伯塔大学哲学教授。

目　录

作者简介 …… (1)

导论 …… 特德·A·沃菲尔德 (1)
第1章　心身问题：一个综述 …… 柯克·路德维希 (5)
第2章　心身问题 …… 威廉·G·莱肯 (56)
第3章　物理主义 …… 安德鲁·梅尔尼克 (76)
第4章　二元论 …… 霍华德·鲁宾逊 (99)
第5章　意识及其在自然中的位置 …… 大卫·J·查默斯 (118)
第6章　思维及其内容：自然化语义学 …… 弗雷德·亚当斯 (163)
第7章　认知建筑学：诸认知表征的结构 …… 肯尼斯·艾扎瓦 (197)
第8章　概念 …… 埃里克·马戈利斯，斯蒂芬·劳伦斯 (217)
第9章　心理因果关系 …… 约翰·海尔 (244)
第10章　民间心理学 …… 斯蒂芬·P·斯蒂克，肖恩·尼科尔斯 (268)
第11章　个体主义 …… 罗伯特·A·威尔逊 (290)
第12章　情绪 …… 保罗·E·格里菲思 (326)
第13章　人工智能与理性的多面性 …… 安迪·克拉克 (348)
第14章　心灵哲学与神经科学 …… 约翰·比克尔 (362)
第15章　人格同一性 …… 埃里克·T·奥尔森 (395)
第16章　自由意志 …… 伦道夫·克拉克 (416)

索引 …… (458)
译后记 …… (482)

导　论

特德·A·沃菲尔德（Ted A. Warfield）

本书是布莱克韦尔哲学指导丛书中的又一卷。[1]它收录了16篇新文章，涵盖了当代心灵哲学研究中很多领域的问题。作者们应邀就他们的专题作出有见地的综述，并以他们自认为合适的方式来谈论这些专题。这使他们在介绍所承担专题的同时，也能自由地对所讨论的问题作出个人的学术贡献。我希望最终的成果能适合不同层次的心灵哲学课程使用。对于想全面了解当代心灵哲学中核心问题的专家或非专业人士来说，本书应该是一个合适的资源。在这篇简短的导论中，我打算解释为什么心灵哲学似乎是一个重要的哲学分支领域的一些理由。我还要解释一下我对人们在心灵哲学内部所发现的巨大差异性之根源的看法。这一讨论将包含对心灵哲学家和一般哲学家所面临的方法论问题作出的一些评论。[2]

哲学家们几乎都会赞成这一主张：心灵哲学在当代哲学中是最活跃、最重要的分支领域之一。至少从20世纪70年代后期起，心灵哲学似乎就占据了这一地位。很多人提出并辩护了下述更强的主张：心灵哲学显然是当代哲学中**最**重要的分支领域。它的地位至少可以归因于两个相关的因素：主题的重要性和领域的多样性。

心理现象在大多数——如果不是所有——人类活动中确实非常重要。仅举几个例子来说，我们的希望、梦想、恐惧、思想和愿望都是我们生活中最重要的组成部分。有的人主张，心性（mentality）对于人性是必不可少的，至少某种心理生活是成为人或者成为完整的人所必需的。还有人主张，人的心性的特定特征（很可能是人的理性）将人与其他有心灵的生物区别开了。不管这些雄心勃勃的主张正确与否，至少心理对于我们的生活极其重要。谁会否认思想、情绪以及其他心理现象处于所有与我们相关的重要之物的核心呢？这个显而易见的事实只是部分地解释了心灵哲学的重要性。这个领域的范围和多样性也使它值得拥有这一地位。[3]

只要翻一翻本书的目录，我们马上就会对这个领域的范围有所了解。[4]除了论述当代心灵哲学的核心问题——例如心理内容、心理因果关系和意识——的文章之外，我们还会发现有些文章将心灵哲学与各种广泛的经验研究结合了起来。在这种具有经验取向的研究所涵盖的领域中，哲学家接触到了诸如情绪、概念等广泛的经验性心理学研究。哲学与神经科学和人工智能的交叉也是当代人很感兴趣的话题。与这种具有经验取向的研究相比，我们也看到一些文章涉及传统的哲学问题，例如心身问题、人格同一性，以及自由意志。这些问题（尤其是后两个）经常被归于当代形而上学的一部分，但它们在传统上是心灵哲学的一部分，因此它们也被收录在本书中。

尽管最初将这些研究分为“传统的”或者“具有经验取向的”，但是不应当把这种区别看成楚河汉界。在研究传统问题的同时对相关的经验研究保持敏感是可能的，在从事具有经验取向的心灵哲学时使用传统的哲学工具——例如某种概念分析——或许是**必要的**。一个人在这个领域所发现的并不是非常准确的方法论分歧。相反，他发现的是不同哲学家所认为的经验研究与心灵哲学在相关程度方面的差异，以及哲学家们努力避免传统哲学分析的程度方面的差异。[5]

只看出现在心灵哲学中的课题和人们为那些计划所采取的不同方案还不足以完全把握这个研究领域的范围和多样性。除了众多的问题和对这些问题所采用的不同方案之外，我们在这个领域中还发现了一系列令人有些吃惊的不同解释对象。从事心灵哲学的哲学家主要感兴趣的或许是理解或者解释**人**心（human mind），或者更恰当地说，人心的一些特征。换句话说，有的人可能对考察（人或其他生物的）“心性”或“精神性”（mindedness）更明显的抽象本质感兴趣。有的人或许还关注**我们**关于人心的**概念**，或者**我们**关于一般心灵的**概念**，他对于我们关于这些东西的概念怎样与这个主题的实在性相联系或多或少有某种特殊的看法。[6]这些（可能的）不同研究目标看起来至少会导致非常不同、形形色色的问题。然而，尽管有这些明显的不同，但这个具有巨大多样性的计划仍相当舒适地处于“心灵哲学”这个大标题的保护之下。

在一个人认识到可以将各种研究目标和不同的方法论混合搭配起来时，心灵哲学的这种多样性就会更加一目了然。有些人的兴趣可能主要是对我们关于人心的概念进行经验研究。换句话说，有些人的兴趣可能是就同一个主题进行广泛的概念研究。这些不同的方法论（再回想一下，这些

不同最好被看成是程度的而非本质的不同）也可以在研究人心的本质或者心性的本质时使用。

有些人可能以为，方法论的争论之所以出现，是因为哲学家们对心灵哲学中的不同课题采取了不同的方案。例如，赞成传统的先验方法论的哲学家可能会质疑这样一些具有经验取向的哲学家，他们声称：要得到关于人心本质的结论，首先得作经验的研究，以此说明他们怎样弥合人心实际所是的样子（the way human minds are）与人心必须所是的样子（the way they must be）之间所存在的明显鸿沟。同样，具有经验取向的心灵哲学家可能要求支持先验方法的哲学家解释：为什么他们认为这些方法能得到关于某些概念（这些概念是那些从事分析的人的概念）的结论。例如，为什么我们应当认为分析我们的心灵概念会对心灵有所揭示？或许，批评会继续下去：我们的心灵概念没有准确地反映心灵的本质。然而，令人遗憾并且吃惊的是，对这些方法论问题的讨论并不常见。[7]幸运的是，这些问题和相关的方法论问题也出现在其他哲学领域，而且在心灵哲学内外，理解和评论不同哲学研究方法的兴趣似乎在与日俱增。[8]

我们不强求本书的作者评论心灵哲学中的方法论问题。他们只是应邀以任何他们认为适当的方式，使用任何他们所选择的用于这项任务的方法论来介绍和讨论所承担的专题。除了思考这些杰出的文章所讨论的一阶的（first-order）哲学问题之外，也请读者们思考与这些讨论相关的方法论和形而上学的问题。这些思考可能有助于我们更好地理解在心灵哲学中所遇到的一些（甚至所有）问题。

【注释】

[1] 编辑这样一本书是不容易的。感谢各位作者对我们这项计划各个阶段遇到的困难所给予的程度不同的耐心和支持。尤其要感谢我的合编者的坚定支持和指导。对于出现在这篇简短导论中的一些问题的有益讨论，我要感谢我的同事利奥波德·施图本伯格（Leopold Stubenberg）和威廉·拉姆齐（William Ramsey）。虽然我未表达对我的雇主圣母玛利亚大学的谢意，但校方爽快地允许我在做这项计划期间使用计算机和打印机。

[2] 本书的开篇是两篇不同的论述心身问题的文章。在介绍这本书时，我压住了写第三篇类似文章的诱惑，而聚焦于一些结构问题和方法论问题。 xii

[3] 这些局部的解释加起来仍未对心灵哲学在当代哲学中的地位作出充分解释。例如，伦理学是极其重要的，也是一个巨大而多样的领域。我不能对心灵哲学的地位作出充分解释。虽然现在有点过时了，但泰勒·伯奇（Tyler Burge）的重要文章《语

言哲学与心灵哲学：1950—1990》（“Philosophy of Language and Mind：1950—1990,” *Philosophical Review*，101 [1992]，pp. 3－51）包含了一些与这个问题有关的有益思想。

[4] 但是实际上没有一本书能涵盖整个领域。另一个对本书有帮助的补充性资源是《布莱克韦尔心灵哲学手册》（*The Blackwell Companion to Philosophy of Mind*，edited by Samuel Guttenplan，Blackwell，1994）。

[5] 同一位哲学家对不同的问题或者在不同时间对同一个问题甚至可能采取不同的一般方法论思路。

[6] 人们很容易就能想象一个人如何作出结论：例如，我们的心灵概念在某种意义上是一个“二元论”概念，但并非由此可以推出二元论是心身问题上的正确立场。

[7] 最近在高度复杂的层次上有一些关于意识的争论，它们包括了一些沿着这些路线的方法论讨论，例如参见大卫·J·查默斯（David J. Chalmers）和弗兰克·杰克逊（Frank Jackson）的《概念分析与还原的解释》（“Conceptual Analysis and Reductive Explanation,” *Philosophical Review*，110 [2001]，315－360）。

[8] 任何希望考察这些问题的人，把迈克尔·R·德保罗和威廉·拉姆齐主编的《反思直觉》（Michael R. DePaul and William Ramsey eds.，*Rethinking Intuition*，Rowman and Littlefield，1998）作为起点都会有收获。

第1章　心身问题：一个综述

柯克·路德维希（Kirk Ludwig）

我曾经说过灵魂不过就是躯体，我还说过躯体不过就是灵魂，而且对于一个人来说，任何事物，包括上帝，都不会比他自身更伟大。

沃尔特·惠特曼

1.1 导　言

理解思想和感觉在自然界中的位置，在一般的自然理解和特殊的自我 1
理解——它们是科学和哲学的主要目标——中处于核心地位。这项计划从古希腊时期开始就引起了科学家和哲学家们的注意，其一般形式是由这个问题给出的："一般来说，心理现象与物理现象之间是什么关系？"人们对正确的回答还没有形成普遍的共识。这对我们理解自然界而言是一道最重要的鸿沟。麻烦在于这个问题给我们提出了另一个难题，对它的每一种**可能的**回答都会产生一些似乎让人难以接受的结果。这个难题传统上被称为"心身问题"（The Mind-Body Problem）。[1]这一章的主要目标是解释这个传统的心身问题是什么、其可能的解决办法是什么，以及在解决的道路上有什么障碍。

我的讨论分两个阶段进行。第一个阶段（1.2—1.4 节）考虑的是进一步阐明我们最初的问题作为解释其可能的解答的前提条件时所具有的重要意义。第二个阶段（1.5—1.6 节）对在我们努力回答我们所描述的问题时出现的一个难题作出解释，并考察能够并且已经提出来的各种解答。

更具体地说，1.2—1.4 节涉及的是如何理解我们最初问题的基本要 2
素。一方面，我们如何识别心理的东西；另一方面，我们如何识别物理的东西，以及我们对它们之间的何种关系感兴趣。1.2 节指出并解释了心理

现象的两个传统标志：意识和意向性（intentionality），讨论了它们是如何产生联系的。1.3 节解释了我们应当对最初问题中的“物理的”作怎样的理解，才不会将任何在心身问题上的传统立场排除在外。1.4 节接着提出了最初问题中的第三个要素，标出了心理现象与物理现象之间可能具有的基本关系，并指出了一些需要特别关注的关系。

1.5—1.6 节涉及对最初问题作出回答时遇到的困难根源给予解释，以及已经提出的各种解答。1.5 节解释了为什么最初问题导致了一个难题，并为心身问题提供了一种准确的形式，这种形式表现为一个由四个命题组成的集合，其中的每个命题单独提出来时似乎是令人信服的，但是把它们放在一起却又不相容。1.6 节对心身问题的各种回答进行了分类，其依据是，这些回答所要否定的是不相容集合中的哪个命题，这一节还简要概述每一类的主要类型以及每一类所产生的一些问题。1.7 节是就我们在理解心灵在自然界中的位置时的困难之源所作的一个简短结论。[2]

1.2 心理之物的标志

意识是心理之物的标志，这种看法至少可以追溯到笛卡尔。[3]意识是我们的心理生活中最为突出的特征。正如威廉·詹姆斯（William James）所说：“每个人首先会断定属于他内部经验的具体事实是这个事实：**某种意识在进行**。”（1910：71）按照这种看法，如果一个状态或者事件（一个物体状态的变化[4]）是**有意识的**，它就是心理的。反过来，状态是根据这些属性——物体拥有它们构成了物体自身的存在——个体化的。

把意识确定成心理之物的标志只是将我们的问题推后了一步。我们现在必须说明：是什么使某物成为有意识的。这是不容易办到的。有两个直接的困难。首先，用 G. E. 穆尔（G. E. Moore）的话说：“一旦我们打算将注意力集中于意识之上，想清楚地看看它是**什么东西**时，意识似乎就消失了：我们面前似乎是空无所有……它似乎是透明的。”（1903：25）其次，即使我们确定了意识的位置，仍不清楚它是否就能用其他的术语理解。就它所提出的一些实质性的看法仍是一些有争议的看法。然而，就现在的目的来说，以所有人都能赞成的方式指出我们的兴趣之所在就够了。那么，我现在的意见不是想提供一种意识分析，而是想让人们注意到这种现象，

并想对它作出描述，这与一个自然主义者完全相同：他通过标明一个成员，或者描述它的观察条件和特征，使人们注意某种昆虫或植物，即任何人在适当位置上都能够确认构成其特征的那些部分。

这样一来，首先，当我们醒着而不是处于无梦的睡眠中时，以及**在**我们做梦**时**，**我们**就是有意识的。当我们有意识时，我们就具有意识状态，我们可以辨别、记忆，也可以忘记这些状态。每种有意识的心理状态都是一种具有意识的模式或方式。关于我们有意识的心理状态的知识，即使在知觉经验中与世界的知识联系在一起时，也与世界的知识不同，因为正如这种可能性所表明的那样，即可能有难以区别但非真实的知觉经验。有意识的心理状态范例包括知觉经验、躯体感觉、本体感受、疼和痒，感觉悲伤或者愤怒、饥饿或者口渴，以及正在发生的思想和愿望。用托马斯·内格尔（Thomas Nagel）的话来说，一个有机体拥有有意识的心理状态，当且仅当“该有机体有成为是什么样子的某种东西”（1979b：166）。相反，人们通常认为，成为一枚脚趾甲、一把椅子或者一片草叶并没有相关意义上的像什么样子的东西。

据说在试图掌握我们在意识模式（或者具有意识的方式）之间所作的这种辨别时，意识状态有一种现象的或质的特征，有意识的心理状态的现象性质经常被称为“感受性质”（qualia）。有时候，人们这样看待和对待感受性质，即它们好像是觉知的对象，就像桌子和椅子是知觉的对象一样。但这是错误的。当一个人觉知自己有意识的心理状态或者它们的现象性质时，所述的对象唯有这个人自身，他觉知到的是该对象的一个特殊的变体，一种它具有意识的方式。同样，当我们看见一个红苹果时，我们看见的仅仅是苹果，而不是与之并列的另一种东西——红（redness），更确切地说，我们是将所看见的苹果表征**为**红的。

我们有意识的心理状态的一个突出特征是：我们对它们具有非推理的知识。当我们有意识时，我们知道我们有意识，我们也知道我们是**怎样**意识的，即我们的意识的模式，但是，当我们有意识时，我们不是从更直接的觉知或者独立认识的东西来推断我们有意识或者我们是怎样意识的。[5]如何说明这种非推理的知识实际是什么非常困难。也很难明白如何把它与我们所认为的有意识的心理状态的质的特征区别开。[6]这种“第一人称的”知识恐怕是**自成一格的**。在我们与自己的有意识心理状态和别人的有意识心理状态的关系中有一种相关的**不对称性**。我们不必**推断**自己是有意识的，但其他人必须推断——通常是从我们的行为中进行推断，而

且他们不可能非推理地认识。其他人对于我们的心理状态充其量只有“第三人称的”知识。意识状态的这些特殊特征与试图回答我们开篇问题时所产生的一些难题是联系在一起的。意识经常被看成是心身问题的核心之谜，被看成是对心理之物作出一种恰当的物理主义理解之主要障碍。[7]

心理之物的另一个传统标志，是弗朗兹·布伦塔诺（Franz Brentano，1955 [1874]，bk 2：第 1 章）首先明确地提出来的，它被称为“意向性”。[8]形容词形式是“意向的”（intentional）。但是这是一个技术术语，不只包含英语中被称为“意图”（intentions）的状态（例如，我再喝一杯咖啡的意图）。其实，意向性是一个状态或者事件使之**关于**（about）或者**指向**（directed at）某种东西的特征。进一步对此作出解释最好是举几个例子。与我写作时所坐的椅子不同，我有各种各样**关于**我自己、我周围的环境以及我的过去和将来的信念。我相信今天结束前我会再喝一杯咖啡，但我的椅子没有相应的信念，其他任何东西都没有。信念是典型的意向状态，它们将世界表征为某种样子，可能对也可能错。这是它们特殊形式的**满足**条件。用约翰·塞尔（John Searle）的贴切用语来说：它们具有心灵到世界的适应指向（mind-to-world direction of fit，1983：第 1 章）。它们应该适应世界。任何具有心灵到世界的适应指向的状态、任何表征状态或者态度，都是一种意向状态（在这种技术的意义上）。错误信念和正确信念一样是意向状态，即使世界上并不存在它们所表征的这种东西供它们指向。尽管不存在独角兽，但我能够**思考**独角兽。表征在没有表征对象时也能存在。在思考意向性时，一般的分歧就是这种意义的“关于性”（aboutness）或“指向性”（directedness）。

除了信念之外，还有一些意向状态具有心灵到世界的适应指向，如期望、推测、确信、意见、怀疑等等。然而，并非所有意向状态都有心灵到世界的适应指向。另一个重要类型的范例是愿望或者需求（want）。我相信我将并且也想马上再喝一杯咖啡。这个愿望也是指向或者关于世界的，而且甚至比在信念的情况下更明显，它尚且不需要世界上有任何对应的东西。但是与信念相反，它的目标不是使它的内容（我马上再喝一杯咖啡）与世界相符合，而是**使世界与它的内容相符合**。它有世界到心灵的适应指向。愿望可能得到满足也可能未得到满足，就像信念可能对也可能错一样。这是它满足条件的特殊形式。任何具有世界到心灵的适应指向的状态都同样是意向状态。

显然，信念和愿望之间可能有共同之处。我相信**我将马上再喝一杯咖啡**，以及我愿望**我将马上再喝一杯咖啡**——这些有共同的内容，而且正是由于它们的内容，各自才是一个意向状态。（内容之间的共同因素——是用一个一般词项表达的——通常被称为“概念”，因此，人们说**咖啡**这个概念是**咖啡是一种饮料**这个信念的内容，也是**咖啡包含咖啡因**这个信念的内容的一个组成部分。）每一个信念的内容都与世界符合或者不符合。信念和愿望的区别在于它们在我们心理系统中所起的作用：它们的目标是改变其内容以使之符合世界（信念），还是改变世界以符合它们的内容（愿望）。这些状态具有我们能用句子表达的内容，与此类似的状态被称作“命题态度”（propositional attitude，这是伯特兰·罗素［Bertrand Russell］提出的一个术语，它位于假定的态度、命题对象之后，是用“that p”这种形式的短语来命名或者表示的，这里的“p”可用一个句子替换）。对命题态度进行个体化，依据的是它们的心理模式（信念、推测、怀疑、愿望、渴望等等）和内容。具有世界到心灵的适应指向的状态是**正面的态度**（pro attitude）或者——如果是否定性的——**负面的态度**（con attitude）。除了愿望和需求之外，还有很多其他种类，例如希望、恐惧、喜欢、反感等等。

还不能确定所有表征内容都完全是命题性的。我们的知觉经验，如我们的视觉、听觉和触觉经验，把我们的环境表征为某种样子，它们可能是**真实的**（正确地表征），也可能是**非真实的**（不正确地表征），就像信念可能对也可能错一样。它们具有心灵到世界的适应指向，从而具有表征内容和意向性。但是还不能确定它们所表征的一切都能从命题上来把握。态度和知觉经验或许可以看成是不同的货币，对于它们没有准确的兑换标准。

可能会存在指向或者关于某种没有完整内容的东西的状态吗？有的人可能害怕蜘蛛却没有任何指向特定蜘蛛的愿望，尽管这种恐惧在某种意义上是指向或者关于蜘蛛的。然而，害怕蜘蛛确实隐含着不接触或者不接近蜘蛛的愿望，这一点加上想起或者感知蜘蛛所引起的特定情绪气氛，就是我们所认为的对蜘蛛的恐惧。无论如何，我们把这类状态也称为意向状态，尽管它们的意向性似乎是以表征的态度（正面的或负面的态度）为基础的，这些态度支撑了它们，或者如我们所说，它们依赖于这些态度。

于是，我们可以说：一个意向状态就是一个具有内容的状态（在我们所描述的意义上），或者是一个依赖于这样一个状态的状态（在刚才指出的意义上）。[9]

这样一来，一个状态是一个心理状态（或事件），当且仅当它是一个

意识状态（或事件）或者是一个意向状态（或事件）。一种东西是一种能思维的东西，当且仅当它有心理状态。

意识状态和意向状态之间是什么关系？如果这两者是无关的，那么我们最初的问题就分解成了两个子问题：一个涉及意识与物理现象的关系，另一个涉及意向性与物理现象的关系。如果这两者不是彼此无关的，那么对总问题的任何回答都肯定会同时解决两个子问题。

有些意向状态显然不是意识状态。你关于澳大利亚位于大洋洲的信念，不是刚才出现的一个有意识的信念（或者一个**正在发生的**信念），那时你没有**想到**它，尽管你相信它，它是一个**倾向性的**（dispositional）而非一个正在发生的信念。这种区别可以推广到所有态度类型，一个愿望可能正在发生，如我现在想喝一杯咖啡的愿望，它也可能是倾向性的，如我在毫无征兆的情况下产生买某本书的愿望。[10]然而，这并没有解决意向的心理状态与有意识的心理状态是否无关的问题。我们把倾向性的心理状态看成意向态度的一个必要条件可能是：在它们的表现属性（manifestation properties）当中，有些正在发生的态度具有相同的模式和内容。在这种情况下，**分而治之**的策略是行不通的，我们无法将理解意向状态和理解意识状态的计划分开，并各自作出单独的处理。[11]

有些有意识的心理状态似乎没有意向性，例如那些愉快或者焦虑的情景（episodes）。尽管它们通常是由我们的信念和愿望引起的，但不能确定它们本身是关于什么东西的。同样，躯体感觉，如痒和疼，似乎有非表征的因素。躯体感觉通常表征的是人体中正发生的东西。头疼表征在头部，脚趾疼表征在脚趾上。但是，尽管疼痛性质本身被看成是（比如说）身体损伤的一种生物学标志，但它就像烟指示燃烧一样，似乎没有任何相关的表征内容。疼痛不**表征**（与指示相反）损伤，而且，尽管我们通常都希望我们的疼痛停止，但希望疼痛停止的愿望——它有表征内容——并不是疼痛本身，正如想换一座更大房子的愿望并不是房子本身一样。[12]

1.3 物理的东西

按照我们最初问题的意图来描述物理现象，并不像看起来那么容易。我们不能说物理现象就是当前物理学所谈论的东西，物理理论不断变化，当前的物理理论可能在未来遭受根本的修改，就像过去的物理理论那样。

心身问题没有随着一时的物理理论而改变，这里至少还有三种选择。

首先是把物理现象描述成最终正确的物理理论所谈论的东西，这里我们把物理理论看成这样一种理论，它向我们说明了事物的基本成分及其属性。其次是依据定义把物理现象看成非心理的现象。我们有理由认为这两者都没有把握我们最初问题的含义。

对心身问题的一种解释是：事物的诸基本成分具有不可还原的心理属性。根据第一种解释，这一立场将被归属于物理主义的一种版本（我们将在1.4节的末尾对此作出准确的描述），因为它认为心理属性在相关意义上就是物理属性。但是常见的看法是：这种主张事物的基本成分具有不可还原的心理属性的观点与物理主义是水火不容的。

对于主张心理现象可以从概念上还原为物理现象的看法，第二种解释一开始就没有袖手旁观。如果物理的东西本身是非心理的，那就说明心理属性真是属于范畴F中的属性，这不过是说明范畴F中关于属性的一个子范畴不是物理属性。但是，我们需要这样一些术语，它们是我们陈述最初的问题，悬置心理属性能否从概念上还原为物理属性这个问题所必需的。（我们在后面再讨论这一点。）

第三种选择是把物理现象看成是我们当前的物理学所例示的一个一般种类。这里，我们旨在描述一类属性，它们包含了自科学革命以来过去的和当前的物理理论所诉诸的属性，但它们很广泛，足以涵盖我们当前用于解释物体动力学方案范围内所诉诸的属性。这种解释未解决我们前两种解释所排除的选择，而且与现代早期以来有关心理现象和物理现象关系的认识发展是一致的。我们不容易说明如何描述这类计划中的属性。有关它们的核心概念是由这些东西决定的，即在17、18世纪被归结为第一性质的大小、形状、运动、数目、固体性、构造，以及这些性质的逻辑结构，还包括根据它们对这些性质（例如，质量和电荷大概属于后一类）的作用而从本质上予以描述的那些属性。[13]仍不能肯定的是，这足以涵盖我们希望囊括的一切。但是可以公正地说，通常哲学家们在提出我们最初的问题时所想到的就是这个物理之物概念，即便对于如何描绘相关类型的属性还没有具体概念。[14]

1.4 心身关系

心理与物理的关系问题，可以在同等意义上被看成是关于心理的和物

理的属性、概念或谓词的问题。一种属性就是物体的一种特征，例如，圆形，或者距离地面三英尺。如我们所述，一个概念就是一个一般词项所表达的不同思想内容中的一种共同因素，我们在思考一个事物的属性时会使用概念。因此，和圆这种属性相对应的是“是圆的”或者“圆”（roundness）这个概念，当我想这个球是圆的，因而认为它有圆这种属性时，我就有一个思想，它包含有“是圆的”这个概念，人们说我把球置于圆的概念之下。谓词表达概念，并被用于向物体归属属性。[15]因此，“是圆的”（在英语中）表达了圆的概念，并被用于归属圆这种属性，我们可以说它挑选了那种属性。对每种属性来说，都有一个关于它的独特概念，反之亦然。多个谓词能表达同一个概念，挑选同一种属性，但是它们必须是同义的。[16]与每个属性类型（例如心理的或物理的）相对应的是概念和谓词的类型。因此，就心理属性和物理属性的关系来说，我们所问的任何问题都可以改写成关于概念或谓词的问题，反之亦然。

在思考心理属性和物理属性的关系时，基本选项可以依据下列三个句子的形式来解释，这里“是 M”表示一个心理谓词，而“是 P”表示一个物理谓词（这可以直接推广到关系词项）。

[A] 对于所有 x，如果 x 是 P，那么 x 是 M
[B] 对于所有 x，如果 x 是 M，那么 x 是 P
[C] 对于所有 x，x 是 M，当且仅当 x 是 P

8 尽管 [C] 等价于 [A] 与 [B] 的合取，但分别陈述仍是有用的。决定心理和物理的关系所依据的是：[A] — [C] 哪种情况是正确或错误的，以及判断的理由是什么。有的人可能认为，根据“必然性”的三种含义之一，即概念的、形而上学的（所谓的）和法则学的（nomological），其中每一个都必然正确或者必然错误。

在讨论心身问题的时候，有两个观念可以用这个框架来描述。第一个是还原的观念，第二个是随附性（supervenience）的观念，每个都可以是概念的、形而上学的或者法则学的。我首先考察概念还原和随附性。

概念的必然性是基于表达概念之上的真理，这是最强的一种必然性。概念上必然的东西在所有形而上学和法则学的可能世界里也都是必然的，但反之并不然。概念真理的知识可以从反思所涉及的概念获得，而且无需依赖于经验（传统上，关于人自己有意识的心理状态的知识被视为经验知识）。人们因此认为它们是能先验地认识的，这样获得的知识就是先验知

识。根据经验所认识的命题是后验地（a posteriori）或者经验地认识的，以此为基础的知识是后验的或经验的知识。概念真理不能根据任何经验的内容来否定，表达（在一种语言 L 中）概念真理的句子是在**分析的意义上为真（在 L 中）**，或者在相同意义上，是**分析的（在 L 中）**（我在下文中省略了这种相对关系）。一个句子是分析的，当且仅当它的真是由关于其成分的真意义陈述衍推的。[17] 例如，“没有一个都柏林的居民住在外地”或者“没有最大的素数”通常都被看成是分析的。[18]

心理属性向物理属性的概念还原，或者相反的情况，是它们之间所能得到的最强关系。（在这种情况下，我们在相同意义上说：心理概念/谓词可以依据物理概念/谓词来分析，反之亦然。）如果一个心理属性能从概念上还原为一个物理属性，那么要满足两个条件：（a）[C] 的情况在概念上是必然的，在这里，“是 M”被一个挑选了心理属性的谓词替换了，而“是 P”被一个挑选了物理属性的（或许很复杂的）谓词替换了；（b）“是 P”所表达的概念在概念上优先于“是 M”所表达的概念，也就是说，要理解“是 M”所表达的概念，我们必须具有“是 P”所表达的概念，但反之并不然（可以把我们建立几何概念的顺序作为一个例证）。第二个从句为下述观念赋予了内容：我们影响了一种还原，因为相比心理概念而言，它对物理概念的要求更基本。心理属性向物理属性的概念还原表明：心理属性就是一种物理属性。这相当于把心理属性与物理属性同一起来，物理属性向心理属性的还原也同样如此。

有的人可能相信，[C] 的情况在概念上是必然的，但不认为心理的东西或者物理的东西能从概念上还原为另一方。在这种情况下，它们的必然关系将借助于另一组既非物理的也非心理的概念来解释，每一方都能根据这些概念来理解。例如，每个三角形都有三条边，这在概念上是必然的，但这两个概念都未提供关于另一方的概念还原。

“随附性”是当代论述心身问题的很多哲学文献所使用的一个专门术语。有人可能怀疑它是讨论心身问题所必需的，但鉴于它在当代的普遍使用，目前的心身问题研究都不该不提到它。一类相关观念已经用它予以表达。虽然它们自身之间仍有强度的差异，但它们一般都想表达比还原论弱的命题，它们只使用充分条件，而非充分必要条件。[19] 人们认为，随附性主张不是要提供解释，而是要依据一种属性对另一种属性的解释**形式**施加限制。

我在这里介绍一类属性的定义，这类属性随附于另一类属性，这个定

义对于陈述我们所称的“物理主义”立场是有用的，而且在我们后面讨论（被称为“功能主义”的）关于心理属性与物理属性关系的立场时也是有用的。首先看“概念随附性”。

> F类属性**在概念上随附于**G类属性，当且仅当对于任何x，如果x有一个来自于F的属性f，那么就有一个来自于G的属性g，以至于x有g，并且如果x有g那么x有f，这**在概念上**是必然的。[20]

一类属性向另一类属性的概念还原蕴含了相互间的概念随附性。但是，一类属性对另一类的随附性并不蕴含着它们之间的可还原性。

我将把“物理主义”描述成这种立场：根据它，**不管物体有什么心理属性，这些心理属性都将在概念上随附于物体所具有的物理属性，而且不管有什么心理规律，物理规律都蕴含这些心理规律**。[21]这会使有的人——他认为没有什么东西具有心理属性，不存在心理规律——被看成物理主义者，而不管他对心理属性和物理属性之间的概念关系持什么看法。[22]这个定义在这里是约定的（stipulative），但它想追踪在论述心身问题的哲学文献中流传广泛的（但不是普遍的）一种用法。[23]这样理解的物理主义是否正确？这个问题标志着在有关心身问题立场上的根本分歧。

我们可以依据概念的必然性和自然律的概念来解释法则学的必然性。一个陈述p是法则学上必然的，当且仅当：如果L，那么p是事实。这在概念上是必然的，这里的“L”代表一个表达所有自然律——不管是不是物理的——的句子（给“L”加上“边界条件”会产生更严格的概念）。我对形而上学的必然性只给出了一种否定的描述，这种必然性在当代心身问题讨论中受到了相当多的关注。我将在1.6节指出，没有一个概念是与这些语境下“形而上学的必然性”这个表达式相对应的，尽管它已被广泛应
10 用。目前，我们可以说：形而上学的必然性应该是一种不能先验地发现的必然性，但是它比法则学的必然性强，比概念的必然性弱。为了得到**形而上学随附性**和**法则学随附性**的对应概念，我们可以用“在形而上学上”或“在法则学上”替换我们上述的“在概念上”。

形而上学还原和法则学还原要求［C］这种形式的双重条件在形而上学上或者在法则学上分别是必然的（但不要求更强），但是还原是不对称的。因此，我们也必须为这种观念提供一种意义，即这个双重条件的一边表达了更基本的属性。实践中的问题是：如何理解心理向物理的形而上学还原或者法则学还原的不对称性。就形而上学的必然性或者法则学的必然

性来说，不存在与概念的优先性（priority）对应的东西。看起来我们最多只能将想要的这种不对称性建立在物理属性的基础之上，物理属性在我们的总解释图式中是基本的。人们通常认为，这意味着物理的东西构成了一个解释上封闭的系统，而心理之物并不如此，这意味着每个事件都能借助物理的前提而不是心理的前提来解释。

1.5 心身问题

一个哲学问题就是我们在思考某个基本问题时难以破解的一个纽结。通常，这包含格外难以理清楚的概念问题。由于哲学问题涉及基本问题，因此我们怎样解决它们对于整个研究领域的理解都有重要意义。一个哲学问题往往可以表现为一组命题，所有命题在最初考察时似乎都是正确的，或者对所有命题都有强有力的根据，但它们放在一起却又不相容，这就是意志自由问题和关于外部世界的怀疑论所产生的方式。如果我们能提出这样一个问题，那就是一个重大的进步。因为使我们的看法恢复一致的方式，决定着解答这个问题的逻辑空间，心身问题就可以这样提出来。因此，历史上和当代有关心理与物理关系的立场，可以依据它们不恢复哪些命题的一致性来进行分类。

这个问题是借助下述四个命题产生的：

1. **实在论**。有的东西有心理属性。

2. **概念的自主性**。心理属性不能从概念上还原为非心理属性，因此没有一个非心理命题蕴含任何心理命题。[24]

3. **成分解释的充分性**。依据其基本成分、成分的非关系属性[25]和成分间的关系[26]，以及它们与作出类似描述的（成分描述）东西的其他基本成分的关系，对一事物的完整描述蕴含它的完整描述，即从一事物的成分描述得出的关于它所有属性的一种解释。

4. **成分的非心理主义**。事物的基本成分本身并没有心理属性。[27]

现在就可以准确地陈述逻辑的难题了。命题（2）—（4）蕴含对（1）的否定。因为如果正确的基础物理学不援引任何心理属性，即（4），并且每种自然现象（即每种现象）都能从一个事物的描述——它依据的是其基本

成分及其排列——演绎出来，即（3），那么，假如没有一个非心理命题蕴含任何心理命题，即（2），我们就能推出：**没有任何东西有心理属性**，这就相当于否定了（1）。

倘若不是由于这个事实，即（1）—（4）中的每一个都对我们有强大的吸引力，那么这个逻辑难题将是容易解决的。

命题（1）看起来**显然**是正确的。我们对自己有意识的心理状态似乎有直接的非推理知识。我们在解释自己的所作所为时，彼此都归属心理状态，并且将我们对其他人将要做什么的预言部分地建立在我们关于他们有什么态度、他们的意识状态是什么的信念基础之上。放弃（1）似乎是不可思议的。

命题（2）是由这样一个躯体的显而易见的可理解性强有力地支持的：它的行为与一个能思维者的行为类似，但它没有我们从自己的观点所觉知到的这种心理生活。我们推测自己的心理状态引起了我们的行为，这种行为是由其他原因引起的，这似乎能想象。的确，似乎可以想象，它确实是由我们身体的这些物理状态引起的，我们有独立的理由认为这些物理状态使我们的身体具有生命却没有伴随的意识合唱。它同样得到了非物质的能思维者（如上帝和天使，甚至无神论者通常也认为它们是可以想象的）明显的可理解性的支持。因此，初看起来，对于具有我们的这种心理生活来说，拥有一个物质的身体似乎在概念上既不必要也不充分。

思想实验要求我们设想一种与事实情况相反的可能情况，并自问，某个事态因此能够得到，这是否只是看起来像有道理的（不只是它能否与自然规律相容）。我们通常都是这样检验概念联系的。例如，我们可以自问，我们能否想象一个物体，它是红色但没有广延，回答是“不能”。我们同样可以问，我们能否想象一个物体，它是红色并有企鹅那样的外观，回答是“能”。这表明：第一个是概念上不可能的——它被其描述中所包含的概念排除了，第二个是概念上可能的——未被所包含的概念排除。可能没有一个人会怀疑这里的结果[28]，但是我们有可能被误导。例如，想象这样一个集合是很容易的：它含所有不含自身的集合，并且只含这样的集合（罗素集合）。因为很容易想象一个不含集合的集合和一个只含集合的集合，因此也似乎很容易想象一个特殊的关于集合的集合，它的元素只是那些不含自身的集合。但是有可能证明这会导致矛盾。把所有不含自身的集合所组成的集合称为“R”，如果 R 是 R 的一个元素，它就没有满足 R 的

12

元素条件（membership condition），从而不是自身的一个元素。但是，如果它不是其自身的一个元素，那么它就满足了元素条件，从而就是自身的一个元素。因此，它是自身的一个元素，当且仅当它**不**是自身的一个元素，这是一个矛盾而且必然错误——不可能有这样的一个集合。[29]因此，我们似乎能想象某种东西，即使在它是不能想象的时候。按照这种观点，对有的人来说可以提出这样的反驳：尽管这些思想实验表面的可理解性支持命题（2），但是我们在从它们得出结论时已经犯了某种错误。[30]

科学在依据控制其成分的规律解释复杂系统的行为方面所取得的成功支持了命题（3）。虽然对于微观现象与宏观现象的关系，我们还有很多东西不理解，但是看起来迄今得到成功运用的技术，都可以推广到复杂系统的这些特征，我们依据其成分的属性还不能完全理解它们——可能心理现象是个例外。命题（3）表达了一种思想，它对我们的这种科学世界观具有强有力的意识形态掌控，这种科学世界观可以理解成一种庞大的机器、一个复杂系统，对它的完整理解可以通过分析其结构以及控制其各部分属性的规律来实现。用威尔逊的话说就是："它已经在实验和逻辑的酸液池中得到了检验，并且再三得到了证明。"（E. O. Wilson，1998：5）这一思想推动着很多科学研究，它依然是我们当代科学世界观的一个核心的方法论原则。就有些物体的特征不能从它们的构成方式来解释而言，我们也可以承认自然含有某种基本的任意性因素。

最后，物理学（迄今为止）在不借助任何心理属性来解释其领域中的现象方面所取得的成功也支持了命题（4）。在粒子物理学的属性清单中，我们发现了质量、电荷、速度、位置、体积、旋转等等，但是没有什么对心理有丝毫暗示，也没有任何东西看起来是解释最小物质的相互作用和动力学所必需的。[31]甚至看起来也很难理解，什么是将心理属性归属给物质的最小成分，这些成分没有任何心理活动的外部标志。

这就是心身问题。命题（1）—（4）似乎都正确。但是，它们不能都正确，因为它们是不相容的。这就是我们最初的问题——一般来说，心理现象与物理现象之间是什么关系？——为什么导致了一个哲学难题的原因。我们有可能给出的每种回答都包含着要在（1）—（4）中否定我们的一个命题，然而，如果单独考虑，我们似乎有合适的理由接受这些命题中的每一个。

1.6 解答的逻辑空间

已提出的心身问题的解答，可以依据它们反对恢复（1）—（4）中的哪一个的一致性来分类。只有四种基本的立场，因为我们追求最低程度的修改。要否定（1），就要接受关于心理的**非实在论**或者**取消主义**（eliminativism）；要否定（2），就要接受关于心理之物的**概念还原论**，它包括中立一元论（neutral monism）、心理物理同一论、功能主义以及外在主义的功能主义（functionalism-cum-externalism）；要否定（3），就要接受**概念的反还原论**，而非本体论的反还原论，中立的突现论和突现的唯物主义就属于这一类；要否定（4），除了接受概念的反还原论之外，还要接受**本体论的反还原论**。这包括各种所谓的“心理粒子理论”（mental particle theories），还包括实体二元论、唯心主义、泛心论、两（或双）方面论（根据某种概念），以及我所说的“特殊粒子理论”。

我按相反的顺序来考察每一种理论，因为这展示了它们的历史发展。我主要讨论现代初期以来对心身问题的看法，尽管事实上除取消主义之外的所有基本立场在古代就都被预言了。[32]

1.6.1 本体论的反还原论

否定命题（4），即事物的基本成分的非心理特征，是历史上最流行的立场。其总的看法可以称作“心理粒子理论”，根据它，事物的某些**基本成分本身**就有心理属性。依据认为心理粒子是只有心理属性，还是既有心理属性又有物理属性来看，这些观点可以进一步分为**纯粹的**心理粒子理论和**混合的**心理粒子理论，另外，依据是所有事物还是仅仅部分事物有心理属性而言，我们还可以再划分（**普遍的**和**受限的**）。

这种理论最突出而且历史上最重要的观点是实体二元论，它要追溯到古代关于灵魂是一种简单实体的观点。[33]实体二元论认为，世界既有物质实体也有心理实体，前者只有物理属性，没有心理属性；后者只有心理属性，没有物理属性。这是一种受限的、纯粹的心理粒子理论。笛卡尔（1985［1641］）是现代早期最杰出的二元论支持者。二元论的吸引力部分地在于它能为不可还原的心理属性在世界上找到一个位置，这个位置似乎在很大程度上可以解释成一个机械系统，它可以还原为它们自身根据其第

一性的质来彻底描述的部分。笛卡尔著述于科学革命的初期，他本人就是这种新“机械论哲学”的主要倡导者，这种哲学的基本假设为现代物理学提供了基本假设。二元论是笛卡尔对机械论哲学提出的在自然界中为心灵寻找位置的问题所作的回答。

笛卡尔对西方哲学传统的发展具有重大的影响，因此有必要简要地回顾一下他对二元论的正式论证，这会为随后的心身问题讨论打好基础。然而，要解释笛卡尔的论证过程，我们首先必须对实体概念作进一步说明。这个概念在17、18世纪的哲学讨论中处于中心地位[34]，它要追溯到亚里士多德的这一描述，即它“既不述说一个主体也不存在于一个主体中”（《范畴篇》1b2-5；1984：4）。这是作为属性载体的实体概念，即经历变化并在变化中保持不变的东西，“实体……在数目上保持单一，能够容受相反的性质……有时白，有时黑”（《范畴篇》4a19-21；1984：7）。这在中世纪哲学（经院哲学，即重新发现亚里士多德的著作所产生的传统）中导致了实体是独立存在物的观点，因为它们和属性相反，而属性被认为只存在于主体之中，不是独立存在的。笛卡尔对实体给出了两种描述，一种是，实体绝对不依赖于一切别的东西。这普及了经院哲学的概念。笛卡尔认为，根据这一概念，上帝是唯一的实体，因为一切东西的存在都依赖于上帝。但是笛卡尔承认，在辅助的意义上实体是属性的载体，他承认，除了上帝之外，有两种根本不同的实体——思维实体和物质实体（《原理》，1644，I.51—52；载1985，第1卷：210），我此后把注意力限制于后一种。笛卡尔实体种类理论的一个核心特征是每个不同的实体种类都有一种主要的个体化属性，这种实体的其他一切属性都是它的一个变体，就物质实体来说，这种属性是**广延**，就思维实体来说，是**思想**（《原理》，1644，I.53—54；载1985，第1卷：210-211）。这种理论的这一特征在引导性讨论中经常被忽视，但它对于正确理解笛卡尔实体二元论论证的说服力是不可或缺的。

每种实体都有一种主要属性，这个学说促使一个实体种类个体化的、本质的属性成为了某种东西的一种基本方式或范畴属性。一种范畴属性就是一种可确定的（determinable）而非确定的（determinate）属性。一种可确定的属性就是一个物体能以不同方式并且一定以某种特定方式具有的一种属性，例如，有颜色。某种东西有颜色，可能是由于它是蓝的、绿的或红的等等，而且如果它有颜色，一定是以某种确定的方式而有颜色的（因此就有术语：“可确定的”、“确定的”）。笛卡尔认为，广延和思想是确

定的属性，它们自身显然不确定其他任何可确定的属性。[35]

由于有这种理论，心身二元论就有一种简单易懂的论证。如果有两种最一般的存在方式，并且有具有它们的事物，那么就能直接推出两种实体
15 的存在。笛卡尔论证说：对于他自己是一种能思维的东西、一种至少能不依赖于其身体而存在的东西，他拥有清楚、明白的概念，对于有形物体是一种唯一有广延的东西、一种至少在没有思维时能存在的东西，他同样拥有清楚、明白的概念，而且，这些概念是完备的，不需要借助任何更一般的种类概念。[36]由此可以推出：思维和广延是范畴属性。从这种实体理论可以推出：能思维的实体和有广延的实体必然是不同的。

这个论证无疑是有效的：如果它的前提正确，它的结论必然正确。但还不太清楚的是，我们是否应当接受它的前提（从而它是否合理，即它除了有效之外是否有正确的前提）。它最弱的前提是这个假设：不同种类的实体必定只有一种范畴属性。我还不明白笛卡尔为什么坚持这一点。实体是属性载体的思想所提供的支持是不够的，甚至深受笛卡尔影响的斯宾诺莎也反驳说：恰恰由于心理属性和物质属性是概念上无关的，因此推断一种实体拥有两种属性是没有什么障碍的（《伦理学》IP10，附释；载斯宾诺莎，1994：90）。正如斯特劳森（P. F. Strawson，1958）所说：我们通常给同一种东西——即人——既归属了物质属性又归属了心理属性：我能走、能睡，也能思维、能感觉。

笛卡尔赞成心理实体与物质实体之间的因果交感论，他以此解释为什么我们四肢的活动与我们想做的事情相一致，以及我们如何能正确地感知我们身体所处的物理环境中的东西。有些哲学家——包括很多笛卡尔同时代的人——反对说：我们不能想象像心灵和身体这样根本不同种类的实体——后者处于空间中，而前者不在——之间的因果作用。（但很难将此看成是一个**概念的**难题；参见贝多［Bedau］，1986。）这产生了一种副现象论，根据它，心理的东西与物理的东西不是因果相关的。对因果交感论的否定加上心理事件与物理事件之间明显的相关性就导致了平行论。根据平行论，心理事件和物理事件是独立发展的，但在某种意义上产生了非因果的相关性，这就像两只钟表的指针，它们独自设定，相差一分钟，由于位置的相关性，它们可能看起来有因果相互作用，但实际上它们并没有。[37]对平行论的解释通常会引证上帝最初对事物所作的安排，从而使得心理的东西和物理的东西能平行发展（前定和谐），或者由于上帝对事件——既有物理事件又有心理事件——的产生进行持续不断的干涉，导致

了相互作用的假象（偶因论）。

除非有理由认为属性载体不能既有不可还原的心理属性又有物理属性，否则笛卡尔的论证最多只能够确定：**可能**有些东西只有心理属性，还可能有些东西只有物理属性，但不能确定实际上有或者一定有。如果我们至多能先验地确定二元论可能是正确的，那么，想要确定二元论是否正确就只能由经验研究来决定。迄今为止，似乎没有很好的经验理由认为二元论是正确的。[38]

唯心主义是二元论的历史继承者，它是无物质实体的二元论。因此，它是一种普遍的、纯粹的心理粒子理论。古典的立场出现在乔治·贝克莱的《人类知识原理》（*A Treatise Concerning the Principles of Human Knowledge*，1710）中，更复杂的现代版本被称为“现象主义”（phenomenalism）。[39]唯心主义的动机通常来自于这种关注，即理解有关普通知觉对象——森林与草地、山和雨、星辰和窗玻璃——知识的可能性。在心灵与世界的关系问题上，笛卡尔主义观点使得我们何以能拥有这种关系的知识问题变成了一个谜：如果我们最初只知道自己的有意识的心理状态（且不管我们只根据理性能知道什么），而心理与物质在概念上是无关的，那么我们似乎就没有理由相信存在一个产生了我们意识经验的物质世界。贝克莱对这个问题的解决，即是否认知觉对象是物质的，而把它们等同于观念的集合（因而是**唯心主义**）。新近的处理办法是将普通常识知识的对象等同于由现象状态构成的逻辑结构。贝克莱甚至还否认我们能理解物质实体。莱布尼兹（1714）同样认为，事物的基本成分，即单子（**单元**，源于希腊文 *monos*），是一种心灵——但是他不认为所有单子都有意识——有关普通事物的谈论要依据单子及其状态来理解（就像大卫·阿姆斯特朗［David Armstrong］所说，根据莱布尼兹的观点，“物质对象是低级灵魂的聚集地”［1968：5］）。康德（1781）有时候也被看成是现象主义者，但这种观点现在并未被普遍接受。我们认为自己有关于世界的知识，我们的一部分世界观似乎是：世界并不依赖于能思维的存在物而存在，这些存在物是世界舞台上的临时演员。

泛心论认为，万物都是心理属性的基本载体（这不只是由于万物与一种基本载体相联系——就像我的椅子有被某个在思考心身关系的人占据这种属性一样）。泛心论有还原的种类也有非还原的种类。它的根源可以追溯到古代（安娜斯［Annas］，1992：43-47），泛心论者的代表既有文艺复兴时期的哲学家，也有 19 世纪杰出的哲学家，包括叔本华、克利福德

(W. K. Clifford)、(一段时期的) 威廉·詹姆斯，以及皮尔士 (C. S. Peirce)。[40]泛心论 (似乎是) 经常与一种修正的形而上学和特定的动机联系在一起，就像唯心主义那样，这是一种还原形式的泛心论。然而，非还原的泛心论接受一种基本的唯物主义本体论，它的动机来自于这一思想，否则，复合物具有心理属性就难以解释了 (一种魔法)。威廉·詹姆斯在其不朽名著《心理学原理》(*Principles of Psychology*，1890) 第 6 章中明确提出了这一论证："进化心理学需要心灵的尘埃。"托马斯·内格尔 (1979a) 最近复活了这个论证 (参见孟席斯 [Menzies]，1988)。[41]泛心论是一种普遍的心理粒子理论，它可能是纯粹的，也可能是混合的。

两方面论应当被看成一个理论家族，而不是单一的学说。统一这个家族的是它们的相似性 (affinity)，它可以用这句口号表达：心理和物理是我们借以理解同一个事物的不同方面。但根据这个理论的不同"版本"，
17 这句口号可以作不同的理解。斯宾诺莎关于思想和广延的平行论学说是两方面论的源头，尽管他自己没有这样描述他的立场。[42]斯宾诺莎认为，存在一种单一的、无限的、永恒的和必然的实体，它具有每种可能的范畴属性，因此它既有广延又有思想。普通事物被 (重新) 看成是世界实体的样式 (变体)。思想和广延是按照平行论论题相联系的，"观念的次序和联系与事物的次序和联系是相同的" (《伦理学》，IIP7；载 1994：119－120)。正如斯宾诺莎在附释中所作的进一步解释："思想的实体和广延的实体是同一种实体，不过时而通过这个属性，时而通过那个属性去了解罢了。同样，广延的一个样式和该样式的观念也是同一的东西，不过由两种方式表达出来罢了。" (1994：119) 这不是一种意思完全清晰的学说，只有理解了斯宾诺莎的形而上学，我们才能理解它，不过，这种形而上学本身就是一个难以解释的问题。然而，这种观念 (即心理和物理是了解一种东西的两种方式) 能在否定斯宾诺莎的形而上学之后继续存在，而且它还启发了很多借助类似语言的观点。

如果我们承认有很多实体，那么两方面论就推断：每个对象或者有些对象可以被看成是心理的或物理的，这取决于我们怎样处理它。用 G. H. 刘易斯的比喻来说，把一种东西理解为心理的或物理的，就像是把一条线看成凸的还是凹的，"曲线上的每一点都有这种凸和凹的对比，然而它始终都是同一条线" (G. H. Lewes，1877；重印于维西，1964：155)。两方面论现在不流行了，这部分由于它不清晰。它不仅仅想成为这种主张，即有的东西既有心理属性也有物理属性，两者都不能从概念上还原为对方

(尽管它有时候是在这种更宽的意义下被使用的)，或者甚至是主张：每个物理之物与某个心理之物之间具有系统的相关性。[43]但是，一般来说，对于它最终是什么似乎没有更多的东西可说，相反，我们必须依靠特定的理论来给予它内容。它不流行，部分是由于一些与细节无关的因素，尤其是我们当前科学世界观的优势地位，根据这种科学世界观，世界上以前并没有能思维的东西，它根据自然规律演进到了现在的状态。

两方面论可能是普遍的、混合的心理粒子理论，也可能是受限的、混合的心理粒子理论。因此，有些两方面论是泛心论的形式，就像在斯宾诺莎那里一样，因为他的确主张一切存在都有心理属性。然而，与这个主导观念相容的是，人们也可以主张：有些东西有两个方面、两种理解它们的方式——心理的和物理的，但是并非所有东西都是如此。[44]

最后，存在我所称的特殊粒子理论，它认为事物的有些基本成分——它们至少可以作出空间定位——具有心理属性，但并非所有成分都有。这是一种受限的、混合的心理粒子理论，它把空间定位看成是一种宽泛的物理属性。就我所知，这不是关于心身问题的传统回答中已经提出的一种观点。[45]

1.6.2 概念的反还原论

否定命题（3）会导致突现论（emergentism）。原则上有两种突现论：中立的突现论和突现的唯物论，这是依据基本成分被看成物理的，还是被看成既非物理的也非心理的来划分的。大多数突现论者都是唯物主义者，因此我聚焦于突现的唯物论。突现的唯物论者认为，只存在有物质的东西，但是有些复合物质的东西——尽管没有一种简单的东西能独立于它们所参与的复合物——具有心理属性，而且这些心理属性不能从概念上还原为拥有它们的复合物的任何物理属性。突现论在历史上是对唯物主义本体论的一种回应，这种本体论否定各种形式的二元论和唯心主义。一般来说，它与19世纪科学的兴起联系在一起，尤其是与进化论的发展相联系。它摒弃了笛卡尔二元论的本体论的反还原论，但保留了笛卡尔二元论概念的反还原论。19世纪末20世纪初的突现论者有T. H. 赫胥黎（T. H. Huxley，“达尔文的斗牛犬”，1901）、塞缪尔·亚历山大（Samuel Alexander，1920）、C·劳埃德·摩根（C. Lloyd Morgan，1923）和C. D. 布罗德（C. D. Broad，1925）等。之所以使用“突现的”这个术语，是因为人们认为宇宙中以前没有任何东西具有心理属性。既然宇宙中的一切都是物质的东西，它们以前没有心理属性，但现在有些东西有心理属性，并

且这些属性不能从概念上还原为物理属性，那么心理属性肯定是以某种方式从某些物质组织中突现出来的，但是这不能从具有心理属性的东西的完整描述——依据的是其物理属性——推导出来。[46]突现论者很重视这一事实，即至少心理之物的有些方面在任何意义上都不是物理现象。这是传统的看法，而且不可否认是一种最初有吸引力的立场。一旦我们摆脱了导致下述看法的混乱——一定存在与物质实体不同的心理实体，后者有不可还原的心理属性，那么我们是物理世界的后来者的看法——是通过自然过程从这些自身完全属于机械学领域的材料产生出来的自然对象——自然会导致突现的唯物主义。

各种突现论都产生于对基础属性与心理属性之间关系的不同看法。传统的突现唯物论者认为，物理状态与心理状态之间有类型的法则相关性（type-type nomic correlation）。这就是主张：对于每种心理属性，获得的某个具有形式［C］的句子都具有法则学必然性的力量。有人可能认为，心理属性只是在法则上随附于物理属性，并不存在类型的相关性。[47]最后，有人或许坚持所谓的“异常一元论”的一个版本，最初提出异常一元论是作为心理事件与物理事件关系的一个论题（戴维森［Davidson］，1980）。它认为，每个心理事件都与一个物理事件个例地同一[48]，但是不存在严格的心理物理规律，从而不存在严格的桥梁法则。[49]但这个论题仍
19 然承认松散的、**非严格的**、法则的随附性或者法则的类型相关性。一个更强的版本甚至否认心理事件类型与物理事件类型之间有松散的法则关系，这种观念可以被改造成这种看法：尽管有些复合对象有心理属性，但物理属性与心理属性之间并没有严格的法则相关性或随附性，或者在更强的版本中，根本没有任何关系。

突现论经常（尤其是现在）与副现象论（epiphenomenalism）联系在一起。[50]副现象论认为，心理属性与任何东西（或者，至少与任何物理的东西）都不因果相关。19 世纪末 20 世纪初的突现论者对于心理之物的因果效力存在意见分歧，有的人（例如摩根和布罗德）认为，不仅存在突现属性，而且存在突现规律，它们控制着突现属性层次上的系统，因此突现属性能够影响低层次事件的发展（下向因果性）。[51]突现论传统中的这一潮流如今已经近乎绝迹了（不过，参见斯佩里［Sperry］，1986）。[52]还有些著名的突现论者认为心理属性完全依赖于物理属性，它们没有因果效力。在一个著名的讨论中，赫胥黎指出意识是“物质变化的直接功能”（1874：141），但他也指出意识对我们身体的运动没有任何影响力，“就像

机车运转的汽笛对机车的机械装置没有影响一样”（1874：140。也可参见霍奇森［Hodgson］，1870；罗曼斯［G. J. Romanes］，1895）。按照这种看法，心理活动是神经活动投下的阴影，由神经活动决定，但它反过来什么也不能决定：用费格尔（1958）贴切的措辞来说，有意识的心理状态是“法则学的悬摆”（nomological danglers）。

20世纪下半叶以前，突现论者相信人体的中枢神经系统的状态与心理状态之间有类型的相关性，这被看成是一个自然规律问题。这些规律不纯粹是物理规律，而是桥梁法则，因为它们的陈述不可还原地既包括心理谓词，又包括物理谓词。副现象论的动机来自这一思想，即如果我们只是从宇宙中减去桥梁法则，那么宇宙仍会像往常一样物理地运行，要解释任何物理事件，我们原则上不需要提及任何非物理的事件或者规律。正如即使我们拆除了汽笛，机车仍会继续行进一样；即使我们去除了灵魂，我们的身体仍会安之若素。[53]下列两个观点的合取就是法则学的还原论：一个观点认为存在这种类型的法则相关性，另一个观点认为物理的东西是一个封闭系统。显然，我们越是不相信法则的类型相关性，我们能为心理属性的因果效力找到一个位置的宣称就越不可信。被感知到的副现象论威胁是物理主义的动机之一。具有讽刺意味的是，有些想给物理主义奠定基础的流行方法，对理解心理属性何以能与它们本该成为的东西因果相关，也提出了一些难题。[54]

1.6.3 概念的还原

20

否定命题（2），就要接受对心理属性的概念还原论。

首先我们简要地考虑一下否定（2）的非物理主义方式。这有两种选择：心理之物能从概念上还原为非物理的东西，或者随附于非物理的东西。虽然后一种立场是一个选项，但是还没有人选择它。然而，中立一元论认为，心理的东西和物理的东西都能依据更基本的东西来理解，它在19世纪末和20世纪上半叶享受了短暂的繁荣。[55]这种看法与威廉·詹姆斯（1904）有关，他认为“纯粹经验”是世界的原材料，心灵和物体应被看成是经验的不同集合，因此，相同的经验可能在一个集合里被看成是思想，而在另一个集合里被看成是与思想有关的物体的一个成分。就像詹姆斯所倡导的，中立一元论反对存在经验主体的看法，它只保留了传统所认为的经验主体的对象，如詹姆斯所说：“那些死抱着意识不放的人，抱着的不过是一个回响，不过是正在消失的‘灵魂’遗留在哲学空气中的微弱

声音而已。”（1904：3—4）恩斯特·马赫（Ernest Mach，1886）持类似看法，而伯特兰·罗素受詹姆斯启发，提出了一种中立一元论，在这种理论中，可感物（sensibilia，或者“感觉”，如罗素在《心的分析》[*The Analysis of Mind*，1921] 中对它的称呼）最初是作为独立于心灵的直接觉知对象引入的（1917），它们发挥着中性材料的作用，心灵和物理的东西是用它们逻辑地建构起来的（1921）。

将这种观点描述成一种唯心主义似乎应当更合适，因为詹姆斯、马赫和罗素用以描述这种中性材料的术语通常和心理现象联系在一起。但是他们认为，人们并不完全认为这种中性材料具有心理特征，而是仅仅在某种配置下考虑它时才认为它有。因此，至少在这种意义上（即中立一元论认为每种基本的东西都既是能参与一系列构成心理之物，也是能参与一系列构成物理之物的东西）将中立一元论描述成一种两方面论似乎是合乎情理的。因此，可以说每种基本的东西都能从物理的方面来看待，也能从心理的方面来看待。然而，由于人们认为关于思想和物质的东西的言论既不能翻译成心理的言论也不能翻译成物理的言论，因此，心理的东西和物理的东西在中立一元论的本体论中都没有基本的地位。[56]相反，两者都与中性材料具有这种关系：即根据唯心主义理论，普通物体与现象经验的关系。正如唯心主义理论不赞成真正的物质实体一样，中立一元论也不支持其基础本体论中有心理或物理的实体，但它对每一种**谈论**都给出了一种解释。

中立一元论有一些理论的价值。它避开了与这种打算相关的难题：将心理的东西还原为物理的东西，或者将物理的东西还原为心理的东西。而且如果取得成功，它就会依据一种基本的东西对各种东西作出基本的、统
21 一的解释，这是唯心主义者和物理主义者的梦想。即便如此，它也不是一种流行的观点。它既未吸引那些认为心理是一个实在的基本特征的人，也未吸引那些梦想着物理学荒漠图景的人。此外，很难详细地阐明这种解释，而且很难理解它所依赖的中性材料的本质。

我们现在转到关于命题（2）的物理主义否定。

20 世纪第一种得以流行的物理主义立场是逻辑行为主义，它部分地受到了第二次世界大战前逻辑实证主义者的证实主义的推动，后者是这种观点：句子的意义要在确证或否证它的经验条件下寻找（这种观点的根源是古典的英国经验论）。[57]逻辑行为主义既有一种较强的形式，也有一种较弱的形式。我将强的形式称作“翻译的行为主义”（translational behaviorism），较弱的形式则是“标准的行为主义”（criterial behaviorism）。翻译

的行为主义认为，每个心理学陈述都能翻译成一个关于身体的实际的和潜在的行为的陈述。相反，标准的行为主义仅仅认为，心理谓词的应用具有从行为上分析的充分条件。

逻辑行为主义早就不流行了。这部分要由为其提供了理论支撑的证实主义的失宠来解释，部分也要由这一事实来解释：不仅令人满意的解释图式还没有被提出来，而且有理由认为它在原则上不可能被提出。尤其麻烦的一个问题是：我们预测具有某种心理状态的某个人所作出的行为表现，依赖于他所具有的其他心理状态。因此，不可能将心理学主张逐一地翻译成行为术语。另外，行为主义与我们关于心理状态是（可能的）行为原因的概念似乎是水火不容的，因为要将关于心理状态的谈论还原为关于行为的谈论，就是要相信，它只是一种更简洁的描述行为的方式。但是，行为不可能引起它自身。[58]

对于行为主义的缺陷，有两种主要的物理主义回应：分析的功能主义和心理物理同一论。虽然心理物理同一论在分析的功能主义之前占主导地位，但首先讨论功能主义是有好处的，因为它是逻辑行为主义自然的继承者，这使我们能够有效地阐明心理物理同一论，它早期的版本里有很多混淆和互相抵触的倾向。

分析的功能主义认为，心理状态能从概念上还原为功能状态，在 1.4 节所定义的意义上，功能状态被认为是在概念上随附于物理状态之上的。[59]这样，心理状态与功能状态的同一实质上会导致这样的**物理主义**——它没有心理向物理的概念还原。在相关意义上，功能状态就是物体的这样一种状态，对它的定义依据的是其与系统的输入、系统的其他功能状态以及系统的输出之间的关系。有些逻辑行为主义者，例如《心的概念》（1949）中的吉尔伯特·赖尔，可以被看成是转到了与此类似的立场（因此，功能主义可以被看成是由行为主义孵化而来的）。功能主义至少部 *22*
分地受到了第二次世界大战之后计算机技术[60]兴起的启发，它在 20 世纪的最早形式，即机器表功能主义，是由希拉里·普特南（1967）提出来的，其灵感直接来自于对有限状态机，即什么是（有限状态的）计算机的理论研究。[61]机器表对系统的描述是依据一系列详尽无遗但相互排斥的输入、一系列可能状态和一系列输出，也依据（对每个可能状态来说）它在接收到某个输入时所转入的状态和由此产生的输出。任何运行了一个程序的计算机的操作都能根据一个机器表来透彻地描述，对于可编程的计算机来说，程序决定着它所例示的机器表（这涉及将一个系统划分为特别令我们

感兴趣的状态）。普特南将有限状态自动装置（即能用一个具有决定性状态转换的有限状态机器表来描述的系统）的概念推广到了概率性有限状态自动装置，这种装置的转换是概率性的。这个计划的一般形式是：一个系统处于某种心理状态，当且仅当它有一个合适的机器表描述，有合适的输入或合适的状态。普特南把他的计划看成是一个**经验**假说，我们通常仿效布洛克（1978）[62]称其为“心理功能主义”（psychofunctionalism）。尽管如此，它只是分析的功能主义的主要灵感之一，很容易被重新看成是一个关于我们心理状态概念的论题。**理论的**或者**因果作用的**功能主义是这个论题的一个变化形式。按照这种观点，我们是从一个嵌入了心理学术语的理论开始的，这些术语所表达的概念被认为是状态的概念，这些状态是由其与别的状态，与这个理论抽象说明的输入和输出关系详细描述的。[63]

功能主义是有吸引力的。它包含着这样一个促进行为主义的思想，即我们的心理状态最终是与对行为的理解紧密相联的，但它之所以这样做，在某种程度上是要将心理状态与行为相区别，并把心理状态看成行为的原因。此外，功能主义承认这种可能性，即存在非物质的能思维之物，因为一个系统具有某种功能组织，这并不取决于它是用什么制造的，而是取决于它对输入和输出的因果力，它只需要在输入和输出之间维持恰当的起中介作用的组织。功能状态是**可多样实现的**（multiply realizable），这包含着一个思想实验——它是心理的东西和物理的东西都具有概念的独立性这一假设的诱因。它在自然界中为心理的东西找到了这样一个位置，表明心理的东西是以物理的东西为基础的，这就是说它表明心理的东西是在概念上随附于物理的东西的，但它并不坚决要求向物理属性的概念还原。因此，它承认心理学语言不同于物理学语言，也同时承认心理状态的实现仅仅要求这样的物体，它们有受物理规律支配的物理属性。功能状态的可多样实现性（初看起来）也使功能主义避开了心理物理同一论所遇到的指责，即坚持认为只有在物理方面与我们类似的东西才能有心理状态，这是难以置信的沙文主义思考。[64]

23 分析的功能主义已经受到了很多批评，但仍很流行，尤其是在除哲学以外对新的认知科学学科作出贡献的领域。对功能主义的第一种反驳是：还没有人提出一种成功的心理概念向功能概念之概念还原，人们或许会说，这可能也是这些功能概念复杂性的标志。对功能主义的第二种反驳建立在这样一些系统——它们在功能上与我们相同一但没有心理状态——明

显的可理解性的基础之上。内德·布洛克的一个思想实验（1978）提供了一个例证。[65]设想一个机器人的躯体是由一个程序驱动的，这个程序例示了某个人的机器表；再设想我们例示这个程序，是通过为每个中国人提供一个带有显示屏的双向无线电设备，它向机器人系统显示当下的输入，还给每个中国人提供一个指示器，它指示系统是否处于他的状态，每个人在他的状态激活时就按下无线电设备上的相应按键，信号被传送到身体，以便作出合适的行为。假设中国人对此非常擅长，以至于我们的机器人和附属设备构成了一个和我们真人在功能上相同一的系统。那么，这个系统现在就变成了一个有智能、有意识的东西吗？大多数人在初次遇到这个思想实验时，都否认我们已创造了一个新人（他会在这项练习停止时死去）。[66]

另一个重要的反驳也归功于内德·布洛克（1978）。功能主义者必须决定怎样说明系统的输入和输出，这会让他们左右为难。如果我们从物理上把自己当成模型来说明输入和输出，就不难描述某个系统，它可能有一个心灵，这个心灵不能引起那些输出，但却会引起别的输出（例如，我们不想先验地排除存在具有智能的水母，或者排除其输入和输出是电磁频谱的不同部分的东西等等）。此外，也很难理解我们怎么能对输入和输出的物理特征施加先验的限制。然而，如果输入和输出只是被看成是不同的，那么我们几乎到处都能找到心灵就不是不可能的，因为似乎可信的是：大多数复杂系统都允许被分解成这样的状态和输入、输出，它们会例示某个据认为对于具有心灵是充分的机器表（例如，世界经济）。

人们还可以提出这样的反驳：想象在经验的性质方面有区别的功能复制品是不费力的。一个用于证明这一点的著名实验就是颠倒光谱思想实验。我们设想有两个人，他们从功能上难以区别，因而从行为上也难以区别，但是设想他们对于环境中物体颜色的经验是彼此颠倒的。例如，当一个人经验的是一个红色的东西时，另一个人经验的却是一个绿色的东西，他们在描述这种东西时都说相同的句子，但每个人看它是不同的。如果这是可以想象的，那么他们的颜色经验就不能从概念上还原为他们的功能组织，因此，功能主义在这些现象性质上是错误的。[67]

另一个困难是：还不能确定功能状态和正确行为能否因果相关。根据定义，功能主义是将心理状态看成行为的**原因**[68]，但这可能是以错误的方式保证了因果联系，因为一个依据其在各种环境中的影响来定义的状态，不可能是那些影响借以产生的这种类型。事件或状态之间的因果关系

是以将它们所属类型联系起来的**偶然的**因果规律为基础的。[69]一个类型与另一个类型是因果相关的（在某些环境下），当且仅当它们是由一个因果规律联系起来（在这些环境下）。然而，一个功能状态与依据它得到的部分定义的输出（类型）之间的关系并不是偶然的，因此，状态类型与输出类型不可能在偶然的因果规律中起合适的重要作用，由此，功能状态类型与依据它们定义的输出不是因果相关的。[70]如果这种推理是正确的，那么分析的功能主义就衍推了与这些输出相关的副现象论。功能主义相对于行为主义的优势是：它使心理状态成为了行为的原因。但问题在于，它这样做，在某种意义上削弱了这种可能性，即那些状态与我们期望它们成为的东西是因果相关的。

更糟的是，似乎完全可以相信，我们确实认为自己的心理状态与我们根据功能分析用以定义心理状态的行为是因果相关的。我们关于心理状态与行为的因果相关性的信念可能是错误的，它要视什么因果规律能成立而定。但是，如果这些信念并不必然错误，那么功能主义就不可能是正确的，因为它排除了**这种可能性**，即我们的心理状态与我们的行为是因果相关的。[71]

现在让我们转到心理物理同一论，即这种观点：心理属性就是物理属性。我从自己认为最可信的心理物理同一论形式着手，它以大卫·刘易斯（David Lewis，1966、1972）所提倡的一种方案为基础。这种方案利用从一种心理学的“民间理论”选取的功能主义状态**描述**来将心理状态与物理状态同一起来。

分析的功能主义认为，心理学的概念和属性就**是**功能的概念和属性。这应该与心理学属性是**由功能描述挑选的**观点相区别。这种观点没有将心理属性还原为功能属性。相反，它将心理术语看成是理论术语。人们认为，理论术语挑选了一些世界上的属性（从而表达了任何与那些属性相关的概念），这些属性实际上起着这个理论在其所适用的系统中给予它们的作用。我们将心理学理论描述成一个句子：“$T(M_1, M_2, \cdots, M_n)$”，这里，“M_1”等表示指称属性的心理学术语。这样，我们就用相应的变量“x_1”、“x_2”等来替换每个这样的术语，并在整个句子之前为每个变量加上一个量词：“存在唯一的 x_1，以至于”（符号化为“$[\exists!\ x_1]$”），等等，以此产生：“$(\exists!\ x_1)(\exists!\ x_2)\cdots(\exists!\ x_n)\ T(x_1, x_2, \cdots, x_n)$”。所挑选的属性“$M_1$”可以作如下描述，这里我们省去了“$T(\cdots)$”之前与“$x_1$”相联的量词：

M_1是唯一的属性 x_1，以至于（∃！x_2）…（∃！x_2）T（x_1，x_2，…，x_n）

根据这个假设，对这种属性的理论描述是由我们身体或者中枢神经系统的物理属性满足的，在运用于人类时它可以推出：$\mathbf{M_1}$**就是**这种物理属性。因此，我们就得到了心理物理同一论。

鉴于我们描述概念、谓词、状态和属性之间关系的方式，如果我们将一种心理状态或属性同一于一种物理状态或属性，就可以推出：相应的心理概念**就是**一个物理概念。因此，心理属性是由功能描述挑选的，此种看法会导致这样的结论，即如果那些描述挑选了物理状态或者属性，那么心理概念就能从概念上还原为物理概念。[72]然而，这不是我们能先验认识的，可能它仅出现于经验研究之后。因为根据这种看法，我们的理论术语所表达的概念受制于这些现象的本质，即我们将这些概念运用于其上的现象。我们的出发点只是关于这些属性的描述，因而实际上只是关于这些属性的概念描述，只有在我们后验地发现这些概念之后，才能用它们先验地推理。

心理物理同一论与功能主义和突现论相比，其优势在于保证了心理属性的因果相关性。没有人怀疑我们的物理状态与我们的行动是因果相关的。由于将心理状态与物理状态相同一，心理物理同一论就使它们的因果相关性确定无疑。有些哲学家认为，由于将心理状态同一于物理状态就能保证它们的因果有效性，而心理状态是因果有效的，因此我们将它们同一起来是有道理的（帕皮诺［Papineau］，1998）。

可是，以上推论也是有代价的。按照这种看法，在进行经验研究之前，我们并不能确定根本不存在心理属性，不存在同我们对它们所做的理论描述相符合的属性。这表明：这种观点和取消主义有共同的假设，即对于某种东西具有我们认为的心理学术语所挑选的属性，我们是不能直接知道的。当与此类似的一种观点与这一主张（即没有一种物理状态或任何别的状态起所需的作用）相结合时，就会导致取消主义。就我们觉得我们有可能发现自己没有任何心理状态这一点不可信——甚至难以理解——而言，我们应该会发现刚才所述的心理物理同一论论证同样是不可信或难以理解的。[73]

心理物理同一论（也被称为“中枢状态唯物主义”）和功能主义一样，也有可追溯到古代世界的先驱。它在第二次世界大战后得到了阿林·普赖斯（Ullin Place，1956）、赫伯特·费格尔（Herbert Feigl，1958）和斯马

特（J. J. C. Smart，1959）强有力的倡导。[74]普赖斯和斯马特认为，感觉在理论上是同一于大脑过程的，正如闪电同一于某种电的释放一样（这可以直接推广到状态；参见阿姆斯特朗，1968）。[75]他们认为这是一种偶然的同一性，因为它是经验的发现。这种立场有时也被称为“主题中立方案”（the topic neutral approach），因为斯马特尤其认为，为了这个目的，
26 即我们没有不可还原的心理属性但仍理解偶然的同一性的可能性，我们用以挑选心理过程（更一般地说是心理状态）的描述——它们在经验上同一于物理的描述——一定不会理睬这个问题：这些心理过程是否是物理的。这一立场招致的很多批评都是缘于这个主张：同一性可能是偶然的（参见克里普克，1980：98－100，144－155）。如果我们谈论事物——就现在来说，是属性——的严格同一性，就没有偶然性的空间，因为必然的同一性存在于万物与其自身之间，而不存在于任何不同的东西之间。我依据刘易斯的方案所提出的观点是对这些早期心理物理同一论的继承。它保留了这种观点：心理属性就是物理属性（根据这个假设，只有物理属性起正确的作用）。但它反对这种观点：同一性是偶然的（假定真有起正确作用的物理属性）。认为理论术语是用来追踪起某些作用的属性，这有助于我们理解：尽管同一性是必然的，但发现同一性何以能是经验的。它也给了下述思想以准确的内容，即挑选心理状态的描述是主题中立的，因为这些描述都是由民间心理学理论所产生的结构给出的。

在这一点上，对形而上学的必然性作个说明是妥当的。这种模态在有关心身问题的当代讨论中是经常使用的。据说，它既不同于法则学的必然性，也不同于概念的必然性，它比前者要强，比后者要弱。它是怎样提出来的？形而上学必然性的范例据认为应该是这种必然性，它来自于涉及**自然种类**（natural kinds）的理论同一关系，就像金与原子序数为 79 的元素的同一关系一样。金是原子序数为 79 的元素，这不是偶然的或者只是自然规律的问题，而是必然的，因为即便在具有不同自然规律的世界上，一切原子序数不是 79 的东西也不算是金。尽管如此，它仍是经验的发现而不是我们完全能先验地知道的东西。但是，由于概念的真理能先验地知道，因此形而上学的必然性一定不同于概念的必然性，论证大概就是这样的。

形而上学的必然性被认识到的用处是：它对论证心理与物理之间的联系（其实是同一性关系）比法则学的联系要强提供了一种方法，同时它也不会被这些思想实验否定，这些思想实验似乎证明心理现象和物理现象是互不依赖的。由于形而上学的必然性据认为并不受概念上可能的东西控

制，而这些思想实验也是如此，因此它们对这个主张并无影响。[76]

如我前面所述，在我看来，还没有一个哲学家用“形而上学的必然性”成功地表达一个适合于这个论证的概念。使我们对“形而上学的必然性”产生怀疑的，首先应当是：对于据说是关于它的主张的根据是什么，我们没有作任何解释。除此之外，还值得怀疑的是：对于这里据说由“形而上学”这个术语所表达的东西，我们已经有了清晰的认识。使我们产生怀疑的第二点应当是：对于这些事实（它们采用形而上学必然性的动机——这种必然性不需要任何一种神秘的新必然性）存在一种简单的解释可用。

我们对刘易斯关于理论同一关系的说明所作的解释提供了答案。根据这种说明，我们把每个理论术语都与它所挑选的属性描述（在如此这般的系统中起着如此这般作用的属性 P）联系了起来。什么属性实际地满足了这个描述是一个经验研究的问题（就像决定哪个人是纽约市市长一样）。然而，如我们所看到的，一个术语所表达的概念决定着它所挑选的属性：它们是相配的一组。因此，要通过经验地发现什么满足了相关的描述来揭示一个理论术语所挑选的属性，同样也要经验地揭示这个术语所表达的概念。在此之前，我们对概念有了一种描述，但它不是直接给予我们的。因此，当发现“是金”挑选了原子序数为 79 的元素时，我们就发现了它所表达的概念，在此之前，我们并不知道它所表达的概念。一旦知道了，我们就能明白：“金是那种原子序数为 79 的元素”，它表达了一个能先验地知道的概念真理。不能先验地知道的不是“金是那种原子序数为 79 的元素”而是“表达原子序数为 79 的元素这个概念的‘金’”。在发现它们所表达的概念之前，我们有资格使用这些自然种类术语，这是由这一事实解释的，即我们认为这些术语是在追踪我们在实践中将它们运用于其上的属性，这些属性**能轻松地解释事物可同一的特征**。最初论证中的错误是将运用自然种类术语的能力与掌握所表达的概念混为一谈：假定我们不知道什么属性被挑选了，我们就同样不知道其表达的是什么概念。我们所知道的只是这种属性应该做的工作，这使我们能用挑选这种属性的术语来发展一种应用实践。

因此，引入“形而上学的必然性”是没有必要的。我们没有理由认为有什么东西与它相符合，也不知道倘若如此会怎样。因此，我们不能指望形而上学的必然性会成为解决心身问题的新途径。[77]

在离开还原论的专题之前，考虑一种混合观点是很重要的，这种观点将功能主义与关于思想内容的外在主义结合了起来。关于心理状态的外在

主义解释强调，从概念上讲，我们与环境中的事物的关系在对心理状态进行个体化时具有重要作用。在功能主义的难题日益增多的同时，一些有影响的论证已经被独立地提了出来，它们提出内容属性是关系属性。[78]根据这些解释，我们所拥有的思想依赖于我们与环境中事物具有实际的和潜在的因果关系。（文献中通常把从关系上进行个体化的状态称为“宽状态”，而把非关系地个体化的状态称为“窄状态”。）各种外在主义观点之间最重要的区别是物理外在主义与社会外在主义之间的区别。物理外在主义认为，思想内容（部分地）是根据与我们物理环境的关系而个体化的；社会外在主义认为，思想内容的个体化，（部分地）是根据我们语言共同体中的其他人使用这些词语——我们打算像他们那样使用——的方式。[79]关于思想内容还原论的外在主义解释通常认为：至少我们关于有内容的心理状态的概念可以还原为功能的和因果的概念，这里我们包括了在确定思想的内容时与外部事物系统的因果关系。

外在主义理论也受到了很多批评，其中有两个值得一提，因为它们和刚才谈到的主题有关。[80]第一个批评是这种异议：如果外在主义正确，那么如果不作经验的研究，我们就不能知道自己思想的内容，但由于我们要进行经验研究首先必须知道自己思想的内容，因此外在主义就让人难以接受地蕴含着我们根本不能知道自己思想的内容。[81]第二个批评与已经指出的功能主义的一个困难有关，它认为内容属性部分地是根据关系属性来个体化的，这恰好预示着它们不适合于解释我们的（物理地描述的）行为。这里的问题不是关系属性不能与某种东西因果相关。对此有明显的反例，例如，某种东西是一颗行星，这可能被引证来解释为什么我相信它是一颗行星。但是，外在主义的困难仅仅要求内容属性最终成为的这种关系属性可能与我们的行为不是因果相关的。因为外在主义理论利用了这种可能性，即（物理地描述的）行为保持不变，是因为一个人非关系的物理状态在他的思想内容发生变化时却保持不变。这似乎表明：关系状态被非关系的物理状态——它们对于解释行为是充分的并且具有独立的必然性——从相关的影响类型中“清除”了。[82]

这种观念，即我们的（至少是有意识的）心理状态是以下状态：它们（a）能够被拥有者但不能被其他人非推理地知道（因此，我们关于它们的概念不是理论概念），（b）它们与别的东西（别的心理事件和状态以及非心理的事件和状态）（可能）是因果相关的，可能被称为笛卡尔主义心灵观念的核心。对于心理之物的还原论方案来说，我们一直在考察的困难都

与这些特征有关。任何物理状态似乎都不能拥有这两个特征。第一个特征使心理物理同一论不可信，而且可以论证关于思想内容的外在主义不可信，第二个特征似乎排除了向这些状态的概念还原，这些状态是根据它们与其他事物的因果关系，或者也是根据它们与环境中事物的关系描述的。

1.6.4 非实在论

最后，我们来看取消论，取消论者是通过否定来寻求解脱。根据取消论，任何东西都没有心理属性。这种立场的主要支持者有保罗·丘奇兰德（1981）和斯蒂芬·斯蒂克（1983），他们主张我们的心理概念是空概念。[83]它们是前科学的或“民间的”行为理论所使用的概念，它们在时机成熟时会被一种使用不同范畴的更完善理论所取代，这些范畴更符合我们的解释兴趣。民间心理学会重蹈诉诸魔力精灵的疾病理论的覆辙，我们的常识概念图式的心理实体也是无知的产物。我们现在必须走向更光明的未来，摆脱在文明的童年时期反复灌输给我们的迷信阴影，为心灵信仰的罪过进行忏悔。

毫不奇怪，取消主义仍是少数人的立场。它有一些优势——如卡尔·波普尔所说：“困难的心身问题会完全消失，这无疑是非常方便的，它使我们免除了解决它的麻烦。”（Karl Popper，1994：8）但它是难以相信的。它必须否定这种观点，即关于我们自己有意识心理状态的知识在认识上先于其他东西的知识，这似乎与对我们怎样通过知觉经验认识周围世界中的东西所作的一种很自然的解释有矛盾。在思考我们提出这种理论时的立场以及解释我们怎样对它进行辩护方面，也有一些困难。因为如果某个人的确主张这个理论是正确的，那么至少有某个人会相信某种东西，即取消主义是正确的，在这种情况下，取消主义就是错误的。困难在于：我们并没有什么词汇来描述接受、拒绝以及支持这些理论，它们没有预设理论家们具有心理状态。取消论者坚持认为，这只是一个实际的困难，但不是他们已经克服的困难。

1.7 结　论

在这里，我们将结束对心身问题及其主要回答的考察，图1—1概述的就是我们考察过的各种立场。

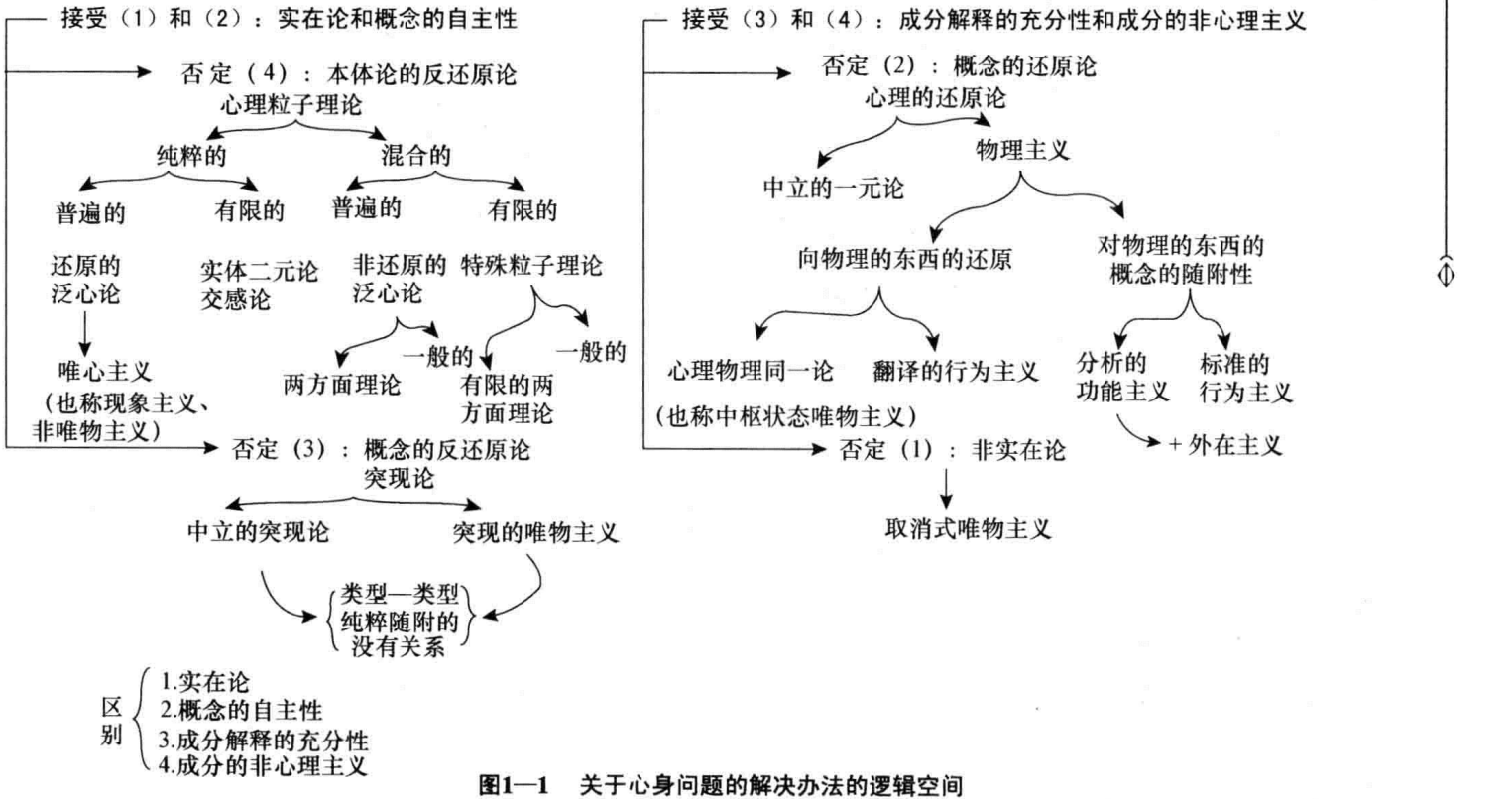


图1—1　关于心身问题的解决办法的逻辑空间

两种基本立场代表着在心身问题上两大根本对立的倾向。我们考察过的一切观点均可归入其中。一种立场主张：心理的东西是实在的基本特征，不可能根据别的特征来解释。这一立场的基本特征是，它同时承认命题（1）和（2），即实在论和概念的自主性。另一立场坚持认为，心理是实在的基本特征这一表象一定是一种错觉，诉诸让我们同时理解世界的明显的非心理现象术语，就能从根本上对我们及我们的一切属性作出完全的理解。这一立场的基本特征是它同时承认命题（3）和（4），即成分解释的充分性和成分的非心理主义。第二种观点受事物的基本成分是物理的（成分的物理主义）这一假设的限制，它相当于物理主义，取消主义是它的一种退化的情况。尽管我们很清楚在回答它时的各种选择，但心身问题没有消失的原因是由于常识与这种势力之间持续不断的斗争。常识支持心理是实在基本特征的观点，而这种势力认为科学的权威意见应是：事实并非如此。我们发现自己在这两极之间被不断地拉来拉去，难以看到我们的心灵并不是物理东西之上的东西，不愿意看到宇宙包含有最终不能根据其基本的、明显的非心理成分解释的东西。

【注释】

[1] 使用“心身问题”这个术语并不是没有歧义。引导我用法的是理解下面这个难题的兴趣：它产生了我们关于心理现象与物理现象关系的哲学和科学的讨论中所发现的各种立场。如果我是正确的，那么就有一个我们能清楚说出来的难题，所有关于心理现象与物理现象关系的立场都可以看作是对它的回答。如果有一个问题值得“心身问题”这个称呼，那就是这个问题。

[2] 在讨论过程中，将采用很多的术语。这部分地使我们能够准确地陈述问题及其可能的解决办法。采用的术语严格说来比做这项工作所需要多，这种多余是想为进一步阅读涉及这个问题的相关文献奠定基础。我将经常提供代表着特定观点或论证的参考文献，这里我列出其中一些，它们共同为心身问题看法的历史和当代的发展提供了一幅相当全面的图画：维西（Vesey，1964）、安德森（Anderson，1964）、奥康纳（O'Connor，1969）、博斯特（Borst，1970）、罗森塔尔（Rosenthal，1971）、布洛克（Block，1980）、埃克尔斯（Eccles，1985）、莱肯（Lycan，1990）、罗森塔尔（Rosenthal，1991）、比克利和勒德洛（Beakley and Ludlow，1992）、沃纳和舒巴卡（Warner and Szubka，1994）、布洛克等（Block et al.，1997）、库尼（Cooney，2000）。罗森塔尔（1991）的书尤其全面，维西（1964）收录了一些在其他文选中找不到的历史资源，安德森（1964）的书收录了早期论述心灵的计算机模型的论文，埃克尔斯（1985）书中的作者主要是科学家，既有哲学的风格又有科学的风格，布洛克等

(1997) 的书尤其奉献了新近的意识研究。

[3] 笛卡尔说："我用'思想'这个术语来理解我们觉知到的正在我们内部发生的一切，就我们对它具有觉知而言。"(1984，第 1 卷：195 [1644：I. 9]) 这符合我们下面描述的意识特征，即对我们的意识模式拥有非推理的知识。笛卡尔还认为：一个状态之所以是一个心理状态，**仅当**它是有意识的。但人们普遍认为，由于后述考虑的理由，这个条件太严格了。

[4] 根据这种有关事件会变化的常识观念，事件是可以确定日期的殊相。它们可能是复杂的，也可能是简单的，我咬我的手指是一个事件，第二次世界大战也是。如果一个物体从 F 变成了非 F，这个事件就是从 F 变成了非 F。如果我们对事件的个体化根据的是它们变化所涉及的物体、时间和属性，那么心理事件是不是物理事件的问题就被还原成了这个问题：心理属性是不是物理属性。

[5] 人们有时认为这太强了。因为（例如）假如一个人预测并料想到某个会引起疼痛的事件发生了，他可能就会由于预期疼痛而错误地认为他感觉到了疼痛。例如，
32 如果某个人被告知一块热得发红的金属块将压在他脖子后面，那么当有人将一个小冰块放在他脖子后面时，他可能就会认为他感觉到了疼痛。然而，这种可能性，即他在这些情况下具有错误的信念，并不表明他不知道他有经验的东西，因为他会纠正他的错误。他很快会意识到他现在和当时都未感觉疼痛，因为他能回想起这种经验是什么样子。这需要知道它在那时所具有的特征，因为一个人不能记住他最初不知道的事情。记忆能保存但不能创造知识。

[6] 对这个问题的讨论，参见布洛克等 (1997) 的书中第 20～24 篇文章。

[7] 参见内格尔 (1979b、1994、1998) 和麦金 (1989、1991、1999)。麦金和内格尔认为：一定有办法理解我们大脑的操作是怎样产生意识的，但我们目前对于这是如何可能的还一无所知。麦金更悲观，因为他认为不管正确的解释是什么，鉴于我们的认知构造，它仍是我们原则上不能理解的解释，而内格尔认为我们总有一天能提出适当的概念。意识是核心难题的观点与心身问题的讨论一样古老。

[8] 这个术语可以追溯到中世纪哲学，它源于拉丁语动词 *intendere*，表示"指向"或"瞄准"。它被用于描述这样的思想对象——它并不**实际地**存在，但具有**意向的内存在**，或者**仅仅意向地存在于思维主体之中**。

[9] 除了我们一直在讨论的这种态度之外，可以认为还有些东西能对事物作出表征，从而具有意向性，例如，一个句子或者一幅肖像画。然而，这些东西具有表征内容，仅仅是由于自主体依据各种规则把它们当成了表征，这是派生的而非本源的意向性（塞尔，1983、1984）。心理状态具有本源的意向性，如果不加限定，我一般是用"意向性"指本源的意向性。

[10] **倾向**是物体的这样一种状态，它是在某些条件下经历某种变化的稳定趋势。水溶性就是盐和糖所拥有的一种简单的倾向状态：当在某个温度和压力范围内被放入不饱和的水中时，它们就会溶解。它所经历的变化描述了一种倾向，这就是它的**表现**

属性，即表现出来的属性。**表现条件**是表现属性得以表现的条件。通常这两者都编码于倾向的名称之中，就像在“水溶性”中那样。倾向性的态度不是简单的倾向，而是吉尔伯特·赖尔所称的“多轨倾向”（1949：43-44），这是指它们是在各种条件下以各种方式表现自身。此外，它们是连锁的倾向：在任意某个态度的表现条件中，有些条件涉及自主体所具有的其他态度。购买某本书的愿望不会表现出来，除非我**相信**我有机会购买它，并且我没有别的置于买这本书的愿望之上的愿望，同时我认为只有排除这个愿望才能满足它们。

[11] 最近很多关于认知活动的理论在其解释中原则上都求助于无意识推理，因而都预设这两者能独立地想象。对这些观点的批评，参见路德维希（1996c）。

[12] 最近，有些哲学家主张：意识状态能根据它们的表征内容作出彻底的描述。例如莱肯（1996）、德雷斯基（1997）和泰（Tye，1997），相反的观点，参见塞尔（1993）、查默斯（1996）和西沃特（Siewert，1998）。关于意识的表征解释的动机经常来自于这一思想，即理解意向状态何以能还原为物理状态，比理解意识何以能还原为物理状态要容易。在我看来，意向性最终可以被理解成一种意识状态，而不是相反的情况，倾向性意向状态的内容来自于它们在意识中的表现。倘若如此，意识与物理东西的关系问题就是基本的问题。

[13] 重要的是，我没有把这里的这类物理属性本身描述成非心理的，但是假定有这些基本属性，它们本身就显然不是心理的。这并没有断定：可以把心理属性分析成第一性的质的逻辑构造，或者是在概念上随附于第一性的质的（参见 1.4 节）。

[14] 参见波伦（Poland，1994：特别是 109-147）和帕皮诺（Papineau，1993：29-32）。

[15] 更适当地说，语言 L 中的一个完全有意义的谓词表达了一个概念并挑选了一种属性。在不同的语言中，同一个词语可能表达不同概念，或者什么概念也未表达。为了简洁我省略了这种相对关系，但是只要我们涉及语言项目与真理、概念和属性的关系，就应当认为它是隐含于其中的。我在很大程度上也忽略了由时态和自然语言中其他语境敏感因素所造成的复杂情况。

[16] 别的属性概念或许有时确实是在这并非为真时被使用了。例如，有人对属性进行个体化可能是根据拥有它们可能的个体集合。那么，两个谓词挑选同一种属性，当且仅当它们必然具有共同的外延，这并不要求同义性（例如，“是三边形”和“是三角形”）。但是，能这样表达的属性同一性论题也能得以表达，而不会有可疑的本体论和无用的术语创新——这是不应该被鼓励的。

[17] 更一般地说，我们会谈到：相对于说话的场合，句子是分析的，因为在自然语言中很多句子所表达的东西都是相对于话语语境的。

[18] 是否有分析陈述、概念真理和能先验地知道的真理？这个问题也是有争议的，但在陈述心身问题时，在这一点上表明立场是不必要的。W. V. 蒯因（W. V. Quine）的《经验论的两个教条》（“Two Dogmas of Empiricism,” 1953）是反分析性论证的名篇，

格赖斯和斯特劳森（Grice and Strawson，1956）的书是早期的一种重要回应。

[19] 通常认为，当前使用的“随附性”是20世纪50年代初由黑尔（R. M. Hare）在伦理理论的语境下提出来的，用以描述伦理属性与自然属性的关系，之后由戴维森（1980）引进了心灵哲学讨论。它最初是在突现论传统中使用的，但是可能并没有非常具体的意义，参见金在权（Kim，1993b：文8）。

[20] 可以对这种阐述作出多种改变。例如，如果我们在双重条件右侧开头加上“在概念上必然的是”，我们就得到了所称的强随附性版本（金在权，1993b：文4），也有更弱的类型。我使用这种阐述，是因为我希望允许从心理到物理的概念还原，尽管可能有一个非物质东西的世界具有心理属性。这是（例如）功能主义悬而未决的一种可能性，同时会给如下思想提供内容，即严格来说，随附性比还原要弱。有时，随附性主张是依据不可辨别性主张来阐述的：F属性随附于G属性，当且仅当在其G属性方面相似的东西，在其F属性方面也必然相似。对这些派生形式及其彼此间关系的进一步讨论，参见金在权（1993b）以及萨维洛斯和亚辛（Savellos and Yalcin，1995）。

[21] 心理规律（包括任何心理物理规律）由物理规律所衍推这个条件，是避开侥幸的唯物主义（lucky materialism）问题所必需的（威特默［Witmer］，2000）。

[22] 这一立场表面看比它实际的样子要强。除了它们是物理的之外，我对物理属性没有施加任何限制，复杂的关系属性可能包含在随附性基础之中。因此，它相当于这种观点：关于世界完备的物理描述蕴含关于它完备的心理解释。

[23] 在更弱的意义上，它被用以表示一种唯物主义本体论，而在更严格的意义上，例如常被逻辑实证主义者用来表示：所有陈述都能翻译成物理学的词汇。

[24] 这里，我用“非心理属性”指这样的属性，它们是用算不上心理的术语进行分类的，因此，这类中有些成员，而且当然是所有基本的（即非复杂的）成员，都不是心理的。这承认心理属性可能是所讨论属性的一个子集，也就是说，（2）宣称不存在这样的属性类型：它们本身不是心理的，心理属性能从概念上还原为它们。

[25] 在现在的语境下，我用非关系属性指一个个体所具有的这样一种属性，它不要求有某个偶然存在的个体，它是与拥有该属性或其任一部分的个体不相同一的，也不要求任何东西是非存在的。例如，是已婚的和是一颗行星就是关系属性，是圆的和是红色的就不是。

[26] 由于成分所构成的整体的突现属性，因此这排除了对它们所具有的属性的诉诸。之所以反对诉诸成分所拥有的属性，是由于它们所构成整体的突现属性。

[27] 对于它们在如下意义上可能具有心理属性并未作出判定，即它们拥有可以衍推某种东西拥有心理属性的关系属性，（例如）因为心理属性与这样一种东西是共存的或是其一部分——这种东西具有不可还原的心理属性但它本身不是事物的一个基本成分。这也没有判定：事物的基本成分具有这样的属性，我们不能把它们看成是宽泛的物理属性，但也不能承认它们是心理属性。因此，成分的非心理主义是一个比成分的物理主义更自由的论题。

[28] 对这些用于揭示什么是必要的和可能的方法的一般辩护，参见比勒（Bealer，1992），最近对概念分析厚达一本书的辩护参见杰克逊（1998）。

[29] 这个悖论是伯特兰·罗素在1901年5月发现的，它在20世纪早期的集合论和数学的基础研究方面起着重要作用。

[30] 意识与意向性的关系问题在这里是很重要的，因为以上所提到的思想实验似乎依赖于我们的这种思考，即有意识的观点在一种与我们的物理和行为方面都相似的东西身上可能是没有的，或者它出现在一种根本没有相关身体的东西身上。如果意向状态和意识状态无关，那么这些思想实验对于心理之物的不可还原性的支持就会被简单地降低。

[31] 然而，例外情况是观察的观念在量子力学中的作用：该怎样认真地看待这一点仍是一个有争议的问题。

[32] 对希腊化时期古代心灵哲学的综述，参见安纳斯（1992）。

[33] 这是古代的一种少数派立场，是由柏拉图提出来的。它在后来西方传统中所具有的重要意义，是通过柏拉图哲学对天主教神学的影响，通过这种影响，它将关于心灵和身体的观念渗透进了西方文化之中。

[34] 对于早期现代哲学中理性主义者所用的实体概念的讨论，参见伍尔豪斯（Woolhouse，1993）。

[35] 这里，我们必须排除像具有一种属性这样的“形式”属性。

[36] 这个论证中的最初步骤是在笛卡尔名著《第一哲学沉思集》（*Meditations on First Philosophy*，1985 [1641]）的第二沉思中提出来的，他在第六沉思中作了结论，也参见《哲学原理》（*Principles of Philosophy*，1985 [1644]：§63）。

[37] 这个类比是莱布尼兹设想的，但是他的基本形而上学反对实体二元论。

[38] 尽管当前二元论不是哲学家或科学家中间的一种流行观点，但是，如果不假思索的话，它无疑仍是最广为人们接受的一种关于心理现象与物理现象关系的观点，因为它是世界上主要宗教的背景形而上学，而且在哲学家和科学家中它并非没有当代的支持者，例如福斯特（Foster，1996）、埃克尔斯（1953：第8章）、波普尔和埃克尔斯（1977：第E7章）。

[39] 20世纪的三大里程碑著作是卡尔纳普（Carnap，1928）、刘易斯（1929）和古德曼（Goodman，1951）。新近的一个支持者是格雷灵（Grayling，1985）。

[40] 详细的资料文献目录可以在爱德华兹（Edwards，1967）论述泛心论的文章之后得到。

[41] 如波普尔和埃克尔斯（1977：14）所说，就还原的唯物论来说，泛心论和别的心理粒子理论表达了这种思想——太阳底下无新事，这是充足理由原则的一种表达方式：无物能产生于无。

[42] 19世纪的两方面论者有沙德沃思·霍奇森（Shadworth Hodgson，1870：特别是第3章）和刘易斯（G. H. Lewes，1877）。

[43] 例如，参见莫顿·普林斯（Morton Prince，1885；重印于维西，1964：187）的讨论。

[44] 或许斯特劳森（1958）关于人的概念比人心或人身的概念更基本的观点，可以被看成是这一种。

[45] 我们应当怎样对像休谟关于自我的“束”理论这样的理论进行分类？根据这种观点，不存在自我这种东西，相反，每个自我都被看成是由一组知觉构成的，这些知觉彼此之间具有合适的关系。这些知觉就像心理原子一样，本质上具有心理的特征。但是，显然不能认为它们处于空间之中。因此，尽管它是一种心理粒子观，但是更像这样的实体二元论，它没有作为能思之物的基本心理实体，但具有构成能思之物的思想。如果我们认为“知觉”本身是非心理的，认为自我和普通物体都是由它们组成的逻辑构造，那么我们就得到了詹姆斯、马赫和罗素所倡导的中立一元论的一个版本的支持（见下文）。

[46] 将这种意义上的突现，与经常在“混沌”理论语境下讨论的受简单规则控制的、复杂系统的高层次组织的突现区别开是很重要的。后面一种属性在概念上随附于
36 控制系统的成分以及这些成分的属性和排列的规则。它们仅在令我们吃惊的意义上才是突现的，因此，在这种意义上，它们作为突现之物的地位是一种功能，我们难以轻松地预言它们。

[47] 这要求我们反对来自表达相关类型的无限长的析取，否则，对于某个心理类型来说，将所有用物理术语陈述的法则上的充分条件拆开，我们就总能得到一个法则上的必要条件。

[48] 个例就是类型的一个实例，例如，前一个句子（题字）中有 4 个字母“a”的个例。个例总是殊相，每一个个例都与它自身相同一。当我们用不同方法挑选同一个东西时，我们就得到了关于个例同一性增进知识的陈述。例如，被告知冥王星是太阳系中最小的行星，可能是增进知识的。严格说来，类型同一性关涉的是属性。此外，每个属性都与它自身相同一，而且与任何不同的东西都不同一。增进知识的类型同一性陈述是用不同方法来挑选属性。我们后面会看到心理之物和物理之物类型同一论的一个例子，它会使这成为一个有趣的经验发现。

[49] 注释 4 中所述的事件概念和异常一元论是不相容的，因为它对事件的个体化是依据作出变化的事件的对象和属性。因此，除非心理属性就是物理属性（根据这种看法它们并不如此），否则没有一个心理事件是与某个物理事件个例同一的。我们还能说出各种更弱的关系，例如，有人可能说：每个心理事件都与一个物理事件同时出现并且出现在同一个对象之中。总之，还不能确定有很多东西都以此为转移。更基本的问题是关于对象和属性而非事件的。

[50] 最初它是一个医学术语，指“潜在病因的症状”或者“次要症状”。

[51] 对于更宽泛的突现论传统中的这个特殊学派的讨论，参见麦克劳林（McLaughlin，1992）。不过要知道，麦克劳林所用的“突现论”比这里所用的意义要

窄，即包括我所说的具有下向因果关系的突现的唯物主义。“突现论”是表达否定（3）的合适术语，我们可以将副现象版本与非副现象版本区别开，后者至少会包括具有下向因果关系的突现论。哎，哲学中术语的变化是一个特色。布罗德提出了“突现的唯物主义”这个术语，他本人并未用它暗示下向因果关系，他根据他所认为的心灵研究的证据，试探性地把它当成了一个经验假说。

[52] 副现象的突现论者与下向因果关系的突现论者之间的这种对比和争论，是对古代亚里士多德追随者之间类似争论的重复（Caston，1997）。

[53] 更多新近的讨论，参见阿姆斯特朗（1968：47）和金在权（1993a、b）。

[54] 对“属性二元论”这个术语作点解释是合适的，这个术语主要出现在最近关于心身问题的文献中。这个标签是在突现论的应用中经常使用的，它也适用于这样的立场，它认为：存在具有心理属性的东西，也存在具有物理属性的东西，这两种属性都是基本属性，它们不能从概念上还原为对方或还原为更基本的东西。（属性二元论与任何主张（1）和（2）以及（3）或者（4）的立场不是共外延的，比如唯心主义接受（1）—（3），但将通常所认为的物理属性还原为了心理属性。）**属性**二元论是一种比**实体**二元论弱的观点，但它是由实体二元论所引起的。“属性二元论”常常被看成是一个受还原论吸引的哲学家所滥用的术语，这些哲学家都有这样的思想：通过交叉使用他们的标签而将属性二元论的支持者与不足信的实体二元论观点结合起来。将“属性二元论”引入哲学词汇不完全是一项愉快的术语创新，而这是它没有突出地包含在我讨论中的一个原因。且不说它与煽动行为（demagoguery）的联系，这个标签错误地指出，至少有两个不能相互还原的属性家族。但是即使撇开现在的问题，也有很多不能相互还原的属性家族（例如，颜色和形状这两种属性）。

[55] 一些可能但无人持有的立场，我在这里就不考察了。如这种观点：心理的东西随附于或者从概念上能还原为非心理的东西，物理的东西反过来随附于或从概念上能还原为心理的东西。

[56] 在这方面，它类似于数学向逻辑学和集合论的还原。我们可以保留自己的旧言语形式，但我们的本体论只包含集合，而不包含另外的数。根据中立一元论，心理的东西和物理的东西与基本实在的关系，类似于根据数学的集合论还原的奇数和偶数与基本实在的关系。每一方都不同一于另一方，都有另一方不可能有的本质属性，但每一方都被解释成由更基本的东西所构成的一种逻辑构造。

[57] 卡尔纳普（1931）和亨普尔（Hempel，1935）提供了初期逻辑行为主义的例子，但两者后来都放弃了早期的立场。赖尔的《心的概念》（*The Concept of Mind*，1949）是一个重要且有影响的行为主义宣言（但是赖尔拒绝接受用于其观点的这个术语），维特根斯坦的《哲学研究》（*Philosophical Investigations*，1950）是标准行为主义的一个重要灵感。例如，参见马尔科姆（Malcolm，1958）。重要的心理学行为主义者是华生（Watson，1925）和斯金纳（Skinner，1974），但他们的行为主义是方法论的行为主义而非逻辑的行为主义。

[58] 参见普特南（1968）。逻辑行为主义似乎遇到了每个还原计划都会遇到的危险，如刘易斯所说："在遇到很难解决的分析问题时，有的人有时可能会通过改变主题来绕过它们。"（C. L. Lewis，1941：225）

[59] 这被认为是一个真实的事实。当然，如果有这种非物理的东西，它们有内部的结构，那么它们也会有相应的功能状态。

[60] 每个时代都有其偏爱的、由其声望高的技术所提供的隐喻。在17、18世纪受偏爱的隐喻是钟表和磨坊，19世纪是蒸汽机，20世纪后半叶又变成了计算机。

[61] 在古代，毕达哥拉斯学派支持这种一般的功能主义思想：具有心灵有赖于身体的某种组织。在《斐多篇》（*Phaedo*）中，苏格拉底在答复西米亚斯的意见时所持的立场之一就是：灵魂之于身体，就像一个弦乐器的调音之于这个弦乐器（Plato，1989：69）。

[62] 理解心理功能主义的经验特征有两种方式。第一种方式，可以将它理解成一种具有这样的桥梁法则的突现论版本，这些法则将功能状态与心理状态联系了起来（参见查默斯，1996：第6章）；第二种方式，可以相信：心理状态与功能状态的同一是一种理论的同一，就像闪电与电的释放的同一一样（雷伊［Rey］1997年的著作支
38 持这种观点适用于意向状态）。如果我在下文中就此所作的说明是正确的，那么这就将一种经验的因素引入了这个发现，但这并不会阻止这种观点——如果正确的话——退化为分析的功能主义。参见下文对同一论的讨论。

[63] 这个一般主题新近的变种是联结主义。联结主义系统是由一组相互连接的单元构成的，这些单元具有激活值，相互的连接决定着某个结点的激活值对与之相接的结点的影响。通过相互的连接，单元可能使其他单元抑制或兴奋，而抑制或兴奋的不同程度则取决于它们自身的激活状态，有些单元可能是指定的输入单元，有些则是输出单元。激活值可能是连续的，因此一个联结主义系统不是一个有限状态机。但它与我们起初关于功能系统非常一般的描述是符合的，因为不同的联结主义系统完全是根据它们的状态与输入和输出以及其他状态的关系来描述的。经典功能主义与联结主义之间的区别在我们这里的讨论层次上是不相干的。

[64] 这个反驳的力量还不清楚。要么心理属性能分析为功能属性，要么不能。如果不能，那么就有这个问题，即：它们是否能分析为物理属性。如果能，问题就会终结，沙文主义的指责就没用了。如果不能，心理状态类型与什么物理状态类型——若有的话——相联系，就是一个经验问题，而且我们关于它的直觉或偏见都是不相干的。然而，在后一种情况下，当我们打算确证或否证这些与我们自己迥然有别的物理系统的主张时，显然会出现困难。

[65] 也参见塞尔（1980）的中文屋思想实验，以及查默斯（1996：第3章），查默斯最近用所谓的怪人思想实验来证明有意识的心理状态之不可还原性。

[66] 普特南谨慎地排除了包含有这些部分的系统，这些部分具有与它们所构成的整体相似的结构。这排除了这种可能性，即布洛克思想实验中的系统构成了一个人。

然而，要理解是什么证明了这种排除是很困难的。因为如果我们的心理概念是功能概念，那么具有合适功能结构的系统是怎样构成的，就应该是无关紧要的。

[67] 更充分的讨论以及对出现在文献中反驳的一些回应，参见查默斯（1996：91-101）。

[68] 功能主义者不一定需要这一点。功能系统有可能依据被给定了输入和输出的状态之间的非因果转换来描述，但是对于可能具有心灵的物体来说，这让它们拥有的机器表描述比功能主义者通常想赞成的要多得多。

[69] 今天晨报标题中报道的事件导致佛罗里达沿海地区大规模的海水泛滥，但它并不是由于从前的那类原因，而是由于3号飓风经过海岸。因果关系是特殊事物、可以确定日期的事件或状态之间的关系，但要解释为什么它们是那些特殊事物之间的关系，我们必须借助于它们的类型。

[70] 有关讨论，参见杰克逊和皮蒂特（Jackson and Pettit，1988）、布洛克（1990）、福多（Fodor，1991）、达迪斯（Dardis，1993）和路德维希（1994a、1998）。

[71] 在这一点上，意向状态是否在概念上独立于意识状态的问题也很重要，因为我们关于心理状态与行为因果相关的信念似乎首先是与有意识的心理状态联系在一起的，而且只有通过它们在意识中的表现条件才与倾向状态联系在一起。因为倾向状态也是根据表现条件定义的，因此与那些条件也不是因果相关的。

[72] 这里给出的解释不同于刘易斯本人的解释。差异可能主要是术语上的，但仍然很重要。刘易斯认为：尽管疼痛与人身上的一种物理属性有理论的同一关系，但是，说尽管某种存在物（例如一个火星人）没有这种在人身上就是疼痛的属性，但他能感觉疼痛，仍是有道理的（刘易斯，1980）。这何以可能？如果我们像我所介绍的那样来理解谓词、概念和属性之间的关系，那这就是不可能的。根据我给出的解释，谓词“感觉疼痛”（is in pain）表达了疼痛概念，并被用于归属感觉疼痛这种属性，而且它完成这些属性中的每一种都是由于它在英语中的意义。可以说，属性就是谓词的意义投在世界上的阴影，概念是它投在我们人类思想中的阴影。如果感觉疼痛这种属性是一种物理属性，那么根据这种观点，疼痛概念就是一个物理概念。然而，刘易斯认为疼痛概念是别的东西。简单地说，刘易斯用“疼痛概念”指由下面这个谓词所表达的概念：“是一种具有属性P的东西，以至于T(P)多半适合于K类存在物。”这里，“T(P)”由合适的心理学理论替换，这种理论用代表疼痛属性的变量来替换“P”。概念适用于一种东西，是由于那种东西具有某种属性，这种属性在输入和输出之间起某种中介作用，或许起那种作用的是不同的属性，而且在不同种类的存在物身上，很可能多半是不同的属性起着那种作用。然而，刘易斯并没有说：感觉疼痛这种属性是用这种形式的谓词归属的属性。相反，刘易斯把实际起这种作用的属性称为感觉疼痛的属性。那么也就承认：在不同种类的存在物身上，一种不同的属性可以是（被称为）感觉疼痛的属性。它显然也承认：如果事物不同，那么我们身上的一种不同属性也许就是（被我们称为）感觉疼痛的属性。然而，刘易斯显然真想这样对待“感觉疼痛”这个谓

词：它似乎归属了起着正确作用的属性。因此，他说：在我们将“感觉疼痛”运用于不同种类的存在物，当我们在不同的可能世界上考虑它时，它是有歧义的。这是由于所归属属性的不同蕴含了谓词意义的不同。似乎我们已经决定说：“是富裕的”（being rich）这种属性的归属是使用“拥有比大多数人多得多的钱”，但富裕（richness）的概念是由“是路德维希偏爱的属性”这个谓词表达的。我将疼痛的概念与谓词“感觉疼痛”联系在一起，从而将它与用它归属的属性匹配在一起。这遵循了传统的路线，并且在这些问题上给我们提供了更清晰的观点。

[73] 有很多反对心理物理同一论和一般物理主义的论证，这些论证都依赖于这些思想实验：它们是设计来证明从一个物体物理构成的详尽描述推导不出它具有什么心理属性。最近很多人讨论的一种论证被称为“知识论证”，20 世纪后半叶运用这个论证的人有米尔（Meehl，1966）、内格尔（1979b）和杰克逊（1982、1986）。莱布尼兹在《单子论》（*The Monadology*，1714：第 16 章）中也给出了这个论证的一个版本：“假若我们设想有一台机器，它的结构使它可以思想、感觉、具有知觉，我们可以设想，把它保持相同的比例放大到我们可以走进去，就像走进一座磨房一样，如果检查
40 它的内部情况，我们将只会发现一些在相互推动的部件，而根本找不出什么东西来解释知觉。”这些论证当然都坚决反对任何版本的心理物理同一论，后者提出：我们可以对心理概念作出一种不切实际的（armchair）分析，以此来确定它们实际地挑选出了我们的神经生理属性。这些论证没有处理这一理论的这些版本，它们认为我们关于普通术语的概念是由描述决定的，可是，这种方案所承担的任务同样艰巨，因为如我们所看到的，它必须承认我们的术语根本不能表达任何概念。

[74] 这种观点本身在 20 世纪以前当然也被讨论过。布罗德对它作出了讨论并摒弃了它（1925：622-623）。刘易斯讨论并对一种同一论形式作出了批评，他的理论是想提出心理术语的描述定义，其所秉承的精神与我所提出的这种理论完全相同（1941：230-231）。斯马特对这些异议的一些回应显然针对的是布罗德和刘易斯早期的讨论。

[75] 他们认为，命题态度能从行为或功能上理解。然而，这一立场很容易就能推广到命题态度自身。

[76] 关于反对诉诸科学本质主义的论证，参见比勒（Bealer，1987、1994）。

[77] 总之，应当注意的是，与这里所使用的所有属性同一性概念相联系的含糊性，同样也与形而上学的必然性相联系。如果打算依据传统来解释它，我们就必须承认自己所发现的是：“水”和“H_2O”表达了相同的概念，而这与假设是相反的。

[78] 这些都起始于克里普克（1980）对专名和自然种类术语的研究，以及希拉里·普特南（1975）在 20 世纪 70 年代初对自然种类术语的研究。最初，这些论证是想证明：各种自然语言术语的意义是由它们与我们环境中的事物和种类的因果关系决定的。然而，由于我们也用这同一些术语描述我们的态度，因此人们很快就明白了：

这些论证也可以用以强调我们思想的内容是相对于我们环境中实际存在的事物和种类来进行个体化的。

[79] 参见普特南（1975）和伯奇（Burge，1979、1982、1986）。不加批判地普遍接受外在主义，在当代心灵哲学的讨论中是一个突出的特征。

[80] 各种困难在路德维希（1992a、1992b、1993a、1993b、1993c、1994a、1994b、1996a、1996b）的著作中得到了讨论。

[81] 这个主题的文献很多。较早的一篇论文是布吕克纳（Brueckner，1986），它提出了这个论题，尤其是在对普特南（1981）作出回应的时候。也参见博格西安（Boghossian，1989、1993）。

[82] 关于对心理因果关系讨论的一个相当全面的回顾，参见杰克逊（1996）。

[83] 早期的支持者是费耶阿本德（1963）和罗蒂（1965、1979）。可能维特根斯坦在《逻辑哲学论》（*Tractatus Logico-Philosophicus*，1921）中也赞成取消主义，但是倘若如此，他所依据的理由比新近的取消主义者要更抽象。取消主义可能是一种现代观点，它在古代哲学中没有代表人物。原子论者留基伯和德谟克利特（Leucippus and Democritus）可能被认为是赞成取消主义的，因为他们认为实在只是由原子和虚空组成的，但是他们没有表现出这种倾向，即否认存在能思想、能进行推理的人，而且德谟克利特似乎想根据他的原子论形而上学来解释心理现象（参见泰勒［Taylor］，1999）。一个人可以是不彻底的取消论者，也可以是彻头彻尾的取消论者。乔治·雷伊赞成关于意向状态的功能主义，却也赞成关于质的状态的取消主义（1997）。

参考文献

Alexander, S. (1920). *Space, Time and Deity: The Gifford Lectures at Glasgow 1916—1918*. London: Macmillan.

Anderson, A. R. (ed.) (1964). *Minds and Machines*. Englewood Cliffs: Prentice Hall.

Annas, J. (1992). *Hellenistic Philosophy of Mind*. Berkeley: University of California.

Aristotle. (1984). *The Complete Works of Aristotle*. Princeton: Princeton University Press.

Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul, Humanities Press.

Ayer, A. J. (ed.) (1959). *Logical Positivism*. New York: Free Press.

Beakley, B. and Ludlow, P. (eds.) (1992). *The Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press.

Bealer, G. (1987). "The Philosophical Limits of Scientific Essentialism." *Philo-*

sophical Perspectives, 1: 289–365.

—— (1992). "The Incoherence of Empiricism." *Proceedings of the Aristotelian Society*, 66: 99–138.

—— (1994). "Mental Properties." *The Journal of Philosophy*, 91: 185–208.

Beckermann, A., Flohr, H. et al. (1992). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. New York: W. de Gruyter.

Bedau, M. (1986). "Cartesian Interaction." *Midwest Studies in Philosophy*, 10: 483–502.

Berkeley, G. (1710). *A Treatise Concerning the Principles of Human Knowledge*. In Michael R. Ayers (ed.), *Philosophical Works*. London: The Guernsey Press, 1975: 61–128.

Block, N. (1978). "Troubles with Functionalism." In C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: University of Minnesota: 261–325. Reprinted in Block (1980).

—— (ed.) (1980). *Readings in the Philosophy of Psychology*. Cambridge, MA: Harvard University Press.

—— (1990). "Can the Mind Change the World?" In G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*. New York: Cambridge University Press: 29–59.

Block, N., Flanagan, O. et al. (eds.) (1997). *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.

Boghossian, P. (1989). "Content and Self–Knowledge." *Philosophical Topics*, 17: 5–26.

—— (1993). "The Transparency of Content." *Philosophical Perspectives*, 8: 33–50.

Borst, C. V. (ed.) (1970). *The Mind-Brain Identity Theory*. London: Macmillan.

Brentano, F. (1955 [1874]). *Psychologie vom Empirischen Standpunkt* (*Psychology from an Empirical Standpoint*). Hamburg: Felix Meiner.

Broad, C. D. (1925). *The Mind and Its Place in Nature*. New York: Harcourt, Brace and Company.

Brueckner, A. (1986). "Brains in a Vat." *The Journal of Philosophy*, 83: 148–167.

Burge, T. (1979). "Individualism and the Mental." *Midwest Studies in Philosophy*, 4: 73–121.

—— (1982). "Other Bodies." In A. Woodfield (ed.), *Thought and Object: Essays on Intentionality*. Oxford: Clarendon Press: 97–120.

—— (1986). "Cartesian Error and the Objectivity of Perception." In P. Pettit (ed.), *Subject, Thought, and Context*. New York: Clarendon Press: 117–136.

Carnap, R. (1928). *The Logical Structure of the World*, trans. Rolf A. George. Berkeley: University of California Press, 1967. First published in German in 1928 under the title Der Logische Aufbau der Welt.

—— (1931). "Die Physikalische Sprache als Universalsprache der Wissenschaft." *Erkenntnis*, 2: 432-465. Reprinted in Ayer (1959) as "Psychology in Physical Language."

Caston, V. (1997). "Epiphenomenalisms, Ancient and Modern." *The Philosophical Review*, 106: 309-364.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Churchland, P. (1981). "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy*, 78: 67-90.

Cooney, B. (ed.) (2000). *The Place of Mind*. Stanford: Wadsworth.

Dardis, A. (1993). "Sunburn: Independence Conditions on Causal Relevance." *Philosophy and Phenomenological Research*, 53: 577-598.

Davidson, D. (1980). "Mental Events." In *Essays on Actions and Events*. New York: Clarendon Press: 207-224.

Descartes, R. (1985 [1641]). *Meditations on First Philosophy*. In *The Philosophical Writings of Descartes*, vol. II, ed. and trans. J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press: 1-62.

—— (1984 [1644]). *The Principles of Philosophy*. In *The Philosophical Writings of Descartes*, vol. I, ed. and trans. J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press: 177-292.

Dretske, F. (1997). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Eccles, J. C. (1953). *The Neurophysiological Basis of Mind; The Principles of Neurophysiology*. Oxford: Clarendon Press.

—— (ed.) (1985). *Mind and Brain: The Many-Faceted Problems*. New York: Paragon House.

Edwards, P. (1967). *The Encyclopedia of Philosophy*. New York: Macmillan.

Feigl, H. (1958). "The Mental and the Physical." In H. Feigl et al., *Concepts, Theories and the Mind-Body Problem*. Minneapolis: University of Minnesota Press: 370-497.

Feyerabend, P. (1963). "Mental Events and the Brain." *The Journal of Philosophy*, 60: 295-296.

Fodor, J. A. (1991). "A Modal Argument for Narrow Content." *The Journal of Philosophy*, 88 (1): 5-26.

Foster, J. (1996). *The Immaterial Self: A Defence of the Cartesian Dualist Con-*

ception of the Mind. London: Routledge.

Grayling, A. C. (1985). *The Refutation of Scepticism*. LaSalle, IL: Open Court.

Grice, H. P. and Strawson, P. F. (1956). "In Defense of a Dogma." *Philosophical Review*, 65: 141–158.

Goodman, N. (1951). *The Structure of Appearance*. Cambridge, MA: Harvard University Press.

Hempel, C. (1935). "The Logical Analysis of Behavior." In Block (1980: 14–23). Originally published in Revue de Synthese, 1935.

43 Hodgson, S. (1870). *The Theory of Practice*. London: Longmans, Green, Reader and Dyer.

Huxley, T. H. (1874). "On the Hypothesis that Animals are Automata, and its History." In Vesey (1964: 134–143). Originally delivered in 1874 to the British Association for the Advancement of Science, Belfast.

—— (1901). *Methods and Results: Essays*. New York: D. Appleton.

Jackson, F. (1982). "Epiphenomenal Qualia." *The Philosophical Quarterly*, 32: 127–136.

—— (1986). "What Mary Didn't Know." *The Journal of Philosophy*, 83: 291–295.

—— (1996). "Mental Causation." *Mind*, 105 (419): 377–413.

—— (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Clarendon Press.

Jackson, F. and Pettit, P. (1988). "Functionalism and Broad Content." *Mind*, 97: 381–400.

James, W. (1950 [1890]). *The Principles of Psychology*. New York: Dover.

—— (1910). *Psychology*. New York: Henry Holt and Co. Page references in the text are given to Block et al. (1997), in which a portion of chapter 11 is reprinted.

—— (1904). "Does Consciousness Exist?" *The Journal of Philosophy, Psychology and Scientific Methods*, 1: 477–491. Repr. in James (1976 [1912]).

—— (1976 [1912]). *Essays in Radical Empiricism*. Cambridge, MA: Harvard University Press.

Kant, I. (1997 [1781]). *The Critique of Pure Reason*. Cambridge: Cambridge University Press. Originally published in 1781, a much revised version was published in 1787.

Kim, J. (1993a). "Mechanism, Purpose, and Explanatory Exclusion." In Kim (1993b: 237–264).

—— (1993b). *Supervenience and Mind*. New York: Cambridge University Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Leibniz, G. W. (1714). *Principles of Philosophy, or, The Monadology*. In *Philosophical Essays*. Indianapolis: Hackett, 1989: 213–224. Originally composed in 1714 for private correspondence as a summary of his philosophical views.

Lewes, G. H. (1877). *The Physical Basis of Mind*. London: Trubner and Co.

Lewis, C. I. (1929). *Mind and the World-order: Outline of a Theory of Knowledge*. New York: Scribner.

—— (1941). "Some Logical Considerations Concerning the Mental." *The Journal of Philosophy*, 38: 225–233.

Lewis, D. (1966). "An Argument for the Identity Theory." *The Journal of Philosophy*, 63: 17–25.

—— (1972). "Psychophysical and Theoretical Identifications." *The Australasian Journal of Philosophy*, 50: 249–258. Repr. in Lewis (1999).

—— (1980). "Mad Pain and Martian Pain." In Block (1980: 216–22).

—— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Ludwig, K. (1992a). "Skepticism and Interpretation." *Philosophy and Phenomenological Research*, 52 (2): 317–339.

—— (1992b). "Brains in a Vat, Subjectivity, and the Causal Theory of Reference." *Journal of Philosophical Research*, 17: 313–345.

—— (1993a). "Direct Reference in Thought and Speech." *Communication and Cognition*, 26 (1), 49–76.

—— (1993b). "Externalism, Naturalism, and Method." In E. Villanueva (ed.), *Naturalism and Normativity*. Atascadero, Ridgeview: 250–264.

—— (1993c). "Dretske on Explaining Behavior." *Acta Analytica*, 811: 111–124.

—— (1994a). "Causal Relevance and Thought Content." *Philosophical Quarterly*, 44: 334–353.

—— (1994b). "First Person Knowledge and Authority." In G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy*. Dordrecht: Kluwer: 367–398.

—— (1996a). "Singular Thought and the Cartesian Theory of Mind." *Nous*, 30 (4): 434–460.

—— (1996b). "Duplicating Thoughts." *Mind and Language*, 11 (1): 92–102.

—— (1996c). "Explaining Why Things Look the Way They Do." In K. Akins (ed.), *Perception*. Oxford: Oxford University Press: 18–60.

—— (1998). "Functionalism and Causal Relevance." *Psyche* 4 (3) (http://psyche.cs.monash.edu.an/v4/psyche-4-03-ludwig.html).

Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

—— (ed.) (1990). *Mind and Cognition: A Reader*. Oxford: Blackwell.

Mach, E. (1897 [1886]). *Ernst Mach: Contributions to the Analysis of Sensations*. Chicago: Open Court.

Malcolm, N. (1958). "Knowledge of Other Minds." *The Journal of Philosophy*, 55: 969-978. Reprinted in Rosenthal (1991).

McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind*, 98: 349-366.

—— (1991). *The Problem of Consciousness*. Oxford: Blackwell.

—— (1999). *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.

McLaughlin, B. (1992). "The Rise and Fall of British Emergentism." In Beckermann (1992).

Meehl, P. E. (1966). "The Compleat Autocerebroscopist: A Thought-Experiment on Professor Feigl's Mind-Body Identity Thesis." In P. K. Feyerabend and G. Maxwell (eds.), *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*. Minneapolis: University of Minnesota Press: 103-180.

Menzies, P. (1988). "Against Causal Reductionism." *Mind*, 97: 551-574.

Moore, G. E. (1959 [1903]). "The Refutation of Idealism." *Philosophical Studies*. Paterson, N. J.: Littlefield, Adams and Co.: 1-30. Originally published in *Mind*, 13: 433-453.

Morgan, C. L. (1923). *Emergent Evolution: The Gifford Lectures*. London: Williams and Norgate.

Nagel, T. (1979a). "Panpsychism." In *Mortal Questions*. Cambridge: Cambridge University Press: 181-195.

—— (1979b). "What Is It Like to Be a Bat?" In *Mortal Questions*. Cambridge: Cambridge University Press: 165-180.

—— (1994). "Consciousness and Objective Reality." In Richard Warner (ed.), *The Mind-Body Problem: A Guide to the Current Debate*. Oxford: Blackwell: 63-68.

—— (1998). "Conceiving the Impossible and the Mind-Body Problem." *Philosophy*, 73 (285): 337-352.

O'Connor, J. (ed.) (1969). *Modern Materialism: Readings on Mind-Body Identity*. New York: Harcourt, Brace and World.

Papineau, D. (1998). "Mind the Gap." *Philosophical Perspectives*, 12: 373-388.

—— (1993). *Philosophical Naturalism*. Oxford: Blackwell.

Place, U. T. (1956). "Is Consciousness a Brain Process?" *British Journal of Psychology*, 45: 243-255.

Plato (1989). *The Collected Dialogues of Plato*. Princeton: Princeton University Press.

Poland, J. S. (1994). *Physicalism, the Philosophical Foundations*. Oxford: Clarendon Press.

Popper, K. (1994). *Knowledge and the Body - Mind Problem: In Defence of Interaction*. London: Routledge.

Popper, K., and Eccles, J. (1977). *The Self and its Brain*. New York: Springer-Verlag.

Prince, M. (1885). *The Nature of Mind and Human Automatism*. Philadelphia: J. B. Lippincott Company.

Putnam, H. (1967). "Psychological Predicates." In W. H. Capitan and D. D. Merrill (eds.), *Art, Mind and Religion*. Pittsburgh: Pittsburgh University Press: 37-48. Reprinted under the title "The Nature of Mental States" in Rosenthal (1991).

—— (1968). "Brains and Behavior." In J. R. Butter (ed.), *Analytic Philosophy*. Oxford: Basil Blackwell: 1-19. Reprinted in Rosenthal (1991).

—— (1975). "The Meaning of 'Meaning'." In *Mind, Language and Reality: Philosophical Papers*. Cambridge: Cambridge University Press: 215-271.

—— (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Quine, W. V. O. (1953). "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, MA: Harvard University Press: 20-46.

Rey, G. (1997). *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. New York: Blackwell.

Romanes, G. J. (1895). *Mind and Motion and Monism*. New York: Longmans Green.

Rorty, R. (1965). "Mind-Body Identity, Privacy, and Categories." *The Review of Metaphysics*, 19: 24-54.

—— (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rosenthal, D. (ed.) (1971). *Materialism and the Mind-Body Problem*. Englewood Cliffs: Prentice-Hall.

—— (1991). *The Nature of Mental States*. New York: Oxford University Press.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson and Company, Ltd.

Russell, B. (1917). "On the Relation of Sense-data to Physics." In *Mysticism and Logic, and Other Essays*. London: Allen and Unwin: 108-131.

—— (1921). *The Analysis of Mind*. London: G. Allen and Unwin.

Savellos, E. and Yalçin, U. (eds.) (1995). *Supervenience: New Essays*. New York: Cambridge University Press.

Searle, J. R. (1980). "Minds, Brains, and Programs." *The Behavioral and Brain Sciences*, 3: 417–424. Reprinted in Rosenthal (1991).

—— (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

—— (1984). "Intentionality and Its Place in Nature." *Dialectica*, 38: 87–100.

—— (1993). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Siewert, C. (1998). *The Significance of Consciousness*. Princeton: Princeton University Press.

Skinner, B. F. (1974). *About Behaviorism*. New York: Knopf.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *The Philosophical Review*, 68: 141–156.

Sperry, R. W. (1986). "Discussion: Macro-Versus Micro-Determination." *Philosophy of Science*, 53: 265–270.

Spinoza, B. (1994). *A Spinoza Reader: The Ethics and Other Works*. Princeton: Princeton University Press.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.

Strawson, P. F. (1958). "Persons." In Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science* 2. Minneapolis: University of Minnesota Press: 330–353.

Taylor, C. C. W. (1999). "The Atomists." A. A. Long (ed.), *The Cambridge Companion to Early Greek Philosophy*. Cambridge: Cambridge University Press: 181–204.

Tye, M. (1997). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.

Vesey, G. N. A. (ed.) (1964). *Body and Mind: Readings in Philosophy*. London: George Allen and Unwin.

Warner, R. and Szubka, T. (eds.) (1994). *The Mind-Body Problem: A Guide to the Current Debate*. Oxford: Basil Blackwell.

Watson, J. B. (1925). *Behaviorism*. New York: W. W. Norton.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge* (1st edn). New York: Knopf.

Witmer, D. G. (2000). "Sufficiency Claims and Physicalism: A Formulation." In C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*. New York: Cambridge University Press: 57–73.

Wittgenstein, L. (1961 [1921]). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul Ltd. First published in 1921 in Annalen der Naturphilosophie.

—— (1950). *Philosophical Investigations*. London: Macmillan.

Woolhouse, R. S. (1993). *Descartes, Spinoza, Leibniz: The Concept of Substance in Seventeenth Century Metaphysics*. London: Routledge.

第2章　心身问题

威廉·G·莱肯（William G. Lycan）

人类，可能还有其他生物，都既有心灵也有身体。但什么是心灵？它与身体，或者与一般的物理之物是什么关系？

2.1　心身二元论

47　中世纪以来，首先对心身问题作出回答的是笛卡尔，他认为：心灵同身体和任何一种物理的东西都完全不同。根据**笛卡尔二元论**，心灵是纯精神性的，根本不占有空间，它既无广延也无居所。根据这种观点，一个有生命的正常人具有两重性，即一个心灵和一个与之相对应的身体（尽管身体能脱离心灵而存在，而心灵也能在其对应的身体毁灭以后存在）。令人不可思议的是，尽管心灵与身体截然不同，但它们相互间却有因果作用：身体的行为在人的心灵中产生感觉、经验和思想；反过来，心理的活动也会导致行动和言语，造成肢体或嘴唇的物理运动。

笛卡尔二元论有强烈的直觉魅力，因为从内部看，我们的心灵在感觉上根本不是物理的东西，我们能够很容易地想象它们能离体（disembodied）存在，甚至能想象它们在没有任何物理世界时存在。事实上，20世纪50年代以前，心灵哲学一直为笛卡尔的“第一人称”观点所主宰。所谓第一人称观点，指的是从内部观察我们自己的观点，除了少数例外，哲学家们都接受下列主张：（1）人对自己的心灵比对身体有更好的认识；（2）从形而上学上说，心灵是身体的驾驶员；（3）即使对这一问题有了令人信服的回答，但关于我们的理智如何能够知道“外部的”日常物质对象存在，至少仍是一个理论上的问题。我们人类主体被囚禁在心灵的剧场中，不过我们也有一些可靠的途径来推测剧场之外发生的事情。

到20世纪中叶，所有这一切都由于两个原因而突然改变了。第一个原因是逻辑实证主义和意义的证实理论日益增大的影响。主体间的可证实性或可检验性变成了科学诚实和语言意义本身两者的标准，特别是，如果心灵要么从科学上来考虑，要么甚至首先作为有意义的描述来考虑，那么，心理归属将不得不被限制在公开的、能从物理上检验的证实条件之中。科学对任何事物都采取主体间的、第三人称的观点，为了科学的目的，研究者不得不放弃传统的第一人称观点，而且他们可以感觉到，这也是为了严肃的形而上学的目的。

第二个原因是出现了很多反对笛卡尔二元论的哲学观点，这些观点都是咄咄逼人的，下面是其中的几种：

1. 非物质的笛卡尔心灵和幽灵般的非物理事件就像鬼魂和灵质（ectoplasm）一样令人不安，它们看上去越来越不符合我们其他方面的物理的和科学的世界图景。对于解释任何众所周知的可观察事实来说，它们都是不需要的，因为神经生理学有望对它们作出彻底的解释，尤其是在我们身体的运动方面。的确，幽灵般的心灵（ghost-minds）在这种解释中没有太大的帮助，因为从未听说过鬼的任何属性会影响公开的物理事件。

2. 从诸如没有心灵的单细胞有机体等原始生物开始，由于通过纯粹物理的变异过程和自然选择，人类已经过千百万年的进化。因此，那种认为除了细胞和物理器官之外，大自然还在某个时刻（以种群遗传的形式）以某种方式创造了非物质的笛卡尔式的心灵的观点是不合逻辑的。依据单个人的受精卵到胚胎再到胎儿、婴儿，最后到一个儿童的发展过程，也可以得出相同的观点。

3. 如果心灵真是非物质的、完全不占有空间，那么它们如何能同空间中的物质对象发生因果作用呢？（笛卡尔本人对此也甚感不安。在一个地方，他提出把重力当作非物质的东西作用于物理身体的模型，但是，重力本质上是占有空间的，虽然它不像身体那样是有形的。）

4. 总之，非物质的实体能引起物理的运动，这似乎不符合物理学的守恒定律，例如关于运动和物质—能量转换的守恒定律，物理能量将不得不消失又重现在人脑之内。

2.2 行为主义

什么会取代二元论？首先，卡尔纳普（Carnap，1932—1933）和赖尔
49 （Ryle，1949）提出：就心理归属来说，明显的证实条件或者测试是行为。在身体可能受到损伤的情况下，我们其他人除了根据你脸部扭曲和呻吟的行为之外，如何能说你感觉疼痛呢？或者除了根据你的言语表白和你躲避防风草的行为之外，又如何能说你相信防风草有害呢？如果检验方法是行为，那么（人们认为）这些归属的意义（或者至少是真正被描述的事实）并不是幽灵般的或难以言传的东西，而是行为。这就是作为一种心灵理论和心理学范式的行为主义。

在理论心理学中，行为主义主要采取一种方法论形式，心理学家们没有正式地提出形而上学的主张。但是在哲学中，行为主义确实（自然地）采用了一种形而上学形式：其中主要是**分析的行为主义**，即认为心理归属仅仅指对环境影响的行为反应。因此，“利奥感觉疼痛”不是指利奥假定的幽灵般的自我，甚至也不是指利奥体内发生的任何事件，而是指利奥实际表现出的脸部扭曲和呻吟的行为，或者指他倾向于作出这样的行为（因为若没有什么事情阻止他，他就会作出这样的行为）。“利奥相信防风草有害”不过是指：当有人问他时，他会同意那个命题；并且当遇到防风草时，他会避开，如此等等。

任何行为主义者都会赞成所谓的图灵测试。对于“机器能思维吗”这个持续多年的流行问题，阿兰·图灵（Alan Turing，1964）回答说，更好的问题应该是：精巧的计算机能否通过一组文字测试达到这样的程度——能骗得受限的观察者（比如说，通过邮件与之通信的一个人）相信它是有感情的人。如果一台机器真的通过了这些测试，那么，不管人们如何迷恋于形而上学分析，“机器是否真的能够思维”这个假定的进一步的问题就会变得毫无意义。除了图灵对机器的文字行为——与非文字的反应相对应——所作的倾向性限制之外，任何行为主义者，不论是心理学的还是哲学的，都同意心理学的差异不会超出行为的测试，实际行为和假设行为都相同的有机体（包括机器在内）在心理学方面也会完全一样。

除了解决主体间证实的方法论问题之外，哲学行为主义也巧妙地避免了针对笛卡尔二元论的很多反对意见，包括上述的四点。它摒弃了非物质

的笛卡尔式自我和幽灵般的非物质事件，并将它们视为形而上学的赘疣而彻底消除。它解决了笛卡尔承认的心身相互作用问题，因为它没有假定非物质的、不占有空间的行为原因。在对物理学或生物学中的笛卡尔实体进行调和方面，它没有提出科学的秘诀，因为它不支持这种调和。因此，它是与笛卡尔的非唯物论相对立的一种**唯物论**观点。

然而，有些理论家深感担忧，他们感觉行为主义在彻底否定内部的、私人的和主观的东西时，也遗漏了一些真实而重要的东西。当说出这种担忧时，行为主义者常常给予嘲讽的回答，将怀疑者们斥为相信幽灵、灵质或复活节兔子的旧式二元论者，行为主义是城中唯一的（即使不太合理）游戏。然而，怀疑者们提出了一些不容置疑的观点来反驳它。首先，诚实的和未被麻醉的人都完全知道：他们经验并能内省到内部真实的心理情景或事件，这些情景或事件既未实际地伴随有典型的行为，也不是关于他们在受到某种刺激时将如何行为这个唯一不变的假定事实。普赖斯（1956）谈到了有意识心理状态的一种“难以消解的残余物”，它与任何特定种类的行为都没有明显的关系，同时参见阿姆斯特朗（1968：第 5 章）和坎贝尔（Campbell，1984）。其次，与图灵测试相反，这似乎是完全可能的，即不管两个人在他们实际的和假定的行为总体上如何相似，他们在心理上是不同的，就像约翰·洛克所假定的“颠倒光谱”的情况那样：它或许是，在你看见一个红色的物体时，你会有这种我在看见一个绿色的物体时所具有的颜色经验，反之亦然。因此，一个生物或许会表现出所有合适的刺激—反应关系，却根本没有心理生活。我们可以设想打造一个“怪人”（zombie）[①] 或者愚笨的机器人，它能以正确的方式行事，但根本不能真正地感觉或思考什么（布洛克和福多，1972；柯克，1974；布洛克，1981；坎贝尔，1984）。第三，分析的行为主义者对心理归属的行为分析，似乎只有在一个人对主体心性的其他部分作出实质性的假设时，才是合适的（齐泽姆［Chisholm］，1957：第 11 章；吉奇［Geach］，1957：8；布洛克，1981）。例如，如果利奥相信防风草有害并且有人给他防风草，那么只要他不想死，他就会避开它们。所以，就行为主义者应当详尽无遗地研究一

① 一般认为，心灵哲学是从海地伏都教的传说借用了“zombie”一词，它指的是通过巫术起死回生的僵尸。这些僵尸在吃了盐后会恢复知觉，如果没有吃盐就不会有任何自我意识。在心灵哲学中，“zombie”有特别的含义，指的是这样的人：他们在外部行为表现乃至内部物理生理结构、过程和活动等方面与正常人完全一样，但却没有任何心理活动，没有心灵。目前，学术界对它有多种译名，如“僵尸”、“怪人”、“无心人”等，这里从众译为“怪人”。——译者注

般的心理之物而言，他们的分析要么是循环的，要么是根本不完整的。

于是，这些问题在二元论者、行为主义者和怀疑者那里陷入了僵局，一直到 20 世纪 50 年代后期，U. T. 普赖斯（1956）和 J. J. C. 斯马特（1959）才提出了一条中间道路、一种调和的折中方案。

2.3 同一论

与行为主义者相反，据普赖斯和斯马特所说，至少有些心理状态和事件归根结底是真正的内部情景。它们并不与外部行为相同一，甚至也不与假定的行为倾向相同一。但是，同二元论者相反，这些情景性的心理项目（items）既不是幽灵般的，也不是非物理的，毋宁说它们是神经生理的。它们都同一于发生在其所有者中枢神经系统中的状态和事件，更准确地说，每个心理状态或事件都在数量上同一于某个这样的神经生理状态或事件。例如，要感觉疼痛，就是要使一个人的 c-纤维或者（可能是）a-纤维在中枢神经系统中激活；要相信花椰菜会毒死你，就是要使一个人的 B_{bk}-纤维激活，如此等等。

通过使心理的东西完全变成物理的东西，这种关于心灵的**同一论**就具有了行为主义的长处而避免针对二元论的异议，但它也明确地容纳了内部
51 的和情景性的事件，而行为主义并不这样做。因为根据同一论，心理的状态和事件实际上发生于其所有者的中枢神经系统之中。（因此，在比笛卡尔所承认的更真实的意义上，它们是内部的。）同一论也彻底地证明了这一观念，即：纵使有机体的所有外部行为都相似，它们在心理方面也能够不同，很显然，因为有机体在传达它们外部的刺激—反应规律时，在神经生理方面是能够不同的，这就为颠倒光谱创造了可能性。当然，一个信念或愿望与通常相伴随的行为之间的关系，是可以被当下的其他心理状态取代的，因为一个 B-神经状态或 D-神经状态与其正常的行为结果之间的关系，是能够被其他从心理上描述的相互作用的神经状态取代的。同一论是摆脱二元论者—行为主义者僵局的理想办法。

此外，对同一论还有一种直接的演绎论证，它是由大卫·刘易斯（1966、1972）和阿姆斯特朗（1968）各自独立提出来的。刘易斯和阿姆斯特朗主张：心理术语是从因果上定义的，根据的是心理项目典型的原因和结果，例如，“疼痛”这个词是指通常由身体损伤所引起的状态，它通

常也导致后退、讨饶、诉苦、欲求停止等等。（阿姆斯特朗自称是通过简单的“概念分析”来建立它的。刘易斯更详细地指出，心理术语是一种常识的“民间理论”的术语，他还同实证主义者一样，认为所有理论术语都是由它们出现于其中的理论预先定义的，那种民间理论此后就被称为了“民间心理学”。）根据定义，如果疼痛是任何占据某个因果位置的状态，而且就像极有可能的那样，如果科学研究会表明那个特定的位置实际上是由某种神经生理状态占据的，那么就可以直接推出：疼痛就是该神经生理状态，这就是所要证明的。疼痛保持了它与行为的概念联系，但与其所有者的内部状态也有一种经验的同一关系。（对这种混合观点的一种超前而复杂的阐述，参见刘易斯，1980。有关的详细讨论，参见布洛克，1978；休梅克，1981；泰，1983；欧文斯，1986。）

请注意，尽管阿姆斯特朗和刘易斯的论证都起始于一种关于心理术语意义的主张，但他们“常识因果关系”版本的同一论本身并不就是这种主张，就像普赖斯和斯马特最初的同一论也存在区别。毋宁说，这四位哲学家都依赖于这一观念：即使在这些术语没有同义词的情况下，事物或属性有时也能同一于“别的”事物或属性，确实存在综合的和后验的同一性这种东西，但它却是真正的同一性。虽然三角形与三边形的同一性仅靠这两个术语的意义来维持，而且只根据理性而无需经验研究就能予以确定，但下述的同一性却是综合的后验同一性的典型例证，它们是能经验地发现的：云与大量小水滴的同一性，水与 H_2O 的同一性，闪电与电的释放的同一性，晨星与金星的同一性，孟德尔基因与 DNA 分子链的同一性，温度与平均分子动能的同一性。同一论同样是本着科学思考的精神提出来的，一个人不可能恰当地提出这样的反驳：心理表达式不是指任何关于大脑或神经激活的东西。

52

因此，二元论者的错误在于认为心理项目是非物理的，但其正确之处在于认为心理项目是内部的情景；而行为主义者的正确之处在于其唯物主义，但其错误在于否定了内部的心理情景。将两者综合起来该会多令人高兴。但是，它是如此完美，以至于很难让人相信它会变成现实。

2.4 机器功能主义

不久以后，希拉里·普特南（Hilary Putnam，1960、1967a、1967b）

和杰里·福多（1968b）就指出了一种独断的含义，即同一论被理解成一种关于心理事项“类型”或种类的理论：心理状态如疼痛在任何时候和任何地方都具有最初指派给它的神经生理描述。例如，如果同一论者将疼痛本身同一于 c-纤维的激活，那么就可以推出：任何一个物种的生物（地球上的或科学虚幻的）都仅当它有 c-纤维并且 c-纤维在激活时，才能够感觉疼痛。但是对任何能感觉疼痛的存在物的生物学所作的这种限制，既不必要也站不住脚。为什么我们应当认为，要想具有能被准确辨认的疼痛，任何有机体都必须用和我们相同的化学材料来制造呢？同一论者对行为主义者的困难作出了过度的反应，而且过于狭隘地集中在人类实际的内部状态的生物学细节，而在这样做时，他们已经陷入了物种沙文主义。

普特南和福多提出了明显的修正：重点并非它是激活的 c-纤维（本身），而是 c-纤维激活正在做的事情，是它们对整个有机体运转的贡献。c-纤维的作用能由任何合适的机械元件执行，只要那种作用被执行了，包含有机体在内的心理学就不会受到影响。因此，感觉疼痛本身不是具有正在激活的 c-纤维，而只是处于这一种或那一种状态，不管这种状态能做怎样的生物化学描述，它在我们所研究的人身上所起的因果作用都同 c-纤维激活所起的相同。我们可能会继续主张，疼痛“个例”（tokens）（特定时刻发生在特定主体身上的个别疼痛实例）严格地同一于那些主体在那些特定时刻的特定神经生理状态——换句话说，同一于碰巧起着合适作用的状态。这就是**个例同一论**，或者“个例”唯物主义或物理主义论题。但是疼痛自身（这个种类、共相或“类型”）只能同一于更抽象的东西：即 c-纤维激活与它们潜在的代替者或代理者所共有的因果或功能作用。心理状态类型不是同一于神经生理类型，而是同一于更抽象的功能作用，它是通过状态个例与有机体的感觉输入、行为反应和其他中间心理状态的因果关系加
53 以说明的。因此，**功能主义**就是这样的学说：使一个心理状态成为它所是的这类状态（疼痛、紫罗兰的气味、关于树袋熊的恶意看法）的，是它独特的功能关系，是它在其主体行为系统中的作用。

普特南将心理状态比作计算机的功能或“逻辑”状态：正如计算机程序能由很多在物理方面不相同的硬件构型中的任何一个来实现或例示一样，心理学“程序”也可以由具有各种生理化学成分的不同有机体实现，这就是不同物种有机体的不同生理状态能实现同一种心理状态类型的原因。每当同一论者的类型同一关系采取这种形式——“处于 M 类型的心理状态之中，就是处于 N 类型的神经生理状态之中”时，普特南的**机器功能**

主义——如我对它的称呼——就宣称：处于 M 中，不过是处于这个或那个在相关的计算机程序（也就是说，在合适的抽象层次上，如果给定总体的输入，这种程序就能居间产生此种生物的总体输出，从而起到这种生物的整体心理学的作用）中起作用 R 的生理状态。这个生理状态之所以能“起着作用 R”，是因为它与物理的输入、输出和别的内部状态具有一组关系，这组关系与计算机程序中所编写的抽象的输入—输出的逻辑状态关系是一一对应的。

这样一来，功能主义者使用了三种不同的描述层次，并将它们运用于同一种基本实在。在一个特定时间对某人大脑中的一个物理状态个例有一种神经生理学描述，然而相对于大脑碰巧在实现的机器程序，它还可能有一种功能描述，如果某个心理状态同它所例示的功能范畴恰好是类型同一的，那么它会进一步产生一种心理描述。因此，毕竟在某种意义上，“心理的东西”与“物理的东西”是不同的。尽管很可能不存在非物理的实体或材料，并且每个心理个例本身都完全是物理的，但心理描述并不是物理描述，感觉疼痛的属性不只是某个神经正在激活的属性。此外，与行为主义和同一论不同，功能主义并不严格要求心灵是物理的，只要那些心灵实现了相关的程序，非物理的心灵也可能为真。

2.5 小人功能主义和其他目的论理论

机器功能主义在很多方面都受到了挑战，这一切共同促成了一种特殊的目的论“功能”概念：我们把一物的功能看成是该物是为了什么（what the thing it for），它的责任是什么（what its job is），它应当做什么（what it is supposed to do）。这样“把功能放回到功能主义中”有如下三个理由（索伯［Sober］，1985）。

首先，机器功能主义者仍然用逻辑实证主义的术语来想象心理解释，将观察到的材料纳于越来越广泛的普遍规律之下。但是福多（1968a）、丹尼特（1978）和卡明斯（Cummins，1983）都支持一种相竞争的心理解释 54
图景，根据这一图景，行为材料被看成是主体心理能力的表现，而对那些能力的解释，则是通过把主体理解成由相互联系的成分所构成的系统。每个成分都是一个“小人”（homunculus），因为它被看成是在包含它的主体内运转的一个小小的自主体或官员（bureaucrat），对它的确认是参照它所执行的功能。各种小人成分相互合作，从而能对刺激作出整体的行为反

应。“小人”本身又被分解成子成分，这些子成分的功能和相互作用同样被用于解释它们所构成的子系统的能力，这样一再分解，一直到子成分的子成分的……子成分被看成神经生理结构。因此，生物系统和机械系统同样具有等级组织。（一辆汽车是通过油箱、输油管、化油器、内燃机、点火系统、传动装置以及能转动的车轮来工作和行驶的。如果有人想了解化油器是怎样工作的，他将被告知化油器有哪些部件，它们是怎样一起将氧气注入汽油的，如此等等。）但是，在这种解释模式中，没有什么是与将材料纳于越来越广泛的普遍概括之下相对应的。

第二个理由是：机器功能主义认为功能的“实现”（realization）、个别物理有机体与（据说它所例示的）抽象程序之间的关系，不过是下列两者之间的一一对应问题。一方面是有机体的全部物理刺激、结构状态和行为，另一方面是程序所定义的输入—状态—输出的功能。但是这个实现标准被认为过于自由了，由于事实上任何东西都能与别的事物有某种一一对应关系，因此，这种一一对应意义上的纯粹“实现”太容易得到了（布洛克，1978；莱肯，1987：第3章）。任何中等大小的物体都有一些组成分子的运动，这些运动碰巧与某个机器程序一一对应。有的理论家提出，通过给实现施加目的论限制来弥补这一缺陷：仅当一个有机体有真正完整的器官，并且它的一个物理状态只是为了有机体才发挥它的功能作用时，这个物理状态才算实现了如此这般的一个功能描述。这里的“为了”（for）是在目的论的意义上说的，“功能”也是在目的论的意义上说的。这个状态必须做着它所做的事情，可以说，这是关于它的生物学目的的一个问题。（机器功能主义是在其备用的数学意义上而非真正的功能意义上使用“功能”的。应当注意，这里所使用的“机器功能主义”这一术语是与最初自由的“实现”概念相联系的，因此施加目的论的限制就是放弃机器功能主义。）

第三，范久里克（Van Gulick，1980）、米利肯（Millikan，1984）、德雷斯基（Dretske，1988）、福多（1990a）等都强有力地主张：目的论对诸如信念和愿望等心理状态的意向性、关于性或者指涉特征所作的任何适当的分析，都必须参照这些状态的心理生物功能。信念、愿望和别的命题态
55 度，如猜测、意欲和希望等，都指向可能或不可能实际得到的事态（例如，共和党候选人将赢得大选），它们都涉及可能或不可能存在的个体（如亚瑟王或福尔摩斯）。弗朗兹·布伦塔诺（Franz Brentano，1973[1874]）在指涉物体或事态——甚至是不存在的物体和事态——的心理现

象和没有这样指涉的物体之间作出了区别。然而，如果心理项目是物理的，就会产生这样的问题：纯物理的实体或状态何以能具有“指涉”或关于非存在的事态或物体的属性。这不是普通的纯粹物体（如砖头）能够具有的属性。根据目的论者，一个神经生理状态应算作一个关于花椰菜会毒死你的信念，尤其是关于花椰菜的信念，仅当该状态在某种意义上将表征花椰菜作为它的一种心理生物功能。如果目的论是解释意向性所必需的，而机器功能主义并不导致目的论，那么机器功能主义对于解释意向性就是不合适的。

所有这些关于目的论和生物学功能的谈论，似乎都预设了物理系统生物的和其他的“结构”状态真具有目的论意义上的功能，至少可以说后一种主张是有争议的。但是，对于目的论功能主义者来说，幸运的是现在有一项蓬勃兴起的事业，它是用自然主义的术语特别是根据原因论来解释生物学目的论。例如，人们可能说，一个特征具有做 F 的功能，这是由于它因过去能做 F 而被选择了，心脏的功能是泵血，因为心脏过去泵血给了它们一种选择优势，并因此使更多具有心脏的动物生存了下来（赖特[Wright]，1973；米利肯，1984）。

之前困扰着行为主义和同一论的难题，功能主义也继承了一些。剩下来的这些障碍主要有两类：感受性质问题和意向性问题。

2.6 感受性质和意识方面的问题

心理状态或事件（尤其是感觉）的感受性质，就是对于该状态或者事件的感觉，即它可内省的“现象特征”，或它自身显现于意识时所表现出来的本质。很多哲学家都反驳道：不管是功能主义的形而上学还是前面提到的任何一种相关学说，都不能“解释意识”或者说明乃至容忍这个概念——处于如此这般的一种心理状态的感觉像什么样子。然而，这些哲学家又说：这些感觉是典型的心理的东西——正是这些感受使得心理状态成为它们所是的心理状态。因此，功能主义一定有根本的错误。

关于意识或感受性质的“这个”问题是大家所熟悉的。当然，它如此熟悉，以至于我们往往忽视了它所包含的最重要的东西，即它有很多名称，因为它具有多样性。感受性质的问题不是独立的问题，对于功能主义，至少已经提出了 11 种截然不同的反对意见（其中有些也普遍地适用于唯物主义观点）。这里只谈其中的一部分： 56

1. 布洛克（1978）和其他论者提出了各种反对功能主义的“怪人”式反例——在这些反例中，有的实体似乎实现了适当的程序，却又缺乏心性的一个关键的质的方面。（一种有代表性的“实体”是一群人，例如按照精心设计的一组指令行动的全体中国人。这群个体看起来不会有什么共同的感觉。）可以想见，功能主义者已经通过这样的争辩作出了答复，即对每个反例来说，要么所提到的群体性实体（group entity）实际上并没有成功实现适当的程序（例如，由于缺乏必要的目的论），要么没有正当的理由否认这种实体确实具有相关的质的状态。

2. 内格尔（1974）和杰克逊（1982）把知识方面的不对称性作为一个一般的反唯物主义论证：我能够知道具有如此这般的感觉是什么样子，仅当我自己有那种感觉。对此，再多客观的、第三人称的科学信念都是不够的。功能主义者在答复时对“视角性”（perspectivalness）提出了各种分析，还对具有一种感觉“是什么样子”作了解释，这些都使得那些东西与功能主义相一致。内格尔和杰克逊进一步论证了存在一种特殊的、内在的视角性事实，即关于“是什么样子”的事实，这种事实是棘手的，而且原则上是物质科学难以描述或解释的。功能主义者回应道：这些论证犯了一个逻辑错误（具体来说，是把莱布尼兹定律用于内涵语境的错误）。有人补充说：无论如何，“知道是什么样子”不过是具有一种能力，它并不涉及任何一种事实。然而与此相反，另一些理论家已经承认存在关于“是什么样子”的事实，但他们又坚持认为，这些事实终究是能够被自然科学解释和预测的。

3. 索尔·克里普克（Saul Kripke，1972）巧妙地利用模态的区别来反对类型的乃至个例的同一性，他认为：除非心理项目必然地同一于神经生理项目——实际上它们并非如此，否则它们就根本不能与神经生理项目相同一。克里普克结尾的推理引起了很多批判性的关注，而且人们甚至还提出了更完善的变化形式，例如，杰克逊（1993）和查默斯（1996）所提出的例子。

4. 杰克逊（1977）等人支持这种主张：在意识中呈现给我们的一些心理个体（mental individuals）本身就具有现象的、质的属性。例如，当一个红色的闪光灯在你面前闪光时，你的视野中会出现一个绿点，一个“余像”，这种东西确实是绿色的，它有相当确定的形状并且在消失之前会存在几秒钟。如果存在这种东西，它们与（健康

的）人类主体的大脑中所发现的任何物理的东西都完全不同。虽然相信真正的绿色余像之类“现象的个体”的传言在哲学家中间已经不得人心了，但它仍有强有力的动机（参见莱肯，1987：83-93）。

这是四种很难对付的反对意见，而且表面上看，每种似乎都是可信的。唯物主义者尤其是功能主义者一定会作出认真的回应，很显然，唯物主义者已经作了充分的回应——莱肯的两本书（1987、1996）对某些最重要的驳难作了阐述。然而，近年来常可看到对占主导地位的唯物主义的反驳，包括一些新二元论观点的复兴，就像在鲁宾逊（[Robinson]，1988）、哈特（Hart，1988）、斯特劳森（1994）和查默斯（1996）的书中看到的。

2.7 意向性方面的问题

这个来自于我们前面所提到的布伦塔诺的问题，是要解释纯物理的实体或状态何以能具有关于或“指涉”非存在的事态属性。标准的功能主义回答是：命题态度具有布伦塔诺所说的特征，是因为实现它们的内部物理状态和事件表征了实际的或可能的事态。它们所表征的东西（它们的内容）至少部分是由它们的功能作用决定的。

这里有两个主要的难题。其一是准确地说明物理项目的表征内容是怎样被决定的，一个神经生理状态是借助什么来准确地表征共和党候选人将赢得大选？对这个一般问题的一种回答是福多所称的**心理语义学**（psychosemantics）。人们已经作出了一些尝试（德雷斯基，1981；米利肯，1984；福多，1987、1990a、1990b、1994），但是没有一种尝试是很可信的。特别是，所有尝试都只适合于包含有实际的、当下存在的物理对象的内容。像数这样的抽象实体、未来的实体（如我希望有一天会有的孩子）以及布伦塔诺的非存在项目都被彻底遗漏了。

第二个难题是：普通命题态度的内容并不随附于它们主体神经系统的状态，但即便该主体头脑中的全部状态也不足以决定它们。似乎让人吃惊的是，普特南（1975）的孪生地球和索引性例子表明：两个人能够每个分子都相似，但他们的信念和愿望仍然不同，这依赖于他们空间和历史环境中的各种因素。因此，我们可以将“窄”属性和“宽”属性加以区别，前者是由主体的内在物理构成决定的属性，后者则不是这样。显而易见，表

征内容是宽的，而功能作用是窄的。那么，命题态度与功能作用，或者就此而言与任何窄描述下的大脑状态怎么能是类型同一的呢？

功能主义者用两种方法回答了第二个难题。第一种方法是对“功能”
58 也作宽的理解，即从历史上和/或参照主体实际环境的特征来说明功能作用。第二种方法只是不再把功能主义作为一种关于内容的特殊说明，它为命题态度提供了某种替代性的心理语义学，但保留了关于态度类型的功能主义。（于是，使一状态成为愿望 P 的是它的功能作用，即使别的东西也能使这个状态成为愿望 P。）

2.8 情　绪

在提到感觉状态和具有意向性内容的心理状态时，我们没有对情绪作详细的说明。由于反对行为主义，心灵理论往往都不会被直接运用于情绪；相反，情绪一般被认为能从概念上分析成更重要或更“核心”的心理状态复合体（complexes），通常是像信念和愿望这样的命题态度（而且由情绪的意向性要追溯到态度的意向性）。阿姆斯特朗（1968：第 8 章，第 3 节）实质上走的就是这条路线，所罗门（Solomon，1977）和戈登（Gordon，1987）同样如此。然而，也有功能主义和情绪方面的文献，可参见雷伊（Rey，1980）和罗蒂（1980）所编文献中收录的其他一些论文。格里菲思（1997）大体上持一种功能主义观点，但他认为“情绪”并不构成一个独立的种类。

2.9 工具主义

同一论者和功能主义者（机器的或目的论的）在将心理状态和事件理解为既内在于人类主体又同时是原因时，便与常识（和当前的认知心理学）合流了。尤其是信念和愿望被认为是由知觉或其他认知事件引起的，它们又接着联手从内部促发了行为。如果阿姆斯特朗或者刘易斯的心灵理论是正确的，那么这种思想就不仅是常识而且是一个概念真理；如果功能主义正确，那么它至少是一个形而上学的事实。

如我们在 2.3 节所看到的，由于都支持内部原因解释，同一论者和功

能主义者就与行为主义者决裂了，因为行为主义者并不把心理项目看成实体，看成内在的东西，或者看成是在任何比单纯的假设更强的意义上的原因。行为主义者要么完全摒弃心理主义用语，要么依据公认的对假设刺激的反应来翻译心理归属。最近，有些哲学家还仿效他们，拒绝接受信念和愿望是内部原因的观念，他们还对它们作出更具操作性和工具性的解释。丹尼特（1978、1987）尤其想否定信念和愿望是人体内部有因果作用的状态，他主张：信念归属和愿望归属仅仅是演算的装置，它们对于他接下来要解释的理由碰巧有预言的用处。他承认，这些归属在客观上往往是正确的，但这不是因为描述了内部的机制。

因此，对于信念和愿望之类的命题态度，丹尼特是一个**工具主义者**。（依据当代的阐释，关于X的“工具主义者”就是这样的理论家，他主张：尽管关于“X”的句子常常是正确的，但它们并不真实地描述一种特殊实体，而只是用于使更熟悉的现象系统化。例如，对于“一般的美国房屋业主”，即白人、男性、正好是2.2个小孩的父亲来说，我们都是工具主义者。）丹尼特说，归属一个“信念”或“愿望”，不是描述物理实在的某个片断，而更像是拨动算盘上的算珠。（应当注意，丹尼特最近转变得更温和了：参见氏著，1991。）

丹尼特提供了四条基本理由以反对常识的内部原因论：

1. 他认为，任何科学总会发现某种独特的内部因果机制，这种机制是所有具有一个特殊信念的可能主体共同具有的，但这是极不可能的。

2. 他将关于人类的信念—愿望解释与关于低等动物、下棋计算机甚至避雷针的信念—愿望解释进行了比较，他认为：(a) 就它们来说，我们没有理由认为信念归属和愿望归属不是纯粹的演算预测装置；(b) 就人类来说，我们同样没有理由认为信念归属和愿望归属不是纯粹的演算预测装置。

3. 丹尼特是从信念归属和愿望归属的证实条件出发进行论证的——这基本上就是从主体在其环境中应当相信和需求的东西出发进行合理推断的问题——他接着大胆地将那些归属的成真条件（truth-makers）完全同一于它们的证实条件，他要求内部原因论者证明：为什么工具主义没有提供所有真实的证据。

4. 他认为：无论如何，如果纯粹的规范性假设（“合理性假设”，

即人们一般会相信他们应当相信的东西，愿望他们应当愿望的东西）是准许一种归属所必需的，那么这种归属本身就不能是对一个简单事态的纯事实描述。

斯蒂克（1981）对丹尼特的工具主义进行了详细的考察和批判（但奇怪的是，随后斯蒂克［1983］又对一种几乎和丹尼特一样会受到抨击的观点进行了辩护，尽管这种观点显然与丹尼特的不同）。丹尼特对斯蒂克作出了反驳，对“合理性假设”作了更清晰的说明。（针对丹尼特的其他批评，是1988年由丹尼特命名的专题论文集《行为与脑科学》的作者们提出来的。）

丹尼特观点的一个近亲是唐纳德·戴维森（1970）的**异常一元论**，因为它也关注合理性假设。与丹尼特的工具主义不同，它赞成个例物理主义，它强调：个别的心理个例是原因，但它根据类似的认识论理由反对这种可能性，即对命题态度能作出某种有意义的唯物主义类型还原。

2.10 取消主义与神经哲学

在否定信念和愿望等命题态度是人们真实的内部原因状态方面，丹尼特的工具主义同常识和哲学传统都决裂了。但是，丹尼特承认——实际上，他急切地强调：信念归属和愿望归属是正确的，而且在客观上仍然是正确的。有的哲学家采取了一种不太调和的、更激进的非常识观点：心理归属归根结底是不正确的，而且是完全错误的。在认为人们能相信和愿望什么方面，而且可能在认为人们有感觉和感受方面，常识是彻底错误的，这种虚无主义主张看起来可能令人不安。

按照标准用语，让我们将这种虚无主义主张称为“取消式唯物主义”（eliminative materialism），或者简称为“取消主义”（eliminativism）。不过重要的是要看到：取消主义者和个例物理主义者之间存在着即使是意料之外但却是普遍的联盟：取消主义者、同一论者和功能主义者都同意，心理项目（如果有的话）是人们真实的内部原因状态。他们的分歧仅仅在于这个经验问题，即人们的任何真实的神经生理状态与“民间心理学”的常识心理范畴是否实际相符。取消主义者赞赏同一论者和功能主义者，是因为他们都明确地愿意接受并运用经验的方法。不仅取消主义者，个例物理

主义者也很鄙视工具主义者的拙劣借口。（但是，取消主义者与工具主义者一样，认为功能主义者是一场白日梦，而功能主义者同意工具主义者，认为心理归属往往正确并且明显是正确的。这三种观点构成了一种并不特别的三角关系。）

保罗·费耶阿本德（Paul Feyerabend，1963a、1963b）的第一个公开主张是：民间心理学的心理范畴绝对捕捉不到物理实在中的任何东西，因此日常的心理归属是错误的。（罗蒂［1965］也采取了声名狼藉的取消主义路线，但他仿效塞拉斯［1963］试图弱化他的虚无主义；莱肯和帕帕斯［Lycan and Pappas，1972］认为，这种弱化只能使罗蒂的立场更加支离破碎。）费耶阿本德没有吸引太多的追随者，这很可能是由于他的观点粗暴地藐视了常识。但取消主义被保罗·丘奇兰德（Paul Churchland，1981）等人复活了，后者对它作了更详细的辩护。

丘奇兰德主要是从民间心理学的贫乏出发进行论证的；他主张：从历史上看，根据科学的理由，当别的原始理论（如炼金术）的表现像民间心理学一样差时，它们就会被抛弃，而且这样做是正确的。P. S. 丘奇兰德以及丘奇兰德和塞耶诺斯基（Sejnowski，1990）都强调神经生物机制相对的科学实在性和因果效力：鉴于神经生理学解释在科学上的杰出成就与民间心理学相对的模糊不清和无法还原形成的巨大反差，为什么我们应当认为——即使是一瞬间，更不用说自动地——民间心理学的老生常谈表达了真理呢？

反对取消主义的理由一目了然。首先，我们认为我们知道存在命题态度，因为我们通过内省就能知道它们。其次，态度不仅对于预测、推理、深思和理解是必不可少的，而且对于获得重要的宏观概括也是必不可少的。如果不提到它们，我们往往不能条理清晰地进行交谈。但是，P. M. 丘奇兰德以及 P. S. 丘奇兰德和塞耶诺斯基的论证又该如何评价呢？

有人可能不同意这种主张，即民间心理学是一种失败的或拙劣的理论；基切尔（Kitcher，1984）以及霍根和伍德沃德（Horgan and Woodward，1985）采取的就是这种态度。还有人不同意更基本的主张，即民间心理学是一种理论。赖尔（1949）和维特根斯坦（1953）在它被清晰地阐述之前就坚定地反对前面的主张。最近的批评者还包括莫顿（Morton，1980）、马尔科姆（Malcolm，1984）、贝克（Baker，1988）、麦克唐纳（McDonough，1991）和威尔克斯（Wilkes，1993）。

参考文献

Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Baker, L. R. (1988). *Saving Belief*. Princeton, NJ: Princeton University Press.

Block, N. J. (1978). "Troubles with Functionalism." In W. Savage (ed.), *Minnesota Studies in the Philosophy of Science, Vol. IX: Perception and Cognition*. Minneapolis: University of Minnesota Press: 261 - 325. Excerpts reprinted in Lycan (1990, 1999).

—— (ed.) (1980). *Readings in Philosophy of Psychology*, 2 vols. Cambridge, MA: Harvard University Press.

—— (1981). "Psychologism and Behaviorism." *Philosophical Review*, 90: 5-43.

Block, N. J. and Fodor, J. A. (1972). "What Psychological States Are Not." *Philosophical Review*, 81: 159-181. Reprinted in Block (1980).

Brentano, F. (1973 [1874]). *Philosophy from an Empirical Standpoint*. London: Routledge and Kegan Paul.

Campbell, K. (1984). *Body and Mind* (2nd edn). Notre Dame, IN: University of Notre Dame Press.

Carnap, R. (1932-1933). "Psychology in Physical Language." *Erkenntnis*, 3: 107-142. Excerpt reprinted in Lycan (1990).

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Chisholm, R. M. (1957). *Perceiving*. Ithaca, NY: Cornell University Press.

Churchland, P. M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78: 67-90. Reprinted in Lycan (1990, 1999).

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: Bradford Books/MIT Press.

Churchland, P. S. and Sejnowski, T. (1990). "Neural Representation and Neural Computation." In Lycan (1990): 224-252. Reprinted in Lycan (1999).

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press/Bradford Books.

62 Davidson, D. (1970). "Mental Events." In L. Foster and J. W. Swanson (eds.), *Experience and Theory*. Amherst, MA: University of Massachusetts Press: 79-101. Reprinted in Block (1980) and in Lycan (1999).

Dennett, D. C. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.

—— (1981). "Making Sense of Ourselves." *Philosophical Topics*, 12: 63-81. Re-

printed in Lycan (1990).

—— (1987). *The Intentional Stance*. Cambridge, MA: Bradford Books/MIT Press.

—— (1988). "Précis of *The Intentional Stance*." *Behavioral and Brain Sciences*, 11: 495–505.

—— (1991). "Real Patterns." *Journal of Philosophy*, 88: 27–51.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books/MIT Press.

—— (1988). *Explaining Behavior*. Cambridge, MA: Bradford Books/MIT Press.

Feyerabend, P. (1963a). "Materialism and the Mind-Body Problem." *Review of Metaphysics*, 17: 49–66.

—— (1963b). "Mental Events and the Brain." *Journal of Philosophy*, 60: 295–296.

Fodor, J. A. (1968a). "The Appeal to Tacit Knowledge in Psychological Explanation." *Journal of Philosophy*, 65: 627–640.

—— (1968b). *Psychological Explanation*. New York, NY: Random House.

—— (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

—— (1990a). "Psychosemantics." In Lycan (1990): 312–337.

—— (1990b). *A Theory of Content*. Cambridge, MA: Bradford Books/MIT Press.

—— (1994). *The Elm and the Expert*. Cambridge, MA: Bradford Books/MIT Press.

Geach, P. (1957). *Mental Acts*. London: Routledge and Kegan Paul.

Gordon, R. M. (1987). *The Structure of Emotions*. Cambridge: Cambridge University Press.

Griffiths, P. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.

Hart, W. D. (1988). *Engines of the Soul*. Cambridge: Cambridge University Press.

Horgan, T. and Woodward, J. (1985). "Folk Psychology is Here to Stay." *Philosophical Review*, 94: 197–226. Reprinted in Lycan (1990, 1999).

Jackson, F. (1977). *Perception*. Cambridge: Cambridge University Press.

—— (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–136. Reprinted in Lycan (1990, 1999).

—— (1993). "Armchair Metaphysics." In J. O'Leary-Hawthorne and M. Michael (eds.), *Philosophy in Mind*. Dordrecht: Kluwer Academic Publishing.

Kirk, R. (1974). "Zombies vs. Materialists." *Aristotelian Society Supplementary Volume*, 48: 135–152.

Kitcher, P. (1984). "In Defense of Intentional Psychology." *Journal of Philosophy*, 81: 89–106.

Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1966). "An Argument for the Identity Theory." *Journal of Philosophy*, 63: 17-25.

—— (1972). "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, 50: 249-258. Reprinted in Block (1980).

—— (1980). "Mad Pain and Martian Pain." In Block (1980).

Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press/Bradford Books.

—— (ed.) (1990). *Mind and Cognition: A Reader*. Oxford: Blackwell.

63 —— (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press/Bradford Books.

—— (ed.) (1999). *Mind and Cognition: An Anthology*. Oxford: Blackwell.

Lycan, W. and Pappas, G. (1972). "What is Eliminative Materialism?" *Australasian Journal of Philosophy*, 50: 149-159.

Malcolm, N. (1984). "Consciousness and Causality." In D. Armstrong and N. Malcolm, *Consciousness and Causality: A Debate on the Nature of Mind*. Oxford: Blackwell.

McDonough, R. (1991). "A Culturalist Account of Folk Psychology." In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press: 263-288.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: Bradford Books/MIT Press.

Morton, A. (1980). *Frames of Mind*. Oxford: Oxford University Press.

Nagel, T. (1974). "What Is It Like to be a Bat?" *Philosophical Review*, 83: 435-450. Reprinted in Block (1980).

Owens, J. (1986). "The Failure of Lewis' Functionalism." *Philosophical Quarterly*, 36: 159-173.

Place, U. T. (1956). "Is Consciousness a Brain Process?" *British Journal of Psychology*, 47: 44-50. Reprinted in Lycan (1990, 1999).

Putnam, H. (1960). "Minds and Machines." In S. Hook (ed.), *Dimensions of Mind*. New York: Collier Books: 136-164.

—— (1967a). "The Mental Life of Some Machines." In H.-N. Castanda (ed.), *Intentionality, Minds, and Perception*. Detroit, MI: Wayne State University Press: 177-200.

—— (1967b). "Psychological Predicates." In W. H. Capitan and D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh, PA: University of Pittsburgh Press: 37-48. Reprinted in Block (1980) under the title "The Nature of Mental States."

—— (1975). "The Meaning of 'Meaning'." In *Philosophical Papers*. Cambridge: Cambridge University Press.

Rey, G. (1980). "Functionalism and the Emotions." In Rorty (1980): 163-195.

Robinson, W. S. (1988). *Brains and People*. Philadelphia, PA: Temple University Press.

Rorty, A. O. (ed.) (1980). *Explaining Emotions*. Berkeley and Los Angeles, CA: University of California Press.

Rorty, R. (1965). "Mind-Body Identity, Privacy, and Categories." *Review of Metaphysics*, 19: 24-54.

Ryle, G. (1949). *The Concept of Mind*. New York, NY: Barnes and Noble.

Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul.

Shoemaker, S. (1981). "Some Varieties of Functionalism." *Philosophical Topics*, 12: 93-119.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review*, 68: 141-156.

Sober, E. (1985). "Panglossian Functionalism and the Philosophy of Mind." *Synthese*, 64: 165-193. Revised excerpt reprinted in Lycan (1990, 1999) under the title "Putting the Function Back Into Functionalism."

Solomon, R. (1977). *The Passions*. New York, NY: Doubleday.

Stich, S. (1981). "Dennett on Intentional Systems." *Philosophical Topics*, 12: 39-62. Reprinted in Lycan (1990, 1999).

—— (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: Bradford Books/MIT Press.

Strawson, G. (1994). *Mental Reality*. Cambridge, MA: Bradford Books/MIT Press.

Turing, A. (1964). "Computing Machinery and Intelligence." In A. R. Anderson (ed.), *Minds and Machines*. Englewood Cliffs, NJ: Prentice-Hall: 4-30.

Tye, M. (1983). "Functionalism and Type Physicalism." *Philosophical Studies*, 44: 161-174.

Van Gulick, R. (1980). "Functionalism, Information, and Content." *Nature and System*, 2: 139-162.

Wilkes, K. (1993). "The Relationship Between Scientific and Common Sense Psychology." In S. Christensen and D. Turner (eds.), *Folk Psychology and the Philosophy of Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates: 144-187.

Wittgenstein, L. (1953). *Philosophical Investigations*, trans. G. E. M. Anscombe. New York, NY: Macmillan.

Wright, L. (1973). "Functions." *Philosophical Review*, 82: 139-168.

第3章　物理主义

安德鲁·梅尔尼克（Andrew Melnyk）

65 当今的大多数心灵哲学家都自称是物理主义者（或唯物主义者），不是这一派，就是那一派。然而，一般来说——即使不是一成不变的——他们都认为：他们关于心灵的物理主义是将非常一般的物理主义论题特别地运用于心理现象，意思是说：在“物理的”任何一种意义上，说心灵是物理的，**一切**都是物理的，这类表述都是正确的。这一章将要考虑的就是这种很一般的物理主义论题（此后称为物理主义）。有的人或许倾向于认为物理主义所导致的重大哲学困惑仅仅出现在心灵哲学之中；但正如我们将看到的，结果证明：即使一个人不考虑这一问题，即能否对过去就很棘手的心理现象（如意识、意向性和合理性）给出能为物理主义者接受的解释，物理主义也提出了很多东西需要考虑。在后文中，我将首先考察对物理主义作出合适表述时所产生的问题，接着考虑证明物理主义的尝试，最后讨论反对物理主义的看法的特征。尽管我希望公正地对待反对我的观点，但有必要提醒读者的是：我的讨论必然会反映基本的哲学承诺，而这些承诺是本领域的其他作者不赞成的。[1]

然而，为了了解背景，请允许我提出这样的哲学问题：物理主义常被合理地看作是对它的一种可能的解答（尽管还有别的解答）。随意地浏览一下大学的课程表我们也能发现，除物理学之外还有很多科学，如气象学、地质学、动物学、生物化学、神经生理学、心理学、社会学、生态学、分子生物学等，更不用说还有像民间心理学和民间物理学这样名义上的科学了。每门科学都有其自身特有的理论词汇，就它确实了解事物而言，它就是用这种词汇来描述特定领域的对象、事件和属性。但是多种科学的存在也引起了各种各样的问题：这多种科学相互之间是怎样联系的？

66 每门科学所特有的对象、事件和属性领域与其他科学特有的领域是怎样联系的？这多种科学是以某种方式描述同一事物的不同方面吗？或者说它们

自身是处理实在的不同片断吗？倘若如此，那么或许除了时空接近之外，这些实在的不同片断的存在是彼此无关吗？或者说有些片断在一些有趣的方面依赖于其他片断吗？这个关于多种科学的问题——如我们对它的称呼——显然是对心身问题的一种概括，至少就被理解成说明民间心理学与科学心理学之间的关系的问题来说，应是这样。[2]

物理主义对这个问题给出了一种回答。对它的**任何**回答，不管是不是物理主义的，都必须对多种科学之间以及它们的多个领域之间的关系作出系统的说明；因此，物理主义必须承担起这项雄心勃勃的计划，即勾画实在总体的图景，就像科学和常识向我们透露的那样。然而，对这个问题的物理主义回答与其他回答的区别在于这一事实，即它对多种科学之间以及其领域之间关系的说明有这样的结果：给物理学及其领域以**优越地位**，赋予物理学和物理之物某种描述和形而上学的优先性。我们很快就会看到这样做的不同方式。但是非物理主义的回答也是可能的。在当前的舆论氛围下，有两种回答尤其值得注意。第一种与传统心身二元论的意图最符合，它实际上主张物理主义**几乎**是正确的：物理主义对非物质科学与物理学之间关系所作的说明，对于**除**民间心理学**之外**的所有非物质科学都是正确的，而非物质科学必须被看成是描述了真实的现象，这些现象和基础物理学所描述的现象一样是基本的，而且一样能保证具有优势地位。第二种非物理主义回答主张物理主义是完全错误的，它断言：盛行于各种科学与名义科学之间的是一种多元的平等主义，因此每门科学在本体论上都和所有别的科学不相上下，而且最终会证明世界根本不是有层次的。第二种回答的倡导者否认心理的东西是物理的，在这个方面它与传统的心身二元论者合流了，但他们会补充说：两者都不是地质学、气象学或微生物学。当下最有影响的反物理主义者似乎都支持第二种回答。[3]

3.1 对物理主义的阐述

“一切都是物理的”，这句物理主义的口号可作为我们再好也不过的出发点。什么样的事物应划入“一切”的范围？这里一个重要的（即使常被忽视的）问题是：物理主义者是否只是想提出关于**具体**（concrete）实体的主张（例如，专门科学所描述的现象），或者说，是否还想提出关于**抽象**实体的主张（例如，像柏拉图主义者所理解的数或命题）。由于缺乏相 67

关文献，请允许我作两点解释。第一点，就具体实体与抽象实体之间存在的原则性区别而言，物理主义者完全可以**规定**：他们的论题范围仅限于具体实体；所产生的论题尽管比不受限制的论题争议性更小，但仍是富有争议的。第二点，在裁定这样一种对物理主义范围的限制是否是令人反感的武断之举时，我们要考虑的关键问题是：就所有**具体**实体都是物理的这种看法来说，不管其理由是什么，它都同样能适用于**抽象**实体吗？例如，有的人可能会考虑已提出的对物理主义的证明——对它们将在后面 3.2 节作出简要描述，他还问：在每种情况下，是否能对它作出修改以得出这一结论：抽象实体是物理的。

对于物理主义的范围，第二个被忽视的问题涉及所宣称的物理实体从属的**范畴**。由此往后，请允许我们假定这些实体是具体的；那么这里的问题就是：这些具体实体是物体、事件、属性、事实，还是别的什么。从直觉上看，只说一切具体**物体**都是物理的，不会表达出物理主义的全部内容；因为它们的**属性**也必须是物理的。但是由于这些属性**本身**不一定是物理的，因此只要它们的所有**实例**是物理的即可，物理主义者可能就应当主张：所有具体物体和属性实例都是物理的。[4]为了简单清晰，我将假定这是正确的。除物体和属性实例之外，假如一个人的本体论还包括事件、状态、过程、条件或者其他东西，再假如这些都不能还原为物体和属性实例，那么（个例）事件、状态、过程等就能够而且也应当被包含于他的物理主义之中。

让我们回到这句口号："一切都是物理的。"我们应当怎样理解"……是物理的"？一种有吸引力的观点是：将一个实体当成物理的，当且仅当它是由达成共识的当代物理学理论中的某个谓词表达的一个种类，这里，翻阅一些最新的物理学教科书就很容易发现达成共识的当代物理学理论的本质；因此，物质实体的样本是这些东西，如电子和夸克，而物理属性的样本是这些属性，如电荷和质量。根据我们之前的假设，由此产生的物理主义学说主张：每个具体物体和属性实例都是由达成共识的当代物理学理论中的某个谓词表达的一个种类。

表面上看，这种物理主义似乎极其违背直觉，因为它明显蕴含了几乎所有由专门科学和常识描述的东西都不存在的意思。因为卷心菜和国王、大使馆和大象、过滤器和价格等都不是由达成共识的当代物理学理论中的一个谓词表达的。当用这种激进的取消主义方式来解释时，这种物理主义的拥护者没必要否认：世界上存在的东西**看起来**的确比他们的物理主义所

支持的要多，但他们希望只借助于物理实体来解释这种错误的表象：他们可能会否认存在大象，尤其会否认某头大象是**存在**的。但是他们将坚持认为：**某种东西**占据了常识所确定的由大象占据的空间——很可能是微观物理粒子的某种时空排列（麦克斯韦［Maxwell］，1968）。 68

然而，实际上，当物理主义被理解成这一主张，即每个具体物体和属性实例都是由达成共识的当代物理学理论中的某个谓词表达的一个种类时，它就根本不能**单独**推导出取消主义的结果；只有同另一个前提相结合，它才有这样的功能。另外的一个前提是：专门科学和常识所描述的各种东西，与由达成共识的当代物理学理论中的谓词所表达的各种东西（在数量上）不是相**同一**的。如今另一个前提被普遍接受了，这主要是由于专门科学和常识所描述的各种东西是**可多样实现的**，也就是说，在物理方面迥然有别的微观物理粒子的组合物却能（或者只能）构成相同专门科学或常识种类的个体。因为假定物理种类 K 中的一个物体和不相容的物理种类 J 中的一个物体都能构成过滤器；那么如果**是一个过滤器**和**是物理种类 K** 是完全相同的属性，就能推出**每个**过滤器都是物理种类 K 中的一个物体；但是由于至少有一个过滤器（物理种类 J 中的一个）不属于物理种类 K，由此就能推出**是一个过滤器**和**是物理种类 K** 不可能是完全相同的属性。然而，这个反对将专门科学种类和常识种类同一于物理种类的论证可能会受到质疑。[5]一方面，有些现象（如现象意识）**实际的**可多样实现的程度还不明朗，而可多样实现的**纯粹可想象性**（mere conceivability）的意义可能会受到这些哲学家的怀疑，他们怀疑可想象性是否是真正的形而上学可能性的可靠向导（乃至怀疑它是够格的向导）。另一方面，如果物理谓词的任意析取陈述了真实的种类和真正物理的种类，那么看起来仅仅将明显的非物理种类与由物理谓词（的合取）的合适**析取**所表达的物理种类同一起来，就能摆脱这种从可多样实现性出发进行的论证。例如，**是一个过滤器**表达的可能是这种物理属性：**是物理种类 K** 或**是物理种类 J**。

然而，可多样实现性论证事实上几乎使所有哲学家都相信：专门科学和名义科学的种类不能与物理种类相同一。如果这是正确的，那么由于我们刚才一直在考虑的这种物理主义构想蕴含着前面所讨论的这些强取消主义结论，因此，近几十年来，有兴趣提出一种全面的物理主义学说的哲学家，都采用了不同的方案。根据这些方案，物理主义者主张：每个具体物体和属性实例都是物理的。这里的“物理的”要么用的是某种**狭**义的意义，要么用的是广义的意义——与狭义的物理之物有某种关系。在这样表

述时，物理主义不一定会否认专门科学和常识所描述的各种东西的存在；就那些东西来说，即使它们不是狭义的物理之物，也可能与狭义的物理之
69 物有正确的关系（不管这种关系最终是什么）。根据这种新方案，物理主义也不必主张专门科学和常识所描述的种类与（狭义的）物理种类是同一的，因为（很可能）这种正确的关系既不必是也的确不蕴含同一性。

按照这个方案，我们应当怎样理解狭义的“物理的”？一种明确的建议是重新使用前面给出的关于**绝对的**“物理的”的说明，并由此声称：某种东西是狭义的物理之物（下文称为“狭义物理的”），当且仅当它是达成共识的当代物理学理论中的某个谓词所表达的一个种类。但是如果这样理解“狭义物理的”，那么，所产生的这种物理主义表述就只有在当代物理学本身既正确又完备的时候才是正确的，而鉴于物理学之错误和疏忽的历史记录，这似乎是极不可能的事情；因此，物理主义表述本身根本不可能正确，这在物理主义听起来当然很糟糕。这个难题如果是真实的话，就会困扰迄今为止所讨论的所有物理主义表述。因此，要么物理主义根本不能得以表述（如有些反物理主义者所说），要么我们应当以别的某种方式来理解“狭义物理的”。但怎样理解呢？说某种东西是狭义物理的，当且仅当它是由**完备的**物理学的某个谓词所表达的一个种类，这是不行的。因为如果对完备的物理学是什么样子不加限制，那么就我们现在所知，研究者就有可能假定笛卡尔式心灵的存在，这和志向远大的物理主义者的意图是相悖的。此外，很难明白，我们**当前**能够取得的科研成果如何能成为以下这种物理主义的证据——对它的陈述诉诸一种我们目前对其内容尚一无所知的完备物理学。因此，对于“狭义物理的”似乎需要第三种理解方式。它也许应当借助当代物理学的一种**适当延伸**（modest extension）的观念，这种东西与当代物理学非常相似，因为我们并非对其内容一无所知，但是它又很灵活，足以经受得住关于当代物理学既不完备又（在有些方面）错误的发现。但是，对于避免困扰着刚才考虑的两种方式的问题来说，是否能找到第三种方式，这仍是一个未解决的问题，直到最近，它在很大程度上还未被考察。[6]

对“狭义物理的”就说这么多。根据我们现在考虑的方案，我们应该怎样从**广**义上理解“物理的”（下文称为“广义物理的”）？难以忽视的一种答案是：某种东西是广义物理的，当且仅当它**随附**于狭义的物理之物。这一回答在过去 20 年中最为流行，它所产生的物理主义表述主张：每个具体物体和属性实例都要么本身是狭义物理的，要么随附于狭义的物理之

物。哲学中的随附性概念（日常的“随附”的用法与其哲学用法没有多少关系）从直觉上可以作这样的解释：心理的东西（例如）随附于物理的东西，当且仅当物理事实被确定，心理事实也会因此被确定；事物的心理方面**不可能**在其物理方面不发生变化的情况下发生变化（并且不仅仅是不变化）。

如今，大量有关随附性概念的文献都充斥着解释怎样更准确地理解随附性主张的意见，随附性物理主义者（supervenience physicalist，这是我们对这些哲学家的称呼，他们希望某种随附性论题在说明物理主义时能起重要作用）必须解决的一个问题涉及的就是一种随附性主张，如果正确理解的话，一种物理主义的表述就应该使用这种主张。经常讨论的一个随附性概念是金在权（Jaegwon Kim）的**强随附性**。[7]根据这种随附性，非狭义的物理属性随附于狭义的物理属性的主张应当这样理解：非狭义的物理属性**强随附**于狭义的物理属性，当且仅当一种东西有一种非狭义的物理属性，那么这种东西必然具有某种狭义的物理属性，以至于具有那种狭义的物理属性的东西也必然具有最初的非狭义的物理属性。然而，根据这种随附性主张，某个物体的所有非狭义的物理属性都是由**那个物体**的狭义的物理属性（甚至可能是由那个物体与它的非狭义的物理属性**同时发生的**狭义的物理属性）决定的。而这个推断与很多关于各种非狭义的物理属性之构成的可信意见似乎是不一致的；例如这种看法：美钞的真实性部分是由它与有关主管当局的（历史）关系构成的。再如这种看法：一种生物实体（例如，心脏）对一种功能的拥有部分地是由其祖先的选择历史构成的，再如心灵哲学家普遍接受的外在主义看法：命题态度的表征内容部分地是由它们与其所有者头脑之外的事态的关系构成的。因此，随附性物理主义者一般都喜欢使用金在权的**整体随附性**（global supervenience）概念。在表达想得到的随附性主张方面，第一个很不成熟的尝试是：任何两个可能世界，如果它们在包含的狭义物理实体和属性实例以及所具有的物理规律方面都完全相似，那么它们在其包含的所有（具体的）实体和属性实例方面也会完全相似。文献中收录有对这样一种主张应该怎样改进的深入讨论。[8]

志向远大的随附性物理主义者必须解决的第二个问题涉及他们的主张所借助的模态性的特征和适当强度。就模态算子表述来说（例如，上述的强随附性主张），问题就是：必然性算子应该表达哪一种必然性。是概念的必然性，形而上学的必然性，法则学的必然性，还是别的东西？就可能

世界的表述来说（例如，上述的整体随附性主张），问题就是：这一主张应该对哪些可能世界进行量化。是几乎所有可能世界，还是那些只与实际世界具有相同自然规律的可能世界，抑或其他可能世界？对这些问题的任何一种恰当的回答，显然都必须在两个极端之间取中庸之道。请让我以整体随附性主张为例来加以解释。

一方面，如果这些主张对几乎**所有**可能世界进行量化，那么它们似乎就是错误的。因为它们蕴含着：每个在狭义的物理方面无法与（比如）我们的世界相区别的世界，也不能在心理方面与我们的世界相区别。但的确有这样一个世界，它难以在狭义的物理方面与我们的世界相区别，却能在心理方面**相区别**，因为除了其狭义的物理内容之外，它还包含有灵质的精灵（ectoplasmic spirits），这些精灵发挥着心灵的恰当功能。这个所谓的“额外之物问题”（problem of extras）还没有公认的解决办法（对已提出的解决办法的批判性讨论，参见威特默［Witmer］，1999）。另一方面，整体随附性的主张不能只对法则学的可能世界（即与实际世界具有完全相同的自然规律的可能世界）进行量化，因为在这种情况下，即使物理主义是错误的，这些随附性主张也可能正确。例如，假定物理主义最终证明是**错误的**，但这仅仅是由于现象属性最终证明是非广义的物理属性，而这些属性却是根据基本的心理物理规律以双重条件的形式与某些物理属性联系了起来；在这种情况下，对法则学的可能世界进行量化的整体随附性主张仍是**正确的**。因为任何难以在狭义的物理方面与实际世界相区别的法则学的可能世界都是这样的世界：在其中基本的心理物理规律是有效的，这些规律的执行还导致了与实际世界所得到的现象属性分布完全相同的现象属性分布。因此，整体随附性的主张一定是对某个可能世界集合进行量化的，这个集合既不同于所有可能世界组成的集合，也不同于由法则学的可能世界组成的集合。在这个集合是什么的问题上，目前还未达成共识。

随附性物理主义者必须解决的第三个问题，涉及随附性主张——不管怎样表述——打算在物理主义的总体表述中所起的确切作用。不同作者对预期的作用看法不同，它有时还非常模糊。有的作者似乎认为，构成整个物理主义的是一种恰当的整体随附性主张。有的则认为至少还需要别的某种主张（大意是：每个物体都是一个狭义的物理对象，否则就是一种狭义的物理对象的时空集合）；但是还不清楚他们究竟把什么看成是充分的。他们可能认为随附性论题表达的是关于**属性**的物理主义，而要表达关于**特定事物**的物理主义则需要别的主张。至少这里还有一个未解决的问题。

然而，即使刚才提到的所有问题以及其他问题都解决了，关于物理主义随附性表述的充分性仍会受到怀疑。在过去十年间，对这些表述的热情持续下降。人们怀疑的内容是：尽管适当的随附性主张可能是物理主义一个逻辑上的必要条件，但它不是逻辑上的充分条件（即使对于关于属性的物理主义）。简单来说，怀疑的根据在于，任何一种被当成物理主义表述的随附性主张，都只是这个主题的一种变化形式，即事物狭义的物理方面**使**它们非狭义的物理方面**成为了必需**。但是对于这种必需（necessitation）如何以及为什么会出现，还没有作出解释——**随附性主张本身所蕴含的**解释。因此，依据随附性主张本身所作的说明，狭义的物理之物对非狭义的物理之物的必需可能构成了一个严酷的模态事实；但是，依据随附性主张本身所作的说明，如果狭义的物理之物对非狭义的物理之物的必需只是一个严酷的模态事实，那么随附性主张**本身**就无法在直觉上产生令人满意的意义，在这种意义上心理之物是狭义物理的。因此，随附性主张对于关乎任何东西的物理主义都不是充分的。[9]

对于随附性物理主义理解“广义物理的”的方式，迄今还没有人能达到随附性主张曾经享有的声望。但是，虽然它的一个替代者直到最近一直都被忽视了，然而它确实存在了几十年（博伊德，1980；莱肯，1981、1987：第4章）。其最主要的想法是：某种东西是广义物理的，当且仅当它是狭义的物理之物**所实现**的一种**功能的**东西。它所产生的物理主义表述主张：每个具体物体和属性实例要么本身都是狭义物理的，要么都是功能的，且由某种狭义的物理之物实现的。因此，**实现物理主义者**（realization physicalists，这是我们对接受这种物理主义表述的哲学家的称呼）主张：所有（实际的）非狭义的物理种类实际上都是功能的种类；但他们不承诺任何功能的或语言学的论题——例如，不承诺这样的论题——宣称非狭义的物理**概念**或**术语**能给予功能定义。（因此，在心灵哲学中，实现物理主义可以看成是心理功能主义的一种泛化形式。）实现物理主义者也不**否认**措辞恰当的随附性主张；事实上，他们可能认为这些主张是物理主义为真的逻辑必要条件。但是，他们坚持认为，**解释**非狭义的物理之物对于狭义的物理之物的随附性（如果它这样随附的话）的是这个事实：非狭义的物理之物是功能的，是由狭义的物理之物实现的。[10]

实现物理主义者所使用的这两个伴生概念，即功能种类的概念和实现的概念，从心灵哲学来看当然不陌生。但是，实现物理主义要求它们承担的重任也表明，甚至当它们用于心灵哲学时，其意义也是既不一致也不清

晰的。因此，实现物理主义者必须说明他们怎样理解这些概念。一种吸引人的方案是将功能种类看成高阶种类：一种功能属性 P 就是具有能起某种作用的属性的属性；一种功能的物体种类 O 就是这种当且仅当存在能起某种作用的一个物体才存在的物体；如此等等。这里提到的作用可能是因果作用，也可能是法则作用，还可能是计算的作用——还可能是任何别的作用，由于起一种作用不过是符合一个能详细说明的条件，而这个条件原则上可以属于任何一类。那么，就可以把实现理解成起作用（role-playing）。如果功能属性 P 是具有能起某种作用的属性的属性，那么，任何起着某种作用的属性 Q 都可以被认为实现了 P。然而，在实现物理主义者拥有一个适合表述物理主义的概念之前，这种理解实现的方案还需要进一步改进。要格外注意的一个问题是：实现物理主义者是否能对**个体**（个例）而非**种类**（类型）作出令人满意的解释？而在这样做时，他们是否必须或者应该坚持这些主张：非狭义物理的个体（物体、属性实例等）与实现它们的狭义物理个体之间具有**同一性**。[11]

既然我们已经回顾了一些想更准确地表述物理主义的尝试，那么我们就能简要地考虑一下**关于**物理主义论题的三个问题：（1）它是偶然的还是
73 必然的？（2）它是先验的还是后验的？（3）它是还原论的吗？

表面看，物理主义似乎是偶然的。因为（在以口号的形式表达时）只有在所有具体之物都是物理的时候，它才正确，即使仅有一个非物理的具体之物时，它也是错误的。但是，是否存在具体的非物理之物（例如，具有外质的幽灵）的确是偶然的。因此，物理主义是偶然的。此外，在将物理主义更准确地表述成以下主张时——所有具体物体和属性实例都属于由达成共识的当代物理学理论中的某个谓词所表达的一个种类——这个结论仍然成立。因为下述情况的确是偶然的，即是否存在这样的具体物体或属性实例，它**不**属于由达成共识的当代物理学理论中的某个谓词所表达的一个种类。但是，当依据随附性或者依据实现来陈述它时，物理主义仍是一个偶然论题吗？

它是偶然论题。无可否认，物理主义的任何一种随附性表述都主张：事物狭义的物理方面**使**它们非狭义的物理方面**成为了必需**，这听起来当然像一个非偶然的主张。另一方面，物理主义的随附性表述必须运用于**现实**世界，它起码暗示了**现实**世界是这样的：在**现实世界**中，事物狭义的物理方面使它们非狭义的物理方面成为了必需，也就是说，任何一个在狭义的物理方面与**现实**世界完全相似的世界，在非狭义的物理方面也与现实世界

完全相似。那么，要保证物理主义的随附性表述在这方面取得成功，一种方式是把它看成一种对**所有**可能世界都毫不例外地进行量化的整体随附性主张。因为如果这种主张对**所有**可能世界进行量化，并断定**任何**两个世界在狭义的物理方面完全相似就会在其他一切方面都完全相似，那么它显然就是说：任何在狭义的物理方面与**现实**世界完全相似的世界，就会在一切方面与现实世界完全相似。那么，在被看成是关于几乎所有可能世界的一种量化时，物理主义的随附性表述就确实表达了一种非偶然的主张，它是否正确并不依赖于现实世界最终是什么样子。然而，为了避免上面所讨论的“额外之物问题”，物理主义的随附性表述进行量化的可能世界应该比所有可能世界**要少**；它应该只对所有符合某个偶然条件 X（不管它是什么）的可能世界进行量化，因此它仅仅主张：任何两个在狭义的物理方面完全相似的 **X 世界**就会在其他一切方面完全相似。但是，在这种情况下，只有在现实世界符合条件 X 时——这是偶然的事情，这一主张才适用于现实世界（即暗示任何在狭义的物理方面与现实世界完全相似的世界，也会在非狭义的物理方面与现实世界完全相似）。因此，严格说来，物理主义的随附性表述肯定不仅包含有一种随附性主张——它对一个受到适当限制的可能世界集合进行量化，而且包含有这种偶然的主张：现实世界实际上就处于那个受限制的集合。因此，一种关于物理主义的可信的随附性表述是一个偶然论题，它是否正确依赖于我们碰巧居住在什么样的世界里。

然而，有些随附性物理主义者保证他们对物理主义的表述适用于现实世界，他们达到这一目的是通过使这些表述明确地指称现实世界；根据这 74
些表述，任何在狭义的物理方面与现实世界完全相似的世界，也会在非狭义的物理方面与现实世界完全相似。[12] 现在假定“P”是对现实世界的一种完备的狭义物理描述，“Q”是对现实世界的一种完备的非狭义的物理描述。根据这个表述，如果物理主义是正确的，那么“如果 P，那么 Q”这个条件就表达了一个必然真理。根据这个表述，物理主义是一个非偶然论题吗？不是。因为“如果 P，那么 Q”这个条件对于物理主义不是逻辑充分的；只有和下述这个**偶然**主张相结合——“P”**是**对现实世界的一种完备的狭义物理描述，“Q”**是**对现实世界的一种完备的非狭义物理描述时，这个条件才是逻辑充分的。“如果 P，那么 Q”这个条件根本不蕴含任何在狭义的物理方面与现实世界完全相似的世界也会在非狭义的物理方面与现实世界完全相似的意思，除非“P”表达现实世界狭义的物理方面，“Q”表达现实世界非狭义的物理方面是（偶然地）正确的。对它可作这样的思

考：给你一种外延的狭义物理的世界描述“S”，一种外延的非狭义物理的世界描述“T”，你了解到（让我们假定，是使用一种先验方法）“如果 S，那么 T”表达了一个必然真理；因而你就能断定物理主义正确吗？显然不能，因为你还不知道“S”和“T”是否对**现实**世界作出了准确描述；而它们是否是准确描述仅是一个偶然的事实。那么，严格说来，这种随附性物理主义——它的表述明确地指称现实世界——就是这个论题：（1）“P”是对现实世界一种完备的狭义物理描述；（2）“Q”是对现实世界的一种完备的非狭义物理描述；（3）“如果 P，那么 Q”表达了一个必然真理。

在借助实现概念予以表述时，物理主义仍然是偶然的。因为在这样表述时，它主张每个具体物体和属性实例都要么本身是狭义物理的，要么是功能的且是由某种狭义的物理之物实现的。如果存在既非狭义物理的又非由狭义的物理之物实现的具体存在物，那么物理主义本身就是错误的。但存在什么具体的东西的确是偶然的；因此，这样表述的物理主义是偶然的。基于前述的理由，我们也不应当由于下述原因否定这个结论：物理主义的实现表述需要一种表述恰当的随附性主张。

我们现在转到物理主义的认识论地位。由于根据任何一种可信的陈述，物理主义论题都是偶然的，因此我们就能有把握地认为它是后验的。现实世界（尽管可能不是别的世界）包含的非狭义物理现象同一于（或者随附于）它所包含的狭义物理现象，或者是由后者实现的，这个论题其实只能通过考察现实世界来确定真伪。尽管如此，一旦发现并说明了事物的实际情况，物理主义者所承诺的特殊模态主张的认识论地位会怎么样？也就是说，这个主张会怎么样：（现在严格地使用明确的描述）事物狭义的物理方面的实际情况是否使它们非狭义的物理方面的实际情况成为了必需？学界对此的看法是不同的。大多数观点认为：这种必需是由于下列两者之间具有后验的必然同一性才成立的：一方面是非狭义的物理之物，另
75 一方面是狭义的物理之物，或者功能的东西、狭义物理地实现的东西。那么，根据这种观点，即使我们已经知道（后验地）现实世界在狭义的物理方面和非狭义的物理方面是什么样子，物理主义是否正确仍然是后验的。然而，根据一种少数派的观点，如果物理主义是正确的，那么某人若对现实世界有一种完备的狭义物理描述，**并且**他拥有所有用于表述非狭义的物理主张的概念，原则上他就能对世界演绎出一种完备的非狭义的物理描述（查默斯，1996；杰克逊，1998）。如果这种观点正确，那么，一个人采取关于现实世界的一种完备的狭义物理描述并使用他所掌握的非狭义物理概

念，就能演绎出一种**据说**适用于现实世界的完备的非狭义物理描述，然后通过比较这种非狭义物理描述和一种通过经验发现的完备的非狭义物理描述，他就能对物理主义**作出检验**。

最后，请允许我们问这个问题：物理主义是否是还原论？即它是否蕴含非狭义的物理之物能还原为狭义的物理之物的证明。根据一种普遍的共识，就我们考虑过的三种物理主义表述来说，第一种表述（被看成非取消论的）是还原论，而第二种（随附性）和第三种（实现）不是还原论；当然，后两种表述的开创者的目的很明确，即要表述非还原论的物理主义版本。[13]但是这一共识肯定有赖于假定对什么**是**能还原为狭义的物理之物的某种解释；而根据这种可还原性解释（来自于恩斯特·内格尔［Ernest Nagel］）——它似乎是人们实际假定的，非狭义的物理之物能还原为狭义的物理之物，当且仅当借助这些额外的前提（即“桥梁原则”）——它们断定非狭义的物理种类与狭义的物理种类（的易处理的析取）具有**同一性**，所有非狭义的物理**规律**都能从狭义的物理规律演绎出来。物理主义的随附性表述和实现表述都不要求这种可还原性，也就是说，不要求非狭义的物理之物还原为狭义的物理之物，因为两者都与这种可多样实现性相一致，即非狭义的物理种类能由**多得数不胜数的**不同狭义物理种类实现。（实际上，第一种表述也不要求这种可还原性，只要允许它以这种方式来摆脱可多样实现性的问题，即将非狭义的物理种类与狭义的物理种类数不胜数甚至无限多的析取同一起来。）

但是，这种对可还原性的新内格尔主义解释是一种重要的哲学主张。如果它不正确会怎样？或者说如果它不是**唯一**正确的（从而对可还原性没有**唯一的**正确解释）会怎样？可还原性的核心思想似乎是：仅当非狭义的物理之物**能**以某种方式依据狭义的物理之物来**解释**时，非狭义的物理之物才能还原为狭义的物理之物。新内格尔主义解释无疑是这一核心思想的一种具体说明（把解释理解成一种派生），但是，其他的说明应该也是可能的，这似乎是可能出现的情况；其他类似说明没有一种应该是唯一正确的，这也似乎是可信的。在这种情况下，物理主义仍是非还原论或者说在一些重要方面仍是非还原论吗？在文献中能找到一些对替代性的可还原性解释的探索（例如，参见沃特斯［Waters］，1990；史密斯，1992；布鲁克斯，1994；梅尔尼克，1995；查默斯，1996；比克尔，1998）。在认为物理主义正确的情况下，也欢迎大家进一步探讨像心理学这样的专门科学是否能享有的各种**自主性**；例如，我们要是知道心理学的**方法论**自主性在

多大程度上需要其**形而上学的**自主性就好了。

3.2 对物理主义的辩护

根据这个假设——物理主义论题能作出令人满意的表述，而且在这样表述时它是后验的——自然会出现这个问题：是否真有什么经验的证据能证明这个论题是正确的。原则上，什么样的非演绎推理策略能提供这种证据？这些策略中的任何一种在执行时能产生物理主义的证据吗？尽管在心灵哲学和其他领域的人们都热衷于物理主义，但这些问题所受到的关注却出乎意料地少。另一方面，它们受到的注意比一些反物理主义言辞的要多。让我们简要回顾一下对物理主义如何被证明的一些建议和那些建议所引起的问题。

一种建议（仿效大卫·刘易斯对心理物理同一论的论证）是：一种有两个前提的论证能用于支持这个结论：某个非狭义的物理种类 N 是与一个狭义的物理种类 P 相同一的（刘易斯，1966；这种概括是在杰克逊 1998 年的著作中提出的）。第一个前提说：N 是起某作用的种类；第二个前提说：P 是起某作用的种类；结论 N＝P 是从同一性的传递性推导出来的。第一个前提能从其据说是先验的概念主张或语言学主张推导出来，即“N”在语义上等价于“起某某作用的种类”；检查狭义的物理种类实际所起的作用，就能经验地发现第二个前提。但是，大多数物理主义者都怀疑这种论证策略是否切实可行，因为事实上专门科学的特有概念或者术语一般（甚至永远）都不会在语义上等价于限定摹状词，就像第一个前提似乎所要求的那样；这种怀疑构成了对关于概念或术语之意义的描述主义理论的普遍怀疑的一部分。目前一个仍未解决的问题是：普遍的刘易斯主义论证策略能否通过这种假设来改进。即专门科学的概念或术语尽管不在语义上等价于限定摹状词，但其指称仍是借助于这些能先验认识的限定摹状词确定的，这些摹状词有这种严格的形式：“**实际地**起某作用的种类。”

第二种建议是：断言非狭义的物理种类与狭义的物理种类之间有同一性的各个结论，可根据一个能推出下述最佳解释的推理来加以论证。这个解释把这样的观察事实当作是它的材料，即任何时候、任何地方只要出现了所讨论的狭义的物理种类的个体，就一定出现非狭义的物理种类的个体。那么，假定我们观察到了非狭义的物理种类 N 和狭义的物理种类 P 在

这种意义上的共同例示，我们便能解释被观察到的这种共同例示的一个假说的确是 N＝P。当然，如果 N＝P，种类 N 和 P 的个体完全能被观察到，那么它们就不可能不被同时观察到。这种同一性假说能被合理地看成是一种比对立假说更好的替代解释，对立的假说断言 N 和 P 不同，并通过假设它们是由一个基本的自然规律联系在一起而解释了它们被观察到的共同例示。同一性假说看起来比这种对立的假说更好，因为在两个不同的方面它都比对立的假说更**经济**：它只假设了一个种类（＝N，＝P），而对立假说假设了两个种类（N **和** P）；而且，它必定视为基本的自然规律的数量比对立假说必定视为基本的自然规律的数量要少（1 个）。因此，被观察到的 N 和 P 的共同例示就提供了 N＝P 的证据（详细阐述，参见希尔［Hill］，1991：第 2 章）。

这种推理模式可能看起来作用有限，因为当今想接受它支持的这种种类同一性主张的物理主义者非常少。但是，它所提出的这种推理能够作出扩展，从而说明如何支持这些物理主义结论而不支持那些断言非狭义的物理种类和狭义的物理种类有同一性的结论。一种显而易见的扩展是：如果作为这种推理之出发点的经验证据是某个非狭义的物理种类 N 与某个**功能的**（而不是狭义物理的）种类 F 被观察到的共同例示，那么这种推理大概就能用于支持这种假说：N＝F，即实现物理主义者显然感觉很合适的一种假说。当然，发现一个非狭义物理种类 N 是一个功能的种类，这还不是发现了 N 是从物理上实现的；那么，由于后一个结论是实现物理主义所需要的，因此对最初的推理路线作进一步扩展就是可取的。假定充当材料的观察是：任何时候、任何地方只要存在狭义的物理种类 P 的一个个体，也一定会有非狭义的物理种类 N 的一个个体。然而，由于可多样实现性，相反情况是观察不到的。这些观察潜在地能用这种假说来解释，即（大致来说）N 同一于某个功能种类 F，并且它实际上是由狭义的物理种类 P 实现的；因为如果 N（＝F）是由 P 实现的，那么 P 对于 N 就是充分的，因此，任何时候、任何地方只要有 P，自然就一定会有 N。如前所述，如果这种假说在经济性方面胜过其对手，那么最初的观察就提供了 N 是功能的且是狭义物理地实现的证据。

这些建议至少提出了两个重要问题。一个是像所提议的推理模式那样依赖于经济性（或简单性）是否合理；这个问题显然依重大的认识论问题的解决情况而定。另一个问题是公认的科学成果能否用来作为这些具有正确前提的推理模式的实际例证；这个问题如今几乎被忽视了。

第三条支持物理主义的经验推理路线是最受注意的推理路线，它推导的过程如下（参见皮科克［Peacocke］，1979：134－143；梅尔尼克，1994）。第一个前提是所谓的狭义物理的**因果封闭性**（或**完备性**）。它声称狭义的物理之物是封闭的，这就是说，一个人不必到狭义物理领域之外去
78 寻找狭义物理现象的充分原因，每个狭义的物理事件都有一个充分的狭义物理原因（就它真的有一种原因而言）。[14]这个前提得到了当代物理学的支持，当代物理学研究了非常多的狭义物理事件，也知道所有狭义的物理事件的解释都不**必然**要借助于非狭义的物理原因。第二个前提是：非狭义的物理事件有狭义的物理结果。非狭义的物理事件无疑有非狭义的物理结果（例如，飓风吹倒了树），但当代物理学向我们保证——如今，科学实在论者都没有真正对此提出质疑——非狭义的物理结果至少有狭义的物理部分。若一个非狭义的物理事件不对一种非狭义的物理结果的某个狭义的物理部分也有某种影响，我们就很难理解它如何能有一种非狭义的物理结果，因此，非狭义的物理事件一定有一些狭义的物理结果。从这两个前提和这个假设（具有狭义的物理结果的非狭义的物理事件甚至对“物理的”作某种广义的理解的情况下也不是物理的）出发，我们可以推出：有些狭义的物理事件**在因果上是多因素决定的**（causally overdetermined）；因为每个具有一种非狭义的物理（根据假设，也是非广义物理的）原因的狭义物理事件也有一种完全不同的狭义物理原因。但是，在这种意义上——这是第三个前提——狭义的物理事件在因果上多因素决定是极不可信的，我们能够合情合理地**抵制**这种假设：具有狭义的物理结果的非狭义的物理事件甚至在对“物理的”作某种广义的理解的情况下也不是物理的，并因此而**承认**它们在对“物理的”作某种广义的理解的情况下是物理的。进一步作枚举—归纳就能得出这个普遍结论：**所有**非狭义的物理事件（而且不仅仅是已知有狭义的物理结果的那些）都是某种广义上的物理事件（也就是说，同一于、随附于或者是由狭义的物理之物实现的）。

这条推理路线引出了很多问题（有关讨论参见米尔斯［Mills］，1996；斯特金［Sturgeon］，1998；威特默，2000；帕皮诺，2001）。抵制物理主义据说会导致这种因果的决定过多真的是坏事吗？倘若如此，原因何在？狭义的物理因果封闭性是这种东西吗——它的证据是不相信物理主义的人能接受的吗？它有反例吗？在任何通常意义的“原因”上，非狭义的物理事件都引起狭义的物理结果，这真的是正确吗？最后，能对这个论证进行修改来容纳狭义的物理领域明显的非决定论特征吗？

对于物理主义可能被如何证明，文献中还收录了其他学者的意见（洛伊尔［Loewer］，1995；帕皮诺，1995；有关批评参见威特默，1998）。但是，由于哲学家们对为物理主义进行辩护的问题的关注少得惊人，因此我觉得一切可能的意见不可能都已提出来了。

3.3 对物理主义的反驳

由于我们所理解的物理主义论题具有全称概括的逻辑形式，因此它原则上会遇到反例，即那些既非狭义物理的也非广义物理的具体物体或属性实例。因此，对物理主义的一种反驳就是以相信这些实体确实存在为基础的。这些实体可能是物理主义者彻底否认其存在的实体（例如，上帝、生命力、魂魄），在这种情况下，提出异议的人就必须提出认为它们确实存在的后验或先验的理由。另外更可信的是，这些实体可能是这类实体（例如，合理的决策、胚胎发育中的情景），它们的存在（至少根据对其存在所需之物的中立解释）是物理主义者根本不能怀疑的，但是对它们既非狭义物理的又非广义物理的描述是有争议的，在这种情况下，提出异议的人就必须为所讨论的实体确实既不是狭义物理的也不是广义物理的提出后验或先验的理由，即要说明没有一个物理主义者会否认人类存在（几乎没有）。但是，如超心理学研究者想系统表明的那样，如果能够证明人类拥有若他们是广义物理的就根本不可能具有的能力（例如，心灵致动力），那么人类就是物理主义的一个反例。在心灵哲学文献中，对物理主义的大多数常见反驳都属于第一种，即假定的反例（参见鲁宾逊，1993）。

在对它们作出回应时，物理主义者自然想对每一案例作出不带个人成见的考察，而我们这里显然可能进行详细考察。但我们应该暂停片刻，转而简要地评论一下某种哲学观点，它可能构成了物理主义者特殊论证的基础，而且它不可能为反对者共享。这种观点实际上是对下述看法的（任何据说是先验理由的）极度怀疑，这种看法是：**要么**某种具体实体存在，**要么**（它的存在仅仅是假定）它既不是狭义物理的也不是广义物理的。根据这种观点，认为任何一种具体实体都存在的理由一定是后验的。同样，认为一个具体实体既不是狭义物理的也不是广义物理的，其理由也一定是后验的，因为它们必定排除了以下可能性：这个实体要么后验地与一个狭义的物理实体相同一，要么后验地与一个功能的但狭义物理地实现的实体相

同一。

因此，对于这些反驳——它们是从关于人们**能想象**什么的前提出发的，物理主义者不可能会受到触动，甚至也不可能去认真地考虑它们[15]。相反，这种反驳真**会**让持有这种观点的物理主义者触动，即证实了有某种无偏见的经验现象，对这种现象的最好解释要求**要么**再假定某种既非狭义物理的也非广义物理的东西，**要么**把它**看成**所有人都承认其存在的既非狭义物理的也非广义物理的东西。这种现象的一个例证是某种人类行为，它确实**不可能**仅仅是广义物理系统运转的产物。然而，物理主义者否认任何这样的现象是实际存在的，根据上个世纪科学的发展，这种否认的可信性对于解释物理主义所享有的流行地位是大有帮助的。[16]

那么，提出第一种反驳的反物理主义者，可能会对多种科学问题作出肯定的回答，根据这种回答，物理主义对于**很多**非狭义的物理之物的广义
80 物理特征的解释是正确的，但是，当然并不是对所有非狭义的物理之物都正确。然而，这些反物理主义者提出了我将要考虑的下述三种反驳，他们可能对多种科学问题作出多元主义、平等主义的回答，根据这些回答，每门科学（或者名义的科学）和其他所有科学都具有相同的本体论地位，因此几乎所有非狭义物理的东西都不是广义物理的。

第一种反驳是：物理主义不可能正确，因为它甚至不能得到恰当的表述。它之所以不能被恰当地表述，原因见于 3.1 节所述的理由——对“狭义物理的”作出定义还没有令人满意的办法。如果这是正确的，那么这种反驳显然对于物理主义是毁灭性的。但是它是否正确仍存在争议，并成了正在进行的研究课题。第二种反驳主张：即使物理主义能作出满意的表述，也没有任何理由能认为它是正确的。[17]（有些哲学家看起来甚至希望借助于某种完全非理性的物理学崇拜来解释物理学的流行。）如果这是正确的，那么，由于科学表面上看无疑有很多种，因此，没有理由不在表面上接受这种表象，并由此认为科学的所有分支都具有相同的形而上学地位。但是，主张不存在物理主义的证据充其量是不成熟的，因为如我们所看到的，物理主义确实存在着有希望的论证路线，而且无论如何，哲学家们对这个问题的研究到目前为止都还很肤浅。此外，如果物理主义仅仅**是**一种偏见，那么它显然会在这些人（我主要指非哲学家）中更为流行——这些人了解凝聚物理学对熟悉的宏观物理现象的解释，了解量子力学对化学的解释，还了解生物化学对细胞生物学的解释，等等。**也许，是一个物理主义者**与在科学上见多识广之间的（公认是不完全的）相关性能从社会

学上作出解释，也能借助于某种系统的推理错误来解释（尽管有人很想理解这种假说），但是，总体上更有可能的似乎是：所讨论的这些人模糊地认识到他们对科学的认识**确实**构成了物理主义的证据，虽然他们还不能说出它究竟是**怎样**构成的。[18]如果这种推测是正确的，那么，哲学家们要想帮助他们解决问题显然还要继续努力。

我要考虑的第三种也是最后一种反驳是：物理主义不可信。因为它不外是说：只有狭义的物理事件才能成为原因，而且只有这些事件的狭义物理属性才是因果相关的，这就是说，只有它们才会对那些事件所具有的结果产生影响（参见洛［Lowe］，1993；莫泽［Moser］，1996）。如果物理主义真的仅仅说了这些，那就很糟糕了；因为非狭义的物理事件（例如决策、地震、化学反应）有时的确能成为原因，而且一个事件（例如，与一把利刃相撞）有时的确能具有它所具有的结果，尽管这个事件的种类，即**与一把利刃相撞**，并不是一个狭义的物理事件种类。而物理主义似乎真的仅仅就是这些主张。因为如果对于每一非狭义的物理结果都有一个对应的充分且基础性的狭义物理现象（如物理主义所要求的），如果所有这些基础性的狭义物理现象都完全是严格按照狭义的物理规律由较早的狭义物理现象引起的，那么狭义的物理现象似乎就完成了所有真正的因果工作，而表面上看到的非狭义的物理因果关系就是一种假象。但是物理主义者几乎不承认这条推理路线——第三种反驳明显依赖于它——能顺利进行；他们可能想对因果关系和因果相关性提出另外一种独立的可信解释，这种解释倾向于认为，如果物理主义正确，那么并非只有狭义的物理事件才能成为原因，或者只有狭义的物理属性才是因果相关的。[19]

【注释】

[1] 有兴趣的读者会在梅尔尼克（2003）的书中找到一种关于我的观点的完整解释。

[2] 心身问题在丘奇兰德（1981）的书中就是通过这些术语作了著名的阐述。

[3] 这种反物理主义者似乎包括古德曼（1978）、普特南（1987）、克兰和梅勒（Crane and Mellor，1990），以及杜普雷（Dupré，1993）。

[4] 一个具体的属性实例就是一种具体属性的一个实例（例如，具有质量的属性）；一个抽象的属性实例就是一种抽象属性的一个实例（例如，可被5整除的属性）。

[5] 一位挑战者是金在权；参见金在权（1998）。

[6] 对于这里的难题和一种可能的解答，参见波伦（Poland，1994：第3章）。在梅尔尼克（1997）中，为了向亨普尔致敬，我将这一难题称为“亨普尔困境”

(Hempel's dilemma)，我认为，尽管文中有这种推理，但对于依据当代物理学来定义"物理的"并没有正当的反驳。可参见戴利（Daly，1998）、蒙特罗（Montero，1999）、克鲁克和吉利特（Crook and Gillett，2001）。另一种方案，参见帕皮诺（2001）；一种批评则可参见威特默和吉利特（2001）。

[7] 参见金在权（1984），这篇论文至少在美国为关于随附性的争论设定了术语。进一步的讨论，参见金在权（1993）、萨维洛斯和亚辛（1995）的书中收录的麦克劳林和波斯特的论文。一篇出色的综述参见霍根（1993）。

[8] 例如，参见赫尔曼和汤普森（Hellman and Thompson，1975）、豪格兰（Haugeland，1982）、霍根（1982、1987）、刘易斯（1983）、波斯特（1987）、杰克逊（1998）。一种密切相关的（因为也是模态的）方案，参见科克（1996）。

[9] 这是我陈述这个问题的方式；参见梅尔尼克（1998、1999）。类似的考虑参见霍根（1993）。

[10] 关于实现物理主义的最充分的解释是梅尔尼克（2003）。也参见波伦（1994），他倡导一种混合的物理主义形式，这种物理主义既加入了随附性因素，也加入了实现因素。

[11] 我将只在这里提到这个论题，即每个个例（例如，个别事件）都与某个狭义的物理个例相同一。尽管它提出时很有名（例如，福多，1974；戴维森，1980），但在最近的物理主义讨论中它起的作用无人知晓，这可能是因为没有一个人认为它对于物理主义是充分的（除非被当成金氏事件［Kim-events］，在这种情况下，它就等价于本文最先讨论的那种不受欢迎的类型同一的物理主义）。戴维森本人提出的无疑是这种与随附性论题结合在一起的论题。

82 [12] 参见查默斯（1996）和杰克逊（1998）。在本文中对当前问题的讨论方面，我曲解了这些作者，尽管并无恶意。为了处理"额外之物问题"，他们不是说"在非狭义的物理方面与现实世界完全相似"，而是说"在**正面的**（positive）非狭义的物理事实方面与现实世界完全相似"。

[13] 这一共识的源头可能是福多（1974）。也参见福多（1997）。

[14] 难以理解的是，另一种主张有时也被称为"物理的因果封闭性"，这种主张是：狭义的物理原因是狭义的物理结果的**唯一**原因。鉴于另一个前提，即：非狭义的物理事件是狭义的物理结果的原因，这后一种主张立即就会导致物理主义；但正由于那个原因，反物理主义者会认为它是用未经证明的假定在论证。它并不为文中的这种封闭性主张所要求。

[15] 对这种观点的辩护可以在我的著作（2001）中找到，这种观点反驳了查默斯（1996）对它提出的质疑。

[16] 我相信，任何否认这样的现象存在，都是符合在斯马特的经典论文（1959）中找到的亲物理主义（pro-physicalist）思想路线**之一**。

[17] 尽管说起来令人不快，但是我担心，声称在文献中找不到支持物理主义的论

证的哲学家，一定会被问到他们在寻找时付出了多大努力。

[18] 这种形式的假设在帕皮诺的著作（2001）中得到了辩护。

[19] 几乎所有关于心理因果关系问题的心灵哲学文献无疑都与此处相关。

参考文献

Bickle, John (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: The MIT Press.

Boyd, Richard (1980). "Materialism Without Reductionism: What Physicalism Does Not Entail." In Ned Block (ed.), *Readings in the Philosophy of Psychology*, vol. 1. London: Methuen: 268–305.

Brooks, D. H. M. (1994). "How To Perform A Reduction." *Philosophy and Phenomenological Research*, 54: 803–814.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Charles, David, and Lennon, Kathleen (1992). *Reduction, Explanation, and Realism*. New York: Oxford University Press.

Churchland, Paul (1981). "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy*, 78: 67–90.

Crane, Tim, and Mellor, D. H. (1990). "There Is No Question Of Physicalism." *Mind*, 90: 185–206.

Crook, Seth and Gillett, Carl (2001). "Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism." *Canadian Journal of Philosophy*, 31: 333–360.

Daly, Chris (1998). "What Are Physical Properties?" *Pacific Philosophical Quarterly*, 79: 196–217.

Davidson, Donald (1980). *Essays on Actions and Events*. New York: Oxford University Press.

Dupré, John (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Fodor, Jerry A. (1974). "Special Sciences, or The Disunity of Science As A Working Hypothesis." *Synthèse*, 28: 97–115.

—— (1997). "Special Sciences: Still Autonomous After All These Years." In James E. Tomberlin (ed.), *Philosophical Perspectives*, 11, *Mind, Causation, and World*. Cambridge, MA: Blackwell: 149–163.

Gillett, Carl and Loewer, Barry (2001). *Physicalism and Its Discontents*. New

York: Cambridge University Press.

Goodman, Nelson (1978). *Ways of Worldmaking*. Sussex: Harvester Press.

Haugeland, John (1982). "Weak Supervenience." *American Philosophical Quarterly*, 19: 93-103.

Hellman, Geoffrey, and Thompson, Frank (1975). "Physicalism: Ontology, Determination, and Reduction." *The Journal of Philosophy*, 72: 551-564.

Hempel, Carl G. (1980). "Comments on Goodman's Ways of Worldmaking." *Synthèse*, 45: 193-199.

Hill, Christopher S. (1991). *Sensations: A Defense of Type Materialism*. New York: Cambridge University Press.

Horgan, Terry (1982). "Supervenience and Microphysics." *Pacific Philosophical Quarterly*, 63: 29-43.

—— (1987). "Supervenient Qualia." *Philosophical Review*, 96: 491-520.

—— (1993). "From Supervenience To Superdupervenience: Meeting the Demands of a Material World." *Mind*, 102: 555-586.

Jackson, Frank (1998). *From Metaphysics To Ethics: A Defence of Conceptual Analysis*. New York: Oxford University Press.

Kim, Jaegwon (1984). "Concepts of Supervenience." *Philosophy and Phenomenological Research*, 45: 153-176. Reprinted in Kim (1993).

—— (1993). *Supervenience and Mind: Selected Philosophical Essays*. New York: Cambridge University Press.

—— (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: The MIT Press.

Kirk, Robert (1996). "Strict Implication, Supervenience, and Physicalism." *Australasian Journal of Philosophy*, 74: 244-257.

Lewis, David (1966). "An Argument for the Identity Theory." *Journal of Philosophy*, 63: 17-25.

—— (1983). "New Work For A Theory Of Universals." *Australasian Journal of Philosophy*, 61: 343-377.

Loewer, Barry (1995). "An Argument for Strong Supervenience." In Savellos and Yalcin (1995): 218-225.

Lowe, E. J. (1993). "The Causal Autonomy of the Mental." *Mind*, 102: 629-644.

Lycan, William G. (1981). "Form, Function, and Feel." *Journal of Philosophy*, 78: 24-50.

—— (1987). *Consciousness*. Cambridge, MA: The MIT Press.

Maxwell, Grover (1968). "Scientific Methodology and the Causal Theory of Percep-

tion." In I. Lakatos and A. Musgrave (eds.), *Problems in the Philosophy of Science*. Holland: North Holland Publishing Company.

McLaughlin, Brian P. (1995). "Varieties of Supervenience." In Savellos and Yalcin (1995): 16-59.

Melnyk, Andrew (1994). "Being A Physicalist: How And (More Importantly) Why." *Philosophical Studies*, 74: 221-241.

—— (1995). "Two Cheers For Reductionism: Or, The Dim Prospects For Non-Reductive Materialism." *Philosophy of Science*, 62: 370-388.

—— (1997). "How To Keep The 'Physical' In Physicalism." *The Journal of Philosophy*, 94: 622-637.

—— (1998). "The Prospects for Kirk's Non-Reductive Physicalism." *Australasian Journal of Philosophy*, 76: 323-332.

—— (1999). "Supercalifragilisticexpialidocious: A Critical Study of Savellos and Yalcin's Supervenience: New Essays." *Nous*, 33: 144-154.

—— (2001). "Physicalism Unfalsified: Chalmers' Inconclusive Conceivability Argument." In Gillett and Loewer (2001): 329-347.

—— (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Mills, Eugene (1996). "Interactionism and Overdetermination." *American Philosophical Quarterly*, 33: 105-117.

Montero, Barbara (1999). "The Body Problem." *Nous*, 33: 183-200.

Moser, Paul K. (1996). "Physicalism and Mental Causes: Contra Papineau." *Analysis*, 56: 263-267.

Papineau, David (1995). "Arguments for Supervenience and Physical Realization." In Savellos and Yalcin (1995): 226-243.

—— (2001). "The Rise of Physicalism." In Gillett and Loewer (2001): 3-36.

Peacocke, Christopher (1979). *Holistic Explanation: Action, Space, Interpretation*. New York: Oxford University Press.

Poland, Jeffrey (1994). *Physicalism: The Philosophical Foundations*. New York: Oxford University Press.

Post, John F. (1987). *The Faces of Existence: An Essay in Nonreductive Metaphysics*. Ithaca, NY: Cornell University Press.

—— (1995). "'Global' Supervenient Determination: Too Permissive?" In Savellos and Yalcin (1995): 73-100.

Putnam, Hilary (1987). *The Many Faces of Realism*. Illinois: Open Court.

Robinson, Howard (1993). *Objections to Physicalism*. New York: Oxford Univer-

sity Press.

Savellos, Elias E. and Yalçin, Ümit D. (eds.) (1995). *Supervenience: New Essays*. New York: Cambridge University Press.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review*, 68: 141-156.

Smith, Peter (1992). "Modest Reductions and the Unity of Science." In Charles and Lennon (1992): 19-43.

Sturgeon, Scott (1998). "Physicalism and Overdetermination." *Mind*, 107: 411-432.

Waters, C. Kenneth (1990). "Why The Anti-Reductionist Consensus Won't Survive: The Case of Classical Mendelian Genetics." In A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990*. East Lansing, MI: Philosophy of Science Association.

Witmer, D. Gene (1998). "What is Wrong with the Manifestability Argument for Supervenience." *Australasian Journal of Philosophy*, 76: 84-89.

—— (1999). "Supervenience Physicalism and the Problem of Extras." *The Southern Journal of Philosophy*, 37: 315-331.

—— (2000). "Locating the Overdetermination Problem." *The British Journal for the Philosophy of Science*, 51: 273-286.

Witmer, D. Gene and Gillett, Carl (2001). "A 'Physical' Need: Physicalism and the Via Negativa." *Analysis*, 61: 302-309.

第4章　二元论

霍华德·鲁宾逊（Howard Robinson）

4.1 导　言

心灵哲学中的二元论是一种主张心灵与身体（或心理状态与物理状态）具有极其不同的本质的学说。这种不同究竟该怎样予以表述却是一个有争论的问题，但一般的倾向是强调：心理的东西所具有的两种属性是不同于物理的东西的。其中之一是意识状态的私人性或主观性，它与物理状态的公共可接近性形成了鲜明对比。另一特点是心理状态具有意向性或“关于性”；物理状态处在时空关系之中，并可互为因果，但不能内在地关于任何事物。物理主义者的主要任务就是用物理术语或相容于物理的术语说明这些属性，而二元论者则否定这种做法。[1] 85

一般认为，二元论表现为两种形式，即实体二元论和捆绑（bundle）二元论。前者首先与笛卡尔的名字连在一起，后者与休谟有关。[2] 对于捆绑二元论者又必须作出进一步的区分。如有些人像休谟一样既不承认心理实体，又不承认物理实体，而把两者都看作是状态、属性或事件的集合（取决于那一理论如何被陈述）。对于另一些人来说，只有心灵可以如此看待，因为身体是实体性实在，而只有心灵才是状态、属性或事件的集合。这一看法促成了心灵的相对贬值以及向**属性**（attribute）理论的转向。根据这一理论，心理状态是物理实体——人体或大脑——的非物理属性。这一理论常被看作是唯物主义的最弱形式或具有最少还原论色彩的形式。之所以说它是唯物主义形式，是因为它主张物质实体是唯一的实体；之所以说它是二元论的一种形式，是因为它承认心理状态和属性具有不可还

原性。

实体二元论和捆绑二元论面临着三个相同的难题。第一个难题是，必须说明我们为什么要做一个二元论者——唯物主义关于心灵的说明为什么
86 没有用？第二个难题是要对非物质心灵的统一性本质作出解释。对于笛卡尔主义者来说，这意味着应对他如何理解非物质实体这一概念作出解释。而对于休谟主义者来说，问题是对那捆绑中的不同要素的关系之本质作出解释，正是由于这种捆绑，诸要素才结合成一个事物。[3]过去并未见对后一任务的成功的解决，休谟也宣称自己完全被那个问题弄得不知所措，因而拒绝作出自己的回答（即使从文本出发，也难以弄清个中原因）。[4]第三个难题就是如何令人满意地说明非物质心灵与物质性身体的关系。这首先意味着必须对它们之间的相互作用作出解释，不然的话，要么会陷入副现象论（即主张心理的东西由物质的东西所产生，而又不能对物理的东西产生影响的一种理论），要么会陷入平行论（即主张心理王国与物理王国并行不悖、相互间不存在因果作用的理论）。

由于我能利用的篇幅有限，因此我不准备讨论上述全部问题，敬请读者谅解。我要做的是尝试以笛卡尔的方式说明：第一，能思主体一定超越物理世界；第二，这种主体在本质上一定是简单的。它们（即是说，我们）更像笛卡尔所相信的非物质实体，而不像休谟式的捆绑在一起的心理事件或状态。因此我要集中探讨的是：我们为什么应成为二元论者，我们为什么应成为笛卡尔派的二元论者。至于如何说明心灵的统一性，以及如何说明我们的心灵与我们的身体的关系，这些都不是这里要讨论的问题。当然，有一例外，那就是要对心灵必然具有的简单性作出说明。[5]

为了完成我为自己设定的第一个任务（即说明能思主体超越于物理世界），我将引入至今我没有涉足过的一种二元论形式，即**谓词二元论**。在关于二元论的讨论中，它常被忽视。它是一种主张心理学谓词或心理主义谓词不能还原为物理主义谓词（在下一节，我再来说明这样说的意思）的理论。在当今，几乎没有哲学家相信这种还原，或认为这种还原对物理主义是必不可少的。谓词二元论只是意义层面的二元论，一般认为，它不具有本体论的意蕴。我要论证的是，这样说是错误的，因为谓词二元论否定还原，因而它是物理主义的一大威胁。专门科学的不可还原性一般意味着：心灵并不是这些科学所关注的物理王国的组成部分。

仅有这个结论还不足以迫使我们接受任何特定形式的二元论。因为心灵即使超越于它所构筑的科学所关心的物理世界，它仍有可能就是休谟所

说的心灵状态或属性之捆绑。即是说，谓词二元论要求我们接受的也许只是属性二元论，而这种二元论并不会像归属（the attribute）理论那样让人远离物理主义。接下来，我会进一步说明的是：事实并不是这样，因为属性二元论不足以说明这样一些方面，即**人格**同一性将自己与物理身体、别的复合实在之同一条件区别开来的那些方面，人格同一性的这些约束条件只能由带有笛卡尔风格的实体二元论来满足。

4.2 谓词二元论之论证

如果物理主义是对的，那么在某种意义上，它就可能原则上用完善的物理学词汇对世界作出全面的描述。用质料而非形式的方式说，**真正**有存在地位的一切属性就是基本物理实在所具有的那些属性。但是，除了用物理学词汇述说这个世界之外，还有许多别的真实谈论这个世界的方式；而且很显然，这个世界所具有的许多属性是不包含在这种物理学之中的。这些高阶的谓词和属性可以出现于别的——或专门的——科学之中，例如化学、生物学、细胞学、认识论、地理学、气象学、心理学以及通常所说的社会科学之中。我们的日常话语更是这样，它常能述说这样的事实，即在我们自然地称作科学的任何学问中没有任何地位的事实。尽管我们预设物理学完善地领悟了本体论的基本层次，但它怎样成为别的所有一切本体论的基础，并让这些层次的话语为真呢?

逻辑实证论对此问题有一个简单的回答，即任何值得尊重的话语层次都可还原为它下面的某个层次，最终又可还原为物理学本身。我们所谈论的那类还原有一种“**强**”形式和一种“**极强**”的形式。根据“**极强**”的形式，专门科学和日常话语中所有值得尊重的陈述，从原则上说都可转译成物理学中的陈述。因此，所有真理最终都能用物理学语言来表述。[6]根据纯粹的“**强**”的形式——这是还原主义讨论中一般会用到的形式，只存在着把高层次科学中的概念和规律与低层次科学乃至最终与物理学中的概念和规律关联的种种规律（可称作“桥梁规律”）。[7]因此，心理学的概念和规律从法则学上关联于某种生物科学的概念和规律，而它们又进一步关联于化学，化学从法则上说又可还原为物理学——在这种意义上，“可还原于”意味着非基本话语中所说的实在和属性，**在类型上同一于**某些基本的构造。例如我们的日常概念“**水**”可还原于化学类型 H_2O，而此化学分子

又是由相同的原子排列构成的。这一模型有助于人们从直观上理解：水之存在以及关于水的科学的真理为何只与物理学本体论中存在的事物相关。

但是在各专门科学，更不用说在日常话语和诸社会科学中，并非一切概念都能符合这种模型。例如气象学中所诉诸的每一场飓风，或地理学中所提及的每一个地壳构造移动，并非都有相同的化学或物理结构。事实上，几乎无法想象任何两个事物能以这种方式同一。同样，并非每一种传染性疾病或每一种癌细胞的扩散都有相同的深层次结构，更不用说每一次通货膨胀或每一场政变。就福多来说，在他的一篇重要论文《专门科学》(1974) 中，他正确地指出：还原主义学说强调的不外是我们的科学上合理的概念必定是**自然种类**概念，像“水”一样，它们的相似性必定是根源于有关的基础。福多还认为，这样说对于我们大多数常用的解释性概念来讲是行不通的，尤其是不适用于心理科学中根据功能主义术语而理解的概念，以及用我们的世俗心理主义词汇所表述的概念。因为所有那些概念都是**可多样实现的**。这意思是说，同一类事物的不同例示在低一级的层次上——即在它们的硬件上——是相当不同的，而且只有运用来自于专门科学中的那些概念，才能看到不同例子所具有的明显相似性。尽管你可以抛弃“水”一词，进而在不丧失传递力量的情况下只说“H_2O”，但是对于“有生命的动物”、“想到埃菲尔铁塔”、“大陆漂移”等来说，你则不能这样做。

福多 (1974) 认为，这对物理主义并不构成威胁，因为一个高阶概念的每一个例示将会同一于用基本物理学术语描述的某种结构，并且只能如此。个例还原主义所说的，不过是物理主义和诸科学的统一性强调类型还原主义是多余的。下面，我想说明的是：它为什么不像表面所显现的那样是错误的。

4.3 谓词二元论为什么能使二元论成为正确的

福多的下述看法是正确的，即同一种主观材料能用不可还原的不同方式予以描述，而且如此描述之后，它仍还可以是那种主观材料。在我看来，他没有注意到的东西是：如此不同的解释构架预设了一种关于那个主观材料的观点 (perspective)，而这材料显然是在那个观点之外的。下面是我的看法的梗概。基于实在论的构想，完善的物理学能在物理实在的终极

结合处将物理实在区分开来，也可能有根据说，任何能从法则学上严格还原于物理学的专门科学，都能在它的结合点而非最小之点上利用这种还原将实在区分开来。相比较而言，一门不能从法则学上还原为物理学的科学不会以这种直接的方式从基础性实在中获得它的合理性，毋宁说，它是由下述两方面结合而成的，一方面是那个世界的客观相似性，另一方面是关于那些让那门科学产生出来的事实的观点和兴趣。如果科学实在论是对的，那么完善的物理学就会不依赖于任何专门的兴趣和关心而告诉人们：世界是怎样的，即**世界怎么**正好就是那个样子。然而，板块构造学告诉你的是：它怎样发源于对大陆发展感兴趣的观点。另外，它还从关心气候的观点出发谈论飓风和冷锋。① 要使这些结构或图式具体化，首先必须在心中根据某种**目的论**对现象作出**选择**。这就是说，这些科学及其所设定的实在从某些理智的观点来说是存在的。一种观点，不管是知觉的还是理智的，总是外在于该观点所关于的那个东西。[8]物理主义的难题是必须说明：说一种关于物理世界的观点就是它之内的某种东西究竟是何意。关于世界的统一的自然主义观点要求的是：这些科学所需要的观察者的观点一定会整合到他所观察到的世界之中。而要将观点和兴趣整合到一个世界之中，又离不开将知觉和意向类型的心理状态整合到物理世界之中。然而，这些状态类型恰恰又是抵制从法则学上——类型上——还原于物理学的典型例子。

当然，也存在着诉诸关于意识的现象学所作的各种著名论证，它们主张：个例还原是行不通的。但也应看到，在一种流行的论证中就没有用到这些东西。[9]关于心灵的个例还原主义即使可能碰到现象学难题，但它是个例而非类型这一事实意味着：它预设了这样一种观点的存在，正是从它出发，人们才看到物理世界产生出了这些事实。由于有这一观点，各门科学才可能成为法则学上不可还原的东西，由于它本身是不可还原的，因此只有从关于物理实在的观点出发，它才能存在（如果它是物理的）。既然这第二个观点与我们试图解释的观点（即心理学或理智的观点）在本质上属于同一类型，因此在这里就没有陷入非恶性后退的可能。

现在，我们就能够理解成熟的还原概念的动机了。一种真正基础性的物理学总想按世界本身所是的样子来再现世界，如果专门科学是可还原的，那么它们的本体论就不会仅被视为看待或解释世界的方式，而同时

① 即冷热空气团的分界处。——译者注

能被理解为关于物理事物的表达式，即是说，可以认为这些本体论“来自于基底”，而非来自表层。专门科学所具有的不可还原性并没有向二元论者提出什么难题，因为二元论者也承认：物理科学的解释工作是根据概念上外在于物理世界的观点而做的某种工作。**如果**物理主义者能对心理学作出还原，那么他就不会为上述工作所困，因为他能“从基底”理解这样一些行为，它们（具有它们内在的、意向的内容）促成了其他科学不可还原的本体论。但是心理学是具有最小还原可能性的科学中的一种。

4.4 关于“观点”的话语合理吗?

想抵制这一论证思路的人可能会否定下述观点，即我们不可能为法则学上不可还原的诸科学提供充分的实在论解释，因为这些科学只是关于实在的一种观点。这些抵制的人可能会论证说：专门科学的基础就是丹尼特（1991）所说的实在中的“真实模型”（real patterns），这些模型像根本的、不可还原的科学中的结构一样是客观的。

这样说显然没有抓住要领。我要否定的，不是专门科学以之为基础的“真实模型”是客观的、真实的，而是像**实在中的这个根本的东西**一样，那些科学需要的是这样一种解释性成分，它具有这些相似性，并能说明它们对某些目的是有利的。

理想的物理学与法则学上不可还原的专门科学之间的关系就像直接现象与**格式塔**现象的关系一样。物理学中的实在类似于圆形物体，后者被当成圆形是用不着额外解释的，而不可还原的专门科学中的那些实在则像一串非连续的、胡乱排列在一个圆圈上的点或标志，只有经过人的观看，这个圆圈才能被当作是圆的。例如两场飓风并非完全一样，而它们又能作为一个类别呈现于对天气感兴趣的某人面前，同样，地球的板块构造也只是由于人们对地球的可居住性感兴趣才存在的。从一种完全对立的观点看，这两类现象可以完全正确地被看作是更基本过程的副产物，而不是构成性的自然类型。世界本身是一种连续的事件流——不应该说它在每一处的结构都是一样的。把某一点当作某种过程的开始或结束，这不过是人们基于某种兴趣或兴奋点观察事物时的任意所为。

4.5 一个令人惊喜的支持者

我关于（大多数）专门科学的论述根据能从阿姆斯特朗（Armstrong，1980：vol. 2；1989）对共相的说明中引申出来。他是一个实在论者，但不是关于所有属性的实在论者，而只是关于那些为专门科学所需的属性的实在论者。人们可能会说，这是把这些属性置于专门科学之中，而我则不这么认为。真正的共相是这样的东西，它的作用是作出一种独特的因果贡献，而不是一个接一个地创造非微观的实在。这些实在对微观基础的因果作用不可能增加任何东西。不管我对于阿姆斯特朗把共相、属性的同一性与它们的因果力关联在一起的尝试持何种保留态度，我始终认为，在这一语境下，下述看法不是没有合理性的，即共相在它的特殊例示中是否有作用是真实的共相在那里是否必要的关键。阿姆斯特朗（1980：19－23；1989：82－4）还断言：不存在析取性共相。[10]这一看法也许能够增加上述看法的力度。任何不能通过桥梁规律关联于物理学的专门科学中的属性并不是严格必然的，而相应的谓词对于作为专门科学之组成部分的解释图式来说则是必然的，但是，对立于共相的谓词是人的思想和言论的产物，因此，它以关于主观材料的心理观点为前提条件。[11]

4.6 非基础层次的选择性与心理学的不可避免性

我愿对这里的问题作进一步的探讨。先来讨论下述主张：如果一个人能理解世界中一切物理的（意即物理学层面上的）细节，而不涉及更高的层次，那么可以说他没有一点遗漏。这里讨论的目的在于表明：在专门科学中，只有心理学不能没有遗漏地被省略掉。① 而这又说明：心理的东西与物理的东西之间存在着根本差异。 91

假设有一个半神半人，他能追踪物理学层面的一切事物，但没法关注事物更宏观的模式。因为他有智能，所以他能以尽可能高的准确性预言一切事物的位置。对此，我们可否认为，他没法让自己关注更宏大的样式正

① 意即将心理学抛弃了，就一定会遗漏某些实在。——译者注

好是他根本性无知的一种表现？或者说他只是不知道某些宏观的模式，因为我们不能弄清细节，因此这些宏观模式对于我们的理智是必不可少的？或者换句话说，它们对于我们来说是一种必要的速记法，但对他而言则不是这样？以为这种人从根本上说是无知的人大概会说：不能关注模式和起作用的规律正好就是无知。假设我们的半神半人能关注这些事物，但既然有按物理学的要求行事的能力，因此他又觉得对它们没有任何兴趣。这里必须论证的是：非基本层次在某一方面**对它们自己而言是很有必要的**，即是目的本身。这个问题类似于目的论解释中不可还原性的问题。假如机械论是对的，目的论解释还能做其他非探索性的事情吗？

正如丹尼特在《真正的相信者》（1987：13－42）一文中讨论过的那样，这种状况对心理学来说是至关重要的。丹尼特论证说：即使是一个无所不能的观察者，他能通过预言构成人的个别的原子行为来预言人的行为，他也不可能不用民间心理学。如果他想理解和他交谈的人所说的话，那么他就需要民间心理学，更为重要的是，他若想理解他本人所做的事情，他更没法撇开民间心理学。因此，民间心理学即使不带有任何基本的本体论碎片，但这个层面的描述仍是不可排除的。丹尼特所持看法的问题在于：他没能解释我们为什么要采取民间心理学的观点。如果我们都只是原子的堆积，那么我们为什么必须以这种特殊的、本体论上非基本的方式看待我们自己呢？诚然，我们真的不能把我们自己看作是人，除非我们采取这样的观点，或者说，我们不能**理解**我们的**行为**。但问题是我们为什么要以这种方式看待我们自己？取消论大概会说：这不过是一种概念保守主义。但是，如果人们拒绝这样的观点，即我们正好被授予了这种看待我们自己的方式，并承认这些范畴的适用性真的带有根本性，那么就必须解释为什么是这样。还原论者认为，这个层次的陈述可能是真的，因为它们是可还原的。但这个事实并不能解释以下问题：在各种可能的非基本的话语层次中，这个层次为什么是不可避免的，为什么不是在必要时才权宜一用的东西？可能有这样的说法，即“我们为什么应把我们自己看作人”这个问题本身就给出了回答，因为“我们”一词的运用已预设了人的观点。但这仍未抓住关键，我们称作“人”的那种物理构造的行为不可能用一种似乎完全（或极其充分）但又缺乏人的观点的方式来理解。从物理主义方面
92 说，在某一特定的层面，不应该存在着“我们”。但即使一个人想以物理主义方式思维，他也没法回避这样的想法：在某一复杂的层次上突现出了某种事物，它既不是一种能从外部以某种方式（否则就会无穷后退）

看待或解释有机体的物质，也不是人们要么可能注意要么可能忽视的那些复杂层次中的一个层次。它在那里出现了，其方式完全不同于别的复杂的（因为别的形式要么是可解释的，要么是不可知的，要么是两者兼备）某物的实现方式，这某物是不容易说明的，因为人们对之可能一无所知。这至少包含着笛卡尔在说**“我思故我在”**时所表达的思想的种子。

事实是，即使我们能作出这样的预言，它可由物理学的无所不知所使然，但我们也没法把人们的理解限制在物理事项的范围之内。我在**思想**，这一笛卡尔主义的事实是绝对的，而不是相对于采取了一种可能的（但又像其他的东西一样的）选择性的话语层次而言的。我们在人格层次上的存在不是实用主义的事实，而是根本的事实。想把它看作是思维或概念化的某种方式的功能是不可能的，因为从更加实用主义的方式上说，它是不可避免的，就此而言，它是基本事实。如果物理主义的本体论是正确的，那么它就不可能是基本的。

4.7 为什么捆绑二元论行不通？

如果上面所说的是对的，那么心灵一定超越于物理世界，而且**事实上**是非物理的。当然，这尚未回答它究竟是一种实体还是仅为状态的集合。我接下来要论证的是：捆绑二元论是不够的，因为这样会使心灵成为一种复合的实在，而只有把它设想成简单的，我们才能涵盖关于人格同一性的某些不可避免的直觉。

主张人格超越时间的同一性不像别的（复合）实体的同一性那样，可以有约定或程度的问题，这有悠久的历史传统，至少可以上溯至里德。当然，对这些论点及其所依赖的直觉的批评从休谟一直延续至帕菲特（D. Parfit）。这种批评让我们感觉到：诸直觉之间存在着无法公断的冲突。我的论证不涉及跨时间的同一性，而只关心关于起源的某些反事实的同一性的各种结论。因此我希望，我的论证能打破围绕历时性的同一性争论所陷入的僵局。我的看法是：尽管各种公开的约定论方式常被用来解决人和物质对象的跨时间难题，进而用于关于肉体之起源的反事实事例之上，但它们不能用于关于人或心灵的近似的反事实之上。

当今，主张个体有其本质属性的观点是值得重视的，尽管并非大家都

承认个体有其本质。克里普克指出：一张用木头制成的特定桌子不可能是用冰制成的。这一看法似乎得到了广泛的认可。对于那些个别桌子的存在来说，至少存着一种必要条件，但对于是否有使该对象是其所是的充分且
93 必要的条件——本质，研究者则莫衷一是（克里普克，1980：39-53）。即使承认那桌子有某些必不可少的属性，但它是否有一种本质则是大有疑问的。我们不妨把有关的句子排列如下：

1. 这张桌子可以用冰做成。
2. 这张桌子可以用别的不同的木材做成。
3. 这张桌子可以用该桌子所用的95%的木材加上5%的其他种类木材做成。

按照这里从（1）到（2）再到（3）的顺序前进，将来到这样的点上，在这里，不可能得到关于下述问题的明确答案，即这张被假想的桌子是否等同于那张现实存在的桌子？因为这张假想的桌子与现实的这张桌子是否“真的”是同一张桌子似乎没有清楚明白的意义，比如说，它其实是用75%的相同材料加上25%的不同材料构成的。在这个事例中，这些才是唯一真的事实；因此，关于它在数量上的同一性问题只能用任何方便的方式去裁决，否则就将成为不可解决的问题。这样一来，我们就来到了反事实事例的边缘，在这里，两个事物是否具有同一性就不是事实问题了。

假设有一个人类个体，他所拥有的血统不同于他事实上所具有的血统，这样一来，相对于直觉来说，这一不同是否影响**他是谁**就成了晦暗不明的问题。这样的例子究竟能说明什么？这可能是有争论的问题，但是一定有这样的事例。如果存在着与琼斯的身体一样的配对物，它来自于相同的卵子，但也有来自同一父亲的不同精子（尽管在遗传上相同），那么拥有如此身体的这个人是不是琼斯也许是一个说不清的问题。有些哲学家可能会认为，有相同的精子是人体具有同一性（进而人格具有同一性）的必要条件，这是自明的。如果是这样，那么可进一步设想：有一个精子的复本，其中的某些分子是不同的。那么这是同一个精子吗？如果人们进一步思索下去，那么就会发现这样一种不确定性，正是它使已产生的身体也具有不确定性，因此一定存在着某种差异。不过自然语言和直觉都无法告诉我们：这个差异是否改变了那个人的身体的同一性。归根结底，我们是否具有相同的身体这一问题并不是一个事实问题。

从某些方面说，人们应怎样描述这些事例的确是有争论的问题。某些哲学家认为，人们可以谈论**大致的同一性**或**部分的同一性**，而别的哲学家则断言：这样的谈论是胡说八道。我这里没有篇幅来讨论这一问题。不过，我想说：一个人被允许怎样使用同一性概念只对人们想描述这些事例的主意有影响，而不会影响存在的事实。有一些事例，它们的结构真的有部分一致，在此结构中，**那个**事实是该事例中唯一的基础性事实，因为关于是否“真的”有相同的对象，再也没有进一步的事实。因此如果是这样的话，那么一定存在着一种**个性**（haecceitas）或**“这一个”**（thisness），它属于每个复杂的物理对象，并将其个体化。即使它确有合理性，我也不认为它是可信的。（关于**个性**由以得到理解的条件，详后。）我的观点是：结构中近似部分的重叠完全不适用于心灵的反事实的同一性，用马德尔（G. Madell）的话说就是：“即使我当下的身体在某个可能世界有它的部分的复制品，我当下的意识也不能是这样。我能想象的任何当下意识状态要么是我的，要么不是。这里不存在程度问题。”（1981：xx）[12]

为什么是这样？试想那个著名的事例，在那里，我们不能肯定那个身体是不是琼斯的身体，是不是琼斯其人，因为这身体是由略为有点变化的精子和相同的卵子所构成的。对它，我们能像对于一个没有意识的对象所说的那样说吗？关于**某物相同**、**某物不同**的说明就是要述说的全部，即结构上的部分一致就是关于该物要说的一切？就琼斯的身体本身而言，这个方案与关于任何别的物理对象的方案所说的是一样的。但假设琼斯用反省的方式问自己：“如果那发生了，那么我是存在的吗？”他能作出的回答至少有三个：

（1）“我要么存在，要么不存在，但我又无法判断。”

（2）“对于我是存在的还是不存在的问题，没有事实的参照，它不过是错误提出的问题。”

（3）“在某些方面或在某种程度上，我是存在的；在某些方面或在某种程度上，我是不存在的。那个可能存在着的造物与我的心灵结构有部分的一致。”

第三个回答类似于我在讨论身体案例时所给出的回答。但是如果把它作为一种关于主观状况的说明，那是没有任何意义的。我们不妨把从有轻微变化的精子中突现出来的那个造物称作“琼斯 2”。部分重叠论难道不是这样的主张吗，即正像“琼斯 2”身体的 85％同一于琼斯一样，他的心理

生活的大约85%也同一于琼斯的心理生活？尽管断言他的心理生活像琼斯——事实上，“琼斯2”可能有像琼斯一样100%的心理生活——是完全合理的，但说他们的一致达到那样的程度，如有相同的心灵，或说他有琼斯85%的心理生活，那则是胡说。假如琼斯和“琼斯2”有完全**实际**相似的生活，那么他们具有的100%中的85%的相似的心理生活究竟是哪一些呢？显然没有理由说：琼斯可以享有“琼斯2”的全部心理生活，但这是以大约85%的方式发生的。很显然，数量上同一的心理方面的重叠这一概念是不能使用的，而使用实际的身体构造的部分重叠这一概念则是没有什么问题的。

这可能让人想到第二种回答。我们可以把“部分重叠”一词用之于琼斯的身体，但心灵或主体是否相同这一问题则没有明确的意义。至于为什么没有意义，那则是没法回答的。假如琼斯认识到他原本是孪生人中的一个，这意思是说，他由之而来的受精卵分开来了，但另外一半不久就死了。他可能产生这样的想法，即如果他是他已死了的那一半，那么他绝不会作为一种有意识的生命而存在，即使某人以及他的生命从内到外都非常
95 相似于他自己。若他是那已逝的另一半，他也许会内疚地感激。如果以为琼斯错误地认为存在着这方面的事实，那是不可思议的。人们该怎样“设法”从有事实根据的事例转向没有事实根据的事例呢？

留给我们的只有第一种选择。从主观的观点看，一定存在着一种绝对的事实。但我们考察过的物理案例却表明：当某物在本质上是复合的时候，事实不可能是上述那个样子。当存在着结构时，结构的部分重叠又不可避免地具有可能性，因此心灵一定是简单的。而只有当心灵是像笛卡尔所说的实体一样的某物时，这才是可能的。

4.8 关于这个结论的两点思考

琼斯没法想象自己与“琼斯2”之间存在的关系。这一点与人格同一性中其他更传统的问题是不同的。而第一反思要探讨的正是这种差异。不同于别的事例，琼斯这个例子与人们所说的**移情距离**（empathetic distance）无关。

假设我的父母移民到了中国，当时我的母亲正怀着我，我生下来不久，父母便离世了。接着，我为中国的养父母收养了，经历了那场革命运

动，最后以毛泽东时代在中国的任何一个外国人都会碰到的培养方式而结束了被抚养的生活。这个人[①]离开子宫后的经验与我的经验没有一点是相同的。但是，一方面这个人似乎显然就是我，另一方面又没法弄清我能够感觉到的与这另一个“我”的移情联系究竟是什么。如果我像琼斯那样问：“这是我吗?”我将在下述两个信念之间摇摆，一是相信，像通常所说的那样，他显然就是我；二是完全不能让自己进入到我以前所处的状态。这种**移情的缺失**在许多旨在怀疑人格同一性之绝对性的理论中发挥着重要的作用。看来，必须作这样的尝试，即对我经历了50年的时间变迁之后是否仍是同一个人，或者说对大脑的损伤是否使我在远非隐喻的意义上变成了“另一个不同的人”投以怀疑的眼光。当然，现实中也有清楚分明的地方，例如某物的确可能存在着程度的问题，因为某些差异性比别的差异性更能移情地予以设想。在所有这些情况之下，我们的直觉对同一性不具有决定性的作用。一个重要的事实是：移情的种种问题在我的论证中并没有什么地位。那个可能作为我的替身活着的孪生地球人，或那个在精子发生了轻微变化的情况下仍生存下来的人，如你喜欢设想的那样，可能享有与我实际相似的心理生活。这足以表明：我所讨论的案例与那些包含有跨时间同一性且成问题的案例之间是存在着差异的。在我的事例中，“相似但不那么相同”这一观念得到了经验的支撑。在他的记忆中，未来的我尽管没有感觉到我现在感觉到的全部，但仍会感觉到很多。而在别的那些成问题的事例中，有意识结构的部分一致显然是合理的，但在反事实的事例中，想象的或移情的距离并没有任何根本性的作用，而且完全置与同一性 *96*
伴随始终的相对性于不顾。

第二，我认为，这个论证可得到它有办法澄清的**个性**概念的支持。就复合的物理性身体来说，要设想**个性**根源于什么，或它怎样关联于该对象的其他特征，那是不可能的，因此，断言身体中存在着这种东西就是纯神秘的叫卖。相比而言，我们在心灵中确实能看到**个性**的一种形式，例如在一定意义之下我们都能理解的主观性。正是因为我们直觉地理解了这种主观性，我们才会感觉到：我们可以清楚地说明，如果某东西碰巧出现了，那么它将正好出现于我们自己之上，否则它就不会出现。而且我们还会感觉到：我们能够理解各种纯粹的反事实，例如我变成了一个古希腊人，甚或一个非人。然而当这些反事实被用于下述纯粹的物体时，例如说这张木

① 即养父母所收养的中国的“我”。——译者注

制的桌子是在那个角落里的另一张桌子，甚或是一个金字塔时，就没有任何直觉的意义。可能有这样的看法，即认为断言我的心灵在另一身体之中，从根本上说是没有任何意义的，当然有表面的意义——似乎是有内容的。就此而言，关于纯粹物体且近似的断言连表面的意义也没有。可见，关于主体的反事实似乎是有意义的，正是这一事实表明：存在着某种不会出现在别的事例中的东西，它正是在**个性**中可看到且能发挥功能作用的东西。只有以笛卡尔的方式理解意识，我们才能对**个性**作出经验的阐释。

4.9 一种质疑

对于上述论证，有的人有时可能会觉得，它作为一种关于我们的心灵**概念**的说明是正确的，但作为关于心灵的实际本质的说明则是不正确的。[13]这就是说，事实解构了我们所持的这一心灵概念。因此我们的概念图式只允许我们承诺笛卡尔式的心灵观。但是我们基于别的根据认为，这一看法是错误的。就现在的情形而言，这与其说是一个精心设计的反对意见，倒不如说是一种令人不安的说法。我将对它的两种阐释方法作出考释。

第一种形式可这样予以表述。如果心灵只是由一种共意识关系关联在一起的心理状态集合，那么现象学对我们而言似乎就是它事实所是的那个样子。因此，我的论证并未证明捆绑理论是错误的，因为即使它是真的，但就我们而言，我们仿佛就成了简单的实体。这与“硬（hard）决定论者”关于自由意志所说的话如出一辙，即是说，我们不可避免地会感觉到我们有自由意志，但此感觉却是错误的。[14]这一论证有两个问题。第一，这对琼斯回答他的问题一点儿用也没有。为了避开问题（1）——他与“琼斯
97 2”要么同一，要么不同一——他必须解释另一些选择中的某一个，这个驳难使他无法完成那个任务。这里的意思是这样的吗，即当琼斯设法想象心理结构的部分一致时，我们的概念妨碍他如此设想，但事实上，这样的事情却是可能的？如果是这样，那么我认为，这是没有道理的。在我看来，真的事实没法被证明。我对前述的另一选择（即不存在那样的事实）所提出的反对意见似也没有什么问题。第二，这个论证未免太自恃、太武断了。捆绑论者与实体主义者面对的悬而未决的问题是：是否存在着让心灵成为经验上的统一体的共意识关系。我的论证赞成的观点是，经验统一体

只与简单实体有关，因此我也赞成这样的观点，即不存在像自立的共意识关系那样的东西。如果捆绑论是对的，如果那种状况事实上是一种不可能的状况，那么就不能简单地断言它对我们来说是相同的。与自由意志的类比尽管有助于说明反对者想达到的结论，但几乎没法证明他是正确的。首先，硬决定者立场的一致性是有争论的。其次，对于我们为什么**一定**能感觉到自由，决定论者能提供说明的理由认为，用对某人自己行为的一种纯粹**预言**来替代他自己的**抉择**实践，在概念上是不可能的。反事实的同一性似乎要么存在，要么不存在，为什么会这样？我们找不到近似的解释。既然在基于移情距离的事例中似乎可能有程度的问题，那么这便显得特别神秘。[15]

有一种完全不同的方式可以完善这种驳难。它涉及我对反事实的利用。反事实是有争论的问题，我不打算讨论它们。我敢明确地说：刘易斯对复制品的分析是虚妄的，因为如果琼斯的问题（即他是否存在）只是想弄清是否存在着具有非常相似于他的状态的复制品，那么就没有什么现象学问题。严格说来，所有的复制品都是独特的对象。不过，我与其他几乎所有哲学家一样，为直接否定了刘易斯的理论而深感欣慰。但是那个责难不是出自这里，而是来自于对反事实采取了非实在论态度的其他研究者。有这样的经验论传统，即否认反事实有真值，并主张它们表示的只是一种策略或态度。如果有关的精子有微妙的不同，那么对于未来将发生的事情来说，我们是找不到什么真实的东西的。

这里不可能对反事实作出深入的讨论，我只能说两点。第一，大多数哲学家都赞成关于反事实的实在论，即是说，反实在论的观点并不那么有理，这个论点已为他们接受。第二，反实在论方案有强和弱两种形式。弱版本只否认反事实有真值：如果 c 出现了，那么 a 或 b 是否会发生，这件事是不存在什么事实根据的。c 有可能出现，且如果有 c，那么要么出现 a，要么出现 b（或某种别的东西），从关于该观点所关心的现实世界的观点来看，这里是不存在任何事实的。而它对我们的论证没有丝毫影响，因为对于什么会发生，我的论证只要求这样的唯一回答，即要么是全部，要么一点儿也不要，而用不着有关它的事实。更强的版本则主张：**“可能是别的样子”**这一概念是我们的思维方式（即我们设想事物的能力）的一种投射，而不是真实存在的某东西。这不是说——像它似乎所是的那样——现实世界是必然的（因为没有别的可能出现的东西），而只是说：这些模态范畴不过是纯粹的投射。即使我们赞成这一点——我并不推荐赞成它——

但这完全无损于我的论证。关于心灵的本质，它还是说出了某种有趣的东西，例如在我们通过投射创造的关于可能性的逻辑空间中，要想以对待物体的方式对待心灵，那是没有任何意义的。我们创造了那个空间这一事实并不意味着：我们在它之内所述说的东西没有反映我们正在谈论的对象之间的真实差异。

4.10 结　论

我在本部分的论证仿用的是笛卡尔的风格。首先，在4.2—4.6节中，我论证说：能思主体一定超越于它（除了做别的以外）所思考的物理世界之上。只有当强还原论为真时，能思主体思维才可能成为那个物理世界的组成部分。接着在4.7—4.9节中，我论证了能思主体一定是简单实体的观点，若不这样看，就会碰到矛盾的反事实。这些论点是相辅相成的，但它们又有逻辑上的独立性，而且第二个论证能独立支撑它的结论。

【注释】

[1] 在这里，我不打算对这些属性作进一步更充分的讨论，因为这主要是唯物主义应关注的问题。关于这些属性，我有详细的解释，关于唯物主义为解决这些问题而提出的策略，我也有简要的说明，请参阅鲁宾逊（1999）。

[2] 笛卡尔的“第六沉思”就是阐释实体二元论的**经典文献**。该理论的现代辩护可看：波普尔和埃克尔斯（1977）、斯温伯恩（Swinburne，1986）以及福斯特（1991）。休谟在他的《人性论》（第1册第4部分第6节）中阐发了他的理论，并对他在该书附录中所述的内容发表了不满言论。有几位现代哲学家根据心理事件之间的关系说明了心灵的统一性，因此可以认为他们倾向于捆绑理论，但他们并不愿意成为二元论者，如帕菲特（1971、1984）是唯物论者，而丹顿（Dainton，2000）对本体论问题持中立态度。

[3] 可以这样理解：属性理论已根据特定捆绑的所有要素对同一大脑的依赖性以及对心灵的统一性作了说明，虽然这可看作是一种对统一性的因果解释，但却不是对统一性的分析。因为仅仅有对同一大脑的依赖性并不能保证意识的有统一性。参阅福斯特（1968）对艾耶尔（Ayer，1963）的回应。

[4] 关于种种质疑，可参阅休谟的《人性论》附录。

99 [5] 具身性（embodiment）即使不是专门的相互作用问题，但我也作过讨论，可参阅鲁宾逊（1989）。

[6] 翻译还原论的事例可参阅亨普尔（1980）和卡尔纳普（1934）。

[7] 这方面的经典资料可参阅恩斯特·内格尔（1961）。

[8] 自从斯金纳承认刺激反应模式无效，进而阐发操作条件反射作用概念之后，心理学中真正的还原主义就开始了它退化的历史。刺激反应模式只需要机械论因果概念，而操作条件反射这一概念则是不可还原的目的论概念。对于学会了怎样得到食物的老鼠的行为，尽管可作出低层次的机械论描述，但把它理解成操作条件反射则一定是目的论的，因为这种描述关心的是老鼠想得到什么，或关心的是老鼠的行为目的。更有甚者，由于行为主义者对发生在事物内部的事情不太感兴趣，因而没法得到关于让学习得以实现的过程之本质的见解。当把这用到心灵哲学之中时，便使还原概念变得模糊不清。心灵哲学的首要目标是消除“机器中的幽灵”——即不可还原的私人或主观的东西。这种还原形式与任何物理科学完全无关。第二种目标是铲除或消解在纯物理科学中没有任何地位的类型概念。操作条件反射能满足第一个要求，而无法满足第二个要求。不同于福多，我在本部分要论证的挑战性的观点是：对于物理主义者来说，第二个目标像第一个一样具有根本性。

[9] 这些东西就是经常被讨论的反物理主义的感受性质责难。例如可参阅杰克逊（1982）、鲁宾逊（1993）。

[10] 阿姆斯特朗赞成合取性共相也有利于下述直觉，即强还原是支持关于专门科学的完全的实在论的。水就是关于某种空间关系中的氧和氢的共相的例示之合取。而这些共相又是更加原子性共相的进一步合取。

[11] 由此自然可得出结论说：如果心理学（它不仅包括科学，而且还包括我们的日常心理主义概念）在一种强意义上是不能还原的，那么它的“属性”就只能是谓词，它的主观材料在一定程度上是由心灵的行为创造的，而这样的心灵又只有在行为完成时才出现。阿姆斯特朗的理论不同于丹尼特的解释理论，因为它除了意图不同之外，还伴有倒退的威胁。

[12] 就我在这里所讨论的课题而言，马德尔的著作是最好的文本。

[13] 布莱克本（S. Blackburn）、帕菲特和法卡斯（K. Farkas）在不同场合对我提出了这一责难。值得注意的是，这个反对意见包含着重要的让步。如果我提出的论证表明：我们在思考我们自己时承诺了笛卡尔关于自我的概念，那么这不是由于某种易于修改的定义。这个论证并不是从关于成为一个主体的某些充分必要条件的逻辑结论中派生出来的，而留有改变那些条件的余地。它的程序是以对有意识主体来说可设想的东西为基础的，而关于与自我有联系的概念不可避免地要运用“康德式”的方式，因此断言它是错误的，不过是一种怀疑主义的虚无主义。而虚无主义不外是主张：我们只能甘于无知、苟且偷生。

[14] 我认为，与自由意志的比较应归功于 K. 法卡斯。

[15] 我在 4.6—4.9 节中陈述的是这种论证的一种更复杂版本，它也抵制这里的驳难。我相信可以作出这样的论证：模糊的谓词不会是本体论上的基本谓词，原则上是 100

可予排除的。反事实条件下的关于所有类型物体的同一性概念就是其中的成员。下述两个论断事实上真的是没有区别的：一个论断认为，如果某事是不同的，那么某种物体就是存在的；另一论断主张，在那些条件下，存在着相同类型的“复制”物体。即使反事实的变化不直接涉及这里所说的对象，但这样说仍是适用的。对于主体来说，这种看法完全是不可行的。假设与事实相反，某人正好在被设想之前说出了那个世界的另一面。基于适用于物体的那一原则，下述两个命题之间就不存在事实上的差别，一个命题是：你仍然是存在的。另一命题是：具有同你一样性质的某人是存在的。正像孪生人的事例所表明的那样，这种差别是真实的，因而不可排除。这个论证的完善版本已无法找到，但某些相关的，涉及模糊性的问题的讨论，则可参阅鲁宾逊(2001)。

参考文献

Armstrong, D. M. (1980). *Universals and Scientific Realism* (2 vols). Cambridge: Cambridge University Press.

—— (1989). *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press.

Ayer, A. J. (1963). “The Concept of a Person.” In *The Concept of a Person and other Essays*. London: Macmillan: 82–128.

Carnap, R. (1934). *The Unity of Science*. London: Kegan Paul.

Dainton, B. (2000). *Stream of Consciousness*. London: Routledge.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

—— (1991). “Real Patterns.” *Journal of Philosophy*, 89: 27–51.

Descartes, R. (1984—1985). *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothof, and D. Murdoch (2 vols). Cambridge: Cambridge University Press.

Fodor, J. (1974). “Special Sciences or the Disunity of Science as a Working Hypothesis.” *Synthese*, 28: 77–115.

Foster, J. (1968). “Psychophysical Causal Relations.” *American Philosophical Quarterly*, 5.

—— (1991). *The Immaterial Self*. London: Routledge.

Hempel, C. G. (1980). “The Logical Analysis of Psychology.” In N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1. London: Methuen: 14–23. (Originally published in French in 1935.)

Hume, D. (1978). *A Treatise on Human Nature*, ed. P. H. Nidditch. Oxford: Clarendon Press.

Jackson, F. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–136.

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.

Madell, G. (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.

Nagel, E. (1961). *The Structure of Science*. London: Routledge and Kegan Paul.

Parfit, D. (1971). "Personal Identity." *Philosophical Review*, 80: 3–27.

—— (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Popper, K. R. and Eccles, J. C. (1977). *The Self and its Brain*. Berlin: Springer International.

Robinson, H. (1989). "A Dualist Account of Embodiment." In Smythies and Beloff (1989): 43–57.

—— (1993). "The Anti-materialist Strategy and the 'Knowledge Argument'." In H. Robinson (ed.), *Objections to Physicalism*. Oxford: Clarendon Press: 159–183.

—— (1999). "Materialism and the Mind-Body Problem." In E. Graig (ed.), *The Routledge Encyclopaedia of Philosophy*. London: Routledge.

—— (2001). "Vagueness, Realism, Language and Thought." In T. Horgan and M. Potric (eds.), *Essays on Vagueness*. Oxford: Oxford University Press.

Smythies, J. R. and Beloff, J. (eds.) (1989). *The Case for Dualism*. Charlottesville: University of Virginia Press.

Swinburne, R. (1986). *The Evolution of the Soul*. Oxford: Clarendon Press.

第5章　意识及其在自然中的位置

大卫·J·查默斯（David J. Chalmers）

5.1 导　言[1]

102 意识是不容易与我们关于自然界的概念融为一体的。根据最常见的自然概念，自然界就是物理世界。但是根据最常见的意识概念，要理解它怎么能成为物理世界的一部分是不容易的。因此，要想为意识在自然秩序中找到一个位置，我们似乎必须要么修改我们的意识概念，要么修改我们的自然概念。

在20世纪哲学中，这种困境首先是被C. D. 布罗德在《心灵及其在自然中的位置》（*The Mind and its Place in Nature*，1925）一书中最尖锐地提出来的。对布罗德来说，心灵现象就是意识现象，核心问题是确定心灵在物理世界中的位置。布罗德关于这个问题的详尽讨论，在对有关心理/物理关系的17种不同观点进行分类时达到了高潮。[2]根据布罗德的分类，一种观点有可能认为心理的东西是非存在的（“虚妄的”），是能还原的，是突现的，或者说是实体的一种基本属性（一种“划分”[differentiating]属性）。物理的东西同样能用这四种方式之一来看待，于是就产生了16种观点。（第17种观点来自布罗德对实体/实体观的划分，他根据的是涉及一个还是两个实体。）最后，还有三种观点：根据这些观点，心性要么是一种物理实体要么是一种中性实体（neutral substance）的突现性质，在后一种情况下，物理的东西可能要么是突现的，要么是虚妄的。

在这一章我接受布罗德的提示，用分而治之的策略来处理意识问题。不过我不采用布罗德的范畴：自20世纪20年代以来，我们对心身问题的理解已经有了进步，而且高兴的是，我们对关键问题有了更好的理解。在我
103 看来，在意识的形而上学方面，最重要的观点几乎能无一遗漏地被分成六

类，我将用从“A类”到“F类”来表示。其中的三类（从A到C）大体包括了还原的观点，即认为意识是一种物理过程，它不要求对物理本体论作出拓展。另外三类（从D到F）大体包括了非还原的观点，根据这些观点，意识本质上涉及到不能还原的东西，因此需要拓展或重新设想物理的本体论。

由于这里的讨论旨在概述形而上学的全景图，因此将会停留在抽象的层面。我不是要从事经验的意识科学或者提出关于意识的具体哲学理论，而是要考察诸意识理论所属的一些一般类型。我在讨论中并不自称是中立的。我认为每种还原的观点都不正确，而每种非还原的观点都有前途。因此，这一章的前半部分可以看成是反驳还原的意识观的一种延伸论证，而后半部分可以看成是对我们由此走向何处的一种探讨。

5.2 问　题

“意识”这个词有很多不同的使用方式。它有时候被用于指辨别刺激、报告信息、监控内部状态或者控制行为的能力。我们可以认为这些现象提出了意识的“容易问题”（easy problems），它们都是重要的现象，对它们还有很多疑问没有解决，但是，对它们作出解释的问题是难题（puzzle）而不是谜题（mystery）。这种观念，即一个物理系统可能在这些意义上是“有意识的”，原则上似乎并没有什么深奥的问题，最终用神经生物学或者计算的术语来解释这些现象也没有明显的阻碍。

意识的困难问题是关于经验的问题。人类有主观的经验：有成为他们的某种感觉，当有成为某个存在者的某种感觉时，我们就可以说：在这个意义上，这种存在物是有意识的——或者说是有现象的意识。当处于一个心理状态有某种感觉时，这个心理状态就是意识状态。意识状态包括知觉经验状态、躯体感觉、心理意象、情绪体验、正在发生的思想，等等。看见一个翠绿的东西、感到剧烈的疼痛、想象埃菲尔铁塔、感到深深的惋惜以及认为某人迟到了，这些都有某种感觉（there is something it is like）。其中每种状态都有一种**现象的特征**，都有**现象的属性**（或感受性质），这些属性描述了处于这种状态下是什么感觉（what it is like to be in the stare）。[3]

毫无疑问，经验与像大脑这样系统中的物理过程是密切相联的。看起来是物理过程产生了经验，这至少是说：它产生了这样一种具有适当物理

属性的物理系统（如大脑），这种系统必然会产生相应的经验状态。但是物理过程怎样以及为什么会产生经验？为什么这些过程不是“秘密”（in the dark）发生的（即未伴随任何经验状态）？这就是意识的核心之谜。

什么使容易问题是容易的？就这些问题来说，任务是要解释某些行为或认知的功能：即是说，解释某种因果作用是怎样在认知系统并最终在行为的产生中发挥出来的。要解释这种功能的执行，只需要说明起相关作用的机制就行了。我们有正当的理由相信神经或计算的机制能够起这些作用。

什么使困难问题是困难的？这里的任务不是要解释行为或认知的功能：即使解释了与意识相邻的所有相关功能——辨别、整合、存取、报告、控制，仍会留下另外一个问题：为什么这些功能的执行会伴随着经验？因此，困难问题似乎是一类不同的问题，需要作出不同的回答。

对困难问题的回答包含着对物理过程与意识之间关系的解释，它要根据自然的原则对物理过程怎样以及为什么与经验状态相联系作出解释。有关意识的**还原解释**对此的解释完全是依据本身根本不借助意识的物理原则。[4]**唯物主义**（或物理主义）的回答是这样的：根据它，意识本身被看成了一种物理过程。**非唯物主义**（或非物理主义）的回答则是这样的：根据它，意识被看成是非物理的（即使是与物理过程密切相关的）。**非还原的**回答是这样的：根据它，意识（或者涉及意识的原则）被看成解释的一个基本成分。

人们自然希望对困难问题有一种唯物主义的回答，对意识有一种还原的解释，就像对别的很多领域中的现象有还原的解释一样。但是，意识似乎抵制唯物主义的解释，这与别的现象是不同的。这种抵制可以被囊括于下面概括的三个相关的反唯物主义论证之中。

5.3 反唯物主义论证

5.3.1 解释论证[5]

第一个论证以容易问题与困难问题之间的不同为基础。如上所述，容易问题涉及的是对行为功能和认知功能的解释，但困难问题并非如此。有人可能认为，根据物理解释的特征，物理解释只解释结构和功能，这里所指的相关结构是时空结构，相关的功能是在产生系统行为时的因果作用。

有人可能像上述解释那样认为：解释了结构和功能并不足以解释意识。倘若如此，就没有一种物理解释能解释意识。

我们可以把这个论证称为**解释论证**（explanatory argument）：

（1）物理的解释至多能解释结构和功能。

（2）解释了结构和功能并不足以解释意识。

——

（3）没有一种物理解释能够解释意识。

如果这个论证正确，那么虽然物理解释能解决容易问题（它们只涉及解释功能），但要解决困难问题还需要更多的东西。对于意识似乎还没有一种还原的解释会成功。如果我们加上这个前提，即不能从物理上作出解释的东西本身就不是物理的东西（这可以被看成是解释论证的另一个结论），那么唯物主义在意识问题上就是错误的，因为自然界不仅仅包括物理世界。

这个论证当然是有争议的。但是在考察各种回应方式之前，考察两个密切相关的论证是有好处的，这两个论证的目的也是要证明唯物主义在意识问题上是错误的。

5.3.2 可想象性论证[6]

根据这个论证，可以想象有一个系统，它在物理方面与有意识的存在物相同一，但它至少缺少这个存在物的一些意识状态。这个系统或许是一个**怪人**（zombie）：即它在物理方面与一个有意识的存在物相同一，但根本没有意识。这个系统也可能是**经验颠倒的**（invert），即最初具有的一些经验被其他的不同经验替换了，或者说是一个**局部的怪人**（partial zombie），即缺少某些经验，这个系统还可能是这些的组合物。从第三人称的观点看，这些系统与正常的有意识的存在物是相同一的：特别是它们的大脑过程与最初存在物的每个分子都是相同的，它们的行为也难以区别。但是，从第一人称的观点看就会有区别了。成为一个经验颠倒的人或一个局部的怪人的感觉，与成为正常人的感觉是不同的。成为一个怪人没有任何感觉。

几乎没有理由相信怪人会存在于现实世界中。但是，很多人认为他们至少是可以想象的：我们可以前后一致地想象怪人，而且再三思考也不会发现这种想法有什么矛盾。在对这种想法进行扩展时，很多人认为它同样

适用于**怪人世界**：即这样一个世界，它在物理方面与我们的世界是同一的，但这个世界中没有意识。类似的看法也适用于感觉颠倒的人以及其他的复制品。

从怪人的可想象性出发，这个论证的支持者推出了怪人的**形而上学的可能性**。怪人或许不是自然可能的：他们或许不能存在于我们的世界中，或者遵循这个世界的自然规律。但是这个论证认为，怪人**本可能**（could have）存在，或许是存在于一个迥然不同的世界中。例如，人们有时提出，如果上帝进行选择的话，他本可能创造一个怪人世界。由此可以推出意识一定是非物理的。如果有一个在形而上学上可能的世界，它在物理方面同一于我们的世界但却没有意识，那么意识一定是我们世界的另一种非物理成分。如果上帝本来可能创造一个怪人世界，那么（如克里普克所说）在创造了我们世界里的物理过程之后，为了保证这个世界有意识，他就必须做更多的工作。

我们可以用最简单的形式将这个论证陈述如下：

（1）存在怪人，这是可以想象的。

（2）如果存在怪人是可以想象的，那么存在怪人就是在形而上学上可能的。

（3）如果存在怪人在形而上学上是可能的，那么意识就是非物理的。

——

（4）意识是非物理的。

这个论证更一般、更准确的版本是借助于P和Q，P是关于世界的所有微观物理真理的合取，Q是关于世界的任意的现象真理。（这里，“∧”表示“并”，“¬”表示“非”。）

（1）P∧¬Q是可以想象的。

（2）如果P∧¬Q是可以想象的，那么P∧¬Q在形而上学上就是可能的。

（3）如果P∧¬Q在形而上学上是可能的，那么唯物主义就是错误的。

——

（4）唯物主义是错误的。

5.3.3 知识论证[7]

根据知识论证（knowledge argument），有些关于意识的事实不能从物理事实演绎而来。有的人可能知道所有的物理事实，他是精通推理的人，却不能基于此知道所有关于意识的事实。

弗兰克·杰克逊（Frank Jackson）对这个论证的权威版本给出了生动的说明。根据这个版本，玛丽是一位神经科学家，她知道与颜色视觉有关的一切物理过程。但是玛丽是在一间黑白房间里长大的（根据另一个版本，她是色盲[8]），她从未经验过红色。因此，尽管她有所有知识，但似乎一些关于颜色视觉非常重要的东西是玛丽所不知道的：她不知道看见红色是什么感觉。甚至完备的物理知识和无限的演绎能力也不能使她知道这种东西。后来，如果她第一次经验红色，她就会学到一种此前不知道的新事实：她会学到看见红色的感觉。

杰克逊的论证版本可以陈述如下（这里的前提涉及玛丽还未经验红色时的知识）：

（1）玛丽知道所有物理事实。

（2）玛丽不知道所有事实。

——

（3）物理事实并未穷尽所有事实。

可以对知识论证作更一般的陈述：

（1）有些关于意识的真理不能从物理的真理中演绎出来。

（2）如果有些关于意识的真理不能从物理的真理中演绎出来，那么唯物主义就是错误的。

——

（3）唯物主义是错误的。

5.3.4 诸论证的外观

这三种论证是密切相关的。它们都是首先确定物理领域与现象领域之间的一个**认识性鸿沟**（epistemic gap）。每种都否认这两个领域之间有某种密切的认识性关系——它涉及我们能知道、想象或能解释的东西的关系。特别是其中的每一种都否认某种从物理真理 P 到现象真理 Q 的**认识性蕴含**（epistemic entailment）：从 P 能演绎出 Q，或者说能依据 P 来解释 Q，或者基于反思性想象 P 来想象 Q。

可能最基本的认识性蕴含是先验蕴含或**暗示**（implication）。根据这个概念，当实质条件句 P ⊐ Q 是先验的时候，P 就暗示 Q；也就是说，当主体能知道如果 P 是事实那么 Q 就是事实时，这种推理的证成就不依赖于经验。上述所有这三个论证都能被看成是构成了反对由 P 到 Q 的先验蕴含的论据。如果一个主体只知道 P，他就不能演绎出 Q（如知识论证所示），或者说如果一个人无需 Q 也能合理地想象 P（如可想象性论证所示），那么 P 似乎就不暗示 Q。我们可以认为解释论证取决于这一主张：从 P 到 Q 的暗示要求对意识作功能分析，而且意识概念不是一个功能概念。

在确定了认识性鸿沟之后，这些论证接着又推出一种本体论鸿沟（ontological gap），这里的本体论与世界上各种事物的本质有关。可想象性论证从可想象性推出了形而上学的可能性；知识论证从可推断性的失败推出了事实方面的不同；而解释论证则从物理解释的失败推出了非物理性。有人可能会说：这些论证从认知性蕴含的失败推导出了本体论蕴含的失败。典型的本体论蕴含是**必然蕴含关系**（necessitation）：如果实质条件句 P ⊃ Q 在形而上学上是必然的，或者说如果 Q 不成立时 P 也能成立在形而上学上是不可能的，那么 P 就必然蕴含 Q。人们普遍同意，唯物主义要求 P 必然蕴含所有真理（可能也有次要的限定条件）。因此，如果存在 P 并不必然蕴含的现象真理 Q，那么唯物主义就是错误的。

我们可以将这些论证称为反唯物主义的**认识性论证**。认识性论证可以说是源于笛卡尔的反唯物主义论证（但这些论证有稍微不同的形式），它们第一次完全的公开亮相是在布罗德的书中，它包括了上述所有三种论证的要素。[9]反唯物主义的认识性论证的一般形式是：

（1）物理的真理与现象的真理之间存在一种认识性鸿沟。

（2）如果物理的真理与现象的真理之间存在一种认识性鸿沟，那么就存在一种本体论鸿沟，并且唯物主义是错误的。

——

（3）唯物主义是错误的。

当然，这种看待事物的方式会让问题过于简单化，会抽象掉论证之间的差异。[10]这对依据暗示和必需关系的准确分析同样有效。然而，这种分析就像一种有用的透镜，透过它我们会看到这些论证的共同之处，同时也可以分析对这些论证的各种回应。

唯物主义者抵制这些认识性论证大致有三种方式。A 类唯物主义者否

认存在相关的认识性鸿沟；B类唯物主义者同意存在难以弥合的认识性鸿沟，但否认有本体论鸿沟；C类唯物主义者同意存在深层次的认识性鸿沟，但认为它最终会弥合。下面我将分别讨论这三种策略。

5.4 A类唯物主义

根据A类唯物主义的意见，物理真理与现象真理之间没有认识性鸿沟；或者至少说任何表面的认识性鸿沟都很容易弥合。根据这种观点，世界上会存在有意识的存在物的复制品，它们没有意识状态或具有颠倒的意识状态，这是难以想象的（至少经考虑后）。按照这种观点，并没有玛丽在其黑白房间里原则上不知道的现象真理（当她离开那个房间时，她至多获得了一种能力）。按照这种观点，经过考虑后，一旦解决了解释各种认知的、行为的和环境的功能等容易问题，就不会存在解决意识这一“困难问题”。[11]

A类唯物主义有时采取取消论形式，认为意识并不存在，并没有现象的真理。有时又采取分析的功能主义或者逻辑行为主义的形式，认为意识是存在的，这里的“意识”概念完全是用功能术语或行为术语定义的（例如，在这里，有意识可能就是对信息有某种存取，并且/或者有某种作出言语报告的倾向）。就我们的目的来说，这两种观点之间的不同可以看成是术语方面的不同。两者都同意：就我们有存取、报告、控制等功能性能力而言，我们是有意识的；两者也都同意：在任何更进一步的（非功能地定义的）意义上，我们不是有意识的。分析的功能主义者认为，像“有意识的”这样的普通术语应当在第一种意义下使用（表达一个功能概念）；而取消主义者认为它们应当在第二种意义下使用。除了现存术语和概念用法方面的术语学分歧之外，这些观点实质上是相同的。

有些哲学家和科学家虽然没有明确地支持取消主义、分析的功能主义等，但他们却是公认的A类唯物主义者。A类唯物主义者的特有特征是秉持下述看法——经过考虑，除了解释各种功能之外，对于意识再也没有任何东西需要解释了——解决了这些就解释了需要解释的一切。相当的功能可能非常微妙而复杂，例如，涉及精细的存取、自我监控、报告、控制能力以及它们的相互作用。它们还可能被认为包括了各种环境关系。对这些功能的解释可能会涉及很多神经生物学的细节。因此，这些观点仍可能是

A 类观点，它们在抵制功能主义时被提出，因为功能主义忽略了生物学或者环境的作用。

有人可能认为，这样一种观点在逻辑空间中是有位置的，它甚至否认了这种关于意识的宽泛的功能主义观点，但它仍认为物理的真理与现象的真理之间没有认识性鸿沟。在实践中，这种观点似乎是没有位置的，其理由我将在 C 类型下讨论。另外，这些观点的实际例证也很少。[12] 因此，我理所当然地会认为：A 类观点认为解释了功能就解释了一切，我会将其他观点划为 C 类，它们认为不存在难以弥合的认识性鸿沟。

显而易见，A 类唯物主义的问题是：它似乎否认了明显的东西。一个没有异议的事实是：我们有存取、控制、报告等各种功能性能力，而且这些现象为意识科学提供了没有异议的解释对象（需要解释的现象）。但是，似乎还有一个事实，即我们有意识，而这种现象似乎提供了另一种解释对象。正是这种解释对象提出了有意思的意识问题。断然否认另外的事实，或者说毫无异议地否认意识除了容易问题之外还有一个困难问题，这就会导致一种极其违反直觉的主张，因为它回避了重要问题的实质。这不是说极其违反直觉的主张总是错误的，但它们必须得到极其有力的论证的支持。因此，关键问题是：有什么令人信服的**论证**能支持这种主张，即经考虑，解释了功能就解释了一切？

A 类唯物主义者经常用类比进行论证。他们指出：在别的科学领域，我们承认解释了各种功能就解释了现象，因此在这里我们同样应该承认。反对者在回应时可能承认在其他领域我们只需要解释功能。例如，在解释生命时，需要解释的现象仅仅是适应、生长、新陈代谢、繁殖等，其他现象甚至都不需要解释。但反对者认为，意识的情况不同，它可能是独一无二的，这恰好是因为有别的东西（即现象的经验）需要解释。A 类唯物主义者要么连表现出来的另一种解释对象也必须予以否认（它似乎要否定显而易见的东西）要么必须承认这种表面的不相似，并进一步对为什么——与表象相反——只有功能需要解释作出实质性论证。

在这一点上，A 类唯物主义者经常强调一种不同的类比，他们认为在过去各个时期，思想家们都认为其他现象也有类似的认识性鸿沟，但这些现象最终能得到物理的解释。例如，丹尼特（1996）指出：活力论者可能认为，除了解释生物的功能之外，生命还有一个“困难问题”，但这是被研究者误导的。

根据对这些情况的考察，这些类比并不支持 A 类唯物主义者。活力论

者通常都或明或暗地**承认**所讨论的生物功能是需要解释的东西。他们的活力论之所以产生，是因为他们认为这些功能（适应、生长、繁殖等）不能作出物理的解释。因此，这与意识的情况是截然不同的。这一不相似在布罗德那里非常清楚。对于生命，布罗德是一位活力论者，他认为这些功能要求一种非机械论的解释。但同时他也认为：生命不同于意识，我们对生命现象拥有的证据只有行为，而且“是活的”只表现出了部分行为。其他活力论者的论点都不太明确，但其中很少有人会认为，除了这些功能之外（在有些情况下，除了意识本身之外）还有什么需要解释。如果活力论者持有这一主张，那么明显的答复就是：没有理由相信存在这样一种解释对象。因此，这里没有类比。[13]

因此，这些根据类比所作的论证对于 A 类唯物主义者而言是无效的。在别的情况下，始终都很清楚的是：除了与意识本身直接相关的东西之外，结构和功能就已穷尽了明显的解释对象。因此，在面对意识时，A 类唯物主义者必须处理另一种明显的解释对象：或者断然否定它，或者给出解决它的实质性论证。

对 A 类唯物主义者的某些论证是间接进行的，即通过指出抵制这种观点会造成令人不快的形而上学或者认识论的结果：例如，这种抵制会导致二元论，或者引起涉及意识的知识难题。[14]反对者要么接受这些结果，要么否认它们是结果。只要这些结果不是完全站不住脚的，那么 A 类唯物主义者要取得进展，就必须用一个实质性的直接论证来补充论证，这种论证反对另外的解释对象。

由于这些直接论证出乎意料的困难，因此对 A 类唯物主义的很多论证最终都在关键的地方预设了这种结论。例如，人们（例如，雷伊，1995）有时认为：没有理由假设感受性质，因为它们不是解释行为所必需的；但这个论证预设了只有行为是需要解释的。反对者认为：感受性质本身就是一种解释对象。同样，丹尼特（1991）把“异己现象学”（heterophenomenology，言语报告）当作为他的意识理论奠定基础的主要材料，他的做法似乎依赖于以下假设：这些报告是需要解释的东西，或者说需要解释的“显现”（seemings）仅仅是作出反应和报告的倾向。

一种辩护 A 类唯物主义的方法是作这样的论证：存在某个中介 X，它使得（i）解释了功能就足以解释 X，（ii）解释了 X 就足以解释意识。这里，一种可能的 X 是表征：人们常常既认为意识状态是表征状态，即表征世界上的东西，又认为我们能用功能术语解释表征。倘若如此，似乎就能

推出这一结论——我们能够用功能术语解释意识。但是，经过检查就会发现，这一论证利用了表征概念中的一种歧义性。一个概念是**功能表征**，根据它，当一个系统对P作出反应并/或产生了适合于P的行为时，我们基本可以说P被表征了。在这种意义上，解释了功能可能就解释了表征，但解释了表征并没有解释意识。还有一个概念是**现象表征**，根据它，当一个系统具有一种像P那样的意识经验时，大致来说P就被表征了。在这种意义上，解释了表征可能就解释了意识，但解释了功能并没有解释表征。总之，功能与现象之间的认识性鸿沟仍和以前一样宽。对于这里可能利用的其他X，如“知觉”或“信息”，也能从中发现类似的歧义性。

对A类唯物主义最有意思的论证也许是这些论证，它们认为我们能对我们关于意识的**信念**作出一种物理解释，例如这种信念：我们有意识，意识是另一种解释对象，以及意识是非物理的。由此可以推断：一旦我们解释了这种信念，我们就足以对这种现象作出解释或者作出辩解（例如，克拉克，2000；丹尼特，即出）。这里应当指出的是：只有在这些信念本身能从功能上进行分析时，它才是有效的；查默斯（2002a）给出了否定这一
112 观点的理由。但即使有人承认意识最终是功能的，这种主张仍会还原为另一种主张：解释了我们谈论意识（等等）的倾向就解释了一切。反对者会否认这一主张：解释了报告的倾向可能会免除接受另一种解释对象的第三人称保证（根据其他人的观察），但它不会免除关键的第一人称保证（来自自己的立场）。尽管如此，这种策略仍值得深入讨论。

在某一点上，A类唯物主义者及其反对者之间的争论通常会归结为直觉，最核心的直觉是：意识（在非功能地定义的意义上）是存在的，或者说有某种东西是需要解释的（除了解释功能之外）。这种主张所获得的支持并非来自于论证，而是来自于一种观察，以及对反论证的辩驳。这种直觉似乎是绝大多数哲学家、科学家以及其他人所共有的；它非常有力，以至于若希望否定它，A类唯物主义者就需要格外有力的论证。最后的结果就是：即便在唯物主义者中间，A类唯物主义者也是明显的少数派。

5.5 B类唯物主义[15]

根据B类唯物主义，物理领域与现象领域之间只有认识性鸿沟，而没有本体论鸿沟。根据这种观点，怪人是可以想象的，但从形而上学上说，

他们又是不可能的。从这一观点看，对于某些现象事实，玛丽仅从她所在的房间内部出发是无法知道的，尽管如此，这些事实与基本的物理实在仍是有关系的（当她离开那个房间时，会以新的方式了解那些旧的事实）。按照这种观点，虽然有一个与容易问题不同的困难问题，但它并不对应一个不同的本体论领域。

最常见的B类唯物主义形式认为，现象状态能够**同**一于某些物理或者功能的状态。据认为，这种同一性在某些方面（尽管可能不是所有方面）类似于水与H_2O或者基因与DNA之间的同一性。[16]这些同一性不是得自于概念分析，而是通过经验发现。概念**水**不同于概念H_2O，但人们发现它们实质上指称相同的东西。根据B类观点，类似的看法也适用于意识：意识概念不同于任何物理或功能的概念，但我们能经验地发现这些概念本质上都指称相同的东西。这样，我们就能对为什么在物理领域与现象领域之间存在一个认识性鸿沟作出解释，同时否认有任何本体论鸿沟。这就产生了一种吸引人的可能性：我们能够承认关于意识的深层次认识性问题，但同时保留唯物主义世界观。

尽管这种观点是有吸引力的，但它很快就会遇到困难。这些困难来自于这一事实，即意识的认识性鸿沟的特征与其他领域的认识性鸿沟的特征似乎是不同的。首先，在上述水、基因等案例中，似乎没有类似的认识性论证。要解释基因，我们只需解释系统为什么能在传递遗传特征中发挥某种功能；要解释水，我们只需解释为什么一种物质有某种客观的结构和行为。假如对世界有一种完备的物理描述，那么玛丽通过演绎出那些具有适当结构和功能的系统，就能演绎出与水和基因相关的一切事实。最后，我们似乎不能前后一贯地想象这样一个世界：它在物理方面与我们自己的世界是同一的，但这个世界上没有水或者说没有基因。因此，关于这个世界的完备的物理事实与关于水和基因的事实之间，也就没有与意识所具有的认识性鸿沟相类似的认识性鸿沟。

（也许只是这些派生于意识的认识性鸿沟的认识性鸿沟。例如，玛丽也许不能从关于世界的物理事实演绎或者解释水的知觉**表象**。但这只是我们所讨论问题的另一个例证，对于B类唯物主义者是没有帮助的。）

因此，就意识来说，情况似乎有些与众不同。我们对此所作的解释是：虽然基因与DNA之间的同一性是经验性的，但它**从认识上说**不是**原始的**（epistemically primitive）：这种同一性本身能从关于世界的完备的物理事实中演绎出来。相反，B类唯物主义者一定会认为：意识与物理或

功能的状态之间的同一关系从认识上说是原始的：这种同一性不能从完备的物理事实中演绎出来。（如果它能演绎出来，A类唯物主义反而就是正确的了。）因此，意识与物理状态之间的同一性是人的世界理论中的原始原则。

这里，有人可能会指出有些事情出错了。在其他地方，这种原始原则只能到基本的物理学规律中去寻找。当然，人们通常认为，这种原始性（primitiveness）——不能从更基本的原则演绎出来——是基本的自然规律的标志。实际上，B类唯物主义者承认这样一个原则具有基本规律的认识性地位，但却给予它一种同一性的本体论地位。反对者认为：这一举动更像是偷窃而不是诚实劳动：在其他地方，同一关系都以解释为基础，而原始原则则被承认是基本的规律。

人们很自然会指出这一规律应当同样适用于这里。如果一个人承认，从认识上说，物理状态与意识之间的这种原始关系是一个基本规律，就能由此推出：意识与任何物理属性都不同，因为基本规律总是联系不同属性的。因此，普通标准会导致本章后半部分要讨论的非还原的观点之一。相反，B类唯物主义者认为，物理状态与现象状态之间被观察到的关系不能用更基本的术语解释，并指出这种关系是同一性。提出这种看法主要是为了维护先验的唯物主义承诺。除非对原始的同一性有独立的论据，否则说好听点，这种看法似乎是特别的、神秘的，说难听点就是前后不连贯的。

114 B类唯物主义者能作出不同的回应。首先，有的人（例如，帕皮诺，1993）指出，同一性不一定要作出解释，因此它们始终是原始的。但是我们已经看到，其他领域的同一性至少是能从更基本的事实中**演绎**出来，因此它们在相关意义上并不是原始的。其次，有的人（例如，布洛克和斯托尔内克［Block and Stalnaker］，1999）指出，甚至与水和基因相关的事实也不能从基本的物理事实中演绎出来。这个问题太复杂，不能在这里详细阐述（一种回应，参见查默斯和杰克逊于2001年出版的合著[17]），但是人们能够注意到，一开始时所概括的认识性论证就指出了意识与其他情况之间是迥然不同的。第三，有的人（例如，布赖恩·洛尔［Brian Loar］，1990/1997）承认，从认识上说，与意识相关的同一性是原始的，因此不同于其他的同一性，但他们想借助于意识概念的独特特征来对这种独特性作出解释。这种回应也许是最有意思的，稍后我再回到这种回应。

B类唯物主义者还能采取另一条路线。有的人可能开始注意到，严格

来说，意识与物理状态之间的**同一性**并不是唯物主义立场所必需的。相反，一个人可以合理地认为，关于意识的唯物主义仅仅要求，是物理状态**使**现象状态**成为了必需**，因为物理状态会在现象状态缺失（或者说不同）的情况下出现，从形而上学上说是不可能的。也就是说，唯物主义要求 P⊃Q 这些蕴含是必然的，这里的 P 是关于世界的完备物理事实，Q 是任意的现象事实。

在这一点上，B 类唯物主义者自然能借助于克里普克（1980）的工作，克里普克指出：有些事实不是先验的但却必然为真。例如，克里普克指出："水是 H_2O" 是必然的——在所有可能世界中都为真——但却不能先验地认识。在这里，B 类唯物主义者可以指出：P⊃Q 可能是一种克里普克式的后验必然性，就像 "水是 H_2O" 一样（尽管如此，也应注意到克里普克本人是否认这一主张的）。倘若如此，我们就会**预料**存在一种认识性鸿沟，因为不存在从 P 到 Q 的先验衍推，但同时也不会有本体论的鸿沟。这样，克里普克的工作似乎恰好就是 B 类唯物主义者所需要的。

这里，之前出现的一些问题会再次出现。有人可能认为，在其他领域中，必然性从认识上说并不是原始的。水与 H_2O 之间的必然关系可能是后验的，但它本身能从关于世界的完备物理描述中演绎出来（人们能演绎出水同一于 H_2O，由此可以推出水必然是 H_2O）。这同样适用于克里普克所讨论的其他必然性。相反，B 类唯物主义者一定会认为，物理状态与意识之间的关系从认识上说是原始的，因为它不能从关于世界的完备物理事实中演绎出来。一个人可以再次指出，这种原始的必然关系是神秘且特别的，这种关系反而应该被看成是基本的自然规律。

我在下一节将对这些必然性存在问题作进一步的讨论。但是，这里应该注意的是，在某种意义上，任何一种 B 类唯物主义立场都放弃了还原解释。即使 B 类唯物主义是正确的，我们也不能用纯物理术语对意识作出像我们对基因等所作的相同解释。相反，我们的解释始终都需要一些从解释上所说的原始原则，用它们来弥合物理与现象之间的鸿沟。根据这种观点，意识理论的**解释**结构和其他领域的唯物主义理论的解释结构是判然有别的，它与下面描述的非还原理论的解释结构非常相似。由于这种观点将这些原则称为同一性（或必然性）而不是规律，因此它能保留唯物主义的形式（letter），但由于它需要原始的桥梁原则，因此它牺牲了唯物主义的很多精神。

5.6 反对 B 类唯物主义的两维论证

如上面所讨论的，B 类唯物主义者认为，怪人世界等假设是可以想象的（P¬Q 中没有矛盾），但它们从形而上学上说又是不可能的。也就是说，P⊃Q 被看成是一种后验的必然性，它类似于“水是 H_2O”这样的后验必然性。更详细地考察克里普克的后验必然性案例，我们就能对这一立场作更深入的分析。这些资料有点技术性（因此单独列为一节），读者第一次阅读时（如果有必要）可以跳过去。

人们常说，在克里普克的案例中，可想象性并不必然伴随着可能性：水不是 H_2O，这是可以想象的（因为认为水不是 H_2O 是融贯的），但水不是 H_2O，这不是可能的。同时，对于一个人所想象的东西似乎有**某种**可能性。当一个人想象水不是 H_2O 时，他想象了一个世界 W（XYZ 世界），在这个世界中，海洋里像水一样的液体不是 H_2O，而是（比如说）XYZ。没有理由怀疑这个 XYZ 世界从形而上学上说是可能的。如果克里普克是对的，那么 XYZ 世界就没有被正确地描述成一个其中水是 XYZ 的世界。尽管如此，这个世界在一个稍有不同的方面仍与“水是 XYZ”之真相关，对此我们可以作如下说明。

一个人可以说，XYZ 世界可能**最后证明**就是现实世界，因为就我们的先验所知来说，现实世界与 XYZ 世界是完全相似的。一个人可能会说，**如果** XYZ 世界最后证明就是现实世界，那么结果就是：水是 XYZ。同样：如果我们假定 XYZ 世界是现实的，我们就应当据此合理地推断：水不是 H_2O。也就是说，XYZ 世界与“水不是 H_2O”之间有很强的**认识性**关系。甚至克里普克也承认：水不是 H_2O，**从认识上说**是**可能的**（在广义上，这并不排除它是先验的）。XYZ 世界是现实的，这种认识上的可能性似乎是水不是 H_2O 这种认识上的可能性的一个具体例证。

这里，我们对世界 W 采取了一种特殊的态度。我们认为 W 是一种认识上的可能性：是世界可以实际所是的一种样子。当我们这样认为时，我们把 W 看成**是现实的**。当我们认为 W 是现实的时候，它就能使给定的一个句子 S 为真或者为假。例如，当认为 XYZ 世界是现实世界时，它就使“水不是 H_2O”为真。这一点表现在这一直觉判断之中：如果 W 最终证明是现实的，那么结果就是水而不是 H_2O，而 W 是现实的这种认识上的可

能性，是水是 H_2O 这种认识上的可能性的一个实例。

相反，一个人也可以把世界 W 看成**是反事实的**。当我们这样认为时，我们承认现实世界的特征已经被确定，我们认为 W 是事物本可能所是而非实际所是的反事实情况。如果克里普克是对的，那么假若像水一样的东西是 XYZ，XYZ 就仍不会是水。因此，当我们把 XYZ 世界看成反事实情况时，它并不会使“水不是 H_2O”为真。在把它看成是反事实之时，我们是根据水是 H_2O 这个现实世界的事实来描述 XYZ 的，而且我们得出了这样的结论：XYZ 不是水，而只是像水一样的东西。这些结果并不矛盾，因为它们只涉及两种不同的考虑和描述可能世界的方式。克里普克的主张考虑的是对世界的**反事实**评价，而上个段落中主张涉及的是对世界的**认识性**评价。

可以用**两维语义学**（two-dimensional semantics）对此进行形式化处理。[18]我们可以说：如果被看成是现实的 W 使 S 为真，那么 W 就证实了 S，而如果被看成是反事实的 W 使 S 为真，那么 W 就**满足**了 S。证实与对世界的认识性评价有关，而满足与对世界的反事实评价有关。与此相应，我们可以把 S 和不同的**内涵**或者说从世界到真值的功能联系起来。S **首要的**（或者认识性的）内涵是这样一种功能，它在世界 W 中为真，当且仅当 W 证实了 S，而**次要的**（或者虚拟的）内涵是这样一种功能，当 W 满足了 S，它就在世界 W 中为真。例如，假定这里的 S 是“水不是 H_2O”，W 是 XYZ 世界，我们就可以说：W 证实了 S，但 W 并未满足 S。我们也可以说：S 的首要内涵在 W 中为真，但 S 的次要内涵在 W 中为假。

记住这一点之后，就能这样说：当一个陈述 S 可以想象时——也就是说，当它的真不能被先验地排除时——那么，就存在某个证实了 S 的世界，或者在相当意义上说，存在某个世界，S 的首要内涵在其中为真。这在直觉上是说得通的：当 S 可以想象时，S 代表了一种认识上的可能性。人们自然会提出：与这些认识上的可能性相对应的是诸特定的世界 W，以至于当这些世界被看成是认识上的可能性时，它们就证实了 S。也就是说，W 是这样的：从直觉上看，如果 W 最终是现实的，它就会证明 S。

这种模型似乎适合克里普克的所有案例。例如，克里普克认为，热是分子的运动，这是一种后验的必然性。因此，在相关意义上，说热不是分子的运动是可以想象的。对应这种可想象情景的是一个世界 W，在其中人们对热的感觉是由与分子运动不同的东西引起的。W 代表了一种认识上的可能性：我们可以说，如果 W 最终是现实的，那么结果就是：热不是分子

的运动。很多别的情况也同样如此。这里我希望说明，克里普克的现象涉及两种思考可能世界的不同方式，其中只有一种构成了诸世界之空间的基础。

117 如果将这一原则用于B类唯物主义，就会立即出现麻烦。如前所述，假如P是关于世界的完备的物理事实，而Q是一种现象事实，并且我们说：当S的真没有被先验地排除时，S是可以想象的。那么，有人就能建立起这样一个论证[19]：

（1）P∧¬Q是可以想象的。

（2）如果P∧¬Q是可以想象的，那么一个世界就证实了P∧¬Q。

（3）如果一个世界证实了P∧¬Q，那么一个世界就满足了P∧¬Q或者说F类一元论是正确的。

（4）如果一个世界满足了P∧¬Q，那么唯物主义就是错误的。

——

（5）唯物主义是错误的，或者F类一元论是正确的。

B类唯物主义者承认前提（1）：要否认这个前提就要接受A类唯物主义。前提（2）是上面所讨论的一般原则的一个实例。前提（4）可以被看成是决定着唯物主义的。至于前提（3），一般来说，一个人不能直接从一个证实了S的世界转移到一个满足了S的世界，就像“水是H_2O”（和XYZ世界）的案例所示。但是，就P∧¬Q而言，对P和Q的本质稍加思索就能把我们引到那个方向上。具体说明如下。

先看Q。这里似乎可信的是：如果W证实了“存在意识”，那么W就满足了“存在意识”，反之亦然。这符合克里普克的观点，即就意识而言，并没有与水本身和纯粹像水一样的东西那种区别相类似的一种区别。从直觉上说，如果W证实了“存在意识”，那么它就包含了某种至少**感觉起来**有意识的东西，而如果某种东西**感觉起来**有意识，那么它**就是**有意识的。一个人能说：我们核心的现象概念之首要和次要的内涵是相同的（参见查默斯，2002a）。由此可以推出：如果世界W证实了¬Q，那么，W就满足了¬Q。（这种主张不是进行这一论证所必需的，但它是可信的，它会使论证变得更简单。）

再看P。一个B类唯物主义者可能想通过以下方式来摆脱这个论证，即主张虽然W证实了P，但它并未满足P。再三考虑后，这种主张要成立，唯一的办法可能是：如果一个世界证实了P，它就必须至少具有现实

物理世界的**结构**。为什么 W 可能未满足 P 呢？唯一的原因是：它缺乏支撑现实世界结构的内在属性。（根据这种观点，一个物理概念的首要内涵选定了任何在给定世界里起某种作用的属性，而次要内涵选定了所有世界上实际的内在属性。）如果 W 中的这种不同是 W 中没有意识的原因，那么由此可以推出：现实世界中的意识并不是由物理学的结构方面必然蕴含的，而是根植于其基本的内在本质。这就是我所说的 F 类一元论或“泛原心论”（panprotopsychism）的立场。F 类一元论是一种很有意思的重要立场，但它比通常所认为的 B 类唯物主义要激进得多，我把它看成是一种不同的立场。后面我再来讨论这种推理和由此所产生的立场。

由此可以推出前提（4）是正确的。如果一个世界证实了 $P \wedge \neg Q$，那么要么一个世界满足了 $P \wedge \neg Q$，要么 F 类一元论是正确的。如果暂时不考虑 F 类一元论，那就可以推出：关于我们世界的物理事实并不必然蕴含现象事实，因此唯物主义是错误的。

这个结论实际上是下列三者的一个结果：（i）$P \wedge \neg Q$ 是可以想象的（在相关意义上）；（ii）当 S 是可以想象的时候，就有一个证实了 S 的世界；（iii）这是某个简单的推理。唯物主义者的回应可能是否定（i），但这仅仅否定了物理与现象之间相关的认识性鸿沟，从而仅仅否定了 B 类唯物主义。我认为，B 类唯物主义者在否定（iii）中所包含的这种推理方面是不大有希望的。因此，B 类唯物主义者唯一的希望是否定核心论题（ii）。[20]

要想达到这一目的，B 类唯物主义者既可以否定证实与满足之间的区别具有融贯性，也可以承认这种区别是融贯的，但即使是在标准的克里普克案例中，否认论题（ii）也是有效的，也有些立场承认论题（ii）在标准的克里普克案例中是有效的，同时否认它在意识这种特殊案例中是有效的。前两种选择值得考察，但我认为它们最终是没有希望的，因为这种区别及这个论题似乎非常适合克里普克式现象。我认为，B 类唯物主义者最终一定会主张：意识是特殊的，在其他地方成立的论题在这里是不成立的。

根据这个观点，这些联系物理领域和现象领域的后验必然性比其他领域的要强得多，因为它们被所有世界证实了。在其他地方，我把这些不同寻常的后验必然性称为**强必然性**。我认为，相信它们存在并没有什么正当的理由。至于这些解释上的原始同一性，它们似乎是以一种特别的方式假设的原始事实，主要目的是为了拯救一种理论，但它们没有得到其他地方

的案例的支持。再者，有人可能认为，这种观点会导致一种潜在的**模态二元论**，它有单独的逻辑和形而上学可能性的原始域；有人可能认为这是难以接受的。

B类唯物主义者作出的最有意思的回应可能是承认强必然性是意识所独有的，并设法根据我们的概念系统的特定特征来解释这种独特性。例如，克里斯托弗·希尔（Christopher Hill，1997）认为，一个人可以从以下事实出发来预言有关意识的解释性鸿沟：物理概念和现象概念有不同的概念作用。而布赖恩·洛尔（1990/1997）是借助这种主张：现象概念是缺乏偶然的显现模式的识别概念。约瑟夫·莱文（Joseph Levine，2000）则认为，现象概念有非归属性的（non-ascriptive）显现模式。但我（查默斯，1999）认为，这些回应都是无效的，对此我有系统的理由。[21]不过，这个方向上其他尝试可能马上就出现了。这在关于意识的形而上学争论中仍是关键的领域之一。

总之，我自己的看法是：几乎没有理由认为，解释上的原始同一性或
119 者说强必然性是存在的。没有什么**单独的**正当理由去相信它们，至于假设它们，最好的理由是辩称为了拯救唯物主义，但在关于唯物主义是否正确的争论语境下，这个理由是难以让人信服的，特别是当有可行的替代者的话。尽管如此，进一步研究支撑这一争论的关键问题，从哲学上说很可能将是富有成果的。

5.7 C类唯物主义

根据C类唯物主义，物理领域和现象领域之间存在一条很深的认识性鸿沟，但原则上它是可以弥合的。根据这个观点，怪人是我们现在能想象的，但在极限情况下它们是不能想象的。根据这一观点，现在玛丽似乎缺乏关于现象的信息，但在极限情况下并不存在她所缺乏的信息；根据这一观点，虽然我们现在不知道怎样用物理术语解决困难问题，但这个问题从理论上说是能够解决的。

这种观点初看起来很有吸引力。它似乎承认我们所面对的这条深深的解释性鸿沟，但它同时又承认这条表面的鸿沟可能仅仅由于我们自身的局限性。这一观点有不同的版本。内格尔（1974）指出，正如前苏格拉底哲学家不理解物质何以能有能量一样，我们也难以理解意识何以能是物理

的，但一场概念革命就会提供相关的理解力。丘奇兰德（1997）指出，即使我们现在难以想象意识何以能是一种物理过程，但这只是我们自身的一种心理学的局限性，随着科学的进一步发展，它是可以克服的。范久利克（Van Gulick，1993）指出，可想象性论证是以假定为论据的论证，因为一旦我们对意识作出了正当的解释，怪人等就不再能想象了。麦金（McGinn，1989）指出，这个问题可能是人类无法解决的，这根源于我们认知能力的深层次的局限性，尽管如此，它原则上会有一种解答。

这种观点的一种表述方式是：怪人等乍看起来是可以想象的（就现在我们具有目前的认知过程而言），但它们**从理想上说**是不能想象的（在理想化的理性反思之下）。或者我们会说，现象事实原则上能从物理事实中演绎出来，但这种可演绎性与复杂的数学真理的可演绎性是相似的：它原则上是能得到的（也许是能先验地得到的），但不是现在的我们所能得到的，这可能是由于必需的推理超出了我们现在的能力范围，或者说可能是由于我们现在还没有掌握所有必需的物理事实。如果事实是这样，那么在我们看来，物理过程和意识之间就存在一条鸿沟，但实质上并没有。

尽管它有吸引力，但我认为C类观点是内在不稳定的。经过考察会发现，它最终要么站不住脚，要么会分化成其他观点。特别是，这种观点似乎一定会分化成A类唯物主义、B类唯物主义、D类二元论或者F类一元论的某一个版本，因此它最终并不是一种不同的选项。

就主张认识性鸿沟在极限情况下可能被弥合来说，一种解释办法是主张：在极限情况下，我们会看到解释了功能就解释了一切，并没有别的待解释对象。下述主张至少是融贯的：我们目前犯了某种概念混淆或者含糊，它使我们相信还有别的待解释对象，这种情况有可能被更好的推理解决。我把这种立场看成是A类唯物主义而非C类唯物主义的一个版本：它显然与标准的A类唯物主义是密切相关的（主要的区别在于我们是否拥有了相关的见解），并且会产生相同的问题。与标准的A类唯物主义相似，这种观点的成败最终取决于（实际的和潜在的）一阶论证的说服力，这些一阶论证取消了其他明显的待解释对象。

一旦A类唯物主义被束之高阁，那么弥合认识性鸿沟的潜在选择就会受到极大的限制。这些限制是建立在物理概念的本质和意识概念的本质之基础上的。基本的问题已经在前文提及。首先，世界的物理描述依据的是结构和动力学；其次，从结构和动力学的事实只能演绎出其他的结构和动力学事实；第三，关于意识的事实并不是结构和动力学的事实。不过，我

们可以对这些步骤逐个作出说明。

首先，关于世界的微观物理描述详细说明了粒子、场和波的空间与时间分布。这些基本系统是由它们的时空属性以及像质量、电荷和量子波函数状态等属性描述的。后面的这些属性是依据这些具有某种抽象结构状态的空间最终定义的（例如，连续变化的实数的空间，或者希尔伯特空间状态的空间），以至于这些状态对其他状态也有某种因果作用。我们可以将时空描述和依据这些形式空间中的属性所作的描述划归到**结构**描述的题目之下。随着时间的流逝，这些系统的状态能依据定义于相关属性之上的动力学原理发生变化。结果就是依据世界的基本时空形式结构以及这种结构的动态进化对世界作出描述。

有些C类唯物主义者认为，我们还没有一种完备的物理学，因此我们不知道这样一种物理学能解释什么。但我们也不一定需要一种完备的物理学；我们只需要这种主张，即物理描述依据的是结构和动力学。这一点在各种物理理论之间是通用的。像相对论、量子力学等新理论能引进新的结构和建基于那些结构之上的新动力学，但这种通用的观点（以及意识的鸿沟）仍保持不变。

C类唯物主义者可能认为，或许存在超出了结构和动力学的新物理理论。但是，即便真的有这种物理解释的特征，我们仍不清楚这种理论可能是什么样的。新的物理属性是由于它们在解释现有物理现象中的潜力而被
121 假定的，它们自身是根据结构和动力学来描述的，而结构和动力学在这里似乎始终能满足需要。一种可能性是：物理学最终也许会诉诸意识自身，而不是借助于新的属性，就像某些理论家认为量子力学所做的那样。这种可能是无法排除的，但它导致了一种观点，根据这种观点，意识本身是不能还原的，因此它被归入非还原的范畴（D类或F类）。

对于借助于“完备的物理学”的思路应当认真对待。这里的想法是：当代物理学是依据抽象的结构和关系来描述其基本属性（如质量和电荷），但它没有解释这些属性的内在本质。根据这种观点，关于世界的完备物理描述也一定描述了为这些结构和关系奠定基础的内在属性；而一旦这些内在属性被使用，物理学就会超越结构和动力学，关于意识的事实可能就是这样产生的。相关的内在属性是我们所不知道的，但它们原则上是能被我们掌握的。这是一种重要的立场，而它恰好是在F类之下将讨论的立场，我到后面再来讨论它。

其次，从结构和动力学出发所作的描述能推导出什么呢？低层次的微

观物理描述必然伴有各种让人吃惊而又有趣的宏观属性，就像化学是从物理学，生物学是从化学中逐步独立出来的一样，或者说就像复杂系统理论中更一般的复杂突现行为一样。但是在所有这些案例中，必然伴有的这些复杂属性却是结构和动力学的：它们描述了复杂的时空结构和那些结构之上的复杂动力行为模式。因此，这些案例支持这个一般的原则：从结构和动力学只能推导出结构和动力学。

C类唯物主义者可能提出，有些事实尽管本身不是结构—动力学的，但它们却是由一种结构—动力学描述所暗示的。人们也许认为，关于**表征**或者**信念**的事实可能就有这种特征。但是如我们之前所提到的，似乎很清楚，这些事实是一种结构—动力学描述所暗示的，其任何意义都涉及表征或信念所默许的功能意义。我们所期待的是：如果涉及这些事实的主张**由于**一种结构—动力学描述成立能够被看成是（根据概念的理由）正确的，那么所涉及的诸概念本身在某个层次上就一定是结构—动力学的。

有人可能认为，存在某个中介概念X，以至于关于X的事实是由于结构—动力学的描述而成立的，而关于意识的事实是由于X而成立的。但是就像在A类唯物主义的情况下那样，要么X能从功能上进行分析（在广义上，不过在这种情况下第二步就是无效的），要么X不能从功能上进行分析（在这种情况下第一步是无效的）。这显然出现在表征的情况中：就功能表征的概念来说，第一步是无效的；而就现象表征的概念来说，第二步是无效的。因此这种策略只有根据歧义性才能有效。

第三，解释或演绎出了复杂结构和动力学就足以能解释或演绎出意识吗？显然不能，理由很简单。玛丽在她的黑白房间里就能知道世界上所有层次的一切时空结构和动力学，但这不会告诉她看见红色是什么感觉。就对任何一个系统复杂的宏观结构描述或者动力学描述来说，一个人可以想象那种描述是在没有意识的情况下被例示的。解释了人的系统结构和动力学只是解决了容易问题，但是并未触及困难问题。反对者要抵制这最后一步就必须主张：解释了结构和动力学**因此**就足以能解释意识。要想这样做，大致能站得住脚的办法只能是接受A类唯物主义，但我们已经把它束之高阁了。

C类唯物主义者可能会提出：结构比动力学（就像A类唯物主义者那样）更值得依赖。这里，时空结构似乎非常没有希望：解释一个系统的大小、形状、位置、运动等显然并不是解释意识。最后一种可能是依靠出现在意识状态自身中的结构。意识状态是有结构的，它既有一种位于复杂的

意识状态之内的内部结构，也有意识状态之间的相似性和差异性的模式。但这种结构是一种独特的**现象**结构，它实质上和出现在物理学中的时空结构迥然不同。复杂的现象状态的结构不是时空结构（尽管它可能涉及时空结构的表征），现象状态之间的相似性和差异性也不是形式的相似性和差异性，而是特定现象特征之间的不同。这反映在以下事实之中：一个人在没有任何相关现象结构的情况下也能想象任何一种时空结构和形式结构；一个人在不知道前者的情况下也能知道后者；等等。因此，认识性鸿沟仍和以前一样宽。

任何一种C类唯物主义策略的基本问题是：从A到B的认识性暗示要求某种**概念联系**，A中所描述的条件依靠这种联系就能满足关于B的概念要求。例如，当一种物理解释蕴含关于生命的事实时，它这样做是由于自身蕴含了关于物理系统的宏观功能——即生命所必需的这种宏观功能——的信息：这里，广义的功能概念提供了这种概念联系。但就意识而言，鉴于物理概念的结构—动力学特征以及意识概念迥然不同的特征，它们是不能获得这种概念联系的。

最后，似乎任何一种C类策略的厄运都是由于熟悉的理由。一旦我们接受意识概念本身不是功能概念，接受关于世界的物理描述是结构—动力学描述，那么在概念上就不容许这一战略是由一种物理描述所暗示的。因此，留下的空间只有主张意识毕竟是一种广义的功能概念（接受A类唯物主义），或者主张物理学中不只是有结构和动力学（接受D类二元论或F类一元论），或者主张唯物主义为真不必然要求从物理学到意识的暗示（接受B类唯物主义）。[22]因此，C类唯物主义者最终并没有独立的活动空间。

123

5.8 插　曲

唯物主义者有别的选择吗？另一种选择是反对这种分类法所依赖的区别。例如，有些哲学家，尤其是蒯因（1951）的追随者，反对概念真理与经验真理、先验与后验或者偶然与必然之间有任何的区别。因此，称职的蒯因主义者或许会反对A类唯物主义与B类唯物主义之间有区别，他会主张：尽管谈论认识性暗示和/或模态的蕴含是没有根据的，但唯物主义是正确的。我们可以将这种观点称为Q类唯物主义。然而，这种观点也会产生相似的问题。有些蒯因主义者认为，解释了功能就解释了一切（丹尼特

可能就是一个例证），倘若如此，那么所有 A 类唯物主义的问题就会出现。有些前因主义者认为，由于物理状态与意识状态之间在本质上存在强同型（isomorphic）关系，因此我们可以假设它们之间具有同一性（保罗·丘奇兰德可能是一个例证），倘若如此，B 类的问题就会出现。还有些人可能会借助未来的新奇解释，倘若如此，C 类唯物主义的问题就会出现。因此，前因主义方案是难以避免相关问题的。

如果不考虑这种观点，那么，对于唯物主义者来说，大致可行的选择看起来只有 A 类唯物主义和 B 类唯物主义了。我认为别的观点要么最终是不稳定的，要么会分化为这些观点之一（或者剩下的三种选择）[23]。在我看来，这些观点——它们否认出现在第一种情况下的待解释对象，接受第二种情况下的原始同一性或强必然性——的代价强有力地表明：这些问题是可以避免的，除非不存在可行的替代观点。

因此，未作解释的问题就是：是否存在可行的替代观点。如果意识不是由物理事实必然蕴含的，那么它一定涉及某种从本体论说的新东西：用克里普克的隐喻来说，在确定了所有物理的事实之后，为了确定所有关于意识的事实，上帝就必须做更多的事情。也就是说，除了由物理理论所描述的特征之外，世界还一定会有一些基本的本体论特征。我们都很熟悉这种观念，即世界的某些特征是基本的：在物理学中，像时空、质量和电荷等特征被当成是基本的，它们不用作进一步解释。如果这些反对唯物主义的论证是正确的，那么物理学现有的这些特征就没有穷尽世界的基本特征。我们必须扩展我们关于世界的基本特征的目录。

这里有两种可能性。第一，意识本身可能就是世界的一个基本特征，就像时空和质量一样。在这种情况下，我们可以说现象属性是基本的。第二，意识本身可能不是基本的，但它是由某种更原始的基本特征 X 必然蕴 124
含的，而 X 本身不是由物理学必然蕴含的。在这种情况下，我们可以把 X 称为**原现象**属性（protophenomenal property），我们可以说原现象属性是基本的。为了方便讨论，我以第一种可能性为代表来进行讨论，这一讨论也同样适用于第二种可能性。无论如何，意识都涉及世界上某种新的和基本的东西。

不过这样一来就会出现新的问题：这些新的基本属性是如何与已知世界的基本属性——即微观物理学所使用的属性——相联系的？一般来说，只要有基本属性就会有基本规律。因此，我们可以预料会有某种将物理属性与现象属性联系起来的基本原则——心理物理规律。和相对论或量子力

学的基本规律一样，这些心理物理规律也不能从更基本的原则演绎出来，我们要将它们当成是原始的。

但是，这些规律有什么特征呢？与此相关的一个担忧是：世界的微观物理方面通常被认为是因果封闭的，因为每个微观物理状态都有一种微观物理的充分原因。那么，基本的现象属性是如何与这个因果封闭的网络相结合的呢？

这里，非还原论者似乎主要有三种选择。第一，他可以否认微观物理的因果封闭性，主张微观物理的动力学中存在因果鸿沟，后者要由独特的现象属性所起的一种因果作用来填充，这就是D类二元论。第二，他可以接受微观物理的因果封闭性并主张：现象属性没有发挥与这一物理网络相关的因果作用。这就是E类二元论。第三，他可以接受微观物理网络是因果封闭的观点，同时相信现象属性是与这个网络结合在一起并且起着一种因果作用，因为它构成了物理之物的内在本质，这就是F类一元论。

下面，我将逐一讨论每种观点。这种讨论在某些方面肯定是推测性的，我也不自称能证实其中的某种观点是正确或是确定无疑的。但我的确想指出：其中任何一种都没有明显的致命缺陷，每一种都值得作进一步研究。

5.9 D类二元论

D类二元论认为：微观物理学不是因果封闭的，现象属性起着一种影响物理世界的因果作用。[24]根据这种观点，它通常被称为**交互作用论**（interactionism），物理状态会引起现象状态，现象状态也会引起物理状态。相应的心理物理规律会在两个方向上同时起作用。根据这种观点，微观物理状态的进化不是由物理原则单独决定的。心理物理原则详细说明了现象状态对物理状态的作用，它们也会起一种不可还原的作用。

有关这种观点的最熟悉版本是笛卡尔的物质二元论（D就表示笛卡尔），根据实体二元论，存在发生相互作用的独立心理和物理的物质或实体。但这种观点与属性二元论也是相容的，根据属性二元论，只存在一种物质或实体，它既有物理的基本属性，也有现象的基本属性，以至于现象属性在影响物理属性方面起着一种不可还原的作用。特别是，这种观点与布罗德那样的“突现论”观点是相容的。根据突现论，现象属性从本体论

上说是物理系统的新属性（是不能仅仅从微观物理属性中演绎出来的），它们对微观物理属性有新的影响（这是不能仅仅从微观物理原则中演绎出来的）。这种观点会涉及心理对微观物理的“下向”（downward）因果关系的基本原则（因此，D也表示下向因果关系）。

人们有时会提出异议，说截然不同的物理状态和心理状态不可能发生相互作用，因为它们之间没有因果关系。但是，从休谟和现代科学得出的一个教训是：任何基本的因果相互作用（包括物理学中所发现的那些）也同样如此。例如，牛顿科学并未揭示出万有引力发生作用所借助的因果关系；相反，相关的规律完全是基本规律。其他物理理论中的基本规律也同样如此。很可能这也同样适用于基本的心理物理规律：不需要存在一种不同于物理属性和心理属性本身的因果关系。

到目前为止，对交互作用论最有影响的反驳是辩称它和物理学水火不容。人们普遍的看法是：科学告诉我们微观物理领域是因果封闭的，因此心理状态没有任何发挥作用的空间。交互作用论者有可能作出各种回应。例如，有的人可能指出，虽然实验研究没有发现这些作用，但也没有完全排除它们。有人可能进一步指出，物理理论允许存在许多基本**力**（目前是四种，但新的**力**被发现始终是可能的），一种与心理领域有关的额外的力将是现行物理理论的一种合理扩展。这些看法都依赖于对物理理论作重大修改，因此它们是无法轻松实现的；但毕竟仍有实现的可能。

到目前为止，对这种反驳最有力的回应是指出：当代物理学绝没有排除交互作用论，而且它还积极地支持这种可能性。根据对量子力学的标准阐述，世界的状态是由一种波函数描述的，根据波函数，物理实体经常处于一种叠加状态之中（例如，处于两个不同位置的叠加之中），但是叠加从未被直接观察到。根据标准的动力学，波函数能按两种方式进化：一种是根据薛定谔方程的线性进化（它倾向于产生叠加状态）；另一种是从叠加状态到非叠加状态的非线性**坍缩**（collapse）。薛定谔进化是决定性的，但坍缩是非决定性的。薛定谔进化是持续进行的，但是根据标准的阐述，坍缩只是偶然发生的，即只在测量时发生。

坍缩动力学为一种交互作用论解释敞开了大门。可以认为，任何物理的非决定论都给非物理作用留下空间，但坍缩原则所做的远远不止于此。 *126*
坍缩应该发生在测量的时候，虽然对于什么是测量还没有公认的定义，但是有一种事件，每个人都同意它是一种测量——有意识的观察者所作的观察。此外，任何纯物理的测量标准似乎都是无效的，因为纯物理的系统受

线性的薛定谔动力学控制。因此，人们自然会提出：测量就是一种有意识的观察，这种有意识的观察会引起坍缩。

这种主张不应该太武断：量子力学并没有把这种关于形势的解释强加于我们，对量子力学还有一些替代性的解释，根据这些解释，不存在坍缩，或者说根据这些解释，测量在坍缩中并没有特殊的作用。[25]然而，量子力学同这样一种解释似乎是完全**相容的**。事实上，有的人可能会认为，假如他设计了得体的物理学规律使有意识的心灵能发挥一种作用，那么他所做的就不会超过标准量子力学的双向动力学：一种原则控制着正常情况下的决定性进化，另一种原则控制着特殊情况下的非决定性进化，这些情况初看起来都与心理联系在一起。

当然，这种对量子力学的解释是有争议的。很多物理学家反对它，就是因为它是二元论的，它赋予了意识一种基本作用。这种反驳不足为奇，但是当我们有独立的理由认为意识可能是基本事实时，它就没有任何力量。这一事实中暗含着某种讽刺意味，即哲学家们主要基于物理的理由来反对交互作用论[26]（它与物理理论是不相容的），而物理学家们则主要基于哲学的理由来反对关于量子力学的交互作用论解释（它是二元论的）。如果结合起来看，这些理由是没有多少说服力的，根据本章其他地方的反唯物主义论证时尤其如此。

若想对这种解释作出评价，就需要对它作出详细阐述。[27]我认为，这种解释最有希望的版本会承认意识状态与整个系统的量子状态是相联系的，与这个额外的约束条件也是相联系的：意识状态（它们不同于物理状态）绝不能被叠加。在像大脑这样有意识的物理系统中，这个系统的物理状态和现象状态在（非叠加的）量子状态中是联系在一起的。这样一来，根据对叠加系统的观察，薛定谔进化在观察时会使被观察的系统与大脑相联系，产生最终的大脑状态的叠加，并因此产生（根据心理物理的相关性）意识状态的叠加。但这样一种叠加不可能出现，因此作为结果出现的一种潜在意识状态以某种方式被选择了（很可能是被现象层次上的非决定性的动力学原则）。结果就是：（根据心理物理的相关性）一种明确的大脑状态和一种被观察对象的明确状态也被选择了。这可能同样适用于意识与大脑中的非意识过程之间的关系：当叠加的非意识过程将要影响意识时，就会有某种选择。这样，意识在物理世界中就产生了一种因果作用。

127 （有意思的是，这种理论或许能从经验上检验。在量子力学中，坍缩理论产生的预测与非坍缩理论的稍有不同，关于坍缩位置的不同假说作出

了各不相同的预测，但这些不同是极其微小的，目前无法测量。如果有一天能做相关的实验，那么有些结果就会成为使我们接受坍缩理论的强有力的理由，这反过来又会成为我们承认意识作用的理由。作为额外的好处，这对于意识的出现甚至能产生一种经验的标准。）

关于这种观点的准确表述，它与更一般的物理理论（例如，相对论和量子场理论）的相容性以及它在哲学上的可行性（例如，这种观点产生了我们以为意识一定会有的因果作用吗）还有其他很多问题。但我们不能断言物理理论会直接排除有交互作用理论的可能性。在提到量子力学这样的特殊理论时，提出这种主张的人经常会感到震惊；但很明显这只是一些不一致的态度。如果物理学应该排除交互作用论，那么我们就需要细心地注意物理理论的细节。

所有这些都表明，交互作用论仍有发展的空间，对交互作用论最常见的反驳都是无力的。当然，这并不是说交互作用论是正确的。物理世界是因果封闭的，这种观点很有吸引力，对于这种假说，即行为不能完全根据物理的原因来解释，几乎从认知科学得不到任何直接的证据。然而，如果我们有独立的理由认为意识是不能还原的，如果我们希望保留意识起一种因果作用的直觉观点，那么这就是需要特别认真对待的观点。

5.10 E类二元论

E类二元论认为：现象属性从本体论上说与物理属性是不同的，现象对物理没有任何作用。[28]这种观点通常被称为**副现象论**（下文称E类）：物理状态能引起现象状态，反之则不然。根据这种观点，心理物理规律只在一个方向上起作用，即从物理的东西到现象的东西。这种观点自然是和物理领域乃因果封闭的观点结合在一起的：这种进一步的主张对于E类二元论并不是必要的，但它为这种观点创造了很多动机。

就像D类二元论一样，E类二元论和物质二元论——具有独特的物理和心理物质或实体——是相容的，它与属性二元论——一种物质或实体具有两种属性——也是相容的。此外，它与像布罗德那样的突现论也是相容的。根据突现论，心理属性从本体论上说是基本实体新的突现属性，但在这种情况下，虽然存在突现性质，但并没有突现的下向因果关系。

E类二元论的提出，通常既尊重了意识又尊重了科学：它同时涵盖了关于意识的反唯物主义论证和物理的因果封闭性。同时，E类二元论通常由于极其违背直觉而受到了抵制。如果E类二元论是正确的，那么现象状态对我们物理地解释的行为就没有任何作用。例如，疼痛感觉在我的手从火焰上移开时就没有起因果作用；我作出决定的经验在我移民到新国度时就没有起什么因果作用；红色的感觉在我产生“我现在正经验红色”的话语时并不起什么因果作用。这些结论经常被认为是明显错误的，或者说至少是难以接受的。

然而，E类二元论者可以答复说：没有什么直接**证据**与他们的观点相冲突。我们的证据只揭示了现象状态与行为之间的正常关系，因此体会某种经验之后通常会有某种行动。在接触这种恒常的联系时，我们会对因果关系产生强烈的**信念**（就像休谟在另一个语境下所指出的）；但它与因果关系的缺失却是相容的。的确，如果副现象论是正确的，我们似乎就有完全相同的证据来证明它，而且我们还会基于相同的理由被引导到这种信念：意识有一种因果作用。因此，换句话说，如果副现象论是一致的且可以接受的，那么这些考虑似乎就没有提供反对它的有效理由。[29]

另一种反驳认为：如果意识是副现象，那么它就不可能是通过自然选择进化的。然而，E类二元论者有一种直截了当的答复。根据E类观点，存在基本的心理物理规律，它们把物理属性与现象属性联系了起来。如果进化选择了适当的物理属性（可能涉及大脑中的物理或信息的构型），那么这些心理物理规律就能保证现象属性也被例示。如果这些规律有正确的形式，那么一个人甚至能认为：就像更复杂的物理系统会被选择一样，更复杂的意识状态也会进化。这样，物理进化会将意识视为一种副产品，带着它同时进化。

对副现象论最有趣的反驳，可能关注的是意识和对意识的表征之间的关系。有人认为意识在我说“我有意识”时没起任何因果作用，这无疑至少是很奇怪的。有的人已经更有力地指出：这排除了任何关于意识的**知识**。人们常常认为，如果一种关于X的信念是合格的知识，那么这个信念就一定是由X以某种形式引起的。但是，如果意识对物理状态没有作用，再如果信念是物理地构成的，那么意识就不能引起信念。即使信念不是物理地构成的，我们也不清楚副现象论如何能容纳意识与信念之间的因果关系。

在回应时，副现象论者可以否认知识始终需要因果关系。有的人可以

根据独立的理由主张：意识与关于意识的信念之间存在更强的关系：意识在**构成**现象概念和现象信念时起着一种作用。例如，一种红的经验在构成关于一个人现在有一种红的经验这种信念时就起着一种作用。倘若如此，那么经验与信念之间就没有因果的距离。有的人可能认为，这种与经验和信念的直接关系会使信念得到证明。如果这是正确的，那么副现象论对意识的知识就不会构成障碍。

一种相关的反驳认为：我的怪人孪生人能产生相同的报告（例如，“我有意识”），这是由相同的机制引起的，而他的报告没有得到证明；倘若如此，我自己的报告就也是未被证明的。有的人在回应时主张：真正证明的载体是信念，我的怪人孪生人和我有不同的信念，涉及不同的概念，这是由于意识在构成我的概念而非怪人的概念时所起的作用。此外，我们产出了同型的报告，这个事实暗示了第三人称的观察者在相信我有意识方面和怪人有意识方面同样未得到证明，但它并未暗示第一人称的证明方面的不同。对于我关于我有意识的信念，第一人称的证明无论如何都不是根据我的报告，而是根据我的经验本身，也即是怪人所没有的经验。

我认为，这里对于副现象论并不存在强有力的反驳。然而，必须承认这种情况至少是奇特且违背直觉的。根据副现象论者的观点，副现象论的这种奇特之处被以下事实强化了：意识与关于意识的报告之间的关系似乎存在一种幸运的巧合。毕竟，如果心理物理规律不依赖于物理的进化，那么就有这样一些可能世界，那里的物理进化和我们的相同，但是心理物理规律迥然不同，从而使报告与经验极不匹配。看起来幸运的是，我们生活在这样一个世界上，它的心理物理规律使两者很好地匹配了起来。副现象论者在回应时可能想提出这样的论据：这些规律在某些方面是最“自然的”，是可以期待的；但这里仍存在重要的举证责任。

总之，我认为副现象论是一种没有致命问题的融贯观点。同时，它又是一种笨拙的观点，产生了一幅断裂的自然图景，根据这幅图景，物理属性和现象属性在自然界中只是非常贫弱地结合在一起。当然它也是一种违背直觉的观点，但很多人发现难以接受它。笨拙和违背直觉比不融贯要好；因此，如果有正当的论证促使我们把副现象论看成最融贯的观点，那么我们就应该认真对待它。但同时，我们也应该认真考察别的观点。

5.11 F类一元论

F类一元论是这种观点，即意识是由基本物理实体的内在属性构成的；
130 也就是说，是由基本物理倾向的范畴基础（categorical bases）构成的。[30]根据这种观点，现象或者原现象的属性位于基础的物理实在层次上，它们在某种意义上是物理实在本身的基础。

这种观点的线索来自于伯特兰·罗素在《物的分析》（*The Analysis of Matter*，1927）一书中对物理学的讨论。罗素指出：物理学是根据它们相互间的关系及其与我们的关系来描述物理的实体和属性的。例如，夸克是根据它与其他物理实体的关系描述的，质量之类的属性是依据相关的倾向作用描述的，如抵制加速的倾向。同时，物理学对这些实体和属性的内在本质未作任何说明。无论在哪里，只要我们有关系和倾向，我们就能期待有一些为这些倾向奠定基础的基本内在属性，它们描述着具有这些关系的实体。[31]但是，物理学对夸克的内在本质或者对起着与质量相类似作用的内在属性保持了沉默。因此，这就是一个形而上学问题，即基本物理系统的内在属性是什么？

同时，还有另一个形而上学问题：现象属性是如何与物理世界相结合的？现象属性似乎就是难以与物理理论的结构/动力学特征相适应的内在属性；可以说，它们是我们唯一具有直接知识的内在属性。罗素的观点是，我们或许能同时解决这两个问题。物理世界的内在属性本身可能就是现象属性，或者说物理世界的内在属性可能不是现象属性，但却构成了现象属性；也就是说，它们可能是原现象属性。倘若如此，那么意识和物理实在就紧密地纠缠在了一起。

这种观点认为，在自然界中，把现象属性和物理属性紧密地结合起来是有希望的。这里，自然是由具有内在（原）现象属性的实体构成的，这些实体在时空中具有多方面的因果关系。就我所知，物理学就是从这些实体之间的诸关系中产生的，而且意识是从它们的内在本质中产生的。作为额外的收获，这种观点与微观物理的因果封闭性及现有的物理规律是完全相容的。这种观点能保留物理理论已有的**结构**，它只是给这种结构增加了一种内在的本质。这种观点承认意识在物理世界中有明显的因果作用，即（原）现象属性是所有物理的因果关系的最终范畴基础。

这种观点与唯物主义和二元论都有共同因素。从一个角度看，它可以被看成是一种唯物主义。如果一个人认为物理术语不是指称倾向属性而是基本的内在属性，那么这些原现象属性就能被看成是物理属性，从而能支持一种唯物主义。从另一个角度看，它可以被看成是一种二元论。这种观点承认现象或原现象的属性从本体论上说是基本的，它支持结构—倾向属性（物理理论中被直接描述的属性）与内在的原现象属性（对意识负责的属性）之间有一种基本的二元性。有人可能指出，虽然这种观点符合唯物主义的形式，但它却也暗含了反唯物主义的精神。

在其原现象的形式下，这种观点可以被看成是一种中立一元论：存在基本的中立属性 X（原现象属性），以至于这些 X 属性同时是构成物理领域（根据它们的关系）和现象领域（根据它们共同的内在本质）的原因。在其现象的形式下，它可以被看成一种唯心论，以至于心理属性构成了物理属性。但这些心理属性不一定是观察者心灵中的心理属性，它们可能必须另外补充因果和时空的属性。有的人也可能把这种观点描述成一种泛心论，即现象属性在基础层次上无处不在；还有人可能会将这种观点的最一般形式称为**泛原心论**，即要么原现象属性，要么现象属性构成了所有物理实在的基础。

对于反唯物主义的怪人论证，F 类一元论者所持的态度可能是诸种态度中的一种。有些 F 类一元论者可能认为，完备的物理描述一定能被扩展，从而涵盖内在的描述，因此他们否认怪人是可以想象的。（我们仅仅认为，我们是在想象一个在物理方面同一的系统，因为我们忽略了内在属性。）有的人可能认为，现有的物理概念通过倾向来指称那些为这些倾向奠定基础的内在属性。倘若如此，那么这些概念就有不同的首要内涵和次要内涵。与此相应，F 类一元论者就能承认可想象性但否认可能性：我们错误地描述了这个被想象的世界，认为它在物理方面是和我们的世界同一的，而此时它实际上仅仅在结构上与我们的世界相同一。[32] 最后，F 类一元论者可能认为，物理概念指称倾向属性，因此怪人既是可以想象的也是可能的，内在属性也不是物理属性。这三种态度之间的区别最终似乎只是术语学而非实质的区别。

至于知识论证，F 类一元论者可能主张：玛丽要想有完备的物理知识，就必须有一种关于世界的描述，这种描述包含有直接描述内在属性的概念；如果她有这种描述（而不是她那些虽然无力却包含倾向概念的描述），她或许因此就能够知道红色是什么样子。就解释论证来说，F 类一元论者

可能认为：涉及内在属性的物理解释能解释的不只是结构和功能。换言之，坚持倾向的物理概念的F类一元论者所作出的回应，与上述另外两种回应之一是相似的。

F类观点确实是推测性的，它初听起来可能还很奇特。很多人都发现这种看法是极其违背直觉的，即认为基础的物理系统具有现象属性。例如，认为现实存在成为一个电子是什么样子的东西。这种观点的原现象版本反对这一主张，但它保留了它的一部分奇特之处：任何对构成意识负责的属性似乎都一定是奇特的、不寻常的属性，是我们在微观物理实在中不
132 可能找到的一种属性。然而，我们还不知道这种奇特之处就能产生强有力的反驳。就像副现象论一样，这种观点与我们的所有证据似乎都一致，也没有直接的证据能反对它。有的人可能认为，如果这种观点是正确的，那么事物表现出来的样子就和它们的实际表现一样。我们已经从现代物理学认识到，世界是一个奇怪的地方：我们不能期望它遵守常识的所有命令。

有人也可能提出反驳，说对于原现象属性可能是什么样子，或者对于它们何以能构成现象属性等，我们还一无所知。这是正确的。但有的人可能指出，这只是我们无知的结果。就熟悉的物理属性来说，否定与现象属性的构成性关系是有原则性理由的（根据物理概念的特征）。这里，并没有这样的原则性理由。至多有无知和关系的阙如。对原现象属性形成一种正面的概念，当然是非常令人高兴的。或许我们能间接地做到这一点，即通过某种从现象属性的特征到它们的基本成分的理论推理；另一种可能是：关于原现象属性之本质的知识超出了我们的能力范围。无论如何，这不是反对这种观点的正确理由。[33]

就此还有一个原则问题，它是由威廉·詹姆斯指出来的（1890：第6章）。我们的现象学中有一种丰富而特殊的结构：它是统一的、有限的、被区分成很多不同的方面，但是它与很多方面都有一种基本的同质性（homogeneity），而且它似乎有单一的经验主体。要想明白大量个别的微观物理系统——每一个都有其自身的原现象属性——的分布何以能以某种方式形成这种丰富而特殊的结构，那是很不容易的。人们不该期望像一堆长短不一、参差不齐的现象学道钉之类的东西吗？

这就是泛心论**组合问题**（combination problem）的一个版本（西加[Seagar]，1995），或者说是斯托尔加（Stoljar，2001）所说的罗素观点中的**结构不匹配**问题（参见福斯特[Foster]，1991：119-130）。要回答它，我们似乎需要更好地理解现象学的**组合**原则：也就是说，正因为这些原

则，现象属性才能从基本的现象属性或者原现象属性组合或构成。我们对物理组合的原则有了很好的理解，但并未真正理解现象组合的原则。这是应当密切关注的一个领域：我认为它很容易成为F类一元论观点中最严重的问题。在这一点上，这个问题能否解决还是一个悬而未决的问题。

有些F类一元论者似乎认为：他们可以通过主张现象属性是**高层次**物理倾向的内在属性（例如，包含在神经状态中的那些属性）来摆脱组合问题，因为它们不一定是由微观物理状态的内在属性构成的（因此，他们可能也否定泛原心论）。但这似乎是站不住脚的：如果低层次网络是因果封闭的，同时高层次的内在属性不是由低层次的内在属性构成的，那么基于熟悉的理由，高层次的内在属性会再次转变成副现象。要接受这一立场，唯一的办法似乎是与否定微观物理的因果封闭性结合起来，主张在微观物理的层次之上存在基本的倾向，它们与其基础一样具有现象属性。但这样一种观点是难以与D类二元论相区别的。[34]因此，一种与众不同的F类一元论一定会直接遇到组合问题。

总之，F类一元论承诺了一种非常综合、非常得体的自然观。还没有人提出任何一种属于这一类的具体理论，我们也不知道这样一种理论能否提得出来。但与此同时，反对这种观点似乎也没有什么令人信服的理由。因此，F类一元论有可能为将来的研究奠定坚实的基础，它可能最终会在自然界中为物理之物和现象提供最好的综合。

5.12 结 论

非还原论者还有别的选择吗？有两种观点可能不直接属于上述诸类型。

首先，有些非唯物论者认为，从本体论上说，现象属性与物理属性截然不同。微观物理学是因果封闭的，但现象属性却对物理的东西起着一种因果作用。出现这种情况的唯一途径是通过一种多因素的因果决定关系（overdetermination）：物理状态因果地决定着行为，而现象状态也同时引起了行为。另一种方法是通过因果的中介：或许，至少在从A到B的微观物理的因果关系的一些实例中，真有一种从A到心灵再到B的因果关系，从而使心灵在进入因果关系时却没有改变这一网络结构。这里可能还有其他的策略。我们可以将这一类称为O类二元论（把决定过多当成一个范

例)。这些观点和E类观点有很多共同的结构(因果封闭的物理世界、独特的现象属性),但避开了副现象论的指责。我们对这些观点的特殊因果设置可能是难以接受的,而且它们和E类观点有一些相同的问题(例如,断裂的自然观、“巧合的”心理物理规律),但是这一类却应当公开地视为一种选择。[35]

其次,有些非唯物论者就是唯心论者(在贝克莱的意义上)。他们认为,物理世界本身是由在进行观察的自主体的意识状态构成的。我们可将这种观点称为I类一元论。它和F类一元论都有这种属性:现象状态在构成物理实在方面起着一定作用,但是根据I类观点,这一现象发生的方式是很不相同的:不是由于具有为每一物理状态提供基础的独立“微观”现象状态,而是由于具有由一种“宏观的”现象心灵整体构成的物理状态。在比上述观点深刻得多的意义上,这种观点似乎是非自然主义的,尤其是它似乎在本质上失去了因果的或解释的封闭性:一旦根据外部世界所作出的自然解释被抛弃了,就必须认定:现象状态之间极其复杂的规律性不能根据更简单的原则来解释。但是,这种观点同样至少是应当予以承认的。

就我的理解来说,非还原论者最好的选择是D类二元论、E类二元论或者F类一元论:即互动论、副现象论或泛原心论。如果我们承认物理之物与现象之间存在认识性鸿沟,如果我们排除了原始的同一性和强必然性,那么我们就会被引导到了这三种观点中的一种。其中每种观点至少都有某种希望,没有一种有明显的致命缺陷。对我来说,我对其中的每一种都有些相信。我认为,在有些方面F类观点最有吸引力,但这种感觉主要还是建立在对其说服力一无所知的美学考虑之上。

如果需要从这三种观点之间作出选择,可能选择人在很大程度上将依赖于特定的理论在这些框架内的发展。尤其是就D类观点和F类观点来说,进一步的理论研究在对这些理论作出评价方面是至关重要的(例如,在解释量子交互作用论或者在理解现象的组合方面)。也可能经验的意识科学会提供一些引导。随着科学的发展,我们会受到引导,能推出为物理状态和现象状态之间相关性奠定基础的简单原则,这些原则可能最终会坚定地指向这些观点中的某一种。例如,如果将微观物理状态与现象或原现象的状态联系起来的简单原则能做解释工作,那么我们可能就有理由支持F类观点;然而,如果这些原则是在更高层次上理解物理世界的,那么我们可能就有理由支持D类或E类观点。如D类观点所示,如果意识对物理

世界有一种特定方式的作用，那么尽管必须付出艰苦努力，但经验研究在原则上是能发现这些作用的。

并非每个人都同意这些观点中的每一种都是可行的。进一步研究就有可能暴露出某些观点中的深层次问题。但是进一步的研究是一定要做的。例如，至今对F类观点的批判性考察还很少；我们已经看到，标准的反D类观点的论证是无足轻重的；与此同时，虽然反E类观点的论证有某种直觉的说服力，但是它们绝没有成为反对这些观点的强有力论据。我猜想，即使进一步的研究暴露出了相关观点的深层次问题，但所有这些观点都被取消也是非常不可能的。

无论如何，这为我们提供了某种关于心身问题的观点。人们经常认为，即使很难理解唯物主义如何是正确的，唯物主义也一定正确，因为替代者是难以接受的。但在我看来，唯物主义至少有三种公开的、可以接受的替代者，其中每一种都与广义的自然主义（即使不是唯物主义的）世界观相容，其中没有一种存在致命的问题。因此在我看来，鉴于这些明显的反唯物主义论证，我们至少应该暂时接受这个结论：其中一种观点是正确的。当然，若要对这一章所讨论的所有观点作出全面的评价，它们都需要更详细的发展，并需要依据所有相关的科学和哲学的发展作出考察。但是从当前情况看，我认为我们有正当的理由认为：意识在自然中占有基础的地位。

【注释】

[1] 本章是对有关意识的形而上学问题所作的一种综述。本章的很多讨论（尤其是第一部分）扼要概括了查默斯（1995、1996、1997）著作中的讨论，尽管它经常采取不同的形式，甚至有时还超出了那里讨论的范围。在参考文献所引证的著作中，我对这里所讨论的很多问题有更详细的论述。

[2] 这种分类出现在布罗德著作的最后一章。在安德鲁·克鲁基（Andrew Chrucky）专门论述布罗德的网站上，他将布罗德的分类形象地改编成一个4×4的矩阵，并通过图示作出了解释，网址是：http://www.ditext.com/broad/mpn14.html#t。

[3] 按照我的用法，感受性质就是那些依据具有它们是什么感觉来描述意识状态的属性。这个定义并未加入任何别的实质性要求，例如感受性质是内在的或者说是非意向的。如果感受性质是内在的或者非意向的，那么这就是一种实质性而非定义性观点（因此，意识的属性是非内在的或者说它们完全是意向的，这种主张就不应当被说成是不存在感受性质）。现象属性也能被看成是个体的（例如，人）而非心理状态的属

性，这些属性描述了在某个时间点成为这些个体是什么感觉的方面；这种不同对于现在的目的并不重要。

[4] 注意，我所用的“还原”的意义比它在日常所用的意义要宽。还原的解释只要求高层次现象能完全根据低层次现象来解释。这与高层次现象在低层次现象上的“可多样实现性”是相容的。例如，消化在生理系统中可能有很多不同的实现方式，但一个人却能根据基本的生理学对系统的消化作出还原的解释。另一个细微之处与这种可能的观点有关：意识能依据一些并不诉诸意识，但它们自身不能从物理上解释的原则来解释。上面的定义认为这种观点既不是还原的也不是非还原的。不论把它归入哪一类都有道理，但我一般把它归入非还原的类型。

[5] 这里所陈述的解释论证的版本是在查默斯（1995）的书中给出的。对解释的相关考虑，参见莱文（1983）对“解释的鸿沟”的论述和内格尔（1974）。也参见载于希尔（Shear，1997）著作中的论文。

[6] 可想象性论证的诸版本是由坎贝尔（1970）、柯克（1974）、克里普克（1980）、比勒（1994）和查默斯（1996）等人提出来的。举足轻重的先驱包括笛卡尔关于离体现象的可想象性论证和莱布尼兹的“磨房”论证。

[7] 知识论证的创始人包括内格尔（1974）、麦克斯韦 1968）、杰克逊（1982）等人。这个论证的前驱出现在布罗德（1925：70-71）关于“数学大天使”（mathematical archangel）——他不能从物理事实演绎出氨的气味——的讨论中，以及费格尔（1967[1958]：64，68，140）关于“火星的超级科学家”的讨论中，这种科学家不知道颜色看起来是什么样子，乐音听起来是什么样子。

[8] 这一思想实验的版本在克努特·诺德比（Knut Nordby）身上有一个真实的标本。克努特·诺德比是一位挪威感觉生物学家，他是一个视杆全色盲患者（他的视网膜中缺乏用于颜色视觉的圆锥细胞），但他却从事颜色视觉的生理学研究。参见诺德比（1990）。

[9] 对于诸种可想象性论证和解释论证中缺乏创见的版本，参见布罗德（1925：614-615）。对于知识论证，参见氏书 70～72 页，布罗德在那里说：即使“数学大天使”也不能从关于原子的微观知识演绎出氨的气味。布罗德反对“机械论”，机械论大致相当于当代的唯物论。从当代的眼光看，布罗德论证中最大的缺陷可能是关于这种可能性的考虑：存在认识性的鸿沟，但不存在本体论的鸿沟。

[10] 对于可想象性论证与知识论证之间关系的讨论，参见查默斯（1996、2002b）。

[11] A 类唯物主义者包括赖尔（1949）、刘易斯（1988）、丹尼特（1991）、德雷斯基（1995）、雷伊（1995）和哈曼（1990）。

[12] 有两种特殊的观点可能值得一提：（1）有些观点（例如，德雷斯基，1995）主张意识不仅涉及功能作用，还涉及和环境中的对象的因果和历史的关系，因此否定了认识性鸿沟，不过也同时否定了功能主义。我把这些观点看成 A 类观点：我们可以

把相关的关系看成是从广义上理解功能作用的延伸，而且双方会出现完全相同的考虑。(2) 有些观点（例如，斯特劳森，2000；斯托尔加，2001）对认识性鸿沟的否定，不是想对意识作功能的分析，而是扩展我们对物理基础的看法，使之涵盖基本的内在属性。这些观点是在 F 类下讨论的（第 5.11 节）。

[13] 丘奇兰德（1996）用另一个类比指出：某个歌德时代的人可能提出了类似的认识性论证，用以反对关于“发光”的还原解释。但是，在认真考察这个类比之后不难看出，在这里可能招致怀疑的其他解释对象只有看见光的经验（参见查默斯，1997）。这一点对 A 类唯物主义是没有帮助的，因为这个解释对象仍未得到解释。

[14] 对于从不受欢迎的形而上学结果出发所作的一种论证，参见怀特（White，1986）。对于从不受欢迎的认识论结果出发所作的论证，参见休梅克（1975）。这些形而上学结果是在本章的后半部分处理的。认识论结果则是在查默斯 2002 年第一部著作中处理的。

[15] B 类唯物主义者包括莱文（1983）、洛尔（1990/1997）、帕皮诺（1993）、泰伊（Tye，1995）、莱肯（1996）、希尔（Hill，1997）、布洛克和斯托尔内克（1999）、佩里（Perry，2001）。

[16] 在某些方面，只要 A 类唯物主义能被看成是派生于赖尔和卡尔纳普的逻辑行为主义，B 类唯物主义就能被看成是派生于普赖斯和斯马特的同一论。然而，这个问题被以下事实弄复杂了，即早期的同一论者支持关于现象属性的“主题中立的”（功能的）分析，他们提出了一种基本的 A 类唯物主义。

[17] 布洛克和斯托尔内克（1999）部分地是通过这种主张来反对可演绎性的：对于像“水”这样的高层次术语，通常并没有用微观物理术语或者说任何其他术语——它们能为从微观物理事实向关于水的事实的先验衍推奠定基础——所作出的明显的概念分析。在答复时，查默斯和杰克逊（2001）称：明显的概念分析不要求先验的衍推，他们有正当的理由相信在这些案例中存在这样的衍推。

[18] 两维语义学的框架源于斯托尔内克（1978）、埃文斯（Evans，1979）和卡普兰（Kaplan，1989）的工作。这些论证中所用的版本有点不同：对于这些不同的讨论，参见查默斯（即出）。

[19] 这是查默斯著作（1996：131-136）中论证较正式一点的版本。它与克里普克的反同一论模态论证是紧密相联的，尽管它们在一些重要的方面是有区别的。核心的前提 2 可以被看成是一种对克里普克的主张进行形式化的方法。克里普克的主张是：只要有“表面的偶然性”，背景中就有某种被错误描述的可能性。这种论证也能被看成是对“双重属性”（dual property）反驳的一个版本进行形式化的方法，这种论证被斯马特（1959）归功于马克斯·布莱克（Max Black），并被杰克逊（1979）和怀特（1986）发展了。这个两维框架在唯物主义问题上的相关运用是由杰克逊（1994）和刘易斯（1994）提供的。

[20] 我这里省略了一些细微之处。其中之一涉及索引词的作用：要处理像“我在

这里”这样的主张，首要内涵就要在中心世界（centered worlds）上定义：这些世界具有作出了标记的个体和时间，这相当于一个人在世界上所处位置的索引性的“定位信息”。然而，这种变化对于B类唯物主义者是没有帮助的。即使我们给P增加了索引性的定位信息I（例如，告诉玛丽她在世界里的位置），Q的认识性鸿沟仍像以前一样宽；因此，P∧I∧¬Q是可以想象的。假如有一个证实了P∧I∧¬Q的中心世界，那么，如上所述，一个人就能看到：要么存在一个满足了P∧¬Q的世界，要么F类一元论是正确的。

[21] 希尔（1997）打算用认知的术语把我们关于意识的模态直觉解释清楚。查默斯（1999）回应说：任何模态直觉都能用认知术语来解释（相似的论证能把我们关于可能存在红色正文形的直觉“解释清楚”），但这并不是想说这个直觉是不正确的。如果这种解释告诉我们关于意识的模态直觉是不可靠的，那么所有的模态直觉也会同样如此。因此，真正需要的不是解释我们关于意识的模态直觉，而是解释为什么这些特殊的直觉不可靠。

洛尔（1990/1997）试图依据现象概念的独特性来提供这样一种解释。他指出：（1）现象概念是识别概念（“**那**类东西”）；（2）它们和其他识别概念一样，能够与在认知方面不同的物理概念分享共同的指称；（3）和其他识别概念不同，它们缺乏偶然的呈现模式（即它们的首要内涵和次要内涵是重合的）。如果（2）和（3）都成立（如果我们假定物理概念也缺乏偶然的呈现模式），那么现象—物理的同一性在上述意义上就是一种强必然性。在回应时，查默斯（1999）认为，（2）和（3）是不能都成立的。其他识别概念与理论概念有共同指称是以它们的偶然呈现模式**为基础**的；如果缺乏这些呈现模式，就没有理由认为这些概念能有共同的指称。因此接受（3）就削弱了对
138 （2）的支持。查默斯（1999）还认为，由于假定了物理属性能非偶然地具有现象的呈现模式，因此洛尔的解释实际上是预设而不是解释了相关的强必然性。

[22] 就前面提到的明显同情C类唯物论的人来说，我认为麦金最终是F类一元论者，内格尔要么是B类唯物论者，要么是F类一元论者，而丘奇兰德要么是B类唯物论者，要么是Q类唯物论者（解释见下文）。

[23] 有的人可能问的是具体的还原观点，如表征主义（它将意识同一于某些表征状态）和高阶思想理论（它将意识同一于高阶思想的对象）。如何对这些观点进行分类取决于给定的理论家如何看待表征或者高阶的状态（例如，能否从功能上定义）以及它们与意识的关系（例如，是概念的关系还是经验的关系）。在表征主义者中间，我认为哈曼（1990）和德雷斯基（1995）是A类唯物主义者，而泰伊（1995）和莱肯（1996）是B类唯物主义者。在高阶思想理论家中，卡拉瑟斯（Carruthers，2000）显然是B类唯物主义者，而罗森塔尔（1997）要么是A类唯物论者，要么是B类唯物论者。人们在原则上也能对这些观点中的每一种持非唯物主义的版本。

[24] D类二元论者包括波普尔和埃克尔斯（1977）、塞拉斯（1981）、斯温伯恩（Swinburne，1986）、福斯特（1991）、霍奇森（1991）和斯塔普（Stapp，1993）。

[25] 无坍缩（no-collapse）解释包括博姆（Bohm）的“隐变量”解释和埃弗里特（Everett）的“多种世界”（或者“多种心灵”）解释。一种不使用测量的坍缩解释是吉拉迪-里米尼-韦伯（Ghirardi-Rimini-Weber）解释（它有随机的偶然坍缩）。这些解释中的每一种都要求对标准的量子力学动力学作出重大修改，而且每一种都有争议，尽管每种也都有其长处（对这些和别的解释的讨论，参见艾伯特，1993）。值得注意的是，除了涉及意识的解释之外，似乎没有一种站得住脚的解释能支持这种权威的主张：坍缩是在测量时发生的。

[26] 在这一点上我和任何人一样也有过失，我曾取消了交互作用论（查默斯，1996），部分原因是为了和物理学相一致。虽然我仍然不是很乐意接受交互论，但我现在认为从物理学出发进行的论证过于轻率了。在上书中我还提到了反对这种观点的另外三种理由。第一，如果意识要对行为作出引人注意的质的影响，就要求它非随机地起作用，而这违背了量子力学的或然性要求。我认为这有些道理，但一个人可能要容忍反应中的非随机性了，或者说他可能会认为，即使意识的随机因果作用也是非常好的。第二，我认为，否定因果封闭性不会产生特别的好处，因为一种有因果封闭性的观点借助F类一元论能达到完全相同的结果。这同样有些道理，但D类观点没有摆脱F类观点的“组合问题”的巨大优势。第三，研究者还不清楚坍缩解释会产生我们原以为意识所具有的这种因果作用。我认为，这是一个需要详细研究的重要问题。

[27] 量子力学的意识—坍缩解释是由魏格纳（Wigner，1961）、霍奇森（1991）和斯塔普（1993）提出来的。只有斯塔普的著作涉及大量细节，他有一种很有意思但有点特别的解释，所采取的解释方向也与上面提到的不同。

[28] E类二元论者包括赫胥黎（1874）、坎贝尔（1970）、杰克逊（1982）和鲁宾逊（1988）。

[29] 有人指责副现象论者持双重标准：在提出反唯物主义的论据时依赖直觉，但在否定意识的因果作用时又与直觉背道而驰。但是，直觉必须基于理性和证据的背景来评价。在反唯物主义的论证中，要否定相关的直觉（尤其是关于其他解释对象的直觉）似乎和能得到的第一人称证据是相矛盾的；但是，经考虑后会发现，否定意识的因果作用似乎又和我们的全部证据，包括第一人称的证据是相容的。

[30] F类一元论的各个版本是由罗素（1927）、费格尔（1967 [1958]）、麦克斯韦（1979）、洛克伍德（1989）、查默斯（1996）、格里芬（Griffin，1998）、斯特劳森（2000）和斯托尔加（2001）提出来的。

[31] 对于所有倾向都有一个范畴基础的论题尚存在哲学的争论。如果这个论题被接受了，那么F类一元论的论据就格外有说服力，因为微观物理的倾向一定会有一个范畴基础，而我们对该基础还没有独立的描述。不过，即使这个论题被否定了，F类一元论仍然可行。在这里，要为内在的属性提供位置，我们只需要这个论题：微观物理的倾向**可能**有一个范畴基础。

[32] 因此，F类一元论是这种“物理主义”，它是从反B类唯物论的两维论证中

所提到的漏洞中产生的。“怪人世界”W满足P的首要内涵而非次要内涵，唯一的方法就是使它和我们的世界有共同的倾向结构但没有共同的基本内在微观物理属性。如果这种区别是造成W中缺乏意识的原因，那么我们世界里的这些内在的微观物理属性就对构成意识负责。麦克斯韦（1979）在答复克里普克的论证时利用了这种漏洞。

注意，这个W一定要么含有一些内在的属性，它们不同于我们世界里的内在属性，要么它根本没有任何内在属性。即使是在关于怪人的物理状态的结构——功能描述之下，主张融贯的内在属性只有原现象属性的F类一元论者最终或许会否定怪人的可想象性——其理由与A类唯物主义者所持理由非常不同。

[33] 麦金（1989）可以被看成是支持F类观点的，但他却否认我们能知道原现象属性的本质。他的论证依赖于这种主张：这些属性无法通过知觉或者通过内省来认识。但这并不排除以下可能性：它们或许能根据某种对（内省的）现象学推出了最佳解释的推理来认识，这要服从（被知觉到的）物理结构的其他限制。

[34] 这样，我们就能看到D类观点和F类观点是密切相联的。我们可以设想：如果D类观点是正确的，的确存在微观物理的因果鸿沟，那么只通过物理的观察就能使我们假设高层次的实体足以填平这些鸿沟——比如说，“心灵粒子”（psychons）——这里的这些高层次实体完全是用结构/倾向术语来描述的。D类观点对此所补充的是以下主张：心灵粒子具有内在的现象本质。D类观点和F类观点的主要区别是：D类观点涉及处于微观物理层次之上的基本因果关系。这与一种更为激进的物理学观点存在联系，但它可能具有摆脱组合问题的优势。

140 [35] O类立场是洛（1996）、米尔斯（Mills，1996）和比勒（即出）所倡导的。

参考文献

Albert，D. Z.（1993）. *Quantum Mechanics and Experience*. Cambridge，MA：Harvard University Press.

Bealer，G.（1994）. “Mental Properties.” *Journal of Philosophy*，91：185–208.

Bealer，G.（forthcoming）. “Mental Causation.”

Block，N. and Stalnaker，R.（1999）. “Conceptual Analysis，Dualism，and the Explanatory Gap.” *Philosophical Review*，108：1–46.

Broad，C. D.（1925）. *The Mind and its Place in Nature*. London：Routledge and Kegan Paul.

Campbell，K. K.（1970）. *Body and Mind*. London：Doubleday.

Carruthers，P.（2000）. *Phenomenal Consciousness：A Naturalistic Theory*. Cambridge：Cambridge University Press.

Chalmers，D. J.（1995）. “Facing up to the Problem of Consciousness.” *Journal of Consciousness Studies*，2：200–219. Reprinted in Shear（1997）. http://consc. net/pa-

pers/ facing. html.

—— (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

—— (1997). "Moving Forward on the Problem of Consciousness." *Journal of Consciousness Studies*, 4: 3-46. Reprinted in Shear (1997). http://consc. net/papers/moving. html.

—— (1999). "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research*, 59: 473-493. http://consc. net/papers/modality. html.

—— (2002a). "The Content and Epistemology of Phenomenal Belief." In Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press. http://consc. net/papers/belief. html.

—— (2002b). "Does Conceivability Entail Possibility?" In T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press. http://consc. net/papers/conceivability. html.

—— (forthcoming). "The Foundations of Two-Dimensional Semantics." http://consc. net/papers/foundations. html.

Chalmers, D. J. and Jackson, F. (2001). "Conceptual Analysis and Reductive Explanation." *Philosophical Review*, 110: 315-361. http://consc. net/papers/analysis. html.

Churchland, P. M. (1996). "The Rediscovery of Light." *Journal of Philosophy*, 93: 211-228.

Churchland, P. S. (1997). "The Hornswoggle Problem." In Shear (1997).

Clark, A. (2000). "A Case Where Access Implies Qualia?" *Analysis*, 60: 30-38.

Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown.

—— (1996). "Facing Backward on the Problem of Consciousness." *Journal of Consciousness Studies*, 3: 4-6.

—— (forthcoming). "The Fantasy of First-Person Science." http://ase. tufts. edu/cogstud/papers/chalmersdeb3dft. htm.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Evans, G. (1979). "Reference and Contingency." *The Monist*, 62: 161-189.

Feigl, H. (1967 [1958]). "The 'Mental' and the 'Physical'." *Minnesota Studies in the Philosophy of Science*, 2: 370-497. Reprinted (with a postscript) as *The "Mental" and the "Physical"*. University of Minnesota Press.

Foster, J. (1991). *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. Oxford: Oxford University Press.

Griffin, D. R. (1998). *Unsnarling the World-Knot: Consciousness, Freedom, and*

the Mind-Body Problem. Berkeley: University of California Press.

Harman, G. (1990). "The Intrinsic Quality of Experience." *Philosophical Perspectives*, 4: 31-52.

Hill, C. S. (1997). "Imaginability, Conceivability, Possibility, and the Mind-Body Problem." *Philosophical Studies*, 87: 61-85.

Hodgson, D. (1991). *The Mind Matters: Consciousness and Choice in a Quantum World*. Oxford: Oxford University Press.

Huxley, T. (1874). "On the Hypothesis that Animals are Automata, and its History." *Fortnightly Review*, 95: 555-580. Reprinted in Collected Essays. London, 1893.

Jackson, F. (1979). "A Note on Physicalism and Heat." *Australasian Journal of Philosophy*, 58: 26-34.

—— (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127-136.

—— (1994). "Finding the Mind in the Natural World." In R. Casati, B. Smith, and G. White (eds.), *Philosophy and the Cognitive Sciences*. Vienna: Holder-Pichler-Tempsky.

James, W. (1890). *The Principles of Psychology*. Henry Holt and Co.

Kaplan, D. (1989). "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Kirk, R. (1974). "Zombies vs Materialists." *Proceedings of the Aristotelian Society* (Supple-mentary Volume), 48: 135-152.

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Levine, J. (1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly*, 64: 354-361.

—— (2000) *Purple Haze: The Puzzle of Conscious Experience*. Cambridge, MA: MIT Press.

Lewis, D. (1988). "What Experience Teaches." *Proceedings of the Russellian Society* (University of Sydney).

—— (1994). "Reduction of Mind." In S. Guttenplan (ed.), *Companion to the Philosophy of Mind*. Oxford: Blackwell.

Loar, B. (1990/1997). "Phenomenal States." *Philosophical Perspectives*, 4: 81-108. Revised edition in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*. Cambridge, MA: MIT Press.

Lockwood, M. (1989). *Mind, Brain, and the Quantum*. Oxford: Oxford University Press.

Lowe, E. J. (1996). *Subjects of Experience*. Cambridge: Cambridge University

Press.

Lycan, W. G. (1996). *Consciousness and Experience*. Harvard, MA: MIT Press.

Maxwell, G. (1979). "Rigid Designators and Mind-Brain Identity." *Minnesota Studies in the Philosophy of Science*, 9: 365-403.

Maxwell, N. (1968). "Understanding Sensations." *Australasian Journal of Philosophy*, 46: 127-145.

McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind*, 98: 349-366.

Mills, E. (1996). "Interactionism and Overdetermination." *American Philosophical Quarterly*, 33: 105-115.

Nagel, T. (1974). "What Is It Like To Be a Bat?" *Philosophical Review*, 83: 435-450.

Nordby, K. (1990). "Vision in a Complete Achromat: A Personal Account." In R. Hess, L. Sharpe, and K. Nordby (eds.), *Night Vision: Basic, Clinical, and Applied Aspects*. Cambridge: Cambridge University Press.

Papineau, D. (1993). "Physicalism, Consciousness, and the Antipathetic Fallacy." *Australasian Journal of Philosophy*, 71: 169-183.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.

Popper, K. and Eccles, J. (1977). *The Self and Its Brain: An Argument for Interactionism*. New York: Springer.

Quine, W. V. (1951). "Two Dogmas of Empiricism." *Philosophical Review*, 60: 20-43.

Rey, G. (1995). "Toward a Projectivist Account of Conscious Experience." In T. Metzinger (ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh.

Robinson, W. S. (1988). *Brains and People: An Essay on Mentality and Its Causal Conditions*. Philadelphia: Temple University Press.

Rosenthal, D. M. (1997). "A Theory of Consciousness." In N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*. Cambridge, MA: MIT Press.

Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson and Co.

Seagar, W. (1995). "Consciousness, Information and Panpsychism." *Journal of Consciousness Studies*, 2.

Sellars, W. (1981). "Is Consciousness Physical?" *The Monist*, 64: 66-90.

Shear, J. (ed.) (1997). *Explaining Consciousness: The Hard Problem*. Cambridge, MA: MIT Press.

Shoemaker, S. (1975). "Functionalism and Qualia." *Philosophical Studies*, 27: 291-315.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review*, 68: 141-156.

Stalnaker, R. (1978). "Assertion." In P. Cole (ed.), *Syntax and Semantics: Pragmatics*, Vol. 9. New York: Academic Press.

Stapp, H. (1993). *Mind, Matter, and Quantum Mechanics*. Berlin: Springer-Verlag.

Stoljar, D. (2001). "Two Conceptions of the Physical." *Philosophy and Phenomenological Research*, 62: 253-281.

Strawson, G. (2000). "Realistic Materialist Monism." In S. Hameroff, A. Kaszniak, and D. Chalmers (eds.), *Toward a Science of Consciousness* Ⅲ. Cambridge, MA: MIT Press.

Swinburne, R. (1986). *The Evolution of the Soul*. Oxford: Oxford University Press.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.

Van Gulick, R. (1993). "Understanding the Phenomenal Mind: Are We All Just Armadillos?" In M. Davies and G. Humphreys (eds.), *Consciousness: Philosophical and Psychological Aspects*. Oxford: Blackwell.

White, S. (1986). "Curse of the Qualia." *Synthese*, 68: 333-368.

Wigner, E. P. (1961). "Remarks on the Mind-Body Question." In I. J. Good (ed.), *The Scientist Speculates*. London: Basic Books.

第6章　思维及其内容：自然化语义学

弗雷德·亚当斯（Fred Adams）

6.1　概　观

众所周知，维特根斯坦提出了这样一个问题：什么使我关于你的思想成为**关于你**的思想？如果真的有关于你的思想，那么我便可以说：你是我思想的**内容**的一部分；你是我的思想所关于的东西的组成部分。

我们能够想到各种各样的事情，如对象（埃菲尔铁塔）、属性（成为一座举世闻名的纪念碑）、事情（塔的建造过程）以及思维本身（认为埃菲尔铁塔是巴黎最著名的纪念碑之一）。这里我不想作详尽的罗列，只想将上述问题拓展成这样的问题："什么使某人**关于 x** 的思想成为关于 **x** 的思想？"我们都知道[1]，我们有关于某些事物的思想。[2]但我们这里感兴趣的是：对它如何发生作出说明。

我们将集中考察**思想**及其内容，不过，信念、愿望、希望、意愿、意向等也常在宽泛意义上被看作是思想。最明显的是，笛卡尔就曾把各种感觉当作是思想的类别，当然人们对它们的考察常常是有区别的。因为它们不是命题态度，且没有真值（尽管它们也可以是真实或不真实的）。很明显，感觉有内容，而且根据某些论述（德雷斯基，1995），感觉和思想在获得其内容时，存在着显而易见的相似性。

自从 20 世纪 70 年代末 80 年代初以来，学界已涌现出了多种将语义学自然化的尝试。尽管各种尝试之间存在着微妙的差异，但它们都坚持这样的观点，即心灵是自然、物理的对象。心灵获得内容的方式也是自然的（或物理的）事情。至少从 20 世纪 70 年代中期以来，外在广义的（externalistic）内容理论便竭力论证说：思想的内容主要依赖于人的环境，（就

大多数思想来说）它并非唯一依赖于头脑内的东西。在地球上的 A1 头脑中，一个物理状态是关于水（H_2O）的思想，而在孪生地球上的孪生地球 A1 的头脑中，完全相同的物理状态可能是关于孪生地球水（XYZ）的思想。这种思想内容上的差异不是根源于地球 A1 或孪生地球 A1（本身就是前者的物理的复制品）任何内在的东西，而是根源于他们各自环境中那种作为水存在的物质。将意义自然化的理论家在意义的外在主义图景中新增加的东西不过是一种机制（mechanism），这种机制能够对外在物理对象怎样与人的大脑（心灵）的内在物理状态相互关联，进而为内在物理状态怎样意指或关于（to be about）外在物理对象提供解释。因此，我们有必要对它作出说明，自然主义的内容理论提供的是自然主义的机制。

意义的机制不可能依赖于意义或内容。因为自然化的目的是将意义自然化，是说明自然界有意义的东西是怎样从无意义的东西中派生出来的。因此我们不能再根据自主体的话语意义或意向去解释思想怎样获得其内容。当然，一旦有内容的思想或者有意义的语言出现了，它们也许能够解释别的意义或内容是怎样产生的。但是为了不致陷入循环解释，我们必须设定某些无意义的、让意义得以实现的东西。对思想内容的自然主义说明必须诉诸能产生思想和内容的机制，同时在解释时又不利用思想或内容，至少从字面上说是这样的。也许，思考自然主义的一种方式就是思考：世界上的第一个心灵如何想到它的第一个想法[3]，什么样的条件才会使这成为可能？

在本章中，我们拟对两种试图将语义学自然化的重要理论进行考察。我们将根据这些理论探讨产生思想内容的机制，然后分析几种重要的反对意见。由于理论和问题太多，我们没法涵盖所有一切重要的理论及问题。但我们在广度上失去的东西，又将在深度上赐还给我们。当然，其中的许多问题也会出现在其他理论之中。

6.2 思维的媒介

思想要获得其内容，不仅有赖于机制，而且离不开媒介。当 A1 想到埃菲尔铁塔在巴黎时，他的思想部分是关于铁塔本身的，部分是关于巴黎的，部分则是关于一个地方与另一个地方在地理学上的关系的。思想怎么可能做这样的事情？怎么可能会受对象、属性和关系的影响？这是有争议

的问题。本书的其他章节探讨的是关于心灵认知构架的种种设想，例如古典主义、联结主义等。正确的观点一定会说明：思想的不同部分怎样热衷于环境的不同部分或特征。而这样做是离不开心灵（大脑）之物理状态的变化的，正是这些变化的状态成为了思想的不同表征性媒介。某种完全单一的东西[4]不可能表征或思考以下结果：埃菲尔铁塔在巴黎。

上述情况得以发生的一种方式是：存在着一种思维语言（LOT），即从结构上反映了公共的自然语言符号系统。我们有充分的理由主张：LOT不是公共的自然语言，因为我们要想学会第一种自然语言，就必须拥有关于一种语言的资源，如形成关于单词和短语意指什么的假说，以及检验这些假说的证明程序（福多，1975）。因此，要学会我们的第一种自然语言，我们必须能够思维。[5]甚至有根据说：我们成人几乎所有思想所用到的都是自然语言——至少有意识的思想是如此（卡莱瑟斯，1996），不管某些思想实际上是不是使用了公共的自然语言的思想[6]，但它们似乎是能够用自然语言来表达的。当 A1 想到埃菲尔铁塔在巴黎时，人们一般都会说：A1 能用英语表达他的思想（就像说英语的人一样）。要想变成现实，其中的一种方式只能是：与思想的表达式中的每一要素相一致，在思想本身中也存在着一种要素，例如存在一种表示定冠词“the”的要素，一种代表埃菲尔铁塔的要素，一种代表巴黎的要素，一种代表所处关系的要素。从根本上来说，如果思维语言是一种具有构成性句法和语义学的符号系统，这里的句法和语义学又同型于自然语言的构成性逻辑句法和语义学（哈曼，1973），用福多的话说是“或多或少地同型”（福多，1981），那么这就可以解释：思想怎么可能用自然语言予以表述。当然，实际问题绝非如此简单，而且关于同型的许多问题至今仍属阙如（福多，1975）。但是，从已保存内容的一个要素到另一个要素，肯定存在着依赖关系和功能。这是显而易见的。而且似乎有把握说：思想是符号系统的组成部分，因为思想有表征性的特征，而这些特征又依赖于思想的结构（哈曼，1973：59）。

思想的媒介正好是以自然语言具有**产生性的方式**而使思想具有产生性的特点。我们可以作这样的思考：1 是小于 2 的正整数，而 2 又是小于 3 的正整数，3 又是小于 4 的正整数……由此你可以形成你的概念。对于具有无限复杂性的思想，我们也可以形成这样的复述和构想。另外，思想的媒介还能使思想具有**系统性**——如果某人能够思考：对象 a 与对象 b 处在关系 R 之中，那么他也能够思考：对象 b 与对象 a 有关系 R。用什么来解释这些特征至今仍有争论，但至少有一个事物正好可以解释它，那就是思

维语言。[7]

相信思维有符号媒介的进一步理由与自然主义纲领是连在一起的。智能行为似乎依赖于我们思想（信念、愿望、意向、感觉等）的内容。例如杰里为了拿到一瓶啤酒而走向了冰箱（德雷斯基，1988），这离不开这样的条件，即杰里走向冰箱的理由在他的心中得到了表征——有某事让他走向冰箱，另外的某事让他去取啤酒。民间心理学（以及认知科学）在这里迫切需要的似乎是思维的媒介，因为它们能够解释杰里为什么走向冰箱，进而为什么会取出啤酒。而这些似乎又离不开引导他整个行为的其他部分之因果要素。

再来看语言行为。如果我说：肯恩高于加里，因为我相信这一点，并想表达出来，那么让我把“肯恩”和“加里”等说出来的因素似乎是各不相同的。意向实在论者和拥护 LOT 的语义学自然主义者总是试图用人的命题态度和其他思想的内容来解释有目的的行为。有的人试图把人的思想内容说成是对于 LOT 中的内在惯用语（或惯用语系列）的计算加工（从
146 字面上理解一定是这个意思）（福多，1975）。根据这种观点，思考 a 是 F，就是处在与思维语言中意指 a 是 F 的符号的计算关系之中。因此表示“a”和“F”的思维符号一定能够引起或解释这样的行为，它是某人理智地作出的，与 a 和 F 有关（假设 a＝特定的啤酒瓶子，F＝为杰里所打开）。如果我们要想用杰里开瓶子的愿望和相信他能如此（开瓶子的方式）做这件事情的信念来解释他开瓶子的行为，那么他的内在思维符号或媒介就一定能凭借其内容而引起该行为（或动作）。[8]

我们已能看到事情是多么的有趣！事实上存在着许多空洞无物的思想，它们所用的名称如“a”或“F”也是空洞无物的（例如祝融星①是很小的，燃素有负重量），在这种情况下，上述观点怎么能说明这些无意义的思想呢？行为怎么可能根据人的思想内容来加以解释呢（亚当斯等，1993）？更有甚者，在上述情况中，人的思想内容究竟是什么（亚当斯和斯特克［Adams and Stecker］，1994；埃弗里特和霍夫韦伯［Everett and Hofweber］，2000）？一个人怎么能真的具有关于“a 并不存在”这种形式的想法？如果 a 和 F 外在于头脑，且是思维媒介“a”和“F”的内容，那么外在的内容（即通常所说的“宽”内容）怎么可能与“a”、“F”可以引起的东西发生因果关联？宽内容怎么可能与对智能行为的解释关联起来

① 指假想的水中的行星。——译者注

(亚当斯等，1990；亚当斯，1991；亚当斯等，1993)?

在这种语境中，我们不难看到标准的弗雷格难题。如果我想到了“a是F”，且a=F，那么在这里，我是想到了“b是F”，还是没有想到？诚然，标准的弗雷格例证为断定存在着思维媒介提供了另一种漂亮的理由。当我想到a是F时，我的行为方式可能不同于我想到b是F，且a=b时的行为方式。这是怎么可能的呢？LOT提供的答案是：思维媒介“a”不同于“b”。即使a=b（但我并不知道这一点），但我的心灵只可能把“a”与“F”连在一起，而不把“b”与“F”连在一起。因此当我的眼睛被蒙住时，伯里·施瓦茨可能进入了房间，我可能相信这一点，但没有向他询问关于杰米·李的情况，或没有向他索要他的亲笔签名。即使伯里·施瓦茨=托利·柯蒂斯（我会向他打听关于杰米·李·柯蒂斯的那个人，或者是我会向他索要亲笔签名的那个人），但是我的心灵并没有将表示托利·柯蒂斯的心理媒介“b”与表示在房间中的心理媒介“F”连在一起。我的心灵中有“Fa”这样的字样（“伯里·施瓦茨在这里”），但没有“Fb”（“托利·柯蒂斯在这里”）（亚当斯和富勒[Adams and Fuller]，1992）。

现在，我们可以假设：思想有媒介，思想还能把外在对象、属性和关系看作是内容（即使不是永远，至少是经常）。我们还可假设：我们的行为关联的是自然原因。下面我们就来考察意义的一些机制。

6.3 自然化

语义学的自然化关心的不过是把思维媒介（符号）和它们的内容关联起来的机制。[9]那种颇具影响的自然化思路至少可以上溯至格赖斯（Grice，1957），后来又得到了斯坦普（Stampe，1977）、德雷斯基（1981、1988）和福多（1987、1990）等人的支持。这里述及的只是其中一些关键人物。自然化就是旨在揭示内容之机制、解释思维对象如何与思维媒介相关联的一种尝试。

我们先从格赖斯的“自然意义”这一概念说起。该概念与“信息”、“指示”（indication）两概念直接有关。它们都是关于属性的相互关系（和依赖性）。在局部的、稳定的环境条件之下，如果具有G属性的事物与具有F属性的事物相互连在一起，并具有法则学的依赖关系，那么某物之上有G发生了，就可能成为某物之上出现了F的自然标记或指示器。烟

(G) 自然意指的是火 (F)。雪地里的脚印指示的是某人已走过了雪地，树的年轮携带的是关于树的年龄的信息，温度计升高指示的是温度在提高。

自然意义（指示或信息）若要出现，那么这些属性的依赖性就必须是局部稳定的。因为一定不会有超因果的决定作用（人为的烟、人为的脚印或树的公母），一定不会有打破这种依赖性的其他因素（树不能生长的季节，温度计中不可知的毛病）。正是因为有说明这些依赖性的需要，德雷斯基[10]才不得不抛弃他早期的公式（“除非有火，否则就没有烟”），而转向信息理论的一种形式（“如果有烟，那么火的概率一定是 1”）。撇开细微的差别不说，自然意义（或指示或信息）从自然化的计划开始时就一直存在着，这是有其合理原因的。如果 A1 头脑中的某物要意指或关于火，那么 A1 就必须有一种思想媒介，它能自然地意指火，就像烟自然地意指火一样。思维媒介本身也许是由知觉机制引起的，而知觉机制又是由对火的感性分辨引起的（或者说在知觉系统中[11]，有自然意指火的符号，进而这些符号又引起了中枢系统中最终意指火的符号）。

上述状况要出现，必须有这样一种同一性，即知识所必需的环境（或生态学）状况与无歧义的内容所必需的环境状况之间的同一性。假设在一种环境中，没有可能根据某物是 G 而知道某物是 F。再假设这是因为在这种环境中，G 从法则学上也关联于（且依赖于）H——即假设 Gs 是交替地由 Fs 或 Hs 所引起的。在这样的条件下，人们怎么可能基于对 Gs 的察觉而建构出一种对 Fs 的察觉机制呢？没有人能够做到这一点。既然 Gs 依赖于 Fs 或 Hs，那么这种察觉就是关于 Fs 或 Hs 的，**而不只是关于 Fs**。在 Gs 稳定地依赖于 Fs 或 Hs（并与之相互关联）的条件下，某物是 G 披露的信息是：某物是 **F 或 H**。不妨把它称之为**析取问题**。就知识而言，这样一种察觉告诉我们的不过是：某物是 F 或 H。这是因为，它可能指示或自然意指的不过是某物是 F 或 H。就思维来说，如果思维内容来自于自然意义，那么析取的思维内容便来自于析取的自然意义。为了避免这种后果，自然化工程便不得不解决析取问题，并对思维符号如何具有无歧义的意义这一问题作出解释。就眼前的事例来说，如果“G”是只有析取自然意义
148 的思维符号，那么人们充其量只能思考关于 Fs 或 Hs 的想法（而不能思考仅仅关于 Fs 的思想）。而且“G”要想成为一种**真正**的思维符号，就必须上升到自然意义之上的层次。我们很快就会明白这一点。因为自然意义要成为产生无歧义的思维内容的机制之组成部分，就必须来自于非析取的自然意义。我们还将看到：某物有非歧义的自然意义正好是这样的东西，即

它有必要根据某物是 G 而知道某物是 F。这就是知识和思想内容之间为什么存在着内在联系的缘故。如果某人的环境不是局部稳定的，以至不足以知道某物是 F（仅仅只根据某种 G），那么在某物是 F（仅仅如此）这一思想的展开中就无法可靠地获得非析取的思想符号“G”。

普特南举了一个关于硬玉（F）和软玉（H）的例子。假设我不知道玉有两种形式，或者我听说过“玉”一词，但不知道它是何意。即使你把硬玉拿给我看，对我来说，硬玉和软玉仍是没有区别的（G），因此我不能根据它们的外观判断我看的究竟是硬玉（F）还是软玉（H）。我也不能对硬玉形成无歧义的思想（“那是玉”）。我的思想既是关于硬玉的，又是关于软玉的，尽管我只有“玉”这一思想符号（这是思维语言中的东西）。

因此，自然意义（指示或信息）是决定思维内容之机制的一种重要因素。当然，思维内容不能完全等同于自然意义，因为指示与思想具有不同的属性。当 A1 看到一瓶特定的啤酒时，他关于啤酒的知觉符号可能自然标志着啤酒的在场，而且这可能使 A1 想到有一瓶啤酒在附近。但是当周围没有啤酒而他希望有啤酒时，他也可能想到啤酒。[12] A1 的知觉机制不可能以这种方式起作用（除非是做梦、幻觉或其他原因），毕竟知觉机制受制于现在发生的东西。而思维不仅能关注此时此地（借助“这里”和“现在”），而且还可以不受当下的限制。这便使思维具有（真实的）知觉和感觉所没有的自由特质。

不幸的是，使我自由思考的同一种认知能力也会使我们错误地自由思考。当 A1 的知觉机制正常地发挥作用时，他不会看到不存在于此地的啤酒。但是当他的思维机制正常地发挥作用时[13]，他则可能想到此处（在冰箱中）有啤酒，其实根本就没有啤酒（他也许糊涂了，或者他信任的某人告诉他：那里有一瓶剩下的啤酒）。这种情况的发生并不能看作是认知的缺陷。另外，心灵对它的直接环境条件的超越也是有利无害的。但是 A1 能错误地思考具有“Fa”形式的某物这一事实向我们表明：**思想内容不是自然意义**。格赖斯将它称作“非自然意义”。我则将它称作**语义内容**，即能被错误标记的内容。如果“Fa”有自然意义，那么 a 一定是 F，但是如果“Fa”有语义内容，那么 a 就没有必要一定是 F。这样一来，问题就出现了：某物怎么可能从自然意义转化为语义性的思维内容？例如烟自然意指火，A1 在想到自己闻到烟味时也许会在心里形成“烟”的符号，但是从语义学上说，“烟”意指的是烟，而不是火。这是如何发生的呢？某物“火”怎么会从指示火（这离不开火的在场）转向从语义上意指火（不

要求火的在场）呢？

6.4 意义的机制

要从自然意义过渡到语义性思维内容，必不可少的是：符号必须成为被授予了（dedicated）对应内容的东西。[14]“F”要想得到 Fs 作为它的语义内容，该符号就必须关联于“成为一个 F”（being an F）这样的属性。不管人们是否真的想到作为 F 的某物，它都必须意指 Fs，因为它是 F，或者说是关于不是 F 的另一物的。事实上，它一定有某物是 F 这样的内容，即使它是为与某物是否是 F 无关的思想所标记的。正是这一点使**稳固的**（robust）和**错误的**标记成为可能，而把思想与知觉及别的只具有自然意义的事项区别开来的东西正是这两种属性。

因此，接下来我们的任务就是阐释**“授予”**（dedication）这一机制。德雷斯基（1981）曾设想，可能存在着一个**学习时期**（learning period），即一个时间阶段，概念正是在这个阶段形成和获得了它的意义的。根据我们这里的需要，我们不妨假设，概念就是思维符号或媒介。德雷斯基的看法是：某人可能是这样获得关于一个 F 的概念（一个“F”）的，即在适宜于察觉到 Fs 的条件下，他看到了 Fs 和非 Fs。如果成为一个 F 的属性是主体可以有选择地得到的最具体的一个信息（采取了数字表征形式），那么主体的“F”标记（或“F”s）便被授予了 Fs，或被锁定于 Fs。正如我们可能表述的那样，“F”s 被 Fs 激活了，而且 Fs 仅仅是像主体学会把 Fs 与非 Fs 区别开来那样，在学习时期被显现出来。如果人们承认思维符号锁定于它的内容，就像孩子的亲热刻在“妈妈”一词中一样，那么学习时期这一观念就完全是有意义的。内容获得的窗口是敞开着的，符号对专门的信息是敏感的，前者可以锁定于后者，当然内容获得窗口的同时（在学习时期）又是关闭的。根据这一观点，学习时期所起的可能正好就是这种作用，它似乎恰到好处地把印象留在小孩的心上。

学习时期的观念充满着种种难题（福多，1990a），例如主张概念获得类似于获得印象的说法是缺乏充分根据的，因为在一定时期之后，窗户是关闭的。即使有这样的窗户，概念基于呈现给学习者的特定对象通过它而形成，但我们也没有根据说：给予学习者的信息由于被呈现出来的事项的属性而被穷尽了。还是以“玉”为例。正如我们设想的那样，既然我不能

把硬玉与软玉区别开来，因此如果我的思维符号“玉”因为我在学习时期看到了硬玉而得到了专门的标记，那么我被赋予的信息就依然是：某物或者是硬玉或者是软玉。因此我的符号“玉”只与**硬玉**或**软玉**连在一起，即使我面前出现的只是硬玉。在学习时期之后，如果向我显示的是软玉，它使我的符号“玉”得到了标记，那么我便不会错误地相信这是硬玉，而会确定无疑地相信：它是硬玉或软玉。即使我没有用这种方式予以表述，但这就是我思想的内容。因此这个例子事实上暴露了学习时期这一方案所存在的问题。它既未解决析取问题，也没有解释发生错误的可能性。

后来，德雷斯基修改了他关于错误表征如何可能发生的说明（1986），并最终形成了一种极为不同的说明（1988），它不仅想说明符号怎么会关联于别的内容，而且还想说明符号所具有的内容怎么能用来解释行为。在考察这一说明之前，我们先来看看福多本人是怎样说明意义的机制的。

6.5 福多论意义的机制

福多（1987、1990a、1994）[15]为符号“X”意指某物X提供了若干充分条件。正因为如此，他的观点便引发了这样的担忧，即他所说的条件并不适用于我们（或任何人）。如果意义还有别的充分条件，那么福多就是十足幸运的（因为他的条件无意于成为必要的）。我将尽可能简化这些问题，因为很显然，如果福多认为这些条件不适用于我们，那么他就不会提供这些条件。因此，我们先不妨承认他所说的条件是要解释我们的思想由以获得其内容的机制。

我们首先要认识到的是，福多也为原始的、非逻辑思想的符号提供了条件。这也许能在一定程度上解释他为什么认为他所说的条件只是意义的充分条件。逻辑符号和某些别的思维符号可能是以其他方式获得其意义的。具有非原始的（分子性的）内容的符号可以通过分解为原子簇而从原始的或原子的符号中延伸出来。某物何时是一种原始的词项？这是一个经验的问题，而且福多是第一个认识到这一点的人。他还试图表明：他的说明在确定它是否适用于通常不被认为是原始（“独角兽”、“球形门拉手”）词项的过程中究竟有多大的作为。

福多的条件经常在变化，而且无论如何，他也没有用下述的形式将它们排列在一起。不过我又相信：这是他的理论的最好表述方式。[16]（这种

表述方式源于福多，1987、1990a、1994。）我们现在考察的就是他的这一理论。它告诉我们："X"意指X，如果：

（1）"Xs引起'X's"是一条规律。

（2）对于所有不等于Xs的Ys来说，如果作为Ys的Ys实际上引起了"X"s，那么Ys引起"X"s非对称性地依赖于Xs引起"X"s。

151 （3）存在着由非X引起的某些"X"s。

（4）（2）中的依赖性是**共时性的**（非历时性的）。

条件（1）表达的是福多关于自然意义（信息、指示）的看法。如果Xs引起"X"s是一条规律，那么一个被标记成符号的"X"就可以指示X。这一关联能否达成将取决于某人的环境以及它的规律，但是这正好让自然意义能在这种意义机制中发挥作用。[17]很显然，这种条件不足以使自然意义上升为语义性内容。因为"X"要成为表示Xs的一种符号，所需要的不只是为Xs所标记。"X"s**要有其内容**，必须被授予Xs，即忠诚且锁定于Xs。

条件（2）的目的是要说明从自然意义到语义内容的飞跃，同时对析取问题提出解答。它相当于德雷斯基所说的学习时期所起的作用，由此便为把"X"s锁定于Xs提供了一种新的机制。我们与其说存在着一扇打开和关闭的窗户，在这里"X"s固定关联于Xs，倒不如说，福多的关联工作就是使"X"s的所有非X-标记从一开始就合规律地依赖于"X"s的X-标记。因此学习时期在这里是多余的。[18]这个条件强调的是：不仅存在着把符号"X"与它的内容关联起来的规律，而且对于有规律地关联于那个符号"X"的任何其他事项来说，还存在着有规律的或相互关联的非对称性依赖性。这种非对称性是这样的：尽管别的事物（Ys）能引起该符号得到标记，但Y→"X"[①] 规律仍依赖于X→"X"规律，而不是相反。对于后者来说，前者是不能成立的。因此非对称性依赖性规律便将该符号锁定于它的内容。

条件（3）说明的是"稳固的"标记是如何可能的。它承认：存在着由非X引起的"X"s。其中一些是根源于虚假的思想内容，正像我在漆黑的夜晚把一匹马错认为是一头牛，进而错误地把它标记为"牛"一样（以为有一头牛在那里）。另一些则是根源于纯粹的联想，正像一个人把在农场看到的动物与牛联系起来，进而把它们标记为"牛"一样（但这不是一

① Y引起"X"。——译者注

种错误的信念)。这些标记无损于“牛”的意义，由于条件（2），“牛”固定关联于牛。

条件（4）是为了对付潜在的问题而构想出来的，这些问题可能源于各种没有意义授予作用的非对称性依赖性（福多，1987：109）。试看巴甫洛夫的条件反射实验。食物引起狗分泌唾液。后来铃的响声也能引起狗的同样反应。极有可能的是：铃声引出的唾液分泌，仅仅是因为食物引起了唾液分泌。然而唾液分泌几乎不会意指食物，充其量它只会自然地意味着食物在那里，但它不是一种思想或一种思想内容，而且它也不适用于说明错误的语义标记。条件（4）使福多拒绝这样的说法[19]，即唾液分泌本身并没有食物在场这样的语义内容，因为它由铃声所引起的依赖性基于它由食物所引起的依赖性是历时性而非共时性的。动物对无条件刺激首先有无条件的反射，随着时间推移，最终一定有对条件刺激的条件反射。福多所作的依赖性是共时性而非历时性的规定抛弃了巴甫洛夫的条件反射，同时也抛弃了其他许多形式的历时性依赖性。

如果福多后来不进一步讨论历史的例示条件（historical instantiation condition，简称 HIC），那么我们至此就到达了福多关于意义机制论述的终点。尽管 HIC 在 1994 年后从福多的作品中消失不见了（福多，1994），但它毕竟曾经出现在了他的作品之中（福多，1990）。可以肯定，他的理论的折中版本强调的恰恰是 HIC。

HIC：某些“X”s 实际上是由 Xs 引起的。

这是一个已被陈述的条件，这一事实使问题变得复杂起来。由于 HIC 已被看作是第 5 个条件，因此（2）—（4）似乎是因果关系实际例示的条件，而不只是反事实的条件（沃菲尔德，1994）。这一点极其重要，它使福多的理论成为历史性论证，因为福多的理论离不开与对象及其属性的实际的、经验的接触。而这又成了一个难题，因为福多想说，“独角兽”这个符号可以固定关联于未被例示的属性，如成为一个独角兽这样的属性（福多，1990a、1991）。而且只有当我们把 HIC 这样的东西也算进来时，条件（4）似乎才是有意义的。没有它，说规律之间的依赖性是历时性的又有什么意义呢（亚当斯和艾扎瓦［Adams and Aizawa］，1993）？规律是没有时间性的，因此，没有 HIC，条件（1）—（2）似乎只是关于反事实的，而不是关于规律例示的（福多，1994）。

在排除关于戴维森（1987）的沼泽人的思想时，或在说明此处或孪生

地球上的“水”的内容之差异时，如果人们感到困惑，那么 HIC 才真的是有意义的。请允许我作一下解释。先看“水”的内容。在杰里那里，思维符号“水”指的是水（我们的水，H_2O）。在孪生地球人杰里那里，思想符号“水”指的是孪生地球水（XYZ）。如果仅只根据条件（1）—（4），那么上述情况怎么可能发生呢？我们会看到存在着一条从 H_2O 到“水”[①]的规律，但也有一条从 XYZ 到“水”的规律。既然杰里和孪生杰里从物理上说是类型同一的，因此同一条规律一定适用于双方，根本就不存在决定无歧义内容的非对称性依赖性规律，而这却有助于引出 HIC 来。因为杰里并未例示 XYZ→“水”这一规律，而孪生杰里也没有例示 H_2O→“水”这一规律。因此，杰里的符号“水”锁定的可能是一个东西，因为它与那类物质有现实的关联，而孪生杰里的符号“水”锁定的可能是另一个东西，因为它与那类物质有现实的因果联系。由于引进了 HIC，因此这一理论至少从表面上说是可以解释宽内容所具有的那些不同的。[20]而这样一来，（2）的依赖性就只适用于被例示的那些规律。

同样，这一理论也能解释戴维森的沼泽人为什么没有思想。他的媒介是没有内容的。即使相同的反事实适用于杰里和沼泽杰里，但沼泽杰里仍未满足历史性条件 HIC，因为沼泽杰里并没有像杰里那样与那些对象及其属性发生因果联系。

不过，即使这一条件是有用的，福多最终还是将它抛弃了，因为他现在（1994）并不认为孪生地球案例是必须提出来的问题。他现在承认：沼
153 泽杰里与杰里有相同的思想。因此福多引人注目的理论放弃了这个条件。在后面，我们再来说明这样做是否明智。下面，我们先考察德雷斯基一个值得注意的观点，然后再说明两种自然化理论存在的问题。

6.6 德雷斯基论意义的机制

德雷斯基解决内容问题的“处方”包含三个相互联系的方面。（i）符号“C”的内容与它的自然意义 F（Fs——作为 F 的诸对象）一定是有联系的。（ii）自然意义（指示、信息）必须转化成语义内容。一定存在着从由知觉获取的信息内容到认知（语义）内容的转化，而后者是以能为信念

① 即 H_2O 引起“水”。——译者注

和愿望在产生行为 M 时加以利用的形式而被编码的。(iii) 对随意行为 M 的因果解释必须诉诸认知状态的内容（根据它们所拥有的内容）。因此，如果符号“C”引起了身体运动 M 是因为“C”的标记指示（自然地意指）的是 Fs，那么“C”就从单纯自然意指 Fs 上升到了拥有“某物是 F”这样的语义内容。

F←指示“C”，且引起→M（因为它指示 F）

尽管福多对内容的历史性说明（根据 HIC）摇摆不定，但德雷斯基的说明则摆脱了这种摇摆性。从根本上来说，他的说明正是历史性的说明。在不同的条件下，由相同物理构成的自然记号（signs）是可以标示不同事物的，并有不同的自然意义。在地球上，A1 的手指印是表示他在场的自然记号或标记；在孪生地球上，同一种物理的手指印指示的是孪生地球人 A1 的在场，而不是 A1 的在场。要想这样，就必须有某种像**生态学边界**[21]的东西，正是它把一种环境中可能的东西与另一环境中可能的东西区分开来。在地球上，A1 的印迹要指示 A1 的在场，这类印迹为孪生地球人 A1 留下来的概率就必须为零（假设他不能从孪生地球来到这里，或者说不会到这里）。也就是说，假如这些印迹出现了，除了 A1 能留下它们，其他人留下的概率一定是零。如果有罪犯学会了伪造这些印迹，那么这些印记就根本不可能有独一无二的自然意义。因此，一个自然记号是否有一种或另一种自然意义，完全取决于该记号出现于其中的生态学条件。由此看来，德雷斯基的理论完全是历史性的。规律无处不在，但是规律的例示并不是无处不在的。因此，你与哪一规律有关，这取决于你身在何处，以及你与你的环境相互联系的历史。在不同的环境之下，物理上同一的思维符号“S”在不同的（但质上相似的）有机体身上可能会得到不同的思想内容。[22]这些符号获得什么内容将取决于它们在各自的生态学环境下可能获得什么样的自然意义。

德雷斯基对析取问题的解答至少有两个组成部分。第一个组成部分已交代过了，符号“C”一开始就一定具有自然意指 Fs（且只意指 Fs）的能力。如果它指示 Fs 或 Gs，那么析取内容就是能得到的唯一语义内容。第 154
二个组成部分就是向语义内容的跳跃。即使“C”s 指示的只是 Fs，但某一符号要得到语义内容，就必须抛弃让它获得自然意义的那个基础。它必须锁定于 Fs。而且要允许稳固甚至是虚假的标记，同时又不伤及它的语义内容。我们已明白诉诸“学习时期”为什么没有什么用处，而且我们还看

到：福多为解决此问题想出的计策是非对称性因果依赖性规律。在福多使用“C”标记的非对称性依赖性的地方，德雷斯基则求助于自然意义的解释相关性。在他看来，在把“C”s锁定于它们的内容（F）时，重要的不仅仅是什么引起了“C”s，而且还有“C”s引起了什么，以及它们为什么会引起它们。

假设地松鼠要活下来就必须能够对Fs（食肉动物）作出分辨。如果Fs在地松鼠身上引起了“C”s，那么“C”s的标记指示的就是Fs。德雷斯基主张：当“C”s最终有指示Fs在场的功能时，“C”s最终就得到了“某物是F”这一内容。这个时候究竟是什么时候呢？每一种食肉动物不仅是一种食肉动物，而且还是一种动物（G），一个物理对象（H），一个有生命的存在（I），等等，总之具有许多属性。因此“C”的标记会指示所有这一切，而不只是Fs。德雷斯基的回答是：当“C”s对Fs的指示（仅仅只）能**解释**那种动物的行为时，“C”s就获得了这样的语义内容：某物是一个食肉动物。因此，正是解释作用的**内涵性**[23]将“C”s锁定于F，而不是锁定于G或H或I。

在德雷斯基看来，行为就是心理状态引起身体运动的复合过程。因此当“C”引起了某种身体运动M（比如说动物钻进洞中的运动）时，该动物的运动就是由它钻进洞中的过程所构成的。它的行为就是它引起的整个过程。因此，动物的行为，如跑进洞中，就是由“C”s引起M（“C”→M）构成的。根本就不存在必须获得指示功能这样的专门行为。动物有时跑进它的洞中（M1），有时因惊恐而呆着不动（M2），有时则溜之大吉。这个说明所说的不外是：因为“C”s有其所指示（自然意指）的东西，因此“C”s**被征用**并进而引起这些运动。该动物必须跟踪Fs，必须在Fs在场时采取适当的行为（以避免被吃掉）。因此当动物被标记的“C”借助“C”的指示（自然意指）而引起了某种适当的运动M（进而采取了相应行为）时，它就有这样的认识：有一捕食者在场。一当“C”的自然意指有一种**解释作用**时，“C”便锁定于它的语义内容F。因此，“C”所获得的指示或分辨捕食者的功能就把它的内容提升至高一级的层次，即语义层次。

当然，“C”可以以错误的或其他稳固的方式得到标记。动物逃进洞里也许是因为它认为有捕食者来了，其实它只是受到了一种声音或影子的惊吓，而声音或影子的出现并不能解释“C”s为什么引起了有关的Ms（不能解释动物的行为）。[24]因此，即使“C”s的语义内容是被错误地或稳固地标记的，它也不会受析取内容的影响。

155 须知，在这里，学习时期其实是多余的，也用不着窗户打开或关闭

时，它需要的是一种确定的功能。在德雷斯基看来，当对指示器的指示的解释作用被确定（或被决定）后，该指示器的功能便被确定下来了。这会随着时间变化而变化吗？是的，它会。[25]一旦功能被维持的起原因作用的条件发生了变化时，任何形式的功能都会变化，而认知系统又可以适应认知自主体的外在环境或内在结构的变化。

根据这种观点，指示器的功能类似于其他的自然功能，如心脏、肾脏或知觉结构的功能。德雷斯基所宠爱的有关自然功能的说明其实是这样的说明，据此，只要做事情 Y 对有机体有正面的作用或利益，进而有助于解释该有机体为什么生存下来，那么 X 便获得了做 Y 的功能。这样一来，具有 Xs 并做了 Y 的有机体便有一种选择。因此，Xs 仍然存在且仍然做 Y 的部分理由便是：一种选择已发生在这些有机体身上了。很显然，这无法解释 X 怎么会一开始就在那里，或者说 X 怎样着手做 Y 这一事情。

毫无疑问，对指示器功能的选择一定存在于有机体的生命周期之内，而不存在于代际之间。德雷斯基认为，这种选择是一种“雇用”或“学习”的生物过程，而这个过程又与自然功能的标准原因论样式相一致（亚当斯，1979；亚当斯和恩科［Enc］，1988；恩科和亚当斯，1998）。

再看第三个难题，它要说明的是：某个层次上的“C”的内容怎么会与解释有机体的行为有关。“C”也许引起了 M，但这并不是因为它的自然意义，“C”的意义也许是没有作用的。为了解决这个问题，德雷斯基把触发性（triggering）原因与结构性（structuring）原因区别开来了。触发性原因可以是这样的东西，即引起了“C”进而正好引起了 M；而结构性原因则可以解释“C”所引起的为什么是 M 而不是 N 那样别的运动。换言之，结构性原因可以解释引起 M 的为什么正好是“C”而不是某种别的大脑状态 D。因此结构性原因突出的是下述**差异**：（a）“C”s 为什么引起 M 或别的东西；（b）“C”s 为什么引起 **M**。无论在哪种情况下，如果是因为“C”的自然意义，那么我们面对的就是一种结构性因果关系的实例，而且根据这一关于意义机制的说明，内容便有其作用。

下面，我们将把非智能的机器猫与智能猫进行比较以说明这一点。[26]两者都可产生相同的运动，但它们的**行为**是不同的。假设两者都正在跟踪一只老鼠，但这不是说两者都在悄悄地靠近，即使机器猫的大脑有相似于猫的大脑结构也是如此。因为在机器猫内部，无论如何不可能有**关于一只老鼠**这样的东西。而那只猫由于有其学习的历史，因此其内部才有关于老鼠的某东西。同时它还有抓老鼠的愿望，并有关于如何才能抓住老鼠的种

种信念。尽管它们的身体运动在物理上是相似的，但说机器猫“悄悄靠近”老鼠肯定有些夸大其词，而且试图解释原因也是目的论的胡思乱想。既然它是无智能的，那么我们对之所能做的充其量只是对“怎样”作出解释。而对于那只猫则可以合理地说：它正悄悄靠近那只老鼠，而且从目的论上对这里的“为什么”作出解释，也完全是有意义的，如可以说它悄悄

接近是为了抓住那只老鼠。在机器猫“大脑”的各种作用中，的确包含有引起它运动的触发性原因。但是既然机器猫的内在状态没有语义内容，因此由它们的内在状态指示或自然意指的东西所决定，它们也就没有产生运动的结构性原因。[27]这样一来，在这里不像在猫身上那样不存在任何智能行为。可见，根据这一观点，语义内容对智能行为的产生及解释有着举足轻重的作用。

这就是我们对福多和德雷斯基关于意义机制的思想所作的基本描述。下面我们将对反对两种说明的种种看法作一番考察。它将有助于我们更深入地认识这些理论的实质，弄清它们的优缺点。

6.7 种种驳难

专名

福多和德雷斯基两人的理论都无意于讨论专名及其内容。因为两种理论只想解释思维符号是怎样锁定于属性的。既然对象拥有大量的属性，因此，除非它们有个别的本质，否则这些理论就不能轻而易举地说明专名的内容。一般认为，专名和指示代词都有其所指，而这又是由使它们得以进入语言（或思维系统）的因果链所决定的。例如亚里士多德的家庭为他命名，用“亚里士多德”一词表示他。与自然语言中的专名相一致的心理符号也是通过这种因果链而获得它的指称的，这个链条可以口耳相传、代代相传。

根据福多的理论，这个问题也许是再容易不过的了，因为他是根据规律来表述他的理论的。“亚里士多德”要意指亚里士多德，从条件（1）中我们会发现：“亚里士多德引起‘亚里士多德’”必须是一条规律。然而，困难接踵而至。因为这一理论要求作为个体的亚里士多德出现于一条规律之中。但规律的特点在于：它涵盖的是各种属性，而不是个体。因此这一理论无法说明专名的内容（亚当斯和艾扎瓦，1994a）。

福多注意到了这一难题，并试图予以解决（1994：118），他设想与

(1) 有关的规律是这样的："成为亚里士多德这一属性→'亚里士多德'"。但即便他努力得到一个 A，这也似乎只能让"亚里士多德"意指一个**属性**而不能意指**亚里士多德**（那个人）（亚当斯和艾扎瓦，1997a）。福多可能会坚持说：对于每一个体来说，都有使其成其所是的一个属性。但如果拥有一些属性如此轻而易举，那么人们为什么都认为个体在规律中没有什么作用呢？你想要多少规律，可能就有多少规律。似乎更有可能的是：在属性和个体之间存在着差异，因此"亚里士多德"之类的专名命名的是个体，而"成为亚里士多德这一属性"之类的短语命名的则是属性。

福多既然只是说明了意义的充分条件，因此如果他的理论不适用于专名，那么问题便没有完结。他本人的想法是：他的理论不适用于指示代词或逻辑专名。关于指称的因果理论，如直接指称理论，对于专名和指示代词也许是足够的（亚当斯和斯特克，1994；亚当斯和富勒，1992）。思想中的专名把专名与每一个体直接关联起来，同时为那个个体提供了思想的命题内容（这内容由该个体、属性和关系或次序所构成）。

德雷斯基的理论也能解释："C"怎么可能意指亚里士多德。我们知道某物必定指示[28]亚里士多德（例如手指印会如此，DNA 也可以）。可以假定，亚里士多德有其特征，正是由于它们，其家里的人才能认出他来。而"亚里士多德"的意义并不是这些特征或属性，它意指的是亚里士多德。但是在那个时空中，亚里士多德所独有的一系列特征可以允许某结构"C"基于这些特征而对亚里士多德的出现极为敏感，进而自然意味着：亚里士多德出现了。很显然，一定存在着一个因果链[29]，它把亚里士多德与关于亚里士多德的知觉关联起来了，把这些知觉与那些为他命名的人心中的"C"关联起来了（德雷斯基，1981：66-67，以及第 6 章）。德雷斯基理论的其余部分谈的是"C"怎样基于自然地意指亚里士多德而引起某种相关的 M。例如相关的 M 就是他的母亲把他称作"亚里士多德"。这便使亚里士多德（那个个体）成了符号"亚里士多德"的思想内容。

未被例示的属性

人们常能思考独角兽和不老泉之类的东西，但它们压根儿就不存在。因此接踵而至的重要问题是：没有得到例示的属性怎么可能是这些思想的语义内容？[30]一种探讨的方式就是追问：思想符号的这些内容是否是复合的，能否分解为意义的原始构成要素。例如，关于独角兽的思想内容可分为解关于马和角的内容。"马"、"角"以及"具有"可以是具有原始内容

的原始符号（如果不是这样，就应对它们作进一步的分解）。这些原始符号像它们的内容一样可以有被例示的属性。

这样一种标准的经验论策略得到了德雷斯基（1981）的继承。很显然，如果独角兽不存在的话，那么符号“独角兽”并不能自然地意指或指称独角兽。因此，有意义的符号只要有复合的、未被例示属性作为它们的内容，就可以分解为其意义的组成部分（具有更简单的被例示属性作为其内容）。应该看到，这种观点不外是主张：不存在这样有意义的原始词项——它们以未被例示的属性作为它们的内容。

作为一个理性主义者，福多对把未被例示属性作为思想的内容这一难题倾注了大量心血。他对分解策略持宽容态度，但又认为把“独角兽”看作原始的思想符号至少也有其可能性。因此可以假设“独角兽”是一个原词（primitive）。让一个有机体关联于一种属性的方式不外是紧盯住它的各种例示。当然，当不存在“从独角兽到‘独角兽’”这样的例示规律时，这样做显然是有点难的。不过由此可推测说，正是基于这样的理由，福多抛弃了 HIC。

158 另一方案主张：这个世界的非独角兽引起“独角兽”，非对称性地依赖于邻近可能世界中的独角兽引起“独角兽”（福多，1991）。这种方案自然没有告诉我们：用作邻近世界的标准究竟是什么（卡明斯，1989；斯特尔尼［Sterelny］，1990；洛尔，1991）？更糟糕的是，它没有说明在邻近的可能世界中，独角兽是怎样引起“独角兽”的？一种可能的推测是：那是因为属性在那些邻近世界中被例示了。如果是这样，那么 HIC 条件在那些世界中似乎也被利用了，因而必须以某种方式把这种条件放回到福多的理论中。

当进行这样的处理时，福多（1991）强调：他总是有办法退却的，并能说：“独角兽”毕竟不是原词，而是复合词项。[31]但他是不情愿这样做的。更有甚者，他的不情愿还引诱别的人（沃利斯［Wallis］，1995）作出这样的尝试，即发明原始词项，以表示法则学上不可例示的属性。假设有一只巨大无比的蚂蚁（a giant ant，可简写为“gant”），由于生物学的理由，它不具有法则学上的可能性——因为它的脚会被它的重量压碎，它的循环系统不允许足够的热量传递出去。因此沃利斯争论说：不存在这样的邻近世界，其中例示了从巨大无比的蚂蚁到“巨大无比的蚂蚁”的规律。如果福多想固执地坚持他的观点，那么他就应该说明：“巨大无比的蚂蚁”之所以关联于“是一只巨大无比的蚂蚁”这样的属性，是因为在邻近世界

中，自然规律不同于我们这里由实到名的规律。在我们的世界中，引起“巨大无比的蚂蚁”的一切东西之所以如此起作用，是因为巨大无比的蚂蚁在那里引起了“巨大无比的蚂蚁”（而不是相反）。为什么会是这样，无疑是一大难题。

福多本人自然会强调说：对于像方的圆（假设“方的圆”是原始词项）这样逻辑上不可能的属性，他一定会坚持分解策略（福多，1998a）。毕竟不存在这样的世界，在那里起作用的是方的圆引起“方的圆”这样的规律。很显然，如果福多真的拒绝 HIC，甚至当他诉诸邻近世界，把它们作为非对称性依赖性规律的基础时也拒绝，那么他才能够说：并不是方的圆、巨大无比的蚂蚁、独角兽分别引起了“方的圆”、“巨大无比的蚂蚁”和“独角兽”，而是这样的，即如果说它们被例示了，那么它们就能起这种作用。不过，福多这种说法也是极为可疑的，因为如果是那样的话，意义的机制便不复存在了。要想这样，不仅要乞求于未被例示，而且要乞求于不可能被例示，然而我们没有理由相信存在这种关于语义机制的形而上学。

再论析取问题

批评者论证说：语义学自然主义者仍未解决析取问题。福多（1990a）使我们认识到，析取问题源自对德雷斯基诉诸学习时期的回应。为了不让他的说明依赖于临时构想出来的学习时期，德雷斯基修改了他的说明。不过，在他关于指示器功能的新说明中，仍然残存着学习因素（福多，1988、1995）。不可否认，的确有这种情况，即在他称作“雇用”（recruitment）的过程中，某种内在结构得到了它的指示功能，进而由此获得了它的表征内容。这不是权宜之计，也非任意所为。不过，要想让它的指示功能成为固定或不变的，则离不开一种结构“C”。正如我们上面所注意到的，德雷斯基认为，指示器功能得以固定的方式与任何自然功能得以固定的方式是相似的。持极端怀疑态度的批评者担心，所有功能属性都是不确定的（恩科，手稿）。别的人则抱怨说：功能的确定性比语义内容所需要的确定性要少得多。

上面所述的是一种反对德雷斯基解决析取问题方案的代表性观点。福多的解决方案也受到了近似的批评，例如早就有对福多诉诸非对称性依赖性的种种责难（丹尼特，1987a、1987b；亚当斯和艾扎瓦，1992）。艾扎瓦和我指出：孪生地球案例对福多来说是一个难题。既然 A1 和他的孪生地球人在物理的每一相关方面都是相同的，因此如果一个规律适用于 A1，

那么它也一定适用于孪生的 A1。进而，如果有一条 H_2O 引起“水”的规律以及一条 XYZ 引起“水”的规律[32]，那么几乎不可能有非对称性依赖性的规律存在。因为根据孪生地球案例的设想，A1 和孪生的 A1 都不能把水与孪生地球水区别开来，因此，一种规律被打破，另一种规律也会被打破。

正如上面所强调的，福多可能会用 HIC 解释说：A1 例示的是第一条规律（即关于水的），而孪生 A1 例示的是第二条规律（即关于 XYZ 的）。没有哪个能同时例示两种规律。因此，这就是 A1 的“水”为什么只意指水、孪生 A1 的“水”为什么只意指孪生地球水的原因（沃菲尔德，1994）。从根本上说，我并不认为这有什么助益（亚当斯和艾扎瓦，1994a、1994b）。无论如何，福多丢弃了 HIC。他的理论不再阻止这样的说法，即 A1 的“水”个例既对称地依赖于关于地球水的规律，又对称性地依赖于关于孪生地球水的规律，基于此便出现了析取意义。

福多（1994）似乎不再为孪生地球案例所困扰，因为各种形而上学的可能性太遥远了，以至用不着担心。福多也许是对的，因为纯粹的可能性太遥远了，以至仿佛是通过生态学边界才能使它们隔开。孪生地球水与 A1 的环境远隔千山万水（反过来，孪生地球人 A1 也是这样）。用认识论文献中的熟悉词语说，这些事例并非是“相关选项”。正如丹尼特和别的许多人（贝克，1989；卡明斯，1989；戈弗雷-史密斯［Godfrey-Smith］，1989；马隆［Maloney］，1990；斯特尔尼，1990；博格霍赛恩［Boghossian］，1991；琼斯等，1991；亚当斯和艾扎瓦，1992、1994a；曼弗雷迪［Manfredi］和萨默菲尔德［Summerfield］，1992；沃利斯，1994）所指出的那样，孪生地球可能不是一种相关选项，但别的事情却是这样（或者可能是这样）。我们可以假定：相似的人或物之上根本不会有形而上学上异常的东西。究竟是什么东西妨碍“X”意指 X 或似 X 的东西呢？

卡明斯（1989）用 Xs 表示老鼠，用 X－类似物表示鼩鼱。许多人很容易基于外表把这两种动物混同起来。不错，的确存在着一条规律，即从老鼠到“老鼠”①，它可满足福多所说的条件（1）；但同时还存在着另一条规律，即从鼩鼱到“鼩鼱”。因此问题在于：第二条规律是否非对称性地依赖于第一条规律？对于这里的非对称性为什么是不可能的有多种解释，而卡明斯一一作了考察。似乎很清楚，就 A1 来说，他的思维符号“老鼠”

① 这条规律的意思是说，“老鼠”这一标记必然是由作为实在的老鼠引起的。——译者注

(mouse，单数）大概对称性地依赖于老鼠（mice，复数）或鼩鼱（复数）。他的思维符号“老鼠”（mouse，单数）可能对应于**老鼠**（mouse，单数）或**鼩鼱**（单数）。[33]很显然，除了“似鼠的外观”之外，还会有别的属性让 A1 在意指老鼠的过程中都是有关的。还可能有一些属性是老鼠有而鼩鼱所没有的，以至于如果老鼠没有它们的属性，那么鼩鼱也不能违反老鼠引起“老鼠”之标记这一规律。这是福多在解释“老鼠”一词对 A1 来说怎样关联于成为老鼠这种属性时必不可少的东西。但根据福多规定的种种条件，说 A1 的“老鼠”符号可能有析取性语义内容至少有表面的合理性。

卡明斯的例子对福多来说可能不算是问题。我将说明其中的原因。因为福多可以肯定地承认：人们的习惯用语或它在思想中的对应物可能具有析取内容。如果 A1 真的错误地把鼩鼱当作老鼠，那么这应是意料之中的事情。根据福多制定的条件，他不需要的东西是：“老鼠”的任何个例都不意指老鼠。为了避免这一点，他便会求助于对语言劳动的划分。既然在把专名引入自然语言的过程中事实上包含对劳动的划分，因此只要各专家能把老鼠与鼩鼱区别开来，那么英语的“mouse”（“老鼠”）一词意指的仍（仅仅）是 mouse（老鼠）。如果 A1 从专家和说英语的人那里学到了他的思维符号“老鼠”，那么便可能有**语义的借入**（borrowing）。当一个人 A 从另一个人 B 那里得到了一个专名，进而 A 的专名基于此而意指 A 的专名所意指的东西时，语义借入就会出现。如果 A1 听到弗朗克谈论澳大利亚的针鼹，但 A1 又从没有见过这种动物，那么他仍能够思考它。他可能想知道它们看上去像什么、吃什么，等等。他关于针鼹的思维符号可能相当贫乏，尽管如此，这一符号仍可以是意指针鼹的。因此，思维符号可以借助进入别的心灵的因果链条而关联于它们的语义内容。我们必须以这种方式来理解卡明斯的论证，即他否认老鼠/鼩鼱这样的事例中存在着专家。这样一来，A1 的思维符号便不会从英文单词中派生出单义的内容。不过，福多仍承认：对于 A1 来说，“老鼠”一词可以关联于老鼠或鼩鼱。甚至如果没有人能把老鼠与鼩鼱区分开来，那么对于任何人来说，它可能关联于某种析取性的东西。但可以肯定的是，这是不对的。人们必须说明的问题是：“老鼠”是单义的，而根据福多的条件（且不是因为语义借入），它又是析取的。

贝克（1989、1991）为了证明福多的理论作出了错误的内容归因，便用 Xs 表示猫，用 Xs一类似物表示机器猫。她设想：杰里先看到了机器猫，后看到了猫，最后杰里又发现自己关于猫的看法（认为它们都不是机器）

是错误的。这里便出现了下述两种规律，即机器猫引起“猫”和猫引起“猫”。杰里的“猫”这一思维符号的内容是什么呢？贝克坚持认为，对于杰里来说，“猫”不能意指猫（我认为，如果我们排除了语义借入的可能性，那么她就是对的）。贝克还主张：“猫”不能具有关于猫这样的语义内容（这里，我也是赞同的）。不管是哪一种内容，都不能满足福多的条件（2）所说的非对称性依赖性条款。贝克进一步认为，杰里的“猫”不可能有**猫**或**机器猫**这样的析取内容，因为如果它有的话，那么杰里后来就不能认识到他关于猫的看法是错误的。但在我看来，而且福多（1991）也是赞成的，这不过是析取内容的一种情况。的确存在着这样的规律，即或者是猫或者是机器猫引起了“猫”，它是关于“猫”的别的一切个例都会非对称性地依赖的东西。根据这一阐释，福多的其他条件是很容易满足的，而且贝克关于福多发现了他对猫的看法有错误的断言也与这一阐释相一致。它其实是一种二阶错误。福多后来的发现是：他以前关于猫的想法是错误
161 的。因为他认识到他的思想的内容是析取性的（在这里他想到，这是非析取性地关于机器猫的）。[34]总之，贝克的例子对福多来说没有构成难题。

曼弗雷迪和萨默菲尔德（1992）另辟新径。他们主张：即使母牛引起“母牛”这一规律被打破，但“母牛”这一思维符号仍能关联于母牛。他们建议我们设想：杰里看到了许多母牛，因而得到了思维符号“母牛”。假设福多的所有条件都被满足了，同时母牛的外观又发生了改变（比如说经过进化或辐射式进化）。他们论证说：这一变化尽管打破了母牛引起“母牛”这一规律，但并没有改变杰里的“母牛”的内容。撇开语义借入，一个似乎合理的回答便是：母牛引起“母牛”这一规律并未被打破，而只是被遮蔽了。只要成为一头母牛的本质没有改变，母牛引起“母牛”这一规律会随着时间变迁在各种不同的现象之中表现自身。毫无疑问，历史上最初的母牛看上去不同于现在的母牛。这些早期母牛中的任何一个都不可能在杰里身上引起符号“母牛”，这一事实并不能证明母牛引起“母牛”这一规律被打破了。除了发生了缓慢而逐渐的变化之外，还有一种突然的变化，为什么是这样呢？杰里不能基于母牛的外表而认识母牛，这对于福多的理论来说并不是什么难题（即使它可能向杰里提出实践上的种种难题）。这似乎也没有向人们展示什么不可克服的障碍。

过剩的意义（语义混杂性）

亚当斯和艾扎瓦（1994a）提出了这样的观点：福多的理论把意义归属

于不应归属的事物之上，如果你愿意的话，可以说是归属了太多的意义。德雷斯基的理论也有这样的困难。艾伦（C. Allen）提出了一个有趣的例子引起了我的注意，它似乎适用于这两个理论。如果语义内容像这个例子所表现的那样是容易得到的，那么根据被自然化的各种理论，它就是无处不存的。弯角羚羊会吃金合欢树的皮。结果这种树会释放弯角羚羊不喜欢的鞣酸。不仅是这样，风还能把这种东西带到别的树上，让别的树也释放鞣酸。如果人去拨弄这种金合欢树的皮，那么也会让它散发鞣酸。如果我们把鞣酸分子看作是符号，那么福多的所有条件都可被满足。例如，弯角羚羊咬出→鞣酸（条件 1）；人的拨弄→鞣酸（条件 3）；规律（3）非对称性地依赖于规律（1），这就是条件 2；那些依赖性是共时性的，即为条件 4。

这也适用于德雷斯基的理论，例如在金合欢树中，有某种结构（C）被觉察到了，进而能自然地意指弯角羚羊。这种结构也有能力引起鞣酸的产生（M）。因此，当用指示弯角羚羊的功能解释金合欢树上的鞣酸之产生时，C 就变成了关联于弯角羚羊的东西。C 的指示功能就是关联于弯角羚羊（不然的话，它们就会大量地毁坏金合欢树林）。

我认为，容易且有效的解决办法就是把两种理论限定在思维语言（LOT，或它的产物）的符号范围之内。[35]在金合欢树上，没有什么东西与 LOT 有关。福多（1990c）原来有这样的要求，但后来放弃了。德雷斯基的理论是为那些具有信念、愿望、意识经验以及全部认知结构的生物而创立的。[36]这样的事例如果是成功的话，那么它们对某些人会产生下述令人震惊的结论：语义内容能在心灵之外存在。[37]

邻近投射

福多和德雷斯基都面临着这样的难题，即一种思维符号为什么只意指它较远处的原因（如母牛）而不意指更靠近的原因（如母牛在视网膜上的投射），此原因在思维与实在之间起着中介的作用（斯特艾尔利，1990；安东尼和莱文［Antony and Levine］，1991）。德雷斯基的解答（1981）是：起作用的是恒常的机制，正是它们让“母牛”意指牛，而不指示母牛的邻近投射。之所以是这样，是因为他相信存在着多种多样的投射（P1 ∨ P2… ∨ Pn），而且“母牛”以数字化形式携带的最特殊的信息恰恰是：母牛出现了，而不是别的信息，如 P1，P2，…，Pn。

须知，“母牛”仍会携带这样的合取信息，即 F（母牛在场）和 P（出

现了某种或别的邻近投射，如 P1 ∨ P2… ∨ Pn）都出现了。我以为，德雷斯基应说他现在（1988）所说的话，在这里，他认为，C 指示的是：在场的是一个捕食者（F），一个动物（G）和一个物理对象（H）。即是说，C 对 Fs 的指示解释了为什么是它引起了有关的 Ms，进而它从语义上意指 Fs，即使它事实上指的是 F、G 和 H。同样，“母牛”可以指示母牛及 P（在这里，P 即有限的析取属性）。“母牛”仍然可引起有关的 Ms，因为它指示母牛，使我们能知觉到母牛，进而想到母牛，而不是邻近的投射。

福多对邻近投射难题的解答依赖于他的条件 3。由母牛引起的全部“母牛”（“cows”）同时也是由母牛引起的“母牛”（“cow”s）[①] 的邻近投射。因此不存在关于“母牛”（“cow” s）的、借助母牛的邻近投射而形成的固定因果关系，而且，“母牛”也不可能意指**母牛的投射**。

对于福多来说，断言不存在关于“母牛”的固定因果关系似乎是错误的，即使对牛的全部知觉非对称性地依赖于母牛的邻近投射也是如此。因为关于母牛的思想引起了“母牛”，而且这似乎非对称性地依赖于引起了“母牛”标记的邻近投射。如果是这样，那么“母牛”的内容就应是**母牛的邻近投射**（亚当斯和艾扎瓦，1997b）。这对于福多来说仍是一个难题。

沼泽人（swampman）

在福多（抛弃了 HIC 以后）和德雷斯基两人身上，我们不难发现一个重要的差别。[38] 再清楚不过的是，根据德雷斯基的观点，沼泽人在得到物
163 质化例示时是没有任何思想内容的，沼泽人的任何符号都没有指称的功能。任何符号都没有语义内容。很显然，沼泽人是否能得到语义内容是值得探讨的，即便根据德雷斯基的看法也是这样。因为沼泽人是否能得到语义内容取决于他的内在中枢状态是否能有自然意义，以及是否能基于这些状态的自然意义（即学习）而持续地引起有关的运动 M。如果可能，那么沼泽人为什么不能得到语义性思想内容？我们最终在原则上就找不到任何理由。

不过，福多（1994）主张：沼泽人从他物质化的那一刻起便有思想。他相信：适用于杰里的意义机制同样也适用于沼泽人杰里。福多对这一主张提出了六点证明，但其中没有一点是充分的（亚当斯和艾扎瓦，1997a）。

① 请注意这里在拼写上的差别，如前一处“母牛”的写法是“cows”，后一处的母牛是“cow”s。——译者注

第一，有一种意义机制对于一切人来说都是再简单也不过的事（“是更具审美性的事”）。我赞成的是将诸理论统一起来，而福多的理论则有它的赘物，因此他不得不分别论述指示词（demonstratives）、名称和逻辑词项。他没法回答：沼泽人为什么不能获得语义内容？第二，他强调：人们可以在 Xs 没有出现的情况下形成标记“X”s（这意思是说拒绝像 HIC 这样的理论）。但不管“X”s 有没有语义内容，这都是真的。“吉枝”（Giz）可以在吉枝不出场的情况下得到标记，但“吉枝”并不指称任何东西。第三，福多关于沼泽人有思想的直觉是非常强的。不错，欧几里得的直觉也很强，以至关于平行线的公设适用于所有的线段和线段外的点。第四，福多认为，对于沼泽人为什么能回答当天的日期是“周三”的唯一解释只能是：他认为今天是周三。不管怎么说，能对此作出圆满解释的只能是：存在于他的信念盒[39]中的句法的而非内容的“今日是周三”。[40]常见的无思想且带有由句法驱动的语言模块的计算机程序足以证明这一点。这种句法对我们来说是有意义的，但对运行的计算机（如果你愿意的话，可称作计算机杰里）却没有意义。第五，福多主张，可以更合理地断言：沼泽人杰里的“水”个例指的是 H_2O，而孪生的沼泽人杰里的“水”个例意指的是 XYZ。对于其根据的最好解释只能是：它们各自都有这样的语义内容。我认为，下述解释同样是好的，即如果沼泽人的思想有内容，那么他们便具有几乎全部相信者（即地球上的人）所具有的那种内容，而孪生沼泽人具有的是（孪生地球上的）几乎全部相信者的那种内容。如果这些沼泽人有思想，那么这些思想就是他们所具有的思想，但事实上他们并没有思想。我认为，这些否证的力量像福多的论证一样强，或者说前者超过了后者。

心灵依赖性

肖普（R. Shope）责难说：德雷斯基的说明用的是解释概念，而解释是依赖于心灵的活动的，因此德雷斯基的说明并不是真正的自然化说明。只有当对解释的利用不可排除时，这样说才是对的。自然化说明离不开内涵性（intensionality，带有一个“s”），以至于因为“C”s 指示 Fs（而不是 Gs，即使它们指示 Gs），“C”s 才引起 Ms。这种内涵性完全能由规律和目的论功能来提供，因此解释活动对心灵的依赖性是不可排除的。

功能不会那样起作用

许多批评者抱怨说：德雷斯基没有提供关于功能的正确理论，或者说

根本就没有关于目的论功能的一致意见。德雷斯基自认为很幸运（1990），因为他根据需要放弃了“功能”一词。不过，即使他替换上了新词，如“指示器功能”，他的说明仍是相同的。

肖普（1999）和戈弗雷-史密斯（1992）也论证说：有机体在得到指示 Fs（一种语义内容 F）的功能时，不必自然意指 Fs 的符号。差一点的东西也可起这种作用。肉食动物也许会引入“C”，让它最好地关联于猎物。这意思是说：“C”意指的是猎物。我认为，存在着这样一个固有的问题，即肖普或戈弗雷-史密斯是否找到了解释上述事例的方法，在这个例子中，“翡翠”似乎从语义上指称的是硬玉或软玉，其明显的理由是：它常常自然地指示这种析取。就我所知，肖普、戈弗雷-史密斯和别的人都主张：这是可以发生的（“F”意指 F，而从不自然地意指 Fs），但他们不能解释这是怎样发生的。只有当他们能作出解释时，他们才能肯定地说：这已经发生了。他们提供的例子与动物有析取内容这一点是一致的，尽管他们得出的结论正好相反。

空洞无物

福多只为内容提供了充分条件这一事实引出了这样的抱怨：尽管他的理论别出心裁，但空洞无物。这样说可能适用于所有没有实际意义的事项，或者说，如果是这样的话，那么它就会产生错误的内容（贝克，1991；西格尔［Seager］，1993；亚当斯和艾扎瓦，1992、1994a）。水能够在珍妮特身上引起“水”这一符号，但致幻的药物、大脑肿块或高烧也可以产生这种效果。既然福多抛弃 HIC，而这些事物在珍妮特心中引起“水”的能力一定非对称性地依赖于水引起“水”这样的规律。但它们是这个样子吗？为什么是这样？“水”是大脑中的构造，它们可独立于内容而被识别（根据福多自己的条件）。[41]因此，大脑中的某物为什么不能独立于那些构造的内容而引起这样一种构造呢？我能用打字机打出“吉枝”，而不管“吉枝”是否有意义。我的大脑中为什么不能有等同于“水”的某东西？根据那个假定，它是可以的，而仅用福多的条件则是不能解释珍妮特的“水”怎样关联于水的。珍妮特和我们的思想都有内容，但不是因为福多的理论所说的那些条件。因此，他的条件是空洞无物的。

超越上述忧虑的一种自然方法就是回到 HIC。事实上，德雷斯基正是这样做的。他说：在某些人身上，“水”这个符号除了由水引起之外，还
165 可能是由某种别的东西引起的（在“水”获得它的语义内容之前）。[42]如

果是这样，对于这些条件下的个体来说，水就不会让水作为它的自然意义存在，它不可能让水成为它的语义内容。完全可能存在着这样稳定的条件，正是它们把人体身上的这些原因与药物、肿块、高烧、XYZ 等等分离开来。福多必须解释的是：其他事物为什么被分离开来了，而如果不用 HIC[43]，则完全没有完成这一工作的机制（沃菲尔德，1994；亚当斯和艾扎瓦，1994b）。福多将不得不说：在水让珍妮特形成了“水”这一符号的世界比病理学原因在其中起作用的世界要切近很多。但是没有一个世界比现实世界更切近。在这个世界，有这样的人，他们在所有别的物理方面都相似于珍妮特，但他们的大脑中完全有可能存在诸如此类的异常原因。因此福多的理论不适用于他们（或珍妮特）。

意义或非对称性：谁先出现？

对于非对称性依赖性是否能产生意义的问题，许多作者深表怀疑（西格尔，1993；吉布森［Gibson］，1996；亚当斯和艾扎瓦，1994a、1994b；沃利斯，1995）。据福多说，非对称性是让意义出现于世界的东西，而不是来自于世界。如果仅仅因为 Xs 起了作用，Ys 便引起了“X”s，那么这一定不是因为关于“X”s 的任何语义学事实。是什么样的机制产生了这样的句法非对称性依赖性？除了 Xs 之外，许多事物为什么不能独立于 Xs 有此作用的事实而引起“X”s？“X”s 在大脑中的例示是某种形式的神经化学事件，因此应该存在着这样的自然原因，它们能够让这样的事件在人的大脑中产生出来（且在各种条件之下）。在地球上，仅仅因为母牛能在我们心灵中让“母牛”这样的符号产生出来（假设“母牛”是未被解释的中枢事件），就得出结论说牛排能引起“母牛”吗？这是为什么？牛排是畜生吗？

在说明这样的非对称性的存在时，福多求助于“专家”，亦即他们使用专名的意图（1990c：115）。但这样的解释自然是无济于事的。因为一个人不能用意义来解释非派生意义之产生。那么，非派生的非对称性是从哪里来的？我的大胆推测是：它是这个样子，即“母牛”意指母牛，“牛排”意指牛排，于是我们把牛排与母牛联系在一起，因此牛排之所以引起“母牛”，仅仅是因为母牛引起了“母牛”。我们不把牛排与“母牛”连在一起，除非我们把“母牛”与母牛连在一起，把牛排与母牛连在一起。这一对非对称性依赖性的解释利用了意义，但它并没有产生意义。除非存在关于这种非对称性依赖性的更好解释，否则福多的理论就会错误地把它们看

作是意义的基础。

6.8 结　论

这些理论尽管优缺点混杂，但它们无疑是我们已有的关于思维内容的最好理论。当然，它们不是这类理论的全部，而只是体现了将内容自然化的基本计划。这两种理论和别的自然化理论之间的差异相对来说是不大的，它们真的并不那么糟糕，尤其是当你思考诸种可能选择时更是如此——但那只是另一个时期的一种计划。[44]

【注释】

[1] 如果德雷斯基（见手稿）是对的，那么要解释我们怎么知道这些事物比解释我们先前就相信它们要难得多。

[2] 在本部分，我坚持的是意向实在论者的观点。这些理论认为，思想和别的心理状态是有内容的，并能根据所拥有的内容对行为（和别的心理状态）作出解释。还有一些观点认为，把内容归属于思想是解释的问题，但拥有内容则不是可作解释工作的问题。

[3] 对此，内容整体论和内容原子论之间是存有争论的。前者主张：各种思想总是串联在一起的——任何心灵不可能只有一个因素。而后者则认为，思想和心灵可以是点状的——例如一个心灵只有一个思想。我们这里的讨论不可能深入到这个争论之中，欲知其详，请参阅福多和莱泼尔（Fodor and LePore，1992）。

[4] 这是把意义与信息区别开来的一种方式。假设一种单一的信号能够提示或告知这样的信息，即埃菲尔铁塔在巴黎。假设我们事先设定：一盏特定的灯亮了，就代表铁塔位于巴黎，因而灯亮了就能够告知那些不知道铁塔位于巴黎的人们这一事实。但是灯亮了并不能构成铁塔位于巴黎这一思想。

[5] 几乎每个人都会利用的一个充分理由是：没有言语能力的婴儿和动物也会思考。当然也有一些人持不同意见（戴维森，1982；卡莱瑟斯，1996）。

[6] 哈曼（Harman，1973）和卡莱瑟斯（1996）主张：大多数的思想是用自然公共语言表达的。福多是持反对意见的人之一（1998b）。

[7] 福多和派利申（1988）以及福多（1998b）著作中的福多夫妇和麦克劳格林、艾扎瓦（1997）对某些关于思维语言假说的系统论证的效力发表了不同意见。

[8] 至少可以说，如果意向实在论者是对的，那么他们一定是这样。

[9] 我希望澄清的是，当我谈论机制时，我关注的不是人类身上的思想之物质基础（中枢传输中的神经元或化学结构），甚至不是知觉的特殊心理物理机制。我谈论的

是信息的种种要求，而不是能满足这些要求的特定物理或心理执行过程。

[10] 德雷斯基最先提出了“终极性（conclusive）理由”这一概念，在这里，作为理由R的东西（它可能类似于史密斯留在枪上的指纹印记）不是事实，除非P（史密斯接触过那只枪）。R成为事实可以让人知道：P是真实的。为了找到关于拥有知识所必需的属性（例如对于知道史密斯接触过那只枪必不可少的属性）之间关系的更准确说明，德雷斯基后来（1981）转向了信息理论。

[11] 关于知觉系统中的知觉符号本身常被用作思维符号或媒介的这一观点，请看巴萨卢（L. Barsalou，1999）。关于不同的意见，请看亚当斯和坎贝尔（1999）。当然，其他同样有价值的评论还有很多。

[12] 关于这一点，福多（1990b）谈得很多，后来他还把它称作“稳固性”（robustness）。

[13] 对于一个人在错误地思维时是否会正确地运用他的认知能力，笛卡尔有不同的看法（见《沉思录》Ⅳ）。

[14] 福多恰当地把这称作“关联于”（或“锁定于”）某一属性（或内容）。

[15] 福多关于意义之条件的思想经常在变化，而且这三本书的看法也有（微妙的）不同。

[16] 下面，我们将解释该理论的另一种典型，它增加了一个条件，并讨论了他先前为什么愿意增加它，后又为什么抛弃了它的理由。关于这一点的更多内容，可参阅亚当斯和艾扎瓦（1994b）。

[17] 福多愿意把他的观点看作是一种“信息”语义学（1994）。

[18] 学习（时期）是没有必要的——这与福多倾向于天赋论是一致的。

[19] 人们可以认为，没有必要作这样的拒绝，因为唾液分泌并不是思维语言中的媒介。但福多并没有把他的理论限定在思维语言中的事项之上。因此，甚至头脑之外的事物也可以有意义。

[20] 是否存在着“或者是水或者是孪生地球水”引起了“水”这种形式的析取规律？或者说，这里是否存在着规律的对称性依赖性？这些事实上仍大有疑问，但我现在不想涉及这些问题。

[21] 我们试想这样一种生态学的边界，它类似于德雷斯基所说的“信道条件”。

[22] 这也许是不言而喻的，但正因为这一点，德雷斯基成了一个经验论者。同一种认知构造“水”可能先天地出现在了A1和孪生A1身上，但既然各个A1有不同的历史，因此他们（用“水”一词）所作的思考便获得了不同的内容。

[23] 在后面，待我们考察肖普的反对意见时再来讨论这一点。

[24] 棍子和石头可以打断你的骨头，而影子和声音对你不能造成什么伤害，每个地松鼠都懂得这个道理。因此，“C”没法被当作影子或声音的指示器来使用。但捕食者就完全不同。

[25] 见德雷斯基（1988：55）。

[26] 如果有智力的猫也是一种机器型的，但它不能思维，那么我便会表示赞同。就我的目的而言，戴维森式的沼泽猫完全可看作是一种无智商的猫。

[27] 因为工程师设计让机器猫触发那些事物，因此在机器猫中，可能存在着使事物产生的物件。但不存在引起事物产生的结构。因为那些结构的作用是告知那只猫：周围有老鼠！进而与信念、愿望结合在一起，由此引起躯体的运动。但机器猫作为无智商的存在是没有信念和愿望的。

[28] 正是在这里，德雷斯基的理论才显得不那么费力气。因为根据他的信息理论解释（1981），对于一个携带着关于亚里士多德的信息的手指印来说，亚里士多德接触过一对象的概率一定是1，因为他的印记留在它上面了。乍一看，这不是说亚里士多德进入了一种规律。在亚里士多德的时代，人们对那些手指印一无所知，但这远不是问题的关键。他的出现可能一直是作为一种手指印而个体化的。因此对个体的属性的分辨即使不附带说明该个体进入一规律之下，也可能让人跟踪关于该个体的信息。

[29] 这就是关于亚里士多德的知觉为什么不能从质上同一于关于孪生地球上亚里士多德的知觉的原因，但它们自然有这样的意义，即在场的亚里士多德（而非孪生的亚里士多德），因为引起它们的是亚里士多德（而非孪生的亚里士多德）。

[30] 空名自然也有这个问题（参阅亚当斯和斯特克，1994；埃弗里特和霍夫韦伯，2000），但这里没有时间讨论这些问题。

[31] 我认为，福多对语义整体论的憎恨能够说明他为什么要规避这种策略（福多和莱泼尔，1992）。

[32] 很显然，一定有两条规律，因为地球上的水引起的是A1的“水”，而孪生地球水只能引起孪生地球人A1的“水”。

168 [33] 只要引进HIC，人们就能作出种种变换，但我将把它们留给读者。

[34] 再者，德雷斯基的理论尽管不承认语义借入的可能性，但基于贝克的例子，他仍得出了与福多相同的结论，即内容是析取的。如果杰里没有基于猫和机器猫的外观把它们区分开来，那么内容一定是析取的。杰里的符号“猫”由以派生出它们的指示功能的心理状态的自然意义本身也是析取的。

[35] 很显然，必须这样做。因为，自然语言的符号存在于心灵之外确有其意义，但它们的意义来自于心理内容。

[36] 如果人们想把具有心灵与语义内容等同起来，那么这自然会产生令人失望的转向。

[37] 我不知道福多对这种可能性会作何反应，但德雷斯基（通过个人通信）曾告诉我，他自始至终都相信：初期视觉系统（及别的可能的东西）的状态都能满足他关于语义内容的条件。关于金合欢树，德雷斯基嘲笑说：“它听起来像有一点令人讨厌的心理生活。”但如果人们想把具有心灵与语义内容等同起来，那么这就完全是一种“心理生活”，不管它是否令人讨厌。

[38] 他们两人都利用HIC否认沼泽人有思想。因为根据假设，沼泽人没有例示

属性与符号之间有关规律的历史。

[39] 根据关于沼泽人这一思想实验的约定，杰里和作为沼泽人的杰里两人头脑中有完全相同的句法对象（在他们的头脑中，句法随附于纯物理状态），但句法对象有可能与属性无关。

[40] 须知：关于内容为什么仍与对行为的解释有关，以及人们为什么不能像斯蒂克那样退回到纯句法理论，我们是可以找到充分理由的，参阅亚当斯等（1990）。

[41] 根据这一理论，语义内容是非对称性依赖性的产物，而不是它的根源。

[42] 如果这种情况在语义内容得到固定之后出现了，那么它就是一种错误的（或相反，健全的）标记。

[43] 即使是利用 HIC，福多的理论仍有同样的问题（亚当斯和斯特克，1994a、1994b）。例如根据 HIC，如果我们只给珍妮特看硬玉，那么她只会例示“翡翠”引起硬玉这一规律。这样一来，任何或对或错地标记为“翡翠”的东西就会非对称性地依赖于硬玉的标记“翡翠”。因此那一理论就会说：因为她没有例示软玉引起“翡翠”，是故“翡翠”意指硬玉。但这似乎是一个古典的事例，在那里，“翡翠”仍有或硬玉或软玉这样的内容，因为这两者是很多的，只是珍妮特不能把它们区分开来。可见，福多的理论与 HIC 一样仍是漏洞百出。

[44] 谢谢艾扎瓦和德雷斯基的指正与忠告。

参考文献

Adams, F. (1979). “A Goal-State Theory of Function Attribution.” *Canadian Journal of Philosophy*, 9: 493–518.

—— (1991). “Causal Contents.” In B. McLaughlin (ed.), *Dretske and His Critics*. Oxford: Basil Blackwell.

Adams, F. and Aizawa, K. (1992). “‘X’ Means X: Semantics Fodor-Style.” *Minds and Machines*, 2: 175–183.

—— (1993). “Fodorian Semantics, Pathologies, and ‘Block’s Problem’.” *Minds and Machines*, 3: 97–104.

—— (1994a). “Fodorian Semantics.” In S. Stich and T. Warfield (eds.), *Mental Representations*. Oxford: Basil Blackwell.

—— (1994b). “‘X’ Means X: Fodor/Warfield Semantics.” *Minds and Machines*, 4: 215–231.

—— (1997a). “Rock Beats Scissors: Historicalism Fights Back.” *Analysis*, 57: 273–281.

—— (1997b). “Fodor’s Asymmetrical Causal Dependency and Proximal Projections.” *The Southern Journal of Philosophy*, 35: 433–437.

Adams, F. and Campbell, K. (1999). "Modality and Abstract Concepts." *Behavioral and Brain Sciences*, 22 (4): 610.

Adams, F. and Enc, B. (1988). "Not Quite By Accident." *Dialogue*, 27: 287-297.

Adams, F. and Fuller, G. (1992). "Names, Contents, and Causes." *Mind & Language*, 7: 205-221.

Adams, F. and Stecker, R. (1994). "Vacuous Singular Terms." *Mind & Language*, 9: 387-401.

Adams, F., Drebushenko, D., Fuller, G., and Stecker, R. (1990). "Narrow Content: Fodor's Folly." *Mind & Language*, 5: 213-229.

Adams, F., Fuller, G., and Stecker, R. (1993). "Thoughts Without Objects." *Mind & Language*, 8: 90-104.

Aizawa, K. (1997). "Explaining Systematicity." *Mind & Language*, 12: 115-136.

Antony, L. and Levine, J. (1991). "The Nomic and the Robust." In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Crtics*. Oxford: Basil Blackwell.

Baker, L. (1989). "On a Causal Theory of Content." *Philosophical Perspectives*, 3: 165-186.

—— (1991). "Has Content Been Naturalized?" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.

Barsalou, L. (1999). "Perceptual Symbol Systems." *Behavioral and Brain Sciences*, 22: 577-660.

Boghossian, P. (1991). "Naturalizing Content." In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.

Carruthers, P. (1996). *Language Thought and Consciousness: An Essay in Philosophical Psychology*. Cambridge: Cambridge University Press.

Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, MA: MIT/Bradford.

Davidson, D. (1982). "Rational Animals." *Dialectica*, 36: 317-327.

—— (1987). "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association*, 60: 441-458.

Dennett, D. (1987a). *The Intentional Stance*. Cambridge, MA: MIT Press.

—— (1987b). "Review of J. Fodor's Psychosemantics." *Journal of Philosophy*, 85: 384-389.

Dretske, F. (1971). "Conclusive Reasons." *Australasian Journal of Philosophy*, 49: 1-22.

—— (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT/

Bradford Press.

—— (1986). "Misrepresentation." In R. Bogdan (ed.), *Belief*. Oxford: Oxford University Press.

—— (1988). *Explaining Behavior: Reason in a World of Causes*. Cambridge, MA: MIT/Bradford Press.

—— (1989). "The Need to Know." In M. Clay and K. Lehrer (eds.), *Knowledge and Skepticsm*. Boulder: Westview Press.

—— (1990). "Replies to Reviewers." *Philosophy and Phenomenological Research*, 50: 819–839.

—— (1995). *Naturalizing the Mind*. Cambridge, MA: MIT/Bradford Press.

—— (manuscript) "How Do You Know You Are Not A Zombie?"

Enc, B. (manuscript). "Indeterminacy of Function Attributions."

Enc, B. and Adams, F. (1998). "Functions and Goal-Directedness." In C. Allen, M. Bekoff, and G. Lauder (eds.), *Nature's Purposes*. Cambridge, MA: MIT/Bradford.

Everett, A. and Hofweber, T. (2000). *Empty Names, Fiction and the Puzzles of Non-Existence*. Stanford: CSLI Publications.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Crowell.

—— (1981). *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT/Bradford.

—— (1987). *Psychosemantics*. Cambridge, MA: MIT/Bradford Press.

—— (1990a). *A Theory of Content and Other Essays*. Cambridge, MA: MIT/Bradford Press.

—— (1990b). "Information and Representation." In P. Hanson (ed.), *Information, Language, and Cognition*. Vancouver: University of British Columbia Press.

—— (1990c). "Psychosemantics or: Where Do Truth Conditions come from?" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Basil Blackwell.

—— (1991). "Replies." In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.

—— (1994). *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, MA: MIT/Bradford Press.

—— (1998a). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

—— (1998b). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT/Bradford Press.

Fodor, J. and LePore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Fodor, J. and Pylyshyn, Z. (1988). "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition*, 28: 3-71.

Gibson, M. (1996). "Asymmetric Dependencies, Ideal Conditions, and Meaning." *Philosophical Psychology*, 9: 235-259.

Godfrey-Smith, P. (1989). "Misinformation." *Canadian Journal of Philosophy*, 19: 533-550.

—— (1992). "Indication and Adaptation." *Synthese*, 92: 283-312.

Grice, H. P. (1957). "Meaning." *Philosophical Review*, 66: 377-388.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Jacob, Pierre (1997). *What Minds Can Do*. Cambridge: Cambridge University Press.

Jones, T., Mulaire, E., and Stich, S. (1991). "Staving Off Catastrophe: A Critical Notice of Jerry Fodor's *Psychosemantics*." *Mind & Language*, 6: 58-82.

Loar, B. (1991). "Can We Explain Intentionality?" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.

Maloney, C. (1990). "Mental Representation." *Philosophy of Science*, 57: 445-448.

Manfredi, P. and Summerfield, D. (1992). "Robustness Without Asymmetry: A Flaw in Fodor's Theory of Content." *Philosophical Studies*, 66: 261-283.

Putnam, H. (1975). "The Meaning of 'Meaning.'" In *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Seager, W. (1993). "Fodor's Theory of Content: Problems and Objections." *Philosophy of Science*, 60: 262-277.

Shope, R. (1999). *The Nature of Meaningfulness: Representing, Powers and Meaning*. Boston: Rowman and Littlefield.

Stampe, D. (1977). "Towards a Causal Theory of Linguistic Representation." In *Midwest Studies in Philosophy*. vol. 2, *Contemporary Perspectives in the Philosophy of Language*, ed. P. French, T. Uehling Jr., and H. Wettstein, Minneapolis: University of Minnesota Press.

Sterelny, K. (1990). *The Representational Theory of Mind*. Oxford: Blackwell.

Wallis, C. (1994). "Representation and the Imperfect Ideal." *Philosophy of Science*, 61: 407-428.

—— (1995). "Asymmetrical Dependence, Representation, and Cognitive Science." *The Southern Journal of Philosophy*, 33: 373-401.

Warfield, T. (1994). "Fodorian Semantics: A Reply to Adams and Aizawa." *Minds and Machines*, 4: 205-214.

第7章　认知建筑学：诸认知表征的结构

肯尼斯·艾扎瓦（Kenneth Aizawa）

虽然认知建筑学的诸理论涉及与认知相关的那些基本结构和基本过程的本质，但是该领域的哲学志趣却已然主要集中在了假设的诸认知表征的结构上。[1]例如，认知建筑学的经典理论主张：

1. 存在着在句法和语义上组合的（combinatorial）诸心理表征，也就是说，要对在句法和语义上的诸原子表征与在句法和语义上的诸分子表征加以区别。

2. 一个分子表征的每一个个例都精确地包含着该分子表征由以被构成的那些表征中的每一个个例。

3. 一个分子表征的意义是由其各部分的意义和这些部分由以被组合起来的方式所决定的。

4. 一个分子表征的句法部分中的每一个都在它发生于其中的任何语境中具有相同的内容。

5. 存在着敏感于这些心理表征之结构的诸计算机制。

对经典主义的一个替代选项是原子表征主义（AR）。AR主张，诸认知表征在句法和语义上全部都是原子的，因此它拒斥（1）—（5）。另一个竞选项是功能组合主义（functional combinatorialism，FC）[2]，它在某些方面拥有相当多的拥护者。FC主张，虽然有诸连接表征，但它们却不是经典主义所断定的那样。更明确的是，FC断言，诸分子表征不过是其原子的（可计算的）应变物（functions），因此这些原子不必完全是它们由之源出的那些分子的诸部分。

本文不指望纵览认知建筑学的整个领域，甚至不指望纵览围绕着诸认知表征的全部争论，而是要集中在杰里·福多（Jerry Fodor）和泽农·派

173 利申（Zenon Pylyshyn）对经典主义的系统性论证上。粗略来看，这些论证很简单。思维的某些特性被经典主义作出了最佳的解释，即推理的系统性、思维的系统性和诸表征的构成性，因此我们有某些可废止的（defeasible）理由来相信经典主义是正确的。对这些论证的纵览有许多目的。第一，它打算表达（address）关于这些争论要解释什么的一系列误解。第二，它会注意这些系统性论证中一个相对未受到正确评价的特性（feature），即存在进行更好解释的某种原则。第三，它会指出这些系统性论证的一般阐述（formulations）何以在事实上并不支持经典主义，尽管经典主义者会有争辩。最后，它会注意到由福多和派利申的批评所提出的另一种系统性论证。该论证具有福多和派利申所预期的解释效力，并且表现出了AR和FC中缺乏的一种经典主义功效。

本章的计划就是要纵览推理的系统性、诸认知表征的系统性、诸表征的构成性等论证，以及一种新的系统性论证。每一个论证都将通过一个待解释项，连同可能的AR和经典的解释一起被引入。完成第一个步骤之后，我们将回顾这些系统性论证何以对FC的一个特定版本——诸认知表征具有哥德尔数的结构这一假设——产生影响。

7.1 推理的系统性

福多和派利申认为，认知具有下述一般特性："具有类似逻辑类型的诸推理，几乎普遍地会引出类似的诸认知能力。例如，你不可能发现一类心理生活，在其中你能够从P&Q&R中推断出P，但却不能从P&Q中推断出P。"（1988：47）另外：

> 这个模棱两可的说法（"几乎普遍地"）旨在排除这样一些情况：在这里，相同逻辑类型的诸推理，由于——比如，其前提的范围——在复杂性上是不同的。从（AvBvCvDvE）和（¬B&¬C&¬D&¬E）中推断出A，与从AvB和¬B中推断出A，这是相同的逻辑类型。但是，如果存在能够把握第二个推理但却不能把握第一个推理的思维，这也并不会令人感到诧异。（1988：28）

那么，"为什么推理能力是系统的"这个问题就浮现出来了。待解释项的许多特性都值得评说。第一，待解释项牵涉到相同逻辑类型的推

理。[3]能够完成（比如）合取排除（conjunction elimination）其中一个实例的一般认知自主体，如果其他条件不变的话，也能够完成另一个实例。能够完成演绎推理（modus ponens）实例的一般认知自主体，也能够完成另一个实例。简单起见，接下来我们只考虑有限系列的合取排除的系统性。第二，待解释项牵涉到逻辑推理中的诸认知能力（capacities）或者诸认知潜能（competences）。这里的待解释项并不坚持认为，从 P&Q 推断出 P 的任何一般的认知自主体，也会从 P&Q&R 推断出 P。因为，这个待解释项牵涉到了推理的能力，而非推理的实际执行。关于人类在推理中表现的实验文献——已经查明了不同的内容效应、频率效应等——事实上并不直接处理福多和派利申的待解释项。[4]第三，这些系统性论证仅仅设定了人类有限的认知潜能。它们并不考虑人类潜能涉及无限的表征能力。福多和派利申说：

> 我们主张把经典建筑学的生产力地位论证看作是悬而未决的。我们将为下述主张提出一种不同的论证：诸心理表征需要一个被明确表达的内在结构。它与生产力论证呈现密切的关系，但它却不需要对无限潜力的理想化。因此，甚至对像联结主义者这样认为诸认知能力的有限性是其建筑学的内在本质的论者，其设定也应当是可以接受的。(1988：36-37)[5]

第四，弄明白经典主义者们在放弃诉诸无限表征能力的观点时，并未由此放弃诉诸潜力/表现的这个区分，这是至关重要的。很明显，经典主义者们相信，人类在推理中的实际表现是由很多能力决定的。[6]其中之一是逻辑推理能力，但是，一个人也必须承认认知能力、注意能力和记忆能力。诚然，任何称职的实验主义者都会承认，为了查明适于逻辑推理的能力，一种实验情境存在的许多特性都必须被控制——诸如影响动机、认知、注意和记忆的那些特性。承认众因素的多样性是承认潜力/表现的区分必不可少的。所以，即便福多和派利申打算进行这个推理的系统性论证，而不依赖于假定存在有一个适于推理的无限能力，他们也并不因而就打算完全抛开潜力/表现的区分。[7]

待解释项就这么多，解释要素又如何呢？这些系统性论证的某些批评者们已经注意到，有可能形成显示出不同形式的系统关系的诸系统。[8]他们由此断言，系统性争论中的真正问题，是关于到底何种系统关系存在于人类思维当中，以及一个给定的认知理论在何种程度上能够产生这些系统

关系。虽然数据符合（data-fit）是理性选择理论中一个重要的因素，但却不是唯一的因素。更重要的是，它不是福多和派利申在这些系统性论证中启用的那一个因素。在这些论证中，这个问题不仅仅是一个适应可用数据的问题，而是以某种在一定程度上重要的方式对之进行说明的问题。福多和派利申关于推理的系统性论证的评论很好地证实了这个观点：

> 在一种心理生活中，如果能够从P&Q&R中推理出P，那么，你也能够从P&Q中推理出P。联结主义者当然能模拟这种心理生活……但是，要注意，**联结主义者同样能够模拟另一种心理生活，在其中你能够得到这些推理中的一个，但不能得到另一个**。在当前的事例中，既然P&Q&R结节（node）和P&Q结节之间没有结构关系……那么，包含着前者的一种心灵也就没有理由应当包含后者，反之亦然。类似地，没有理由你不获得简化**约翰爱玛丽和比尔恨玛丽**这一前提的诸心灵，但不简化别的前提的诸心灵；或者简化带有1、3或者5合取支诸前提的诸心灵，但不简化带有2、4或者6合取支诸前提的诸心灵；抑或，就此而言，简化在周二被获得的诸前提的诸心灵……事实上，联结主义者的这种建筑学对这些可能性的分析是**完全不偏不倚的**。（1988：47-48）

与单纯的数据符合相比，这些系统性论证中存在更多的问题：福多和派利申稍后的一段文字进一步佐证了上述观点：

> （一个联结主义者）描述一个网络以便支持表征着aRb的一个动态数组，当且仅当它支持表征着bRa的一个动态数组，这无疑是有可能的……问题在于，虽然这个建筑学容许这一做法，但它同样容许（一个联结主义者）描述另一个网络，以便其支持表征着aRb的一个动态数组，当且仅当其支持表征zSq的一个动态数组；或者，就此而言，当且仅当其支持表征着最后的莫西干人的一个动态数组。这个建筑学看起来对这些选择绝对是不偏不倚的。（1990：202）

很明显，与简单地覆盖这些数据相比，解释思维中的这些系统关系要冒更大风险。

很多批评者都对前面的这些段落作出了回应，指出福多等人提出这一观点在方法论上的缺陷。[9]虽然在这个表述中存在真正的缺陷，但归根到底，福多等人似乎意识到了某种具有科学含义的东西，意识到了科学哲学

家们会乐于去分析的某种东西，意识到了认知科学家们应当更加注意的某种东西。[10]因篇幅所限，这些论点只能用科学史上一个著名案例来佐证。在《物种起源》中，查尔斯·达尔文注意到了形态学、分类学、胚胎学和生物地理学中的规则性（regularities），对此他采用了比创世说更好的进化论来加以说明。[11]虽然进化论和创世论都有对这些公认规则的说明，但是进化论说明并不依赖像创世论说明所用的那种武断假设。例如，一个实例涉及两栖类脊椎动物的生物地理学：

> 圣樊尚很久以前就曾说过，大洋上点缀着许多岛屿，但从未发现有两栖动物（蛙、蟾蜍、蝾螈）。我曾煞费苦心地企图证实这种说法，发现除了新西兰、新喀里多尼亚、安达曼诸岛，或者还有所罗门群岛和塞舌尔以外，这种说法是对的。但是我曾经说过，新西兰和新喀里多尼亚是否应该被列为海洋岛还可怀疑；至于安达曼岛、所罗门群岛以及塞舌尔是否应该列为海洋岛就更值得怀疑了。那么多的真正海洋岛上都没有蛙、蟾蜍和蝾螈，是不能用海洋岛的物理条件来解释的；诚然，岛屿似乎特别适于这类动物的生长：甲为蛙曾经被带进马德拉、亚速尔和毛里求斯去，它们在那里大量繁殖，以致成为可厌之物。但是因为这类动物和它们的卵遇到海水就立刻死亡（据我们所知道的，有一个印度的物种是例外），它们很难输送过海，所以我们可以知道它们为什么无法存在于真正的海洋岛上。但是，它们为什么不在那里被创造出来呢？按照创世论就很难解释了。

根据进化论，两栖类动物首先出现在大陆上，但因为海水会杀死它们，因而妨碍了它们的跨洋迁移，一个人会发现（几乎毫无例外）海岛上没有两栖动物。根据创世论，上帝依据某个计划在这个星球上散布生命。问题在于，上帝的计划本能够在这些海岛上安置两栖类就像不在这些海岛上安置两栖类一样容易。对后面这一断言的证明就是：自然主义者们已经注意到，人类将两栖类动物输送到马德拉、亚速尔和毛里求斯，并使之很好地生存下去，这是有可能的。

创世论和进化论都具有可以被视为中心假设和辅助假设的因素。创世论中的中心假设当然是物种为神创之产物，进化论的中心假设当然是物种为变异演化之产物。这些理论所提供说明中的不同之处在于它们所诉诸的辅助假设。进化说明依赖着独立于正被考虑的解释任务且被确证的辅助假设。进化说明认为，大陆比海岛更古老，这是地质侵蚀观察所确证的一个

事实。进化说明还认为，咸水构成了两栖动物的一个迁移障碍，这是易为简单实验所确证的事实。相反，关于上帝计划散布生命形态的创世论假设，是没有被独立确证的。除非一个人已经先行判定上帝事实上确实根据一个计划分别创造了生物体，否则上帝创世计划的本质似乎就是难以理解的。在此意义上，创世论者依靠的是一个武断假设。

凭借对系统性论证中这一解释标准的粗略描述，我们能够想到 AR 关于合取排除的系统性可能会说什么。原子表征主义者会假定一系列在句法上的原子表征 {α，β，γ}，在这里：

> α 意指约翰爱玛丽，并且比尔爱玛丽，并且爱丽丝爱玛丽；
> β 意指约翰爱玛丽，并且比尔爱玛丽；
> γ 意指约翰爱玛丽。

177 用以从约翰爱玛丽，并且比尔爱玛丽，并且爱丽丝爱玛丽这一前提中，以及从约翰爱玛丽，并且比尔爱玛丽这一前提中推断出约翰爱玛丽的一个系统，可能具有类似图灵机的程序：

(P1) $s_0 \alpha \gamma s_1$
$s_0 \beta \gamma s_1$

如此这个程序使得：如果这个系统处于状态 s_0 扫描了一个 α 或者一个 β，那么它就会打印一个 γ，并且进入状态 s_1。但是，这样一个系统可能恰好具有以下程序：

(P2) $s_0 \alpha \gamma s_1$

(P1) 容许一个系统从 P&Q 以及从 P&Q&R 中推出 P，而在这里 (P2) 只容许一个系统从 P&Q 中推出 P。那么，回想一下福多和派利申关于这个可能性一定会说的话，AR 肯定能够模拟一种心理生活——在其中一个自主体能够既从 P&Q 中又从 P&Q&R 中推理出 P；但它同样也能够模拟另一种心理生活——在其中，你能够获得这些推理中的一个，而不能够获得另一个。当然，一个人能够为关于原子心理表征之存在的这个中心 AR 假说增加一个类似 (P1) 而非 (P2) 程序的辅助假说。但是，在这里我们有了一个会招致反对的辅助，这个辅助不可能独立于 AR 的真理而得到确证，正像关于上帝之创世计划的创世论假说不可能独立于创世论之真伪而得到确证一样。

对合取排除的系统性的经典说明，发端于 {α，β，γ，&，b} 这样一

些句法上的原子符号，在这里，α 意指约翰爱玛丽，β 意指比尔爱玛丽，γ 意指爱丽丝爱玛丽，& 意指合取，而 b 是一个无意义的空位符号。[12]一个类似图灵机的程序，能够使得系统推断出它的第一个合取支而不管在这个整体中是否还有两个或者更多的合取，这个程序就是：

（P3）	$s_0 \alpha \beta s_0$	$s_2 \alpha b s_3$
	$s_0 \beta R s_0$	$s_2 \beta b s_3$
	$s_0 \gamma R s_0$	$s_2 \gamma b s_3$
	$s_0 \& b s_1$	$s_2 \& b s_3$
	$s_1 b R s_2$	$s_3 b R s_2$

左侧这一栏中的指令指引这个系统扫描磁条上的第一个符号，并且抹去第一个“&”，同时，第二栏中的这些指令指引它从这个磁条剩余部分中抹去这些非空位的符号。这个方法存在的明显问题是，虽然有使合取排除之系统性发生的诸经典程序，诸如（P3），但是也存在不使合取排除之系统性发生的、满足了经典主义者要求的别的诸程序。换句话说，在经典主义中，一个人能够很容易掌握使合取排除之系统性发生的一个程序，同样也很容易掌握不会如此这般的别的程序。当然，一个人可以为经典主义增加某个其他辅助假设，以论证心灵的这个程序类似于（P3），这并非不可操作，但是这个辅助假说似乎难以企及独立于经典主义之真伪的确证。经典主义者不能比创世论者独立地确证其关于上帝的创世计划更好地独立确证这个辅助。再者，与原子表征主义者确证其关于心灵的计算机程序之本质的辅助相比，经典主义者也无法更好地独立确证这个辅助。因此，结果就是经典主义和 AR 都缺少一个对于推理性的充分说明。

7.2 认知表征的系统性

对这个论证，福多和派利申宣称，在一般的认知自主体中，具有某些思维的一种能力与具有别的思维的另一种能力在本质上是联结在一起的。这个本质的联结根据思维能力之间的两类心理层面上的依存（dependencies）而被清楚地表达出来。一方面，假如一个一般的认知自主体缺乏某些思维能力，那么这个自主体也会缺乏另一些思维能力。[13]另一方面，假如一个一般的认知自主体具有某些思维的能力，那么这个自主体也

就会因此具有另一些思维的能力。应当注意，一般认知自主体的这个特性并非在逻辑和概念上都是必不可少的。具有某些思维的能力对掌握任何其他思维可能完全没有影响，在此意义上，心灵可能完全是点状的。所以，需要解释的是为什么一般的认知自主体具有系统而非点状的心灵。

作为进一步澄清这些公认待解释项的一种方式，考虑一下经典主义者指望在哪里发现这些依存。对于一个一般的认知自主体，一个人指望在约翰爱玛丽这个思维的能力和玛丽爱约翰这个思维的能力之间发现一个内在的联结。如果一个一般的认知自主体缺乏能力以具有约翰爱玛丽这个思维的话，那么这个自主体也会缺乏能力以具有玛丽爱约翰这个思维。这是关于众思维间的诸本质联结的观念的一个方面，以下是另一个方面。假如一个一般的认知自主体能够想到约翰爱玛丽，那么这个认知自主体也就能够想到玛丽爱约翰。要牢记，认知表征的系统性论证中有待解释的就是众思维间的诸系统关系的真实存在，而非这些关系所在之处。揭示出这些依存所在之处，仅是为思维是系统的这一主张增添某种直观可信性的一个说明步骤而已。再一次，我们想要知道为什么一些思维与另一些思维联结在一起，而非根本不与任何东西联结在一起。

179 那么，AR会如何试图解释思维的这一系统性呢？AR可能会推测，心灵这个计算机程序之结构与一个思维相对应的表征的丧失，会导致与一个或者更多个另外的思维相对应的表征的丧失。此外，这个程序会使得与某些思维相对应的诸表征的增加，导致与另外的思维相对应的诸表征的增加。因为诸表征（部分）是思维的潜在基础，所以不同的表征间具有的这种联结会（部分地）构成这些对应思维的能力间的联结。

与思维系统性的AR说明有关的问题，与对推理系统性的AR说明的问题在本质上是相同的。让我们假定，如果仅仅出于论证的缘故，一个人事实上能够为一个计算机编程，使得标志着各种不同表征的能力之间的依存产生出来。假定存在显示出诸表征之间依存的一类AR计算机程序，那么，很明显也会存在着虽然满足AR的这些条件但却不会使诸表征之间的依存产生出来的计算机程序。一个人或许会说，一个AR计算机程序能够很容易地成为系统程序，就像它能够很容易地不变成系统程序一样。当然，一个人能够为AR增加一些辅助假设，其大意为心灵的计算机程序是如此这般，以至于使诸思维之间的这个依存产生出来，而非不能这

样。但是，对于缺乏确证 AR 的独立确证来说，这些假设似乎还是难以企及。

虽然我们明白为什么 AR 不解释思维的系统性，但我们还是要考虑经典主义是否能够符合这一相同的解释标准。假定存在一组句法上的原子表征，Γ=｛约翰，简，玛丽，利萨，爱，恨｝。同时存在组合了 Γ 中的这些原子，以至于产生了 Γ_1^* 组中的诸计算机程序。Γ_1^* =

｛约翰爱约翰	约翰爱简	约翰爱玛丽	约翰爱利萨
简爱约翰	简爱简	简爱玛丽	简爱利萨
玛丽爱约翰	玛丽爱简	玛丽爱玛丽	玛丽爱利萨
利萨爱约翰	利萨爱简	利萨爱玛丽	利萨爱利萨
约翰恨约翰	约翰恨简	约翰恨玛丽	约翰恨利萨
简恨约翰	简恨简	简恨玛丽	简恨利萨
玛丽恨约翰	玛丽恨简	玛丽恨玛丽	玛丽恨利萨
利萨恨约翰	利萨恨简	利萨恨玛丽	利萨恨利萨｝

但是，也有组合 Γ 中的原子产生 Γ_2^* =｛约翰爱玛丽，简恨利萨｝这一组中的诸计算机程序。Γ_1^* 是系统的，Γ_2^* 则不是。在 Γ_1^* 中这些表征之间存在依存，而在 Γ_2^* 的这些表征之间则没有。那么，如果一个人有一个经典的表征系统，那么，一个人就能够很容易地掌握一套系统的表征，就像他也可以不掌握一样容易。因此，经典主义者会愿意增加一些辅助假设，以便 Γ 形成类似 Γ_1^* 的一个组，而非类似 Γ_2^* 的一个组。然而，一如既往的是，我们缺乏对这个辅助的独立确证。所以，再一次声明，即便 AR 缺乏一个对思维系统性令人满意的说明，经典主义在这方面也并没有更好的办法。

7.3 诸表征的构成性

诸认知表征的系统性是关于某些思维依赖着另一些思维的问题。诸表征的构成性则与思维的另外一个属性有关：可能目前正在发生的诸思维在语义上是相关的。简单地说，诸思维意味着这些相同对象的相同属性和关系。因此，上一节业已指出，我们应该指望在哪里找到介于诸思维之间（即介于在语义上相关的这些思维之间）的本质联结。现在，这个语义相

关性被转换成了一个待解释项。

福多和派利申（1988：41）认为，系统性和构成性密切相关，并且最好把它们看作是一个单独现象的两个方面。如果真是这样的话，那么系统性和构成性在逻辑上就是截然不同的属性。这样，作为一个心理学事实，如果一个人丧失了这个能力以获得约翰爱玛丽这一思维，那么，一个人也就因此丧失了获得亚里士多德是一个航运巨头这一思维，这在逻辑上是可能的。作为一个心理学事实，如果一个人有这个能力以获得约翰爱玛丽并且玛丽爱她自己这些思维，那么，一个人也就会有这个能力以获得亚里士多德是一个航运巨头这一思维，这在逻辑上也是可能的。诸认知自主体是系统的而非构成的，这一发现也许会令人极度困惑，但是，这个发现却是可能的。另一方面，具有相互之间不依赖但在内容上相关的诸思维也是可能的。一个人能够拥有符合——约翰爱玛丽，玛丽爱约翰，约翰恨玛丽，并且玛丽恨约翰——这些思维的能力，而这一能力的丧失并不会导致任何别的能力的丧失。此外，一个人能够拥有符合——约翰爱玛丽，玛丽爱约翰，约翰恨玛丽——这些思维的能力，而同时不拥有符合玛丽恨约翰这一思维的能力。诸思维之间的依赖性在逻辑上并不暗示这些思维之间的内容关系，而且这些思维之间的内容关系在逻辑上并不暗示诸思维之间的依赖性。

那么，AR 可能会如何解释（对一个一般的认知自主体来说）一组可能目前正在发生的思维在内容上是相关的这一公认事实呢？AR 会说，约翰爱玛丽这一思维牵涉到意指约翰爱玛丽的一个句法原子 α，而且玛丽爱约翰这一思维牵涉到意指玛丽爱约翰的一个句法原子 β。现在，虽然很可能会有满足这个描述的计算机程序，但是也明显会有不满足这个描述的计算机程序。计算机很容易被写入类似的程序，就像它很容易不被
181 写入这样的程序一样。因此，仅仅 AR 并不导致思维的构成性。此外，要是 AR 不增加一个辅助假设（其大意为心灵的计算机程序是如此这般），以至于如果自主体无法能够掌控 α 的话，它也就不能够掌控 β，那么，情况就没有改观。这个辅助完全属于没有 AR 之真就不可能得到确证的那种。

对诸表征构成性的这个经典说明借助了如下假设：诸思维牵涉到一系列句法原子，以及将这些句法原子组合成句法分子的某种方式。然而，存在有构建分子的诸方式，一系列的句法原子 Σ＝{约翰，简，爱，恨} 能够被联合起来形成一系列的字符串 Σ_1^*＝

{约翰爱约翰　　约翰爱简　　简爱约翰　　简爱简
约翰恨约翰　　约翰恨简　　简爱恨翰　　简恨简}

或者，它能够被组合起来形成一系列的字符串 Σ_2^* ={约翰爱约翰，简恨简}。

我们已经列举的这一系列经典主义者的假设，它并不导致诸思维之间存在有内容关系。因此，经典主义必须求助于一个辅助假设，其大意为 Σ 被联合在一起以形成类似于 Σ_1^* 系列中而非类似于 Σ_2^* 系列中的句法条目。但是，这样一个辅助并未被独立地确证，结果是留下了对诸表征构成性的一个并非货真价实的解释。我们再一次发现，无论是 AR 还是经典主义都缺少对思维构成性的解释。

7.4 别的系统性论证

到现在为止，我们已经有了关于下述企图之基本缺陷的三个例证：即企图使用系统性论证来为一个经典的认知表征系统所提供的假设进行辩护。所有当前的企图都依赖着在某种意义上武断的诸辅助假设。这或许表明，这些系统性论证所求助的这个解释标准高得不切实际。[14]本节的论证则是要表明，这个标准并不是太高。它会坚持主张福多和派利申著作中所暗示的这个解释标准，但又会求助于别的系统性论证中的另外一个标准。当然，该方法并不会向我们显示如何解释福多和派利申已经引入的这些系统关系，但是它会提供某个可废止的理由以使我们相信存在一个组合的（combinatorial）思维语言。

在上一节中，我们注意到表征系统性和构成性的可分离性。我们能够将其中一个而同时不将另一个看作是一个心理事实。然而，这里是另一个心理事实。如果一个一般的认知自主体具有一个系统的心灵，那么它也就具有一个构成的心灵。这又是为什么？另外，这些相互依赖的思维在内容 *182*
上是相关的，这又是为什么？我们已经注意到，对这些独立的特性，AR 和经典主义都没有一个适当的说明，但是用一个去指引另一个会怎样呢？

经典主义似乎具有对这个同现（co-occurrence）的合适说明。鉴于经典主义说明系统性所需要的配备条件，要是没有别的假定，一个人就要有

必要的配备以说明构成性。因此，对系统性的一个经典说明依赖以下假定：存在着组合起来以形成句法上的诸分子表征的句法上的诸原子表征，并且这些原子表征满足语义构成性原则。一些思维依赖另一些思维的原因是它们具有共同的原子或者分子表征。因此，约翰爱玛丽这一思维依赖玛丽爱约翰这一思维的原因就是它们共有：(a) 一个原子表征“约翰”(这意味着约翰处于“—爱玛丽”和“玛丽爱—”这两个语境当中)；(b) 一个原子表征“玛丽”(这意味着玛丽处于“约翰爱—”和“—爱约翰”这两个语境当中)；(c) 一个原子表征“爱”(这意味着爱处于“约翰—玛丽”和“玛丽—约翰”这两个语境当中)；(d) 一个共同的语法结构。鉴于对约翰爱玛丽思维和玛丽爱约翰思维的相互依赖性的说明，诸思维是内容相关的事实就用不着额外的假定了。为说明系统性才需要的这一系列经典假定必须伴有构成性。

相比之下，对于联结思维的系统性和思维的构成性，AR 并没有令人满意的方法。这主要是因为一个句法的原子表征的内容完全不依赖另一个句法的原子表征的内容。鉴于 AR 为说明诸思维的这个相互依赖性所需要的相互依赖的思维没有理由同时在内容上是相关的，即便 AR 论者能够达成这样的假设，即心灵的程序是如此这般的，以至于具有它们各自内容的两个句法条目 α 和 β 是相互依赖的，它也需要关于 α 和 β 特定语义内容的一个辅助假设，以证明 α 和 β 在内容上也是相关的。然而，这样一个额外假设只会是这样一种假设，即独立于关于心理表征的 AR 系统的假设不能够被确证。

这种解释论证的效力在一个事例中得到证实。古代天文学家已经观察到，大体上外地行星火星、木星和土星自西向东穿过一些恒星。然而，这个一般趋势受到一个逆行期的周期性干扰，这个逆行期使这些行星在它们向东的运动中减速、停止，并在再一次减速、停止并最终重新开始正常的向东运动之前，一度以向西逆行的方式移动。托勒密体系的天文学家们意识到这些不规则性，并且能够为它们提供一个基本正确的模型。这个基本观念是要存在诸如火星一样在一个本轮上盘旋的一颗行星，然后这个本轮
183 沿轨道飞行在一个均轮的末端。通过小心调节这个本轮和均轮的旋转的相对大小和相对速度，就有可能大致生成这些行星已被观察到的运动。对逆行运动的哥白尼式说明在根本上是不同的。根据哥白尼，逆行运动仅仅是由于地球赶上一颗行星（它们都绕太阳旋转）而引起的相对运动。哥白尼式说明被证明远优于托勒密式说明的地方就在于它对逆向运动特殊特性解

释的能力上：当这颗行星处在太阳的对立面时，它们总会发生。无论何时，当一颗行星处在它向东的逆行运动的正中间时，就可以发现它和太阳180度的分离开来。通过对均轮系统上本轮特性的巧妙操控，托勒密天文学能够为逆行运动的这个特性提供说明，但是哥白尼系统展示了更详尽的事实，而无需任何额外的假设。哥白尼系统提出了逆行运动的本质，其结果必然就是逆行运动会反相发生。逆行运动的哥白尼式说明的这些必要元素足以解释反相发生的逆行运动。托勒密式的说明则没有这个效力。

7.5 功能组合主义能够解释思维中的这些系统关系吗？

到此为止，我们已经考虑过了一系列的系统性论证如何对经典主义和AR产生影响。然而，对这些论证的一种回应已经断言认知牵涉到第三种形式的表征主义，即非经典的FC。根据FC，诸分子表征不过是其原子表征的（可计算的）应变物。这些原子不必是它们源自于其中的这些分子的精确的时空部分。这个观念的一个缘起是保罗·斯莫伦斯基（Paul Smolensky）的张量积理论。另一个是通过哥德尔数。事实上，哥德尔数经常被引用来表明在功能上组合的诸表征之完璧无瑕的科学声望。[15]福多和迈克劳林（1990）及福多（1996）关于保罗·斯莫伦斯基的理论已经提出了大量技术和概念上的问题，最终将这个讨论引向了我们在此没有时间探讨的方向和深度上去。然而，这给了我们一个机会来探索一条更稳妥的批评的界限。我们能够坚守福多和派利申（1988）所暗示的这个解释标准以表明，体现在哥德尔数中的这种在功能上组合的表征并不能解释思维中的系统关系。

假定，我们试着用哥德尔数来解释一个认知自主体如何能够从约翰爱玛丽，并且约翰爱简，以及约翰爱玛丽，并且约翰爱简，并且约翰爱爱丽丝中推出约翰爱玛丽的论证。这个哥德尔数的故事可能会从下述这些原子表征开始：

“1”意指约翰爱玛丽；

“2”意指约翰爱简；

“3”意指约翰爱爱丽丝；

“4”意指和。

这将会构成一个分子表征的n个原子表征，给了我们一个序列的n个数字。因此，要表征约翰爱玛丽并且玛丽爱约翰这个命题，我们就会使用<1,4，2>这个序列，而要表征约翰爱玛丽并且约翰爱简并且约翰爱爱丽丝这个序列，我们就会使用<1，4，2，4，3>这个序列。这个序列的n个数字为我们提供了前面n个质数的幂，它们随后被乘在一起，以便完成我们的哥德尔表征。所以，我们有了：

“1”意指约翰爱玛丽；

“144”（$=1*2^4*3^2$）意指约翰爱玛丽，并且约翰爱简；

“30870000”（$=1*2^4*3^2*5^4*7^3$）意指约翰爱玛丽，并且约翰爱简，并且约翰爱爱丽丝。

这里这个表征系统是非经典的，因为一个给定的句法分子的每一个个例，诸如“30870000”，不需要精确地包含用以构建它的这些句法原子的每一个个例，即“1”、“2”、“3”或者“4”的个例。[16]为了从这些案例中获得合取排除的这个系统性，我们仅仅假设，存在有对“144”和“30870000”作出反应以产生一个“1”的计算机程序。

这里的问题是我们已经预料到的。宽泛地讲，产生一个既对“144”又对“30870000”作出反应并写出“1”的计算机程序，就像产生一个对“144”而不对“30870000”作出反应并写出“1”的计算机程序一样容易。然而，增加某个假设——大意是这个程序确实产生“144”和“30870000”——将会是徒劳无功的，因为，只要没有对这个系统把哥德尔数作为认知表征这一假设的确证，那么这样一个假设就不可能得到确证。

思维的系统性又怎样呢？假定我们有一系列命题：

约翰爱约翰	约翰爱玛丽	约翰爱简
玛丽爱约翰	玛丽爱玛丽	玛丽爱简
简爱约翰	简爱玛丽	简爱简

我们用常见的阿拉伯数字开始建立对这些命题的一个哥德尔数表征，并且赋予它们下述的语义解释：

“1”意指约翰，

“2”意指玛丽，

“3”意指简，并且

“4”意指爱。

这里我们有了我们的哥德尔数系统原子表征。接下来我们要把每个命题与一个序列的数联系在一起。这样，约翰爱玛丽这个命题就与<1，4，2>这个序列联系在一起，而且约翰爱简这个命题就与<1，4，3>这个序列联系在一起。我们得到了被这个序列中的 n 个数字（numerals）所表征的 n 个数（numbers），并把它们作为 n 个质数的幂。这 n 个质数的乘积产生了另一个数，这个数的阿拉伯十进位表征能够被我们作为我们的命题的表征。因此，我们得到了这个三个数字序列<1，4，2>（它与约翰爱玛丽联系在一起），并且将之应用于三个质数以给予我们肆仟零伍拾这一数字（$=2*3^4*5^2$），在阿拉伯数字符号中写作“4050①”。类似地，我们得到了三个数字的这个序列<1，4，3>（它与约翰爱简这个命题联系在一起），并且用它结合三个质数，以便约翰爱简与“20250”这个数（$=2*3^4*5^3$）联系在一起。根据这个安排，我们用下述数字表征我们的一系列命题：

810（$=2*3^4*5$）表征约翰爱约翰
4050②（$=2*3^4*5^2$）表征约翰爱玛丽
20250（$=2*3^4*5^3$）表征约翰爱简
1620（$=2^2*3^4*5$）表征玛丽爱简
8100（$=2^2*3^4*5^2$）表征玛丽爱玛丽
40500（$=2^2*3^4*5^3$）表征玛丽爱简
3240（$=2^3*3^4*5$）表征简爱约翰
16200（$=2^3*3^4*5^2$）表征简爱玛丽
81000（$=2^3*3^4*5^3$）表征简爱简

在对由此产生的表征进行探查时，我们发现，表征间的相互依存产生了诸思维之间的系统关系。也就是说，鉴于约翰爱简这个表征和简爱约翰这个表征都依靠在这些地方有一个“0”的能力，我们能够获悉，约翰爱简的这个能力和简爱约翰的这个能力之间会有一个依存。

然而，从前面的讨论中我们应已获悉，上述内容仅仅表明哥德尔数能够显示诸思维之间的一个依赖。它并不表明哥德尔数能够解释诸思维之间的相互依赖性。我们不得不考虑在这个说明中是否有某个武断的辅助假设。此外，正如我们可能会预期到的，确实有。潜在于上述这个系统的一

①② 原文中的计算结果 2025 是一处明显的计算错误，所以更正为 4050。——译者注

个假设是，这些哥德尔数（即带幂的这些质数的乘积）通过常见的基数十
186 进制来表达。凭借这个假设以及对代替这些原子表征的诸数字的选择，代替这些命题的某些哥德尔数被证明具有共同的元素，因此这些分子表征之间具有依存。然而，下述这一假定却是武断的：即心灵为这些质数的乘积使用一个基数为十的表征系统。我们可以提出福多和派利申会提出的问题：在一个哥德尔数系统的框架内，一个人能很容易地使用一个基数十的表征系统，就像始终不使用它一样。或者我们可以说，表征的一个哥德尔数系统是一个基数十系统——这一假设是无法独立于当前的解释挑战而被确证的。一个可选择的假设是，这些哥德尔数发生在一个十万进制的系统中，在这个系统中十万的原子符号里没有任何在句法上共同的东西。那么，这个系统中的一系列原子数字可能是类似于｛0，1，2，3，4，5，6，7，8，9，a，A，b，B，…，z，Z，…，♥，◆，♣，♠｝的东西。在这样一个系统中，代替我们命题的这些数字没有共同的元素，因此表征着我们这一组命题的任何数字都不会有相互依赖，从而这些相应的思维之间也没有相互依赖。[17]

其次，对诸可能思维间这些内容关系的哥德尔数说明有什么毛病呢？在本质上，这个经典说明犯的是同一个毛病：一个“哥德尔数语法”很容易产生一系列在内容上相关的分子表征，同样也很容易不产生这样的表征。我们再一次列举一系列的原子表征｛1，2，3，4｝，在这里

“1”意指约翰，
“2”意指玛丽，
“3”意指简，
“4”意指爱，
“5”意指恨。

这一系列的原子表征能够产生具有以下内容的分子表征：

约翰爱约翰	约翰爱简	约翰爱玛丽	约翰爱利萨
简爱约翰	简爱简	简爱玛丽	简爱利萨
玛丽爱约翰	玛丽爱简	玛丽爱玛丽	玛丽爱利萨
利萨爱约翰	利萨爱简	利萨爱玛丽	利萨爱利萨
约翰恨约翰	约翰恨简	约翰恨玛丽	约翰恨利萨
简恨约翰	简恨简	简恨玛丽	简恨利萨
玛丽恨约翰	玛丽恨简	玛丽恨玛丽	玛丽恨利萨
利萨恨约翰	利萨恨简	利萨恨玛丽	利萨恨利萨

它同样很容易产生带有这些内容的诸表征：

约翰爱玛丽　　简恨利萨

要牢记，哥德尔数必定容许以某种方式在一个给定系列的原子表征之上产生不同系列的分子表征。出于认知理论的目的，必定有某个原则或者假设使得“一个哥德尔数语法”并不产生带有诸如“约翰约翰约翰，约翰约翰爱和约翰爱爱”这种“内容”的诸表征。无论关于哥德尔数的理论是否作出假设，那都会是一个使得对思维中这些内容关系的说明崩溃的、成问题的假设。这个假设将会是一个没有独立确证的假设。所以，哥德尔数就像经典主义一样，不能解释思维的系统性。

这使我们想起了上述研究的系统性这一不太常见的特性，即系统性和系统相关性的共现。这里哥德尔数的情况怎样呢？应该很清楚，虽然可能产生在系统上相关诸表征之间的依存，但是也有可能产生在系统上不相关的诸表征之间的依存。“20250”这个表征（它在我们上面的事例中表征约翰爱简）在本质上是与“3240”（它在我们上面的事例中表征简爱约翰）这个表征联结在一起的，但是这个联结不依赖于“20250”和“3240”的数字内容。为了说明诸思维之间的内在联结，一个哥德尔数理论所假定的东西，并不暗示着诸思维之间必定有语义关系，因此也并不解释这个关系。

7.6 结　论

除了澄清这些论证中相对次要的任务，本章还有别的更重要的目标。第一，它更多地注意了下述事实，即这些系统性论证牵涉到在相互竞争的诸解释间进行选择的某个原则。关于这个原则有很多东西需要讲一下，但是在本质上它好像关系到解释必须避免即兴的辅助假设。第二，在辩护经典主义者的论点（即诸如原子表征主义和功能组合主义等理论并不解释思维中的这些系统性关系）时，本章极力主张甚至经典主义也未能解释思维中的这些系统关系。第三，本章阐述了由福多和派利申提出的另一种系统性论证。这一论证具有福多和派利申预期的解释优势，并表现出了 AR 和 FC 中不具备的经典主义效力。

【注释】

[1] 会令行为主义者感到振奋的一种传统语义取消主义存在于布鲁克（Brooks，1977）和范戈尔德（van Gelder，1977）的表征取消主义之中。已经引起哲学家兴趣的另一个重要的研究领域是“心灵的模块性”（参见福多，1983；卡米诺夫-史密斯[Karmiloff-Smith]，1992）。然而，对心灵模块性的任何充分讨论都肯定是另一章的主题。

[2] 例如，参见斯莫伦斯基（1995）、卡明斯（1996）、霍根和廷森（Horgan and Tienson，1996）。

[3] 见卡明斯（1996：612），在这里这个观点似乎被漏掉了。

[4] 范戈尔德和尼克莱森（van Gelder and Niklasson，1994）连同卡明斯（1996）都高估了人类推论文献对推理系统性论证的意义。

[5] 同上：37，38，40。在这里施加给系统性论证的限制希望给予有限认知能力一种假设。

[6] 毕竟，这是注释5观点的一部分。

[7] 范戈尔德和尼克莱森（1994）连同卡明斯（1996）似乎都漏掉了这个观点。

[8] 参见范戈尔德和尼克莱森（1994）、卡明斯（1996）、哈德利和海沃德（Hadley and Hayward，1996）。

[9] 例如，见斯莫伦斯基（1995）、卡明斯（1996）和哈德利（1997）关于必需的待解释项和原则性解释的讨论。

[10] 在此处无人为之辩护的主张，在艾扎瓦（准备出版）那里得到了辩护。

[11] 艾扎瓦（1997a、1997b）研究了其他的例证。

[12] 这里的α、β和γ简化了句子，而非一阶逻辑中的式子。这仍是一个经典的说明，因为组合的表达式、结构敏感性（structure sensitivity）、构成性原则等仍然都处在符合规则的位置。让α、β和γ表征句子仅仅简化了这个讨论。

[13] 这个反事实的效力不在于说明，如果一个人做了移除一种思维的脑损伤手术，那么至少另一种思维也会因此被移除。这样一个待解释项大概是一个可落实的事实（implementational fact），而不是一个可被纯心理层面的理论解释。福多等人针对的这种依赖一定要被理解为一种纯心理层面的依赖。

[14] 哈德利（Hadley，1997）对我们阐明系统性论证所凭借的这个方法作出了这一反应。

[15] 例如，参见范戈尔德（1990）。

[16] 当然，“30870000”确实包含“3”的一个个例，这个“3”是它由之源出的这些原子中的一个，但这只是偶然的。

[17] 在此作一点澄清是适当的。回想一下，对于系统性论证，我们假定仅仅有限的一组思维牵涉进来。要注意，对任何一组有限的思维，都会有用以表示这些哥德尔

数的某个基数，这个基数不会导致这些思维间的相互依存。有鉴于此，建议这个哥德尔数不能够解释这些思维间的相互依存。

参考文献

Aizawa, K. (1997a). "Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward." *Minds and Machines*, 7: 39-55.

—— (1997b). "Explaining Systematicity." *Mind and Language*, 12: 115-136.

—— (in preparation). *The Systematicity Arguments*.

Brooks, R. (1997). "Intelligence without Representation." In J. Haugeland (ed.), *Mind Design II*. Cambridge, MA: MIT Press: 395-420.

Cummins, R. (1996). "Systematicity." *Journal of Philosophy*, 93: 591-614.

Darwin, C. (1859). *The Origin of Species*. London: John Murray.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

—— (1996). "Connectionism and the Problem of Systematicity (continued): Why Smolensky's Solution Still Doesn't Work." *Cognition*, 62: 109-119.

Fodor, J. and McLaughlin, B. (1990). "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work." *Cognition*, 35: 183-204.

Fodor, J. and Pylyshyn, Z. (1988). "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition*, 28: 3-71.

Hadley, R. (1997). "Explaining Systematicity: A Reply to Kenneth Aizawa." *Minds and Machines*, 7: 571-579.

Hadley, R. and Hayward, M. (1996). "Strong Semantic Systematicity from Hebbian Connectionist Learning." *Minds and Machines*, 7: 1-37.

Horgan, T. and Tienson, J. (1996). *Connectionism and the Philosophy of Psychology*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Niklasson, L. and van Gelder, T. (1994). "On Being Systematically Connectionist." *Mind and Language*, 9: 288-302.

Smolensky, P. (1995). "Reply: Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Cognitive Architecture." In C. MacDonald and G. MacDonald (eds.), *Connectionism: Debates on Psychological Explanation*: 223-290.

van Gelder, T. (1990). "Compositionality: A Connectionist Variation on a Classical Theme." *Cognitive Science*, 14: 355-384.

van Gelder, T. (1997). "Dynamics and Cognition." In J. Haugeland (ed.), *Mind*

Design II. Cambridge, MA: MIT Press: 421-450.

van Gelder, T. and Niklasson, L. F. (1994). "Classicism and Cognitive Architecture." *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Atlanta, GA: 905-909.

第8章 概 念

埃里克·马戈利斯，斯蒂芬·劳伦斯
(Eric Margolis and Stephen Laurence)

人类心灵有不可思议的表征能力。我们没有受限于只思考此时此地， 190
正如我们未受限于只思考涉及我们最直接需要的对象和属性。相反，我们能够思考在时间或空间上远离我们的诸事件（例如，亚伯拉罕·林肯、半人马座阿尔法星），并且专注于从直接感觉经验中高度抽象出来的诸事件（例如，民主、圆周率）。我们甚至能够思考从来不曾或者绝对不会存在于真实世界中的诸事件（例如，圣诞老人、独角兽和燃素）。哲学史的中心问题之一就曾是：我们如何能够实现这一切？我们何以能够在思维中将这个世界再现给我们自身？在回答这个问题时，哲学家和心理学家们常常认为，我们的思维能力建基于我们的概念能力之中。诸思维都具有成分和部件，即概念。[1]结果，全部科学、文学和艺术——还有日常思维——都能够被看作是源自于人类概念系统那令人震撼的表达力。

鉴于诸概念对于理解认知之本质所具有的基础作用，要想提供关于诸概念的一个理论而忽视关于心灵的众多基础问题是不可能的。事实上，对于划分切入心灵研究的不同方法甚至不同世界观来说，概念理论已经变成了一个焦点。例如，它相合于这样的问题：是否真的有思维？语义属性是否与对人类行动的研究相关？[2]同样地，它处在关于哲学是不是一项先验事业这一分歧的根源上。不用说，我们不会在这里讨论所有问题。为了使论述集中并易于控制，对于一般在心灵哲学且特别在概念理论中仍然有争论的问题，有必要进行某些前提设定。[3]

过去50年间，凭借众多重要并持久的成果，概念理论已经成为哲学和心理学研究中最有活力的领域之一。接下来，我们将概述许多最有影响的 191
理论，要着眼于对它们进行区分的关键问题——概念结构问题。[4]我们会争论说，概念结构现行的各种不同类型中，没有一个是完全令人满意的。

这使我们要重新思考概念结构的本质本身，并且要区分几种在范畴上不同的类型结构。

8.1 定义的结构

自古以来，对概念之本质的学说创建就已经被以概念的**经典理论**著称的一个说明所主导。这个说明曾一度相当强势，以至于直到 20 世纪 70 年代，真正的可选择项才首次得以阐明。此外，虽然这些可选择的理论在某些方面完全不同于这个经典说明，但是它们都深受经典说明的恩泽。事实上，即便声称现存所有的概念理论大体上都是对经典理论及其缺陷的反动，也算不得夸大其词。[5]所以，领会经典理论及其缺陷的促成因素，对于理解关于概念本质的工作是必不可少的。

根据经典理论，概念是复杂的心理表征，这些表征的结构通常[6]会编码其自身应用（application）之充要条件的一系列规范（specification）。[7]例如，考虑一下**单身汉**这个概念。**单身汉**实际上是一个复杂的心理表征，该表征的主要成分是**未婚的**和**男人**。只有某个东西满足了这些成分概念的每一个，它才能被归入**单身汉**，或者处于**单身汉**的外延之下。再比如，举另一个事例，知识这个概念可以被分解为被辨明的真信念。这样，只有某个东西是被辨明了的真信念的一个实例，那么，它就能被归入知识这个概念。[8]

有很多东西使这个简单并且在直观上诱人的理论受到欢迎。该理论的很多效力和精确性都源自于如下事实：它能够提供对大量关键心理现象的说明，即天衣无缝地契合刚刚对这个指称限定的处理。例如，范畴化是涉及概念形成的所有过程中最基本的东西之一。我们的大多数高阶认知能力——更不要说我们自身的生存——都依赖于我们迅速而且可靠地确定我们周围环境中的不同对象属于哪个范畴的能力。这种经典理论的说明是自然而且非常有说服力的。例如，在范畴化某物之为一只鸟时，就是一个人存取和分解**鸟**这个概念，并且核对其成分是否适用于被考虑对象。如果每一种成分都适用，那么这个对象就被视为一只鸟；如果至少一种关键成分不适合，那么这个对象就不是一只鸟。经典理论为概念学习提供了一个同样有力的说明。概念学习的过程以与范畴化非常类似的方式运作，但是这两种过程却是完全相反的。也就是说，要获得一个概念，一个人要从该概

念的诸成分出发，并且根据一个人的经验，集合（assemble）这些成分。按照这个观点，学习就是一种结构操作。一个人首先具有某些概念，并将这些概念合在一起形成新颖的、复杂的诸概念。简而言之，经典理论为指称限定、范畴化和学习提供了一个精彩的统一说明。[9]

虽然经典理论很有魅力，但它在今天却鲜有拥护者了，因为它面临着许多极有挑战性的异议。在本节的剩余部分，我们会简要地回顾这些异议中的一部分，以显示出诸种竞争理论背后的某些动机，并突出随后会涉及的许多主题。

或许，对经典理论最紧迫的质疑在于现在完全没有无争议的定义实例。如果经典理论是一种新研究纲领的一部分，这就不会成为这样一个问题。但事实是，尽管有超过两千年的密集且持续的哲学分析，但却少有（如果有的话）一个概念能够被说成是只有被定义了才可行的案例。事实上，这个研究纲领的失败是众所周知的。

举一个人所皆知的事例，可以考虑一下不久前我们为**知识**这个概念所引用过的定义。这个提议是，**知识**能够被分析为**被辨明的真信念**。虽然这个定义听起来似乎是有道理的，但它却容易招致一系列强有力的反例，而首先注意到这些反例的是埃德蒙·格蒂尔（Edmund Gettier）。下面这个事例是根据丹西（Dancy，1985）改写的。亨利正在关注温布尔登的男单比赛。他打开电视机观看决赛，看到麦肯罗（McEnroe）战胜了康纳斯（Connors）。于是，亨利相信麦肯罗赢得了比赛，并且他完全有理由推断麦肯罗是今年的冠军。但是，亨利并不知道的是，由于电视网络摄像机的一个问题，这场比赛不能即时播放，正在播放的其实是去年比赛的录像。不过，在今年的比赛上，麦肯罗重现了去年的表演，在决赛中击败了康纳斯。所以，亨利的信念——麦肯罗是今年的冠军是正确的，也是被辨明的，但是很少有人愿意说亨利知道麦肯罗是今年的冠军。

类似的问题不仅仅只是在哲学上令人感兴趣的诸概念。正如维特根斯坦在《哲学研究》中的精彩质疑：日常概念似乎并不比哲学概念更可定义。维特根斯坦的主要事例之一是**游戏**，对这个概念，他考虑了许多在起初看似有道理的定义，但这些定义最终都招致了毁灭性的反例。甚至哲学的常备事例，**单身汉**也不是没有问题的。教皇是一个单身汉吗？一个长期处在同性关系中，与其爱人生活在一起又自称同性恋的男人又如何呢？两者都是未婚男人的案例，但两者似乎都不是单身汉。

经典理论的辩护者们会回应说，虽然定义确实难以获得，但这并不必

然意味着它就根本不存在。或许定义是隐藏的，因此不易为反省所获（例如，雷伊，1993；克里斯托弗·皮科克［Christopher Peacocke］，1998）。然而，通常的感受是，定义之所以如此难以寻找，最有可能的原因仅仅是，它根本就不存在。

193 经典理论的另一个问题是，由于其对定义的承诺，它也就承诺了分析/综合区别的一种形式——经蒯因的著名批判之后，这种区别被许多哲学家认为是大有问题的。蒯因之批判的一个方面集中围绕他的这一观点，即确证是带有整体论性质的。确证牵涉到了诸如简单性、保守性、总体一致性和类似性等因素。此外，既然确证依赖诸辅助假设，那么当一个理论上的主张碰到了不一致的数据时，一个人不能够预先断定需要被抛弃的是这个主张而非一些辅助假设。所有这些似乎表明，我们还没有到达科学研究领域内的真理的先验通道。再者，我们预先并不知道科学的影响力有多远。或许它看起来像一个概念必然性的东西（并因此像是分析的，且免于修正），实际上可能被证明是人们被其自身理论想象的缺乏所误导罢了。

然而，要注意，如果一个概念有了一个定义，那么这个定义会强烈地抑制科学中的理论发展，并且会对我们关于世界有能力发现的东西施加一种先在的限制。例如，如果对**直线**的恰当分析是**两点之间的最短距离**，那么，一个人似乎就是不可能发现一条直线未必总是两点之间的最短距离。而且，如果对**猫**的恰当分析是**（某某类型的）动物**，那么一个人就不可能发现猫不是动物——这样一些定义似乎被看作是非常有道理而且无可挑剔的。然而，正如希拉里·普特南（1962）所指出的，情况并非这么简单。空间是非欧几里得的这一发现之后，我们现在能够明白，第一个定义实际上是错误的。而通过一些科幻小说，我们能够明白，（至少看起来似乎）第二个也是错误的（或许猫实际上是火星人所控制的机器人，根本就不是动物）。但是，如果直线和**猫**具有经典理论所主张的定义，那么这些后来的发现就将被完全禁止。它们就根本不可能出现。像这样的诸事例威胁到了经典理论的真正基础。一个定义可能看似把握了一个概念的结构，但是这个表象可能仅仅只是随后的发现就会帮助我们看穿的一种假象。[10]

联系到类似的诸事例，一个人会发现反对定义的其他考虑要素——特别是，索尔·克里普克（Saul Kripke）和希拉里·普特南关于名称的语义学和自然种类术语的重要著作（特别参见，克里普克，1972/1980；普特南，1970、1975）。克里普克和普特南的研究对象是指称的描述理论，根据该理论，由于知道了识别一个名称或者种类术语的指称的一个描述，所

以某人能够使用该名称或者种类术语。然而，要注意的是，经典理论只是这个描述理论的一种形式，它只在概念而非语词的层面上有效。基于这一理由，普特南和克里普克的所有论证都是关于它的评价。他们的论证之一是详述了蒯因的下述论点：我们能够对特定物质的种类有所发现；而该发现揭示出我们关于该种类的本质的理解是错误的。与此密切相关的是忽视问题：如果人们关于一个种类的某些属性的认识是错误的，那么他们常常也会忽视体现该种类真正本质的那些属性。对金子的统一性具有决定性的要素被证明是它的原子数，而不是它的颜色、重量。与此类似，黑死病的决定性特性，是它的细菌源，而不是与之有关联的寒战、发烧或者恶心，也不是肯定与之没有联系的罪业（尽管现实中存在的一种信念认为，瘟疫是天谴的一种形式）。这里值得强调的是，这样的忽视并不妨碍人们掌握**金子**或者**瘟疫**的概念。如果它妨碍了，那么人们就不可能真的对瘟疫的原因产生分歧；他们总会各执一词地结束讨论。

对经典理论产生不利影响的这些哲学考虑令人印象深刻。但是，它的烦恼并没有到此结束。经典理论也面对着大量基于这些哲学考虑而产生的令人畏缩的问题。

或许这些问题中最突出的是，在明显被设计用以对概念的心理复杂性进行测试的实验情境中，定义未能展现出优势（可参见金切［Kintsch］，1974；J. D. 福多等，1975；J. A. 福多等，1980）。例如，如果**使确信**被分析为**引起……相信**（根据标准的经典分析），那么，一个人就会预期，与**相信**相比，**使确信**将被扣上一个更大的处理负担；毕竟，**使确信**应该是将**相信**作为其的一个成分。然而，这种效果从未在实验室中呈现出来。不但在过程研究中的定义未能显现出来，而且在词汇的获得中也不存在它们的迹象（凯里，1982）。当然，这些实验可能不够精细，或者关于为什么定义未能具有可发觉的心理效果可能还有别的解释，这总是有可能的。但是，它肯定无助于经典理论的问题，即定义不在实验中呈现出来。

然而，反对经典理论的最有力的心理学论证是基于所谓的**典型性效应**。典型性效应是许多与下述事实有关的心理现象：人们乐意根据子范畴相对于一个给定的范畴有多么典型或者多么具有代表性来评定这个子范畴。例如，有人倾向于说，知更鸟与鸡相比，是**鸟**这个范畴的更好范例；也就是说，他们认为知更鸟是更“典型”的**鸟**。这个结果本身可能并不很令人感兴趣。使得典型性判断至关重要的是下述事实：典型性判断追踪了其他许多有意思的心理变元（若希望复核，可参见罗施［Rosch］，1978；

史密斯和麦丁［Medin］，1981；为更严格的复核，参见巴萨卢［Barsalou］，1987）。

埃莉诺·博仕和卡罗恩·梅尔维斯（Eleanor Bosch and Carolyn Mervis，1975）发现，当主体被要求列出与一给定范畴及其从属物联系在一起的诸属性时，此列中属性的分布是为独立的典型性等级所预示的。一个从属物被评价得越是典型，它与这个相同范畴中的其他范例共有的属性就越多。例如，知更鸟被认为具有许多与其他鸟相同的属性，因此知更鸟相应地被判断为高典型性的鸟；相反，鸡被认为较少具有与其他鸟共同的
195 属性，因此鸡被判断为低典型性的鸟。另一个发现是，典型性在范畴化中有直接反映。在主体们被要求判断一个 X 是否是一个 Y 的诸案例中，典型性的独立量度预示了正确的肯定回答的速度。与正确回应“一只鸡是一只鸟吗”相比，他们会更快地正确回应“一只知更鸟是一只鸟吗”，错误率也由典型性所预示。相对于目标范畴（Y）的取样（X）越典型，错误就越少。典型性也和词类获得以及许多别的现象有关，例如，在人们为一个给定范畴提供范例所用到的序列中，更典型的事项会更早被说到。总之，典型性效应似乎渗透到了概念生活的方方面面。难怪心理学家们已经要求一种关于概念的理论对这些数据作出公正的评价。

正是在这一背景下，多数心理学家已经放弃了经典理论。问题就在于，经典理论与这些现象中的任何一个都根本毫不相干。我们上面所勾画的这个范畴化和概念获得的经典模式，无法预示这些效应中的任何一个，而且对它们进行折中的企图很快就遭遇了更多的问题。此外，正如我们在下一节将看到的，存在可替代的诸概念理论为整个系列的典型性效应提供自然并且充分解释的说明。

经典理论面临着一连串有力的哲学和心理学质疑。定义很难获得，它不具有任何心理效应，不能解释已知关于概念的、最重要的心理事实中的任何一个，而且公然违抗蕴因关于分析/综合区分的批评，没有能力去解释一个概念的所指是如何被确定的。结果，这种想法遂成潮流：尽管经典理论有相当大的吸引力，但它不值得拯救。

8.2 盖然的结构

20 世纪 70 年代见证了一个关于概念的新理论之发展，作为对经典理

论的一种替代，该理论获得了广泛的支持。这个新的理论，即**原型理论**，抛弃了原先的观念：一个概念的内在结构提供了关于该概念的一个定义。[11]相反，原型理论采用了对概念结构的一个盖然处理。根据原型理论，多数词汇概念都是复杂的心理表征，这些心理表征的结构所编码的，并非是对诸充要条件下定义，而是这些概念的外延中的诸事项**倾向于具有**的诸条件。所以，与经典理论不同，某一对象若希望处在一概念的外延之中，它便不必满足编码在该概念的结构中的每一个属性，而只要满足充分数量的属性。

下文即将提到，原型理论的优势之一在于，它不要求概念具有定义。人们阐述定义曾有很多的困难，这对于原型理论不是问题。根据原型理论，概念基本上没有定义结构，而只具有原型结构。基于这个理由，定义从不曾出现在心理过程的研究当中，也就不奇怪了。事实上，正是在我们转向经验心理学的数据时，原型理论变得特别诱人。该理论一般被理解为把范畴化看作是一个特性匹配的过程，在这里，一个范例或者个体在相似性上与一目标范畴作比较。只要有足够的特性匹配，它们就被视作是类似的，并且人们也由此得以判断这个事项归入这个范畴。对类似性的依赖为对类型性现象的解释提供了资源（参见史密斯，1995），在解释前一个人只需要设定，类型性判断也是由非常相同的过程形成的。换句话说，知更鸟被判断成是比鸡更典型的鸟的原因就在于，**知更鸟**与**鸟**共有更多的特性，知更鸟在类似性对比过程中评级更高。

还要考虑一下罗施和梅尔维斯的发现：类型性判断追踪一个概念与一个上位范畴的其他范例的共有特性的总数。此外，原型理论对于为什么这一现象会发生有一个自然的解释。其原因就在于，这些主体所列出的从属范畴之间共同的属性符合这个上位概念的某些特性。也就是说，它们刻画了这个上位概念的结构。结果，与其平级从属物（fellow subordinates）共有许多特性的诸概念，会自动地与这个上位共有许多特性。回到鸟这个概念的事例，其要点在于，被诸如**知更鸟**、**麻雀**、**鸵鸟**、**鹰**等这些交叉范畴共同提到的这些属性，是被**鸟**这个结构所编码的这些真正的属性。因为**知更鸟**具有许多相同的结构因素，而**鸡**具有的则较少，所以知更鸟被判断为比鸡更典型的鸟。

简而言之，原型理论具有巨大的心理学优势。难怪心理学界把这个理论奉为对经典理论的替代选择。但是，原型理论并非没有它的问题，至少我们希望一种概念理论能对某些问题作出令人满意的透彻理解。为简明起

见，我们只提到三点。

第一个问题是，原型理论就像经典理论一样存在忽视问题和错误问题。这个问题仍然是，人们能够掌握一个概念，然而却具有关于该概念的外延中那些事项的错误信息，或者缺乏足够数量的正确信息以准确地将这些事项识别出来。此外，类型在处理指称限定问题时糟糕至极。以**外祖母**这个概念为例，原型的外祖母们是有灰白头发的女人们，她们有皱巴巴的皮肤、戴着眼镜等等。然而，我们都知道，有些人没有呈现出这些特点，但却是外祖母；而有人确实呈现了这些特点，却不是外祖母。《道菲尔太太》（*Mrs. Doubtfir*）中的罗宾·威廉斯（Robin Williams）的角色可能看起来像一个外祖母，但是蒂娜·特纳（Tina Turner）才是一个真正的外祖母。①

第二个问题是，很多概念完全缺乏原型。就某些合成概念来说，这一问题尤为明显。正如福多指出的："可能会有原型的**外祖母们**玛丽·沃什（Mary Worth）而且可能会有原型的**外祖母们的诸属性**（**好的、老的**玛丽·沃什）。但是肯定没有**乔叟的外祖母们**的诸原型属性，而且没有**其大多数外孙都与牙医婚配的外祖母们**的诸原型属性。"（1981：297；亦见福多，1988）

第三个问题是，原型似乎并不按照一个构成语义学（compositional semantics）的原理来创作（见福多，1988；福多和莱泼尔，1996）。这个困难在于，根据对概念系统何以拥有生产能力（即对我们何以有能力容纳无限数量的概念）的标准说明，概念必定具有一个构成语义学。福多用**宠物鱼**这个概念来阐明这个论证。这个**宠物**原型编码了与狗和猫联系在一起的诸属性，而**鱼**原型编码了与鲑鱼这样的动物联系在一起的诸属性，但是，**宠物鱼**原型却编码了与金鱼和其他带颜色的小鱼联系在一起的诸属性。所以，很难看出符合**宠物鱼**的这个原型何以能够从符合**宠物**和**鱼**的原型中被形塑出来。

这三个批评一起对原型理论构成了严峻的威胁。然而，原型论者仍有回旋的余地。所有这三个反对意见预设的都是原型论者必定强调某一概念

① 《道菲尔太太》，中文译名常为《窈窕奶爸》，是罗宾·威廉斯出演的一部电影。片中罗宾·威廉斯饰演的角色在与妻子离婚后，思子心切，遂男扮女装，化身为一名65岁的老太太并应聘为自己孩子的奶妈。蒂娜·特纳，1939年生，美国摇滚明星，被誉为"摇滚教母"、"永不衰老的性感天使"。——译者注

的结构被其原型**尽述**了。但是，原型论者完全能够抛开这个限制，他们能够坚持主张，一个概念的原型是其结构的决定性部分。不过，相比于其原型，一个概念可以有更多的东西。

事实上，为了处理我们这三个批评中的头一个，即原型不适于限定指称的问题，许多原型论者都已经提出了恰好遵循这些线索的诸理论。根据**双重理论**（Dual Theory），一个概念有两个类型的结构，一类构成这个概念的“核心”，第二类构成这个概念的“识别程序”（奥舍森［Osherson］和史密斯［Smith］，1981；史密斯等，1984；兰多［Landau］，1982）。原型应局限于识别程序，它们负责说明快速的范畴化过程和所有类型性效应。另一方面，核心应具有某个其他类型的结构，以说明指称限定，并对我们最看重的范畴化判断负责——这里阙如的观点是，核心呈现了经典的结构。[12]

双重理论通过它对概念核心的承诺来处理第一个反对意见。如果原型不能限定指称，这很好，因为根据假设，核心承担了这个作用。它处理第二个反对意见是通过补充以下内容：某些概念缺乏原型，但是这并不妨碍任何人掌握这些概念；他们只需要把握这些概念的核心。最后，它处理第三个反对意见是通过坚持主张这一命题：只要诸概念核心按照构成语义学来组合，概念系统的生产力就会被建立起来，并且诸如**宠物鱼**的事例也不会对这个可能性产生不利影响。

虽然这些回应都是不无价值的，但是要注意，它们能起作用，主要是 *198*
通过将原型结构隔离于许多理论作用——而概念结构一开始就是为这些理论作用而引入的。结果，双重理论对与概念核心联系在一起的概念结构产生了很大影响。就另一结构应该是**经典**结构而言，双重理论继承了经典理论存在的大多数问题。例如，双重理论面临忽视问题和错误问题，它不得不克服蕴因对分析/综合区分的异议，它不得不应对少有真的定义的实例这一困难，等等。简言之，双重理论可能稍微拓展了逻辑空间，但如果没有对概念核心的充分说明，它并不比经典理论或者原型理论好多少。

8.3 理论结构

尽管有这些困难，双重理论却继续享有广泛的支持。我们猜想这是因

为人们有一种感觉，即心理学已经找到一种方法来切断双重理论与经典理论的残余联系。关键是，概念核心应该根据**理论论**来理解（例如，参见基林［Keil］，1994）。诸概念被套叠于在某些重要方面类似科学理论的诸种心理结构当中，而且它们牵涉到满足这个描述性内容——由诸概念在它们各自的心理理论中具有的这些作用所提供的描述性内容——的这些事件（例如，见凯利［Carey］，1985；墨菲［Murphy］和梅迪［Medin］，1985；高普尼克［Gopnik］和梅尔佐夫［Meltzoff］，1977）。[13]一个心理结构要成为类似的理论，就必须体现出一种解释图式，也就是一个思考者在试图弄清楚某事件在一个范畴化过程中的意思时所采用的一系列原则或者规则。这些理论的实例包括，所谓的常识心理学、常识物理学和常识生物学——普通人在解释心理事件、物理事件和生物事件时所凭借的一系列原则。[14]

理论论的主要优势之一是它所支持的范畴化模型。很多心理学家就已经对先前关于概念的诸理论表达了不满，原因在于它们未能并入人们对实在思考的倾向——梅迪和奥托尼（Ortony）称之为**心理本质主义**的一种观点。根据心理实在论，人们在对某些种类的范畴成员进行资格界定时，倾向于和一个实例之展示出某些可观察的属性相比，该成员更主要依赖于这个事项所具有的内在结构或者某个别的“隐含”属性（或许包括关系属性和历史属性）。理论论很容易涵盖心理实在论，因为理论论让人们在作出某些范畴决策时诉诸一个在心理上被呈现的理论。人们会疑问这个事项是否具有恰当的隐含属性，而非匆匆地浏览已检查过的一系列属性。这并不是说，理论论要求人们对遗传学和化学具有详细的理解。人们甚至不必具有关于这个属性特有本质的成熟观点。正如梅迪和奥托尼指出的，人们可能不过是具有“一个本质占位符”（1989：184）。这表明，不同的人在考虑某一物种时，会把不同的信息描述为具有一个唯一的本质。在某些情况下，他们可能具有关于这个本质的细致看法，但在多数时候，他们会具有一个粗略的看法，例如，对基因构造是什么的理解——即便他们不会描述特殊的基因属性，而且对遗传学所知甚少。

理论论最适合解释我们看重的范畴化活动。在这些情况下，重要的不是一个对象的全部知觉属性，而是被认为对其本质来说必不可少的那些属性。与此同时，理论论很不适合去解释我们更迅速的范畴化判断——在这里，概念配置经受着时间和资源上的双重压力。而且一般而言，理论论很少与类型性效应有联系；就像经典理论一样，它对为什么一些范例比另一

些范例更典型，以及为什么类型性与差异如此巨大的变元相关联完全无能为力。另一方面，如果理论论与原型理论相结合，作为结果的这个双重理论版本似乎就会有相当大的前途。与理论结构相随的核心似乎是对与经典结构相随的核心的一大改进。

不幸的是，这个修订过的双重理论仍然面临许多严峻的困难。我们会提到特别与理论论之作为概念核心的一个说明联系在一起的两个困难。第一个问题已经被提过了，所以用不着大惊小怪（指称确定的问题）；第二个问题则是新问题（稳定性问题）。

指称限定的问题从几个方面影响理论论。首先，我们已经看到，理论论者特色鲜明地允许人们能够掌握颇为粗略的理论，在这里，符合一个概念的本质空位符包含的信息相对较少。然而，要注意，倘若这是真的，那么概念最可能编码不充分的信息以识别一个正确并且确定的外延。如果人们不把除关于基因天赋的一些空洞观念之外的符合猫和狗的某种本质描述出来的话，那么，**猫**和**狗**这些概念就将套叠在搜寻相同之处的诸理论当中。靠这些缺乏活力的理论不可能把猫概念和狗概念拆分开来。

另一方面，人们可能已经发展了足够多的理论来对许多概念进行区别，然而，这却伴随着可能将不正确的信息并入他们的理论的危险。回到我们先前的事例，有人可能会主张，瘟疫是由天谴引起的，或者疾病本身牵涉到了恶灵附体。但是，相信这些事件的某个人应该仍然有能力接受和我们完全相同的概念——**瘟疫**。他们具有完全相同的概念以便搞清楚这个观念——我们关于疾病的本质和原因与他们理解的不一致，这是必不可少的。忽视和错误对于理论论和经典理论同样都是成问题的。

两个人是否正在利用同样的概念？[15]这是一个困难问题。我们认为，很多论者都会坚持主张，尽管周围的信念中存在一个不同，但**完全相同的概念**仍可能出现，不过这种想法是完全不恰当的。我对他们提出可供选择的建议是，人们只需要具有**类似的诸概念**。关键在于，信念的不同确实产生了不同的概念，这是不成问题的，因为两个概念在（它们被相同的心理归纳所包摄的）内容上仍然是足够类似的——而且，或许这是真正要紧的东西。

虽然这种处置方式可能很吸引人，但它实际上充满了困难。问题在于，当内容类似性的观念得到阐明时，它一般预设了一个内容同一性的先天性观念（福多和莱伯雷，1992）。例如，度量内容类似性的一个公共策略，是依据两个概念共有成分的总数。如果它们在很多成分中都交叠，那

么，它们就可以被说成是具有类似内容（例如，见史密斯等，1984）。但是，要注意到，这个建议只有在假定这些共有的、重叠的成分是一样的东西时才会起作用。所以，内容类似性的观念是被不正当地建立在它应当要取代的那个观念之上的。

既然这个问题的域还没有被并入哲学和心理学的范畴当中，那么就值得我们去探索别的一些业已提及的解决方案。例如，考虑一下埃里克·罗曼德（Eric Lormand，1996）的一个建议。罗曼德主张，甚至关于内容的一个完全带有整体论性质的理论，也不一定具有稳定性上的困难；换句话说，甚至对于一个宣称整个信念系统中的任何改变都会改变每一个单独信念的内容的理论，稳定性也不应成为一个问题。罗曼德宣称，建立稳定性的诀窍就在于一个给定符号具有**多义性**这个观念。它的每一个意义都是根据它的因果/推理链条中的一个子集而被赋予的。罗曼德称这些子集为**单元**，并要我们把一个单元看作是“对那个表征可接受用途的一个独立测试”（1996：57）。也即是说，他认为带有整体论性质的表征系统能够容许内容的稳定性，因为，在这个系统呈现出变化时，一个概念的某些意义会发生改变，某些则不会发生改变。就它使其单元的某些意义保持原样而言，它保持了这些意义。

不幸的是，这个建议并不起作用。因为罗曼德的单元就是表征本身，它们是这个系统中带有整体论性质且决定着每一个概念的内容的网络之一部分。结果，当推理网络中其他那些意义改变时，套叠在任何单元中的每一个概念都会改变其意义。而且，如果它们改变了自身的意义，那么它们就不可能成为别的概念稳定性的基础了（马戈利斯和劳伦斯［Margolis and Laurence］，1988）。

保罗·丘奇兰德（1988）曾提出过一个不同的解决方案。他曾一度创立了一种研究心理内容的方法，并称之为**状态空间语义学**。状态空间语义学是（就神经网络而言的）关于内容的一个理论，在神经网络中，内容应
201 是具有整体论性质的。一个激活向量的内容——即在这样一个网络中，横跨一个结点集合的一个激活的样式——应该是由它在这个更大的网络结构中所处的位置决定的。因为这个位置将会随着网络中众多其他结点的位置的转移而转移，所以状态空间语义学在实现内容稳定性时，会具有相当大的困难。结果，丘奇兰德很快就拒斥了内容同一性，转而支持内容类似性。

在较早的著作中，丘奇兰德采用了与史密斯等人（1984）非常类似的

一个模型，即想象一个联结主义者的网络带有一系列的输入结点、输出结点和中间状态的一组所谓隐藏结点（hidden nodes），他们把这些隐藏结点看作是指定了有内容的维度，那么，有多少隐藏的结点，我们就能够构造出多少维的一个语义空间。在这里，空间中的诸点对应着横跨这些隐藏结点的激活的样式。简便起见，假定只有三个隐藏结点，因而产生的这个语义空间将会是一个立方体，这个立方体的每一根轴对应一个特定的隐藏结点及其激活的水平。在丘奇兰德早期的处理中，内容类似性被理解为在一个类别的空间中的相对紧密度（closeness）。但是，这种方法陷入了和史密斯等人的说明相同的问题。它解释内容的类似性仅仅是通过预设了一个先天概念的同一性观念（即适用于这个空间且构成维度的观念）而实现的。

鉴于这个困难，丘奇兰德随后提出了对内容类似性的一个新说明。在这个新的模型中，丘奇兰德提出：

> 在激活空间中的一个点，需要一个专门的语义内容，该语义内容不是作为这个点相对于这个空间的构成轴的位置的应变物，而是作为（1）这个点相对于**这个空间中**所有**其他有内容的点**的空间位置的应变物，以及（2）这个点与**这个外部环境**稳定并且客观的**宏观特性**的因果关系的应变物。（1988：8）

丘奇兰德告诉我们，这个新的立场“构成了对福多和莱泼尔之挑战的一个决定性的回应”（1988：5），同时提供了对概念类似性的一个行得通的、带有整体论性质的说明。

但事实上，丘奇兰德的新说明还远不是对之前挑战的一个决定性回应，它实际上并未作出任何改进。他对内容的第一个限定——相对于这个空间中其他有内容的点的空间位置——立即碰到了一个严重的困难。假定两个网络中的某些结点有完全相同的相对位置，那么单凭这一点还不足以确定它们的内容；因为人们完全有理由质疑，为什么随便哪个网络中的任何特定的结点具有它所具有的这个特定内容（而且不是别的某个内容）。例如，丘奇兰德把一个类型的网络描述成是表征了截然不同的家族，就好像是输入的照片，它一定会抽取四个原代表性的外观。但是，是什么因素使得这些网络结点把家族和外观描述成对立于种类广泛的潜在对象中的任何一个？为了回应这个问题，丘奇兰德只能求助于他对内容的第二个限定中的资源——与环境特性的因果关系。然而，这个回应的问题就在于，这不再是理论论的一个版本了。相反，它利用了下一节我们要讨论的关于内 202

容的原子理论。这个结点与其周围诸结点的关系被证明与其内容毫无瓜葛；对内容起作用的只是一个与这个环境的诸特性可靠的因果连接的存在。[16]当然，这些可靠的连接提供了稳定性，但这仅是因为它们认购了一个内容同一性理论：只有两个结点被连接于这个相同的环境特性时，它们才具有相同的内容。丘奇兰德能够具有类似内容的概念并没什么大惊小怪，因为他擅自取用对内容的一致性的一个独立说明，尽管他的措辞恰恰相反。[17]

稳定性被证明是对概念理论的一个粗鲁的约束。对理论论而言，这意味着诸心理理论存在运作不良的核心。它们具有和原型理论涉及到指称时同样多的麻烦，而且它们在保障稳定性时尤为糟糕。如果概念的双重理论中的一个版本获得成功了，那么，其核心看起来将不会具有经典结构或者理论结构的版本。

8.4 无结构概念

我们已经看到，关于概念结构的主要观点都是成问题的。考虑到这些困难，许多论者一直倡导探索词汇的概念不具有任何结构的可能性，这是一种被称为**概念原子论**的观点（例如，见福多，1988；莱斯利，2000；米利肯，1988、2000）。概念原子论的中心论点是，一个概念的内容不是由它与任何别的特定概念的关系所决定的；相反，它是由心灵—世界的一种关系，即符号与它所表征的东西之间的一个因果或者历史关系决定的。毫不奇怪，原子论者在克里普克和普特南对自然种类术语的处理中找到了灵感，只是自然种类术语是用来处理范围更广的语义事项，并且首先对概念系统而非语言负责。

对一个原子论者来说，最困难的任务是要提供一种对理应决定了概念内容的心灵—世界关系的充分详细的说明。一般策略是根据协变的观念来解释内容（我们在丘奇兰德对稳定性的处理中的不正当操作那里见过这个概念）。要点是，一个概念表征它在因果上与之协变的东西。例如，如果概念 D 作为狗的出现的一个可靠的因果后果而被预测，那么，按照当前的说明，这个符号将会表达**狗**这一属性，并成为**狗**这一概念。然而，要注意，这个简单的说明并不得体。原因就在于，各种各样的其他事件完全有可能导致 D 这个符号的诸预测。例如，这可能作为知觉错误的一种结果发

生。在一个黑暗的夜晚，你可能撞上了一只狐狸，并把它误认为是一条追赶上你汽车的狗。

原子论有许多用以将这些非狗案例排除出去的资源。一是增加更深一层的条件，即一个概念在**理想条件下**会表征它将与之协变的东西（容许在条件不理想时，非狗引起狗［DOG］这个可能性；参见斯坦普，1977）。另一种方法是声称，一个概念会表征它具有与之协变的**功能**的东西（容许这个概念或者产生此概念的系统在非狗的情况下不正确地发挥功能的可能性；例如，德雷斯基，1995；米利肯，1984、1993）。虽然如此，还有可能说：**狗/DOG** 依赖在某种意义上比非狗却像**狗/DOG** 依赖更基本。例如，前者的依赖可能会控制后者是否依赖，但是反过来则不行（福多，1990）。

虽然这些策略中的每一个都有其各自的困难，但是我们想要聚焦于关于原子论的更一般的诸问题，即不与任何特殊的原子理论细节绑定在一起的问题。我们会提到三个异议。

第一个异议牵涉到概念的解释作用。多数理论都把一个概念的解释潜力与其结构联系在一起。这在我们已经回顾过的其他理论中是很明显的。例如，原型理论通过概念结构来解释各种各样的心理现象——范畴化、类型性判断、使用功效等。然而，与概念原子论有关的问题是，它宣称概念没有结构，这样它们似乎就不可能真正地解释任何东西。那么，它们的好处是什么呢？

第二个异议表现为如下忧虑：即概念原子论承诺了一个令人难以置信的天赋程度。事实上，杰里·福多这位原子论最出众的辩护者，通过辩护这一主张——实质上所有词汇概念都是天赋的，包括像**汽化器**和**夸克**这样不太可能的候选项——已经明确地给出了这个联系。正如福多所说的，一个概念能够被习得的唯一方式就是通过一个构造的过程，在这里该概念从其成分开始被组合在一起。因为原子论坚持主张词汇概念没有结构，所以它们必定全都是天赋的（福多，1981）。但是，如果**汽化器**是天赋的，那么明显就是有什么物质出错了；或许原子论本身就是这一物质。

第三个异议是原子理论对概念的个体化处理太过粗略。因为它们把内容还原成一个表征与它所表征的东西之间的一个因果关系或者历史关系，所以，诸概念相比于它们所识别的这些事项似乎并没有被更细致地个体化。这确实并不够好。水这个概念与 H_2O 的概念并不是相同的东西——有些人可能具有一个而不具有另一个——但是，它们识别了完全相同的属性。或者，举一个更极端的例子，**独角兽**这个概念与**人首马身怪**这个概念

不是相同的东西，然而因为它们都是空概念，所以人们似乎会将它们识别为完全相同的东西，即空无。所以，很难明白一个原子论概念如何能够将这些概念梳理开。

让我们按颠倒的顺序来考量这些异议。毫无疑问，实现一种条理细致的个体化对原子论是真正有利害关系的，但是原子论者确实具有一些他们能够支配的资源。例如，关于空概念，他们能够坚持主张，决定着协变关系的这个概念是诸属性之间的一个原子关系。这个主张之所以会有帮助，是因为诸属性之间能够有原子关系从表面（即便这些原子关系未被例示）看起来是真实的。关于别的事例，原子论者能够区分共指称的（co-referential）概念，其方法是强调这些概念之一的确是复杂的，而且其复杂性不容争辩。大概这就是他们如何处理**水**/H_2O这个案例——通过坚持主张除了别的东西，H_2O这个概念并入了**氢**的概念（福多，1990）。当然，还有别的一些有挑战的案例，但这些策略都不奏效。在这里，我们想到了在法则论上表达了共外延属性的几对原始概念（例如，**买/卖、追/逃、扩展/定型**），这些被证明是最困难的案例，因为对它们进行区分的自然解决方案宣称，它们与不同的决定内容的推论联系在一起。很难说原子论者是否有可选择的解决方案。

但是，让我们转向对原子论的另一个异议，面对该异议，原子论者所能辩解的回旋余地更小。如果原子论宣称，词汇概念没有结构，那么词汇概念就一定都是先天的吗？而且，如果词汇概念没有结构，那么为什么它们在解释上并不是无用的呢？

福多对彻底的概念先天论的论证已经在心灵哲学中掀起了相当的波澜，结果不同学派的论者都不再讨论任何被认为与这个论点关系密切的原理了。[18]结果，这个论证并没有获得它应得的那种仔细的、有判断力的审视。我们相信，原子论已经被不公正地承载了福多的强先天论，而事实上，我们完全有可能令人满意地说明新的原始概念何以能够以一种与概念原子论相协调的方式而被获得。这里关键是**持续机制**的概念。持续机制是认购了决定一个概念之内容的心灵—世界关系的机制。持续机制是一种类或者另一种类的典型推理机制，因为人们明显缺乏他们能够表征的大多数有关这些属性的转换器。然而，重要的是，这些推理机制不必产生任何分析论（analyticities）或者一个概念所具有的任何语义结构，因为人们掌握概念不需要特定的推理。因此，这样的推理机制是与概念原子论完全一致的。

现在我们能够明白为什么原子论不承诺激进的概念先天论了。原子论者应该承认：有关如何获得一个概念，应该根据——鉴于正确的概念理论，某人如何得以处在一种满足该理论的心灵状态当中——这个更精练的问题来被构造（马戈利斯，1998；劳伦斯和马戈利斯，2002）。根据关于内容的原子论论述，这应根据我们所拥有的适当的持续机制来理解。所以，有关获得的问题仅仅是持续机制如何被组合的问题。而且，在这里，原子论者能说的东西是很多的，它们对于概念没有任何结构这一主张有一致意见。这里有很多东西都与概念没有结构这一主张相一致。例如，我们已经详细研究过的一种持续机制就支持对自然种类概念的占有（马戈利斯，1998；劳伦斯和马戈利斯，即将出版）。这个模型依据的是我们所谓的**基于特征群的持续机制**，这个机制并入了关于一个类型的高度指示性的知觉信息，以及把某物看作是这个相同类型之一员的倾向，因为它共有了与这个范畴的典型实例相同的构成性的隐藏属性（而且不必分享与之相同的知觉属性）。这个暗示结果是，人们有根据他们的经验来组合基于特征群的持续机制的一般倾向。那么，这样一个机制建立了原子论者们所说的构成内容的心灵—世界的关系，并且有能力递送一系列非构造概念的更大环境输入。因为这个机制尊重一个人的经验的特征——获得物继续下去，通过信息的收集、储存和处理，以产生追踪着这个概念的外延中的诸事件的一个表征——所以，我们认为，称这个机制是一种学习模型的说法是公正的。

最后转向原子论使得概念在解释上毫无作用的这一指责。对原子论者来说，最好的策略是宣称，常常通过一个概念的结构而被说明的解释作用，实际上并不一定要直接根据这个概念的本质来解释。要点是原子论者应能够求助于碰巧与该概念联系在一起的信息，也就是说，即便其他这些概念不构成 C，原子论者也能够利用一个概念 C 所具有的与其他概念的关系。这可能看起来是激进的一步，但实质上，为了解释某些概念参与了其中的推理的事实，任何概念理论都会这样做。或许你在年幼时被狗吓到过，并且这一事件使你相信狗是危险的，那么，这个信念或许会很好地解释你对狗的很多行为。然而，一个经典理论的信奉者不大可能认为——狗是危险的——是狗这个定义的一部分。所有的概念理论都宣称，一个概念与其他概念的某些联系是由这个概念的同一性构成的，有些则不是。在作出这种区分后，有时候会出现这样的情况：一个概念如何被配置（deployed）将会反映它的非构成性关系。原子论者常常完全把这个立场推至

极限，并且宣称事实总是如此。思维中一个概念的作用常常会反映其非构成性关系，因为构成一个概念的，不是它与任何其他特殊概念的关系，而是它如何在因果上（或者历史上）与世界上的其他事件相关联。然而，一个人会奇怪，原子论者是否已经离题太远了。一个概念被配置的方法中**没有任何**一个为其本质所解释，真的会是这样吗？

8.5 重新思考概念结构

下述主张有一些令人不安的东西：概念的解释功能是由其伴随发生的关系决定的。再考虑一下类型性效应，类型性效应在其心理输入中如此普遍并且如此丰富，以至于它们构成了所有概念理论的核心待解释项之一。确实，心理学家们陆陆续续地抛弃经典理论，正是因为经典理论未能说明这些效应。然而，要注意，概念原子论在其处理类型效应的能力方面与经典理论并无不同。通过坚持主张概念没有结构，原子论者们承诺了如下这个见解：一个概念的本质无论如何对其在类型性效应中的作用没有影响。当然，这并不意味着必须否认类型性效应的存在。令人费解的是，牵涉到概念的一些最重要的心理学数据到头来与概念的本质完全无关。

与此同时，有强大的压力在弱化人们对概念没有任何结构的这种原子论中心论断的支持。尤其是根据概念结构解释指称限定的所有企图都陷入了难以逃脱的困境。经典理论、原型理论和理论论都沦为忽视问题和错误问题的牺牲品，而且每个理论还都有其自身独有的困难。

摆脱这个绝境的出路在于两个相关联的洞察（insights）。第一个洞察是，概念能够具有多结构。因此，在最初的双重理论中，概念被认为是核心和识别程序。第二个洞察没那么明显，但事实上却具有决定性的意义。那就是，概念能够具有在范畴上对非常不同的解释功能负责的不同类型的结构。[19]出于不同的动机，双重理论含蓄地承认它与核心和识别程序相关联这一可能性。但是，一旦这个观点被挑明，并且一旦它用完全通俗的话语表达出来，全新的一系列理论可能性就浮现出来了。

最直接的影响是，双重理论承认解释指称的功能可能不得不从概念的某些其他功能中梳理出来。这会将一个概念所具有的其他结构类型从一个沉重的负担中解放出来，并且至关重要的是，这暗示着并非所有的概念结构都是限定指称的结构。迈出了这一步，一个人就能够查问存在哪些别的

概念结构类型，以及它们所负责的特定功能。

我们认为，至少存在四种核心的结构类型：

构成的指称限定结构 这个结构凭借一个构成语义学促成了一个概念的内容和指称。这个类型的结构常见于经典理论。是否任何词汇概念都具有这个类型的结构将取决于某些分析性问题、忽视问题和错误问题是否能被找到，以及定义是否能够真正被发现。然而，诸如**棕色狗**这样的短语概念具有这种类型的结构差不多是毋庸置疑的。**棕色狗**由**棕色的**和**狗**组成，并且其指称在构成上是由其成分的这些指称属性决定的：只有某物是棕色的并且是一只狗，它才会被归入**棕色狗**。

无语义的结构 这个结构不促成一个概念的内容，而是有意义地促成了概念的其他一些在理论上重要的解释功能。虽然双重理论关于这方面的意义并不是显而易见的，但是，把双重理论对原型的承诺看作是对无语义结构的一个承诺，似乎也是有道理的。

无指称的语义结构 这个结构促成了一个概念的内容，但是与指称的结果相脱离。虽然我们对概念的意义或者内容的讨论已经集中在它们指称的属性上，但是很可能并没有穷尽概念所具有的语义属性。除了别的东西之外，这个类型的结构还牵涉到所谓的窄内容。[20]

持续机制结构 这个结构通过把一个在理论上有意义的持续机制考虑进来，从而直接促成了一个概念的内容。持续机制结构决定了一个概念的指称属性，但凭借的并不是一个构成语义学。此外，这个类型的结构支持（直接地）决定着一个概念的内容的心灵—世界关系。

这四种不同类型的结构指向了一系列新的可供探索的理论选项。通过说明，我们将简明地勾勒出对下述僵局的解决办法，即概念原子论和在解释概念最突出的行为时需要诉诸概念结构这一压力之间的僵局。

如果我们回顾双重理论，那么我们将发现，它所面对的主要问题是以它对概念核心的处理为焦点。我们已经看到，在这个方面，定义结构和理论结构两者都同样是成问题的。两者都不特别适合指称限定。并且，在任

何情况下，定义都已经被证明是很难描述的，而理论结构有其稳定性的困难。不过，要注意的是，现在有了一个关于核心的可选择的说明。鉴于我们刚刚描述出的这四种类型的概念结构之间的不同，概念原子论最好不要根据词汇概念完全没有结构这一总括性断言来理解，而是要根据词汇概念没有**构成的指称限定的结构**这一主张来分析。这就创造了概念的核心可以是原子的可能性。

确实，原子似乎完全适合担任与概念核心联系在一起的解释角色。如果核心是原子的，那么一个人就不必担忧概念是不可定义的这一事实。原子论暗示着概念是不可定义的。与此类似，如果核心是原子的，那么一个人就不必担忧其稳定性。原子论暗示着，只要限定指称的心灵—世界关系保持不变，一个概念与其他概念的关系就能够随个人的心意而改变。原子核心也能够解释概念的生产力：复杂概念是经由原子核心的这个经典的构成性而产生的。与核心联系在一起的唯一解释作用（原子似乎与之有纠纷）就是说明我们最看重的关于范畴成员资格的判断。然而，对于一个概念核心理论来说，首先它应是一个合理的必需品，这一点并不清楚。如果
208 蒯因关于分析性的著作表明了什么东西，那就是人们最看重的这个类别的判断是带有整体论性质的，所以整体信息能够因每个被单独使用的概念而被分解开，这并不是似是而非的。卸下了这最后一个必需品之后，就有了一个很好的案例可用以确认核心应当是原子的。

与此同时，这个种类的模型消除了原子论在心理上是无法解释的异议。我们能够同意原子论者的意见，即词汇概念一般缺乏指称限定结构，但这并不意味着我们必须说概念完全是无法构建的。例如，原型和持续机制可以很好地成为一个概念结构的一部分，只不过这个结构并不直接限定其指称；指称是由暗示着核心的心灵—世界关系确定的，而原型（和别的类型的结构）则被留下来解释其他东西。而且，作为众多要素中的一个，原型确实解释了很多其他东西。鉴于其巨大的心理意义，如果有什么区别存在的话，原型应被视为部分地构成了概念。

概念是心理种类。正如我们明白的，最好的概念理论是重视概念的心理特征的理论。这样做的方法是，采用一个承认不同种类概念结构的理论，同时通过断言概念具有原子核心，从而把这些概念结构结合在一起。无论如何，概念结构本身的本质都是值得关注的。以概念结构为出发点并使这些结构分离开来，由此明确表达出不同的解释作用，这在概念研究中展开了一系列未经探索但却非常有潜力的理论选择。

【注释】

本文完全是协力完成，作者排名不分先后。

[1] 关于思维本质的观点并非完全无可置疑。然而，我们很难明白，要是没有到达一个被建构的表征系统的优越通道，有限的世界造物如何能够有能力容纳人们可资利用的丰富思维。即便我们按相对简单的思维来说，这些思维的总数也绝对是一个天文数字。例如，有10^{18}个简单的数字陈述与小于100万的数有关。这比宇宙开始至今的时间总数（以秒计）还多，而且比人类大脑中的神经元总数多100万倍。一个关于思维的理论何以可能适应这些事实，而无需假定一个被建构的表征系统，在这个表征系统中，这些相同的元素——概念——能够在一个被建构的集合内以不同的排列展现？无论如何，如果一个理论真的宣称思维没有成分，那么或许最好的补充说明就是，根据这个理论，不存在任何诸如概念这样的东西。

[2] 我们认定，思维和概念具有语义属性，而且这些属性中最重要的是它们的真理论属性。我们把它看作是对关于概念的一个理论——例如，**狗**这个概念意指狗——的一个重要约束。

[3] 仍然值得注意的是，我们讨论的这些理论能够略加修改以适应在这些基本问题上采取不同立场的可选择框架。

[4] 对此处观点更详细的考察和发展，见劳伦斯和马戈利斯（1999：前言）。亦见史密斯和梅迪（1981）。

[5] 对这个限定的主要理由在于，根据经典理论，有些概念原本即没有结构而它们是其他所有概念得以形成的原始概念。经典论者对一个原始概念的指称如何被确定语焉不详。但是，应归功于英国经验论者的这个最值得重视的说明是，原始概念表达了感觉属性，并且也很容易指涉这些属性，因为它们通过感官传感器在因果上与这些属性相连接。

[6] 关于概念理论的著作已经越来越变得跨学科了，而且我们将要讨论的这些理论，很多都带有跨学科（特别是哲学和心理学）调配的思想和动机的印记。与这些研究相符合，我们把概念看作是心理表征（并因此成为所谓的心理特称命题），因为这个看法把心理上各种各样的待解释项搞得最清楚，而这些待解释项——甚至在哲学圈内——确实已经为建立关于概念的理论施加了相当的压力。但读者应注意，这并不是为学界普遍接受的看法，而且很多哲学家坚持主张把概念分析成一种抽象实体。然而，将概念视作抽象的论者，也对关于概念结构的问题深感兴趣。只不过，被考虑的这个结构被认为是抽象实体的结构。例如，见皮科克（1992）和比拉（Bealer，1982）。

[7] 正如此处事例所指示的，经典理论（和甚至我们将要讨论的所有理论）首先是关于和自然语言中的语词——称为词汇概念——相对应的概念的本质的一个理论。这是因为，对概念感兴趣的论者认为，与自然语言短语或者句子相对应的这些表征是人为构建的。

[8] 经典理论的动机绝不限于这些优点。例如，有利于该理论的另一个有影响的

论点是，该理论有能力解释我们的直觉——某些陈述或者论证是正确的，即便从表面判断，但它们未能表达逻辑真值，例如，因为“约翰是一个单身汉”，所以“约翰是未婚的”（例如卡茨［Katz］，1972）。

［9］经典论者对分析性观念的辩护乏善可陈。例如，克里斯托弗·皮科克在论述概念的开创性著作（Christopher Peacocke，1992）中直接赞同了经典理论，尤其是其对定义的承诺，然而，皮科克对与分析性联系在一起问题却未加注意，只是在一处脚注中声明，他承诺将分析/综合区分的某个版本（见 244 页，注释 7）。对分析性观念的一个罕见的经典辩护，尤其针对当前的这些考虑，见卡茨（1997）。

［10］在最极端的情况下，人们几乎完全不知道任何信息。例如，普特南评论说，他不能把榆树和山毛榉区别开来，它们对他而言都只是树。然而，可以认为，他仍然具有分别指涉榆树和山毛榉的两个不同概念的能力。如果这个指称机制不得不是一个内在化的定义，那么这种现象就不会成为可能。

［11］我们称之为“原型理论”的东西，是对一组范围广泛的理论的一个理想化的说法，即一个从众多不同的细节中抽象来的东西。这符合我们介绍的这些理论中的任何一个，虽然就原型理论来说差异或许更明显。对于某些不同种类的讨论，见史密斯和梅迪（1981）。

［12］双重理论不应混同于哲学中所谓的**二因素论**。虽然它们有类似之处，但是双重理论和二因素论处理不同的问题。二因素论主要牵涉到内容有区别的两个不同类型或者方面。一个因素说明随附于一个人的或者由分子对分子的复制所共有的内容的所有方面（“窄内容”）；另一个因素说明超越前者、牵涉到这个人与其环境关系的内容的诸方面（“宽内容”）。结果，双重理论中这两个类型的结构对二因素论中这两个方面的内容进行了交叉分类。

［13］根据理论论，一个概念的结构是由它与牵连在一个套叠理论中的其他概念的关系构成的。要注意，在这个说明中，一个概念的结构不能够根据部分/整体的关系来理解。出于这个理由，我们已经区分了概念结构的两种模型（劳伦斯和马戈利斯，1999）。第一个是容量模型，即宣称一个概念 C1 被包含在另一个概念 C2 的结构当中，只有 C1 完全被包含在 C2 当中（即是 C2 的一个固有部分）。第二是推理模型，即宣称 C1 被包含在 C2 的结构当中，只有 C1 与 C2 处于一个特许的推理关系当中。正如在这个描述中应该明确的，理论论必须根据推理模型来构建，但是经典理论和原型理论依赖支持经典结构和原型结构之假定的这个确切动机，能够根据两者中的任何一个来构建。

［14］近年来这些特定领域已经成为强交叉学科研究的对象。关于常识心理学，见戴维斯和斯通（Davies and Stone，1995a、1995b），卡拉瑟斯（1996）；关于常识物理学，见斯派克（Spelke，1990），巴亚热昂（Baillargeon，1993），许和凯利（Xu and Carey，1996）；关于常识生物学，见梅迪和阿纯（Medin and Atran，1999）。

［15］或者就此而言，这同一个人过去是否在使用相同的概念？

[16] 丘奇兰德的模型最多表明心理**过程**何以可能具有整体论性质。因为它们牵涉到横跨一个网络中大量相关结点的激活模式，所以它们是具有整体论性质的。但这并不意味着这个网络中的**语义学**具有整体论性质。

[17] 应当注意，在这些问题上丘奇兰德多少有些游移不定，尽管他常常忽视了确认他观点上的变化。例如，除了文中提到的这些立场外，丘奇兰德还试图主张内容类似性是“顺流过程”的类型性的一个问题（特别见 1996：276）：

> 对于断定个体间或者文化间认知内容的相同性如此重要的，正是向量这个顺流方面的计算作用。一个人或者一种文化足够可信地把小猫们与环境区别开来，但是若在各个方面都把它们看作是码头老鼠的一个变元形式，那么这样一个人或者一种文化，必定会被归属到与我们的“小猫”概念大大不同的某个“小猫”的概念。另一方面。若一个外星人或者外星种族对小猫们的预期和行为与我们如出一辙，那么，这个外星人或外星种族就必定会被归属到这同一个“小猫”概念，即便他们可能主要借助嗅觉和高频声呐定向把小猫们与他们的前额区别开来。

除了使得他的“状态空间语义学”与这个状态空间没有任何关系之外，这个立场还受困于和丘奇兰德第一个立场完全一样的问题，即它为确定**小猫**向量的内容的这些“顺流”状态预先假定了一个内容同一性的观念。

[18] 例如，见丘奇兰德（1986）和普特南（1988）。

[19] 这两个论点是彼此相连的，因为可以预期，要是一个概念在范畴上具有可属于不同类型的多结构就好了。

[20] 窄内容的本质是有争议的，但是主要观点认为，窄内容是为分子对分子的复制所共有的，即便它们存在于不同的环境当中。在某些二因素论中（见注释 12），一个概念的窄内容是由其推理作用决定的——非常类似于在概念结构的理论论说明中的一个观点。不同之处在于，根据某种二因素论，一个概念的推理作用不应决定其指称。

参考文献

Baillargeon, R. (1993). “The Object Concept Revisited: New Directions in the Investigation of Infants' Physical Knowledge.” In C. Granrund (ed.), *Visual Perception and Cognition in Infancy*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bealer, G. (1982). *Quality and Concept*. Oxford: Clarendon Press.

Barsalou, L. (1987). “The Instability of Graded Structure: Implications for the Nature of Concepts.” In U. Neisser (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. New York: Cambridge University Press.

Carey, S. (1982). “Semantic Development: The State of the Art.” In E. Wanner

and L. Gleitman (eds.), *Language Acquisition: The State of the Art*. New York: Cambridge University Press.

—— (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

Carruthers, P. (ed.) (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1996). "Fodor and LePore: State-Space Semantics and Meaning Holism." In R. McCauley (ed.), *The Churchlands and Their Critics*. Cambridge, MA: Blackwell.

—— (1998). "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/LePore Challenge Answered." *Journal of Philosophy*, XCV (1): 5-32.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: The MIT Press.

Dancy, J. (1985). *Introduction to Contemporary Epistemology*. Cambridge, MA: Blackwell.

Davies, M. and Stone, T. (eds.) (1995a). *Folk Psychology*. Oxford: Blackwell.

—— (eds.) (1995b). *Mental Simulation*. Oxford: Blackwell.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Fodor, J. D., Fodor, J. A., and Garrett, M. (1975). "The Psychological Unreality of Semantic Representations." *Linguistic Inquiry*, 6: 515-532.

Fodor, J. A. (1981). "The Present Status of the Innateness Controversy." In *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

—— (1981/1990). "Psychosemantics; or, Where Do Truth Conditions Come From?" In N. G. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell.

—— (1990). "A Theory of Content, II: The Theory." In *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

—— (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

Fodor, J. A., Garrett, M., Walker, E., and Parkes, C. (1980). "Against Definitions." *Cognition*, 8: 263-367.

Fodor, J. A. and LePore, E. (1992). *Holism: A Shopper's Guide*. Cambridge, MA: Basil Blackwell.

—— (1996). "The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes." *Cognition*, 58: 253-270.

Gettier, E. (1963). "Is Justified True Belief Knowledge?" *Analysis*, 23: 121-123.

Gopnik, A. and Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cam-

bridge, MA: MIT Press.

Katz, J. (1972). *Semantic Theory*. New York: Harper and Row.

—— (1997). "Analyticity, Necessity, and the Epistemology of Semantics." *Philosophy and Phenomenological Research*, LVII: 1-28.

Keil, F. (1994). "Explanation, Association, and the Acquisition of Word Meaning." In L. Gleitman and B. Landau (eds.), *The Acquisition of the Lexicon*, Cambridge, MA: MIT Press.

Kintsch, W. (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Landau, B. (1982). "Will the Real Grandmother Please Stand Up? The Psychological Reality of Dual Meaning Representations." *Journal of Psycholinguistic Research*, 11 (1): 47-62.

Laurence, S. and Margolis, E. (1999). "Concepts and Cognitive Science." In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.

—— (2002). "Radical Concept Nativism." *Cognition*, 86 (1): 22-55.

—— (in preparation). *The Building Blocks of Thought*.

Leslie, A. (2002). "How to Acquire a 'Representational Theory of Mind'." In D. Sperber and S. Davis (eds.), *Metarepresentations. Vancouver Studies in Cognitive Science*, vol. 10. Oxford: Oxford University Press.

Lormand, E. (1996). "How to Be a Meaning Holist." *Journal of Philosophy*, XCIII: 51-73.

Margolis, E. (1998). "How to Acquire a Concept." *Mind and Language*, 13 (3): 347-369.

Margolis, E., and Laurence, S. (1998). "Multiple Meanings and the Stability of Content." *Journal of Philosophy*, XCV (5): 255-263.

Medin, D. and Atran, S. (1999). *Folkbiology*. Cambridge, Mass: MIT Press.

Medin, D. and Ortony, A. (1989). "Psychological Essentialism." In S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*. New York: Cambridge University Press.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

—— (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

—— (1998). "A Common Structure for Concepts of Individuals, Stuffs, and Real Kinds: More Mama, More Milk, and More Mouse." *Behavioral and Brain Sciences*, 21: 55-65.

—— (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. New York: Cambridge University Press.

Murphy, G. and Medin, D. (1985). "The Role of Theories in Conceptual Coherence." *Psychological Review*, 92 (3): 289-316.

Osherson, D. and Smith, E. (1981). "On the Adequacy of Prototype Theory as a Theory of Concepts." *Cognition*, 9: 35-58.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

—— (1998). "Implicit Conceptions, Understanding and Rationality." In E. Villanueva (ed.), *Philosophical Issues*, *9*: *Concepts*. Atascadero, CA: Ridgeview Publishing Company.

Putnam, H. (1962). "The Analytic and the Synthetic." In H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 3. Minneapolis: University of Minnesota Press.

—— (1970). "Is Semantics Possible?" In H. Kiefer and M. Munitz (eds.), *Languages, Belief and Metaphysics*. New York: State University of New York Press: 50-63.

—— (1975). "The Meaning of 'Meaning'." In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.

—— (1988). *Representation and Reality*. Cambridge, Mass.: MIT Press.

Quine, W. (1951/1980). "Two Dogmas of Empiricism." In *From a Logical Point of View: Nine Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press: 20-46.

Rey, G. (1993). "The Unavailability of What We Mean: A Reply to Quine, Fodor and LePore." In J. A. Fodor and E. LePore (eds.), *Holism: A Consumer Update*. Atlanta: Rodopi BV: 61-101.

Rosch, E. (1978). "Principles of Categorization." In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E. and Mervis, C. (1975). "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology*, 7: 573-605.

Smith, E. (1995). "Concepts and Categorization." In E. Smith and D. Osherson (eds.), *Thinking: An Invitation to Cognitive Science*, vol. 3. Second Edition. Cambridge, MA: MIT Press.

Smith, E. and Medin, D. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, E., Medin, D., and Rips, L. (1984). "A Psychological Approach to Concepts: Comments on Rey's 'Concepts and Stereotypes'." *Cognition*, 17: 265-274.

Spelke, E. (1990). "Principles of Object Perception." *Cognitive Science*, 14: 29-56.

Stampe, D. (1977). "Towards a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press.

Wittgenstein, L. (1953/1958). *Philosophical Investigations*, trans. E. Anscombe. 3rd edition. Oxford: Blackwell.

Xu, F. and Carey, S. (1996). "Infants' Metaphysics: The Case of Numerical Identity." *Cognitive Psychology*, 30: 111-153.

第9章　心理因果关系

约翰·海尔（John Heil）

9.1　笛卡尔背景

214 笛卡尔的研究奠定了现代心身关系讨论的基调。根据笛卡尔所说，心与身是两种不同的**实体**。（在这种语境下，实体不应当被看作是某种东西，比如说某种可能弄脏你衣服或刺穿你鞋子的东西，而应当被看作是一种特殊的对象或实在：你窗外的那棵树，一块卵石，你的右耳。）笛卡尔认为，身体是具有空间广延性的实体，但不能感觉和思考；心灵与之相反，是无广延但可以思考和感觉的实体。

通过思考心与物的特征，你或许也会形成这样的观点。从表面上看，它们似乎有很大的不同。心灵的各种状态所表现出来的特征，似乎是外在于物理领域的：当你的肘部被撞时，当你闻到泥炭的气味时，当你听到蚊子在你脑袋边嗡嗡叫时，你所具有的感觉，在性质上似乎不同于任何属于物理世界的东西。这些体验的**各种原因**无一例外的全都是物理事件。然而，这些原因的心理结果，即心理结果的各种**显现**（appearances），似乎包含了物理世界之外的一些性质。就物理的身体而言，它们所表现出来的特征，似乎绝对不是心理的。一块石头有独特的大小、形状、质量以及确定的空间位置。感觉和思想则与之相反，显然没有这些特征。疼痛可能是强烈的，却不可能有三英寸长；你关于即将来临的假期的预想，也不可能有质量。

诚然，我们说思考发生于我们大脑**之内**，脚趾的疼痛发生于脚趾**上**。这种说法表明，心灵的各种状态至少是有空间位置的。不过，说疼痛或思想有空间位置的意义，显然不同于说物理对象有空间位置的意义。笛卡尔
215 清楚地知道**幻肢痛感**（phantom pain）这种症状：它显然是发生在截肢上

的疼痛。这种症状表明：当你把疼痛描述为发生在你的脚趾上时，你真正描述的是一种特定类型的感觉：**如脚趾上的疼痛**感，即脚趾上的疼痛这一类型的感觉。这种（或这些）感觉可能发生于并且的确也曾发生于那些脚趾已被截除的自主体身上。

心理事项和物理事项在另一方面似乎也存在着不同。你的思想和感觉是私密的，别人可以猜测或推断出你的所思所感，但是，只有你才有到达你的思想和感觉的“直接”通道。我和你肩并肩站在一起，我们可以观察到同一棵树或同一个人，我可以观察到你正在形成某一思想，或正在经历着某一疼痛，但是，我显然不能像你一样，直面你的思想和疼痛。我的经验不是关于你的疼痛，而是关于你的疼痛对你所起的作用和你的行为。

像这样的一些思考促使我们追随笛卡尔，并将感觉和思想置于物理世界之外。就笛卡尔而言，这就意味着，心理实体是不同于物理实体的实在，它们本身就具有各种各样的独特特征，而心理性质一定是心理实体的性质。笛卡尔宣称，我们看作是心理属性的东西，其实是思想的样式，即成为一种能思实体的各种方式。与之相反，物理属性是广延的样式，即在空间中具有广延或占有空间的各种方式。不过，一旦我们信奉这种图式，就会遇到这样的问题：心理实体和物理实体如何**可能**发生因果相互作用？波希米亚（Bohemia）的伊丽莎白（Elizabeth）公主在致笛卡尔的信中即称：“对我而言，把移动身体和被身体移动的能力归之于非物质的身体，比把物质和广延归之于灵魂要容易得多。”（肯尼［Kenny］，1970：140）

若要理解笛卡尔所面对的这种困难，我们就要注意到：心理实体和物理实体之间的因果相互作用显然迫使我们放弃物理世界在因果上是自主的（causally autonomous）这一观点。物理学将物理世界视作一个封闭的系统。这个系统内的事件（或许和“边界条件”一起）只可能反映系统内的其他事件。假设这些事件从根本上说是基本粒子的运动，那么这些运动只可能受到其他粒子运动的影响。如果我们设想一个非物理的实在与一个物理系统发生因果相互作用，我们就应该不得不赞同，粒子的运动并不是由其他粒子的运动所引起的。[1]

我们并不是要将所诉诸的这种因果自主性视作反对心与物之间可能有因果关联的先天证据。相反，它提醒了我们，非物理的心灵与纯物理世界之间有因果相互作用的预想（prospect）不仅迫使我们重新思考基本自然规律的特征，还迫使我们拓展关于世界总的说来是由什么构成的观念。我

们应该期待发现这样的一些粒子，它们的行为方式不可能只根据支配粒子间关系的各种规律来说明。不过，这一点并非一定蕴含这样的可能性，即非物质的原因**违反**了自然规律。自然规律不同于法律规则，是不可违反的。一片树叶缓缓地飘落至地，并没有违反万有引力定律。当然，若要解释这片树叶的行为，需诉诸某个系统的各种复杂特征，这个系统包括了这片下落的树叶、地球以及这片树叶据以下落的流动的大气。对各种非物理原因的说明之所以会使这种因果图式复杂化，并不是因为容许违反物理规律，而是因为要补充说明一些不可预知的因果因素。

伊丽莎白公主所表达的这种担忧，反映了笛卡尔同时代人的普遍忧虑。如果心灵是空间上无广延的实在，那么它们如何可能影响到一个空间上有广延的世界？非空间的事物如何可能支配有空间的事物？**因果机制**的本质是什么？为了理解这些担忧的力量，我们思考一下力学因果关系中的一个简单例子：一个转动的齿轮咬住第二个齿轮，从而使第二个齿轮转动起来。在这个例子中，我们有一个独特的因果机制：我们能够理解第二个齿轮的转动是如何由第一个齿轮的转动引发的。现在假设第一个齿轮被一个非空间的实在所代替，那么，这样一个实在如何可能咬住第二个齿轮呢？

当然，转动的齿轮所能提供的只是一个容易想象的因果链的例子，我们还可以思考一下磁铁作用于铁屑或月亮对潮汐的影响。在这两个例子中，我们不能观察到任何类似于原因和结果之间的机械链接的东西，然而我们也不会认为这样的一些例子有什么值得担忧的。这部分是因为我们非常熟悉这些现象，部分则因为我们已经接受了如下观点，即当一些物体存在于同一个**场**内，即便相距甚远，也能彼此影响。磁铁创造的就是磁场，铁屑受到了这种场的各种特征的影响。地球和月亮改变了将二者包含在内的引力场的界限，并通过这种引力场影响场中的其他各种物体。心灵作用于身体的方式或许并不是把身体呼来唤去，而是通过创造或影响场的界限，从而反过来影响存在于场中的身体的行为。

你可能还担忧，场有一个确定的位置，而笛卡尔的心灵则完全是非空间的。某个不存在于这里也不存在于任何别的地方的东西，如何可能对存在于某一特定空间区域中的场的特征负责？不过，这样发问可能就误解了心灵无广延的观点。一个点没有广延，却有一个确定的空间位置。因而一个根本就没有广延的非物质实体，也有可能存在于空间中或从一个地方移动到另一个地方，从而影响附近的空间区域。

不过，如果这一切真的是这样，那么我们不得不将心灵实体纳入到构成我们世界的基本实在之中。这样做至少需要我们采纳一些支配各种基本成分的附属性规律。心物之间的这种关联或许表明，我们正在放弃心物之间的区别，我们实际上正在将心纳入到物之中。

对于这种观点，我们最好应该提醒我们自己，心与物之间有着巨大的鸿沟。与一般对象的各种性质相比，人们很难为有意识的经验的各种性质找到一个位置。台球的硬度、形状和质量看起来一点儿也不像疼痛的体验或芒果的味道。不仅心理现象与物理现象在性质上各不相同，就连各种物理现象之间也有天壤之别。当然，心灵的各种状态的性质也绝对不会与物理性质有交叉。

人们能够理解，比如说台球的性质是如何可能根源于台球的构成粒子的各种特征的。但是，要理解心理性质如何可能根源于构成我们（或者就此而言构成我们大脑）的那些粒子的各种特征，则完全是另外一个问题。在这种情况下，我们所面对的似乎是约瑟夫·莱文（1983）所谓的“解释鸿沟”（explanatory gap）。即使我们接受了心灵依赖于大脑这一众所周知的观念，我们也不可能清楚地知道这种依赖关系的本质。心物之间的关系几乎完全是神秘的。

9.2 意向性

关于心理性质之因果作用的各种难题，是心理因果关系问题的一个组成部分。第二个难题更难解决，最好是逐步予以解决。这里所说的难题源于下述事实，即心灵的许多状态呈现出表征内容。（哲学家们称心灵的这些状态为**意向状态**。）当你碰伤了你的脚趾，你就会体验到一种有特殊性质的经验。或许进而你形成这样的一种思想，对此，你可以用英语说：“I've stubbed my toe!”（“我碰伤了我的脚趾!”）这种思想不同于你的疼痛感，它是表征性的。[2]

我们暂且将对心理性质的最初担忧搁置起来，思考一下唐纳德·戴维森（Donald Davidson）为接受心灵的各种意向状态（参见戴维森，1970、1974）所付出的具有深远影响的努力。戴维森对心理事件与物理事件之间的关系所作的说明，被认为是标准的个例同一论。他认为，尽管心理属性或类型不能还原为物理类型（即是说，前者既不能根据后者来分析，也不

能同一于后者），但是每一个心理**个例**都同一于某一物理个例。[3]（我们假设）你在午夜感觉到疼痛，同一于你的身体在午夜发生的某一个物理事件（其实可能是神经事件），然而我们不可能将关于疼痛的话语转译为关于神经的话语。戴维森并没有诉诸人们熟知的关于心理类型的“可多样实现性”论证，不过人们认为这些论证也支持了他的观点。（我不久将会讨论可多样实现性。）根据戴维森（心理“随附于”物理）的观点，物理上无差别的各个自主体在心理上也一定没有差别，但是这并不意味着处于同一种心理状态的各个自主体在物理上也一定没有差别。你和章鱼或许都处于疼痛之中，但你的物理状况与章鱼的物理状况有显著的不同。

戴维森希望通过诉诸个例同一论来解决心理因果关系问题。如果每一个（独特的、个例）心理事件同一于某一个（独特的、个例）物理事件，并且如果每个物理事件毫无疑问地既是原因又是结果，那么心理事件也可
218 能既是原因也是结果。如果心理属性或类型不能还原为物理属性或类型，那么心理事件如何可能同一于物理事件呢？戴维森的观点是，一个事件如果有心理的描述便是心理事件；与此相同，一个事件如果有物理的描述便是物理事件。而且，同一个事件（比如说发生在你脑中的一个事件）既可以有心理的描述（“感觉到疼痛”），又可以满足物理的描述（“是一个N类型的神经事件”）。不过，我们用来将心理状态归属给自主体的那些原则，在很大程度上不同于将神经状态归属给他们的那些原则。这就意味着，虽然人们每次都可以根据自主体处于独特的物理状态之中而将心理状态（正确）有效地归属于他，但是，人们也绝不可能用神经学的语词来重构关于心理状态的话语。的确，在将一些心理语词运用于不同的自主体的过程中，我们并不知道这些自主体的何种复杂特征能够符合这些语词。情况一般就是如此。如果我正确地将头痛归属给你，我这样做是基于你的行为（言行和举止）。但是，使我做出正确归属的东西，不是你的行为，而是你大脑的某种复杂状态，不过，这种状态可能被我们完全忽视了。

戴维森的论点，即心理语词不能还原为物理语词，可以通过一个类比来阐明。每当击球手击出一个双杀球时，这个双杀球便由一系列的物理事件构成。但是，不能因此得出结论说，我们可以用物理运动的精确系列来重新定义“双杀球”。的确是这样，尽管事实是如果一个独特的物理系列构成了一个双杀球，那么任何一个在物理上没有差别的系列也就成为了一个双杀球。（因此成为一个双杀球“随附于”各种物理系列。）

尽管戴维森提出的解决心理因果关系问题的方案影响深远，但一直受

到广泛的抨击。一般来说，这些抨击都采取了如下形式。假设我们承认了个例同一论：每一个心理事件都同一于某一个物理事件。同时我们还承认，每一个这样的物理事件在因果上都是毋庸置疑的。那么，假设你今晚午夜感觉到头痛同一于你那时的神经状态 N，再假设你所处的神经状态 N 导致了一个独特的身体运动（你伸手拿一瓶阿司匹林），似乎我们就可以这样问：你伸手拿阿司匹林，是由于你感觉到了疼痛，还是由于你处于状态 N 之中？（这个问题有时也可以这样表述：一个事件引起了某种身体运动，它是作为［qua］疼痛还是作为神经状态 N 引起这种身体运动的？）

试看一个类似的例子。马克·麦圭尔（Mark McGwire）在某个赛季中所打出的第 65 个本垒球击中了格斯（Gus）的头部，从而导致他脑震荡。击中格斯的这一球是麦圭尔所打的第 65 个本垒球，但是这个球是麦圭尔所打的第 65 个本垒球与它有这样的一种物理结果之间没有关联。（理解这一点的方式就是注意到，任何一个物体只要与这个球有相同质量和速度，就会产生完全相同的结果。）这里的麻烦在于：心理状态可能与此类似。心理事件之所以可能出现于因果相互作用中，并不是由于它是心理的（不是作为心理的），而是由于它们的那些物理特征，即由纯物理描述方法所说明的那些特征。

9.2.1 心灵的“宽”状态

你或许会怀疑这个例子。毕竟，垒球由马克·麦圭尔击出的事实并不是这个球的内在的（“内置的”）属性，而是这个球由于与别的事物处于某种特定的关系之中（它由马克·麦圭尔击出）而具有的一种特征。我们很难理解，一个对象的任何一个关系特征如何可能影响这个对象的因果能力。你因头痛而服用的阿司匹林药片，可能是拜耳公司在某年 5 月份生产的第一百万个药片，但是，当阿司匹林在你的血液中发挥效应时，药片的这个特征不起任何作用。一般而言，似乎只有对象的内在的（“内置的”）特征可以影响它的因果能力。你可能承认，对象的关系属性与它所做的或能做的事情无关，但是你肯定很想知道，这一点与心理因果关系有何关联。根据很多哲学家的看法，答案是：全部有关联。

20 世纪的心灵哲学家，从维特根斯坦到普特南和柏奇，所延续的传统是，拒绝笛卡尔式的心灵图式，即心灵是自足的实在，它经历着各种感觉，拥有种种思想，并且控制着身体。感觉或许可以被理解为内在于自主体的状态和过程。然而，心灵的意向状态，如信念、愿望和意图等等，被

认为包含了一种不可取消的关系成分。

这一论点也可以通过下述方式来阐明。设想两个内在无差别的自主体，他们处于不同的环境之中，其中的一个——韦恩（Wayne），生活在地球上。当韦恩有一些思想并以“瓶子中装满了水”将其表述出来时，他的思想是关于水的。韦恩的孪生兄弟，德韦恩（Dwayne），在内在方面与韦恩极其相似（德韦恩是韦恩“分子对分子的复制品”）。德韦恩居住的星球在每一个物理细节上都极其类似于地球，除了一个很重要的例外情况——在德韦恩的星球上（德韦恩也称之为“地球”，我们称之为“孪生地球”），河流中、浴缸中以及制冰盘中的东西，虽然也被**称为**“水”，但根本就不是水，而是 **XYZ**，一种十分不同的化学物，**XYZ** 从表面上看非常像水：它的外观、感觉、味道和其他特性，都与地球上的普通水一样。当德韦恩有这样一个思想并以“瓶子中装满了水”将其表述出来时，他的思想并不是关于**水**的（水毕竟是 H_2O），而是关于 **XYZ** 的，即关于**孪生**水的（一种有着不同化学成分的无色透明液体）。

这里的主导思想是，思想的**内容**不仅依赖于自主体内在的各种特征，还十分重要地依赖于它们所处的语境。如果信念、愿望或意图的本质是其有一个独特的内容（如果信念 B_1 与信念 B_2 在内容上不同，那么 $B_1 \neq B_2$），那么信念、愿望和意图（即通常所说的意向状态）就依赖于自主体的语境。

假设真的是这样，让我们再回到戴维森。假设有这样的一种情况，一个自主体 A 处于一种给定的神经状态 N 之中，并且这个状态 N 同一于信念 B（就是说，我们之所以说 A 具有 B，是因为他处于神经状态 N 之中）。
220 再假设，N 引起了某种身体运动 M。A 的信念 B 就是 N，但是由于 N 是 B 部分地是因为 A 的语境，因此我们就很难理解：N 作为一种信念（与作为某种神经状态相对）如何与 M 的发生有任何关联。N 是 A 的信念，但这只是因为 A **在这个语境中**有信念 B。从根本上讲，这种关于 N 的纯外在事实为什么应该与 N 所引起的结果有任何关联呢？

一张美钞是一种独特的物理对象。它之所以是美钞取决于多方面的广义的语境因素：比如这张美钞有某种因果历史，它是由美国财政部印制的。这些语境因素尽管对美钞之为美钞来说是必要的，但在这张美钞被塞入自动贩卖机的运作过程中，没有发挥任何作用。与一个特定对象有关的事件导致这台机器配发了一块糖果，这里所说的这个对象**就是**一张美钞。但是这个对象是一张美钞的事实与机器的运作没有任何关系。

这一结论似乎是不可避免的。即使戴维森是对的，即每一个心理事件同一于某个（毫无疑问有因果作用的）物理事件，我们似乎也不能由此得出结论说，事件之所以有物理结果，是由于它是心理的（作为心理的）。如果我们承认意向特征是与语境有关的，那么至少那些与意向性有关的事件就是如此。

这把我们带到了什么地方？我们已揭示了两种与心理因果关系有关的忧虑。一种忧虑与心理性质有关，这些性质似乎不与物理事件发生关系。第二种忧虑不那么明显，关注的是心灵的各种意向状态——信念、愿望以及意图等；这些心灵状态似乎有一些不可还原的语境成分或关系成分，这些成分使它们都没有资格成为物理结果（例如自主体的身体运动）的候选原因。不过，无论是哪一种情况，心灵对物理世界的作用都很难得到全面的诠释。

但是，作出这样的结论或许太过草率。我们还得研究当今关于心灵的最有影响的说明，即功能主义。功能主义的目的在于为穿越与心身因果相互作用有关的问题丛林指明一条路径。

9.3 功能主义

功能主义有许多来源，但作为一种关于心灵的明确构想，则可以追溯到希拉里·普特南发表的论文《心理谓词》（1967）。[4]功能主义者坚持认为，心灵的各种状态是这些状态被归属的生物所具有的功能状态。关于功能状态的概念最好根据功能特征这一观念来理解。什么是打蛋器？打蛋器是一种设备，它的功能就是用来搅拌鸡蛋。打蛋器可以表现为许许多多的形式，它可以是由金属网制成的搅拌器，也可以是由金属或塑料制成的手摇式设备，还可以是闪闪发光的固态绞碎机（Cuisinart）。每一个这样的设备都可以当作是打蛋器的独特的体现（或“实现”）。每一种设备之所以都可以当作打蛋器，是因为它们都实现了独特的功能：它们都将未搅拌的鸡蛋作为输入，并以搅拌好的鸡蛋作为输出。

心理状态如何可能是功能状态？思考一下疼痛的例子。你感觉到了疼痛，就是你处于具有特定类型的因果作用的状态之中，例如，疼痛产生于组织损伤或疾病，又引起了各种各样的身体反应。而疼痛也属于心理结果，当你碰伤了你的脚趾时，由于你的脚趾碰撞了某个物件，你就会进入

一种状态，即你揉你的脚趾作为反应，并且由此形成了你的脚趾受伤了这一信念和想采取适当的医疗措施的愿望。

如果以这种方式看，那么，当你感觉到疼痛，就是你处于一种具有特定类型的原因和结果的状态中。这一状态通过独特的生理状态在你身上“实现”。就这种状态实现了你的疼痛来说，至关重要的不是这种状态的内在构造，而是它具有适当的功能作用这一事实。其他种类的生物，如章鱼或爬行动物，可能有完全不同的身体构造，但也能够进入那些有相似的因果作用的状态：它们产生于组织损伤，从而导致厌恶反应。可以说这些状态便在这些物种上实现了疼痛。假设我们遇到一个来自遥远星系的生命，它是硅基“生物”。这种生物是否也会感觉到疼痛？功能主义者认为，我们应当倾向作出肯定的回答，因为我们能证明，外星生物也有能力进入这样一些状态，它们在独特的原因和结果方面类似于我们的疼痛状态。如果这些外星生物的身体受到伤害时，我们若是只基于它们的身体构造与地球物种不同就否认它们经受着疼痛，那将显然是粗暴的。

因此我们判定，疼痛（和通常所说的心灵状态）是功能状态。这些状态的特征不是由它们的内在构造赋予的，而是它们所具有的适当的因果作用赋予的。这种观点似乎一劳永逸地解决了心理因果关系问题。如果心理状态是功能状态，而这些状态的本质归结为物理原因和物理结果的各种模式，那么似乎心与物之间的“鸿沟”就并不存在。心理状态从根本上讲是由于引起它们的原因和由它们导致的结果而形成的。

9.3.1 可多样实现性

然而问题并不是这么清楚。功能主义者并没有将心理状态同一于其所有者的物理状态。相反，功能主义者将这些物理状态当作心理状态的**实现者**。疼痛是通过某种物理状态在你身上实现的，但是其他生物（或其他可能生物）的疼痛是通过完全不同类型的物理状态来实现的。心理状态就是这样可多样实现的。你之所以处于心理状态 S 中，是由于你处于物理状态 P_1 中；章鱼之所以处于状态 S 中，是由于它处于非常不同的物理状态 P_2 中；阿尔法·肯陶洛斯[①]之所以处于状态 S 中，是由于它处于物理状态 P_3 中。P_1、P_2 和 P_3 是 S 完全不同类型的物理实现者。疼痛不能同一于这些类型中的任何一种，即便还没有排除其他类型的实现者。功能主义者认为，

① 希腊神话中的人面马身怪物。——译者注

使疼痛之为疼痛的东西，不是实现疼痛的那种状态的物理特征，而是那种状态有适当的因果作用的事实。

这里不是讨论功能主义利弊得失的地方（参见第 1 章）。不过，我们可以分析功能主义者的心理因果关系问题是如何产生的。一种功能状态并不同一于实现它的那些物理状态。功能主义者通过以下方式表述这种观点，即将功能状态描述为“高层次”状态：生物之所以处于这些高层次状态之中，是因为它们处于特定的低层次的实现状态之中。[5]但是，人们很难理解，高层次状态如何可能有低层次的物理结果。如果你感觉到疼痛（我们假定，它之所以在你身上实现，是因为你有某种特定的物理状态 P_1），那么，看起来好像是，是你的疼痛的物理实现者 P_1 导致了任何与疼痛有关的物理反应。这种疼痛本身似乎只不过是漂浮在它的物理实现者之上，因而毫无作为。这种情况可在图 9—1 中阐明。（H_1 是高层次状态，例如你感觉到的疼痛。P_1 是这个状态的低层次实现者。P_2 是某个物理结果，例如你服用阿司匹林。t_1 和 t_2 反映时间的推移。）

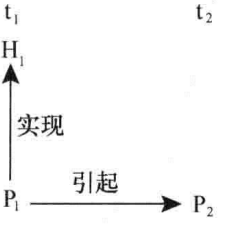


图 9—1

这一难题不仅与功能主义有关，还与任何这一类的心性（mentality）说明有关，它们将心理状态当作“高层次状态”，而这些状态由不同于它们的低层次物理状态实现。戴维森虽然不是功能主义者，但是很多人发现旁观者可以很自然地认为他赞同下述思想，即心理状态是高层次状态，它们随附于但不能还原为低层次物理状态。的确，很多有着不同信念的哲学家一直对这种思想着迷，即心理属性尽管在某种程度上依赖于物理属性，但不能因此就可以还原为物理属性。可多样实现性论证似乎已经证明，虽然心灵的各种状态可以由物理状态多样实现，但是心理状态却不能还原为物理状态。如果心理不能还原为物理，那么，我们要么假定心理状态是某种在它们的物理实现者“之外或之上”的东西，要么（更极端地）假定心理状态根本不存在，有的只不过是物理状态和物理事件。

后一种观点，即**取消主义**（参见第2章），给大多数人的印象是成事不足，败事有余。当然，也有人可能认为，我们具有心理状态是有待解释的事实，而不是有待取消的选项。取消主义者列举了一些实例，在这些实例中，有些实在（热质、燃素以及以太是三个经常被提到的例子）由科学理
223 论所假定，随之又与这些理论一同被取消了。这些实例并不能很好地适用于我们关于心性的直接体验。心理状态及其属性不是像热质、燃素或以太一样的理论假设，而是一种直接经验的对象（或更恰当地说，是各种直接经验的对象）。在思想中没有疼痛，没有忧伤或快乐，也没有思想，只有神经事件，这种说法的意义可能是什么呢？

就心灵的各种状态而言，取消主义似乎前景渺茫，它是一种令人绝望的举动，即试图通过取消心理状态而只保留物理世界，来回答心灵的各种状态如何与物理世界发生关联。其他的可能选择——心灵是高层次实在这一观点和还原论——本身似乎也存在严重的问题。我们早已注意到，还原论显然与可多样实现性的可能性相左。撇开那种担忧不谈，许多理论家认为，还原论是一种隐秘的取消主义，关于心灵的这种构想来源于科学是万物的尺度这一愚昧看法。我们应该承认（正如这些理论家所认为的那样），心理状态是一种不可还原的高层次现象，我们应该按照它们自己的要求来对待它们。如果在物理世界的因果结构中很难调和这些现象的作用，那么，我们不应该去怀疑这些现象，而应该放弃以下这种“科学的”观念，即所有真正的因果关系都可以还原为基本的物理过程（参见波斯特［Post］，1991；杜普雷［Dupré］，1993；波伦［Poland］，1994）。

9.4 实在的层次

我们面对的这三种选择之所以难以接受，与其说是因为事物的本质，或许不如说是因为表述这些选择时所使用的概念框架（正如管理学文献所建议的那样）。如果我们想要在理解心灵在自然界中的地位方面取得进展，就需要“跳出固有思维模式”。

首先思考一下，心灵状态的实在论者可能主张什么。一般而言，如果说你是一个关于某个现象或某个领域的现象的实在论者，也即是认为你相信所说的那类现象独立于你关于它们的思想。我们中的大部分人，是关于

桌子、椅子、山脉和星系的实在论者，是关于鬼、女巫和燃素的非实在论者。一些哲学家支持关于价值的实在论，认为对象是可估值的，也即是说它完全独立于我们对它们的评价而存在。另一些哲学家则是关于价值的非实在论者，他们更倾向于认为，对象的价值以某种方式依赖于评价者对它的态度。对于心灵状态的实在论者来说，什么是必要的？随着论证的逐渐明晰，对这个有点模糊的问题的回答方式能对心理因果关系问题在其中得以提出和解决的概念框架产生明显的影响。

哲学家被训练成了以独特的方式来思考实在论的人。如果你认为鬼、夸克或以太独立存在于你关于它们的思想，那么你就是一个关于它们的实在论者。由于追求精确性，哲学家更喜欢以我们对语词或谓词的态度来描述实在论：

> （P）如果你认为语词“F”指示了一个能被（正确）运用它的所有对象所共享的属性，那么你就是一个关于F的实在论者。[6]

例如，如果你认为，“是红色的”指示了某些对象所具有的、能被正确运用了它的所有对象所共享的属性，那么你就是一个关于颜色的实在论者。

我们首先注意到（P）超越了下述这种不会有人反对的主张，即当“是红色的”运用到一个对象上时，它之所以能运用到这个对象上，是因为这个对象具有某个属性。这个主张不会有人反对，是因为它没有像（P）一样暗指运用“是红色的”的每一个对象都具有同一个属性（可能是属性是红色的）。为了理解这种不同，思考两个红色的球。一个球是深红色的（crimson），另一个球是鲜红色的（scarlet）。这两个球（由于各自的颜色）都符合谓词“是红色的”。但是，很容易让人怀疑的是，是否存在某一个属性——如是红色的——被所有的球**包括**那个具有某种独特的红色色调的球所共有。这两个球都是红色的因此是错的吗？只有一个哲学家可能这么说，也只有一个哲学家接受了P。[7]更理所当然的说法是，“是红色的”之所以能够（正确而准确地）运用到某些对象上，是因为这些对象具有了（可能是无穷）一系列被我们归类为红色色调的颜色中的任何一个。

我的意见是，类似的观点也适用于各种类型的心理语词，这些语词被认为涉及可多样实现的属性和状态。例如，人们之所以认为“感觉到疼痛”能适用于各种各样的生物，是因为这些生物处于不同类型的状态

中。[8]虽然这些状态不同，但它们是相干相似的（pertinently similar）。如果功能主义者是正确的，那么，这些状态在引起它们的各种事件和由它们引起的各种事件的方面就是相似的。如果功能主义者是错误的，即如果感觉到疼痛不只是一种特定类型的功能状态，那么一个生物之所以符合“感觉到疼痛”，是因为它所处的状态与运用了“感觉到疼痛”的其他生物所处的状态是相干相似的。这里的关键词是“相似的”。当你和章鱼都感觉到疼痛，你们有相似但并非完全一样的状态。我们没有必要像功能主义者那样假设你和章鱼都有某个另外的高层次状态，在这个状态上，你和章鱼完全相似，并且这个状态不仅直接地符合疼痛这个谓词，还可以在你和章鱼身上有不同的实现。

如果这是正确的，那么我们至少有可能解决一个关于心理因果关系的重要问题：高层次状态或属性是如何可能有低层次的结果？我的意见是，高层次状态和属性是哲学家们伪造出来的东西，是由于他们偷偷地接受了类似于原则（P）的某种东西。在不放弃关于推定的高层次事项的实在论的前提下，我们能够拒绝高层次状态和属性。“感觉到疼痛”或许不仅可以正确而准确地运用于你的身上，还可以运用于章鱼上，甚至还可以运用于阿尔法·肯陶洛斯上。它之所以运用于你、章鱼和阿尔法·肯陶洛斯上，不是因为你们分享了一个独一无二的高层次属性，而是因为你、章鱼和阿尔法·肯陶洛斯具有一些相似但并非完全一样的属性。

由于受到许多反还原论者的反对，因此我提出的观点属于本体论还原论，而非还原论者的还原论。我暂不用这样假设，我们可以将关于心理状态的话语转译为关于生物状态的话语，甚至可以转译为对与基本粒子有关的那些状态的讨论。我在这里提出的观点也不是某种形式的取消主义。特定的心理语词之所以可以正确而又准确地运用于不同的生物上，是因为这些生物具有许许多多相干相似的物理状态中的任何一个。无论是在哪一个个别的例子中，一个确定的物理状态都符合这个语词。根据这种观点，需要取消的是所谓的高层次状态或属性，因为它是一种纯粹的哲学假设。

对层次图式的消解解决了心理因果关系问题的一个构成部分。不过，它没有提供一个彻底的解决方案。至少还剩下两个问题需要我们解决，一是由语境确定的“宽”心理状态问题，另一个是感受性质问题。

9.5 因果关系与心灵的宽状态

回忆一下这种思想：意识状态如信念、愿望和意图等随语境的变化而变化。[9]关于意向性概念的一个结论就是，一旦语境不同，内在无差别的各个自主体（“分子复制品”）就可能有完全不同的思想。这种观点，与对象的因果力完全随着它的内在构造的改变而改变的观点结合到一起，可以推导出如下观点，即心灵状态在内容上的不同不会产生因果上的不同。这就等于说，由于你的心理状态的内容取决于那些外在于你的那些因素（一般而言，取决于你的语境），因而这些状态的内容——关于它们的东西——不会对你的行为产生影响。但是，你的所思所想所愿的确会对你所做的事情产生影响。

各种选择再次摆在我们面前。第一种选择有些牵强，你的内在属性决定了你所做的事情（例如，决定你如何对外来刺激作出反应）。你具有特定的信念和愿望的确是取决于你所具有的内在状态和所处的语境，但这一事实在因果上是无关紧要的。这种观点仅仅将取消主义的冲动推向一个新的领域。在任何一种情况下，我们很难相信取消主义承认了失败。

第二种选择与功能主义者诉诸实在的层次性相类似，它假设我们用关于**解释**的话语替代关于**因果关系**的话语。我们常常通过诉诸心灵的各种意向状态来描述并解释彼此之间的行为（尤其是非人类生物的行为）。你为什么来餐厅？因为你想来点奶酪，并且相信餐厅里有奶酪。随即你就形成了去餐厅的意向。基于这种意向，你来到了餐厅。我们似乎乐意将这些解释当作基本的解释予以接受。我们关于因果关系的构想就是建立在我们对这种解释的理解的基础上。如果真是这样，那么，通过诉诸对因果关系本质的形而上学忧虑来削弱一种成功解释实践的做法就没有任何用处。在成功解释对象的行为中，原因是我们所要诉诸的东西。设想我们会发现成功的解释一般不需要因果基础则犯了概念错误。[10]

如果你承认心理因果关系问题是一个重要的形而上学问题，那么你就不会着迷于下述所谓的解决方案，即试图用关于解释的话语来替代关于因果关系的话语。如此替换实即本末倒置，这似乎是诉诸未经证实的论据来反对因果解释（从概念上说）先于因果关系这一观念。两种选择在我们面前僵持着，其中的每一种都依赖于这样的前提——由此出发可推出对另一

种的否定。我们该如何打破这一僵局呢?

我们碰到的关于心理因果关系的各种忧虑，是根植于我们关于因果关系的根深蒂固的形而上学信念。几乎没有理论家会怀疑这种实践，即根据自主体的心灵的各种状态来解释自主体的理智行为。第一个问题是，我们如何可能使这种实践顺应那些似乎是毫无根据的实践——包括这种实践，即只根据身体的内在物理属性来解释身体的行为。几乎没有哲学家会由于如下理由而被诱使拒绝考虑对蛋糕的破碎所作的精细的物理说明：蛋糕破碎了，是因为莉莲（Lilian）使劲地关烤箱门。对此我们均有一个完全可接受的日常解释，我相信心理因果关系也是如此。的确，任何一种观点如果推断出对理智行为所作的日常解释完全是错误的，它就会有某种可疑之处。但是这并不意味着这些解释有了坚实的基础。只要我们坚持，我们就有希望找到能顺应关于心理因果关系的说明。这样一来，我们就可以发现修改我们思考理智行为和推动这种行为的意向状态的方式很容易。

这样一来，对于心灵的（由语境决定的）“宽”状态以及它们的因果有效性，我们可以说什么呢？第一，我们不应仓促地抛弃下述观点，即对象的因果力（因而它所做的或它能做的）完全取决于它的内在构造。[11]第二，我们应该更小心地对待关于意向性的语境模型。从根本上讲，心理状态的可投射性特征也许可以根据自主体的内在构造来说明。这种观点可能承认，我们的思想所关于的东西——意即这些思想所指向的东西——取决于这个世界独立于自主体的方式。虽然这种关系问题可以进入对心理状态的描述，但它却不会让我们作出如下的设想，即这些心理状态本身是由自主体和外在因素之间的关系构成的。[12]

9.6 感受性质

许多读者此刻已经失去了耐心。关于心理因果关系的深重忧虑——一
227 般而言，就是心灵说明的困难问题（大卫·查默斯［1966］提出的所谓“困难问题”）——所担忧的是：如何使有意识经验的性质（即所谓的感受性质）与因果自足的物理世界一致。在这里请允许我只关注由感受性引起的两个与说明心理因果关系有关的难题。

假设你正在凝视着杰弗逊纪念碑旁那些开满鲜花的樱桃树，那么你有一个生动但难以用言语表达的视觉经验。现在假设有这样一个科学家，他

相信经验是发生于大脑之中。当你正在经历着这种视觉经验时，他就会仔细地检查你的大脑。这个科学家看到了令人生厌的灰白质，通过更仔细的检查（并借助昂贵的仪器），他观察到了精细的神经活动：细胞放电，沿着轴突的化学反应，等等。这些活动可能对应于你的视觉经验。你的经验有一个确定的质的特征，但是这个科学家的观察并没有揭示出这一点，它所揭示的只是令人生厌的神经性质。如果你个人经验的性质不在你的大脑中，那么它又在哪里呢？或许它们在物理世界之外。

这种推理显然混淆了经验之各种性质的区别。[13]你的经验是关于粉红色的花朵、白色的大理石和波光粼粼的河水。你感知到的粉红色、白色和波光粼粼的特征，是你感知到的那些对象的性质，而不是你个人经验的性质。当你感知到一个球时，球形是这个球的特征，而不是你经验的特征。如果你要坚持认为，当一个科学家检查你的大脑并观察你的经验时，他应该能观察到红色、白色、波光粼粼或球形的东西，那么你就错了。如果你脑中的神经事件**就是**你的经验，那么这些神经事件不必有这样的一些性质，它们属于你正在观察的那些对象。让我们假设，这个正在检查你大脑的科学家正在观察着你的经验的发生。但是这个科学家的经验不必与你的经验相似；你的经验是关于樱桃树的，这个科学家观察到的是某种完全不同的东西：比如，你关于樱桃树的**经验**。

这并不是说经验没有性质（任何东西都有性质），而是说我们一定要小心地区分经验的性质和被感知到的对象的性质。当我们作出这种区分时，我们就知道了，经验的性质根本不同于大脑的性质，难道不是这样吗？当我们思考这个问题时，我们很可能忘记，当我们将大脑描述为灰白的、糊状的东西时，我们正在描述大脑向我们显现的方式：**大脑就像我们观察它们时的那样**。我们没有理由认为，我们关于大脑的经验的性质应该类似于关于大脑之外的那些对象（例如樱桃树）的经验的性质。当这样考虑时，你关于樱桃树和杰弗逊纪念碑的经验就是发生在你大脑中的事件，这一经验的性质就是那种神经事件的性质，这种观点没有明显的问题。这就意味着，如果你大脑中的这些性质的改变会影响你行为的观点不存在特别的问题，那么，有意识的性质就不会引起特殊的心理因果关系问题。

许多哲学家可能会有异议。他们认为，性质在因果上是无效的。当击出的一个棒球导致了脑震荡时，与之相关是这个棒球的因果力（倾向性），而不是它的性质。这种观点是建立在这样一种实践的基础上，即将倾向属性（这些属性被认为赋予它们的所有者以因果力）和范畴属性（纯性质的

非因果属性）区别开来。[14]这种观点有一个几乎没被人注意到的问题，即在日常经验面前，这种区分显然是行不通的。你喜欢看樱花，是因为它们可感知的视觉性质；你喜欢吃冰激凌，是因为它的味觉性质。一个艺术家选择一种特殊的创作环境，部分地是因为这种环境的性质。在所有这些例子中，看起来完全像是自主体正在对这些对象的各种性质作出反应。我们如何可能使这些常识性的见解与属性不是因果的就是性质的属性观相一致呢？

或许我们应该拒绝将属性分为不相容且互不相容的两类：倾向性属性与范畴性属性。所有已拒绝这么做的哲学家，通常都试图将一类属性还原为另一类属性，比如认为所有属性都是具有倾向性的（休梅克，1980）。一种更具吸引力的可能性是，每一个属性（我这里想到的属性是具体对象的自然属性而非抽象对象的自然属性）既是倾向性的又是性质性的。这种观点有时也可以这样说：属性同时有倾向和性质两个**方面**。不过，用关于两方面的话语来处理心理属性也会导致这种想法，即每一个属性可能是两种不同的属性，一种是性质属性，另一种是倾向性的（或者说每一个属性可能是由这两种不同的属性所构成的）。我倾向同意 C. B. 马丁（Martin）的观点，即认为性质和倾向是严格同一的（参见马丁，1977；马丁和海尔，1999；海尔，1988：第 6 章）。想一想一个特殊的球，这个球的球形是这个球所具有的特殊的性质，正是由于这种特殊的性质，球才倾向于滚动。性质和倾向不只是共变，而且它们是同一个属性，只不过是思考和描述的方式不同。

假设这类观点是正确的，那么，我们就可以得出结论说，如果我们愿意支持这样的可能性，即有意识的经验从根本上讲就是物理事件，那么，经验的这些性质如何在物理世界中发挥因果作用，就没什么特别神秘的了。当然，这里有一个很重要的附带条件，许多理论家不会盲目地接受这个附带条件。我与其在这儿试图为这些观点辩护，不如将它们运用于一个让人迷惑的特殊问题上，并观察这些观点会产生什么样的结论。

9.7 怪　人

事先声明一下：关于怪人的哲学概念与人们普遍接受的关于怪人的构想有着很大的不同。[15]哲学家所说的怪人并不是一种说活不活、说死不

死、要想存在需吸人血的“未死之人”。哲学家的怪人与普通人在任何方面没有差别，除了一点：哲学家的怪人是没有意识经验的。你可能有一个怪人复制品。这个复制品的行为可能与你完全一样，它也会表现出你所有的偏好和偏见。的确，没有哪一个观察者可以将你和你的怪人孪生物区别开来。当审问你的怪人孪生物时，它会否认它是个怪人（的确，如果你相信你有意识，你的怪人孪生物也会相信它有意识）。你的孪生物有时会抱怨它头痛，有时会说它爱吃巧克力。对于任何一个观察者（甚至是配备了大脑扫描仪的神经科学家）来说，你和你的孪生物之间的不同的确都是不可见的。

怪人在思想实验中的作用就是抨击功能主义。这种思想浅显易懂。你和你的怪人孪生物在功能上是无差别的。不过，你拥有意识，而你的孪生物，尽管当你说它没有意识会表示抗议，但“大脑里面不会一片空白”。如果这些情况是可信的（即便它们不必有**现实的**可能，即“在现实世界中有可能”，也至少有**逻辑的**可能，正如你一次就可以跳过高楼大厦在逻辑上是可能的），那么，似乎功能主义就不能为心灵的本质提供一个完备的说明。功能主义抛开了心灵的一个重要特征：意识。

一些功能主义者之所以否认怪人的可能性，主要是基于这样的理由：任何存在物只要有适当的功能结构就会有心灵。不过，这就相当于说，拥有心灵只需要具有适当的功能结构。这就是问题的关键所在。

近年来人们广泛地讨论了大卫·查默斯 1966 年的著作，在这本著作中，他采取了一种不同的策略。面对怪人的可能性，查默斯仍然为功能主义辩护。他争辩道，虽然怪人在逻辑上是可能的，但在自然界中是不可能的。他认为，在现实世界中，一个生物如果具有一个与你相同的功能结构，就会有意识。意识由于具有某些适用于它的不可还原性的自然规律而从功能结构中“突现”出来。只有在那个缺少这些基本规律的世界中，怪人的存在才有可能。[16]

查默斯的观点既想将意识纳入到物理世界之中，又想同时说明在关于物理过程的一般说明中为何找不到意识的位置的原因。有意识的经验是有意识的自主体特征，但是，由于这些经验的各种性质仅仅是功能系统突现出来的副产品，所以不能直接作用于物理过程。（因此怪人是可能的。）

这种观点要求自然规律必须能够将有意识的经验的各种属性——意识的性质，即感受性——与这些经验所属生物的功能属性关联起来。查默斯接受了功能主义者的如下观点，即功能属性是高层次属性，一些自主体具

有这些属性是因为他们也具有某种低层次的属性（参见上面的9.3节）。例如，你有特定的功能属性F，是由于你具有神经属性P_1；一条章鱼有功能属性F，是由于它具有一个十分不同的神经属性P_2；阿尔法·肯陶洛斯有功能属性F，是由于有一个完全不同的神经属性P_3。F的这类物理实现者（P_1，P_2，P_3，…）是无限的，这一点对于功能主义来说非常重要。功能
230 属性绝对不可能还原为物理属性。（如果简单设想功能属性是这些物理属性的析取或可能是无限析取，那么我们就会把关于属性的这种观念搞乱。）这就形成了对基本的自然规律的生动描画。有意识的属性从功能属性中突现出来需要通过基本的自然规律将一些简单的意识性的属性与各种物理属性的无限析取关联起来，这些物理属性是从其中突现出来的功能属性的实现者。这些规律可能真的非常奇怪，它们与我们迄今为止在基础物理学中碰到的任何规律都不相同。

撇开这些担忧不谈，让我们考虑一下这种关于心性构想的意义，即关于我们前面已描述的关于怪人的可能性的意义。我一直反对功能主义者的心理状态是高层次状态这一论点及其必然的推论，即心理属性是高层次属性——自主体具有这些属性，是因为他们具有实现前者的低层次属性。功能主义者关于心理属性是可多样实现的观点，最好是根据下述观点来理解，即心理谓词（例如“感觉到疼痛”）之所以可以适用于各种不同的自主体，不是因为这些自主体分享了一个独一无二的、可多样实现的高层次属性，而是因为在一个庞大的、不那么规范的**相似**属性家族中，这些自主体具有其中的某一个属性。不同的自主体由于具有因果上相似的属性而具有心理属性，所以心理属性是“可投射的”（例如它们在解释自主体的行为中发挥着作用）。

这些属性除了把“因果力”赋予它们的所有者之外，还将性质性的本质赋予自主体。这里不存在性质属性和倾向属性这两种不同的属性，只存在思考属性本身的不同方式。如果这种关于属性的“同一论”是正确的，那么我们就绝对不可能单独改变倾向或性质。如果怪人在倾向上是你的精确复制品（因此也是功能上的复制品），那么它一定也是性质上的复制品。这种结论，如果再加上你的心理状态就是物理状态（例如你的神经系统的状态）这一观念，就会得出另外一种结论，即怪人是不可能的——（基于我们这个世界的自然规律）怪人不但是不可能的，而且是绝对的不可能的。可能会有这样的生物，虽然在某些功能方面与我们相似，但在有意识的经验方面与我们不同（或者甚至根本就没有有意识的经验）。但是不可

能有这样的生物，即它们一方面具有我们**所有**的物理属性，同时在有意识的经验方面与我们不同。

你可能不为这种结论所动。毕竟这种结论依赖于某些基本的而又有争议的哲学命题。但是，它至少打破了在讨论意识的“困难问题”时经常出现的那种令人厌倦的基调。诚然，还有另一些问题和困惑有待我们去解决，但哲学无法承担使困难问题变得更加困难的责任。

9.8 结　论

笛卡尔关于心理因果关系的忧虑，源于心灵是非物理的实体这一思想。问题便由此产生了：某种非物理的东西如何可能有物理的结果？（就此而言，物理的东西如何可能有非物理的结果？）这就是古老的心身问题，现在被称作是心理因果关系问题。

哲学家大多已放弃了笛卡尔的二元论。不过，属性二元——认为存在着心理和物理两种不同的属性——则被广泛地接受了，结果又出现了一大堆问题，这些问题像那些由坚持笛卡尔主义所导致的问题一样难以完全解决。如果对感觉和思想的说明需要假设心理属性的存在，那么，这些属性与物理过程，尤其是与产生理智行为的物理过程，究竟有什么关系呢？我已设想，有一种心理图景离不开把心理属性看作是一类特殊的非物理属性的预设，因而它是由我们没有必要接受的某些哲学命题引申出来的。现在是替换这些命题的时候了。在从事这一工作时，我们可以看到超越缠绕着心身争论许多问题的办法。毫无疑问，困难依然存在。在哲学中，当我们有可能解决自设的那些障碍时，我们应对由此实现的进步保持满意。就心理因果关系而言，亦是如此。

【注释】

感谢戴维森学院资助我2001年至2002年的专职研究，感谢莫纳什大学哲学院的热情招待及为我提供的惬意的哲学研究环境。我最想感谢的是C.B.马丁，他以本体论上最严肃的态度对待本体论的严肃命题。

[1] 笛卡尔（1596—1650）逝世于牛顿（1642—1727）创作他最著名的物理学著作之前。笛卡尔认为心灵能够影响粒子，其方式不是通过推动这些粒子的运动，而是通过改变它们的运动方向。物理系统中的运动以这种方式得以保存。不过，牛顿定律

要求动量守恒，如果粒子运动方向的改变有一个非物质的原因，那么就违反了动量守恒。

[2] 一些哲学家希望将经验的各种性质同化到表征中，将感觉经验看作纯表征性的（例如，哈曼，1990；德雷斯基，1995；泰伊，1995；莱肯，1996）。虽然我一直怀疑任何一种这样的观点，但我在这里要说的是，没有什么东西需要接受或被拒绝。如果你的确接受了它，那么你对心理性质在物理世界中的位置的担忧，就会被对表征状态的因果意义的担忧所取代。

[3] 考虑一下这个文本框——Saginaw Saginaw，这个文本框中有几个词？好吧，你可能说，这个文本框中包含了一个词的两次出现或例示。哲学家可能说，这个文本框包含了同一个**类型**的两个个例。

[4] 亚里士多德接受了一种功能主义，在生物科学和社会科学中，功能解释有一个漫长的历史（参见温奇［Winch］，1958）。鉴于以下所述，或许值得注意的是，普特南随后（并且正如我将要表明的那样不恰当地）将这篇论文重新命名为《心理状态的本质》。

[5] 参见布洛克1980年对两种功能主义的讨论：功能同一论（就是我一直所称的
232 功能主义）和功能指示理论（functional specifier theory）（这种观点可参见刘易斯，1966；阿姆斯特朗，1968；斯马特，1959）。根据后一种观点，你感觉到疼痛就是你处于一种特殊的物理状态之中。根据前一种观点，你的物理状态（由于上述提到的原因）只不过是你疼痛的**实现者**。有时可以将其表述为：心理状态是功能作用，而不是功能作用的占有者。当今大多数功能主义者信奉功能主义的功能同一这一概念，因此我将省略对这里的功能指示形式的讨论。功能指示理论虽然避免了这些问题，即将功能状态同一于高层次状态所产生的那些问题，但其付出的代价是牺牲了功能主义的核心论点，即功能状态有多样化的实现方式——这大概是功能同一论者所坚持的。

[6] 参见博戈西安（Boghossian，1990：161）。这种阐明适用于“描述谓词”（指那些用来对属性进行归属的语词，例如“是红色的”或“是聪明的”）。“实体性的谓词”（这些谓词用来对各种对象进行归类，例如“是一匹马”或“是金子”）需要一种稍微不同的阐释。它们之间的区别是否很大尚属富有争议的问题。为了保持这种讨论的简单性，我将忽略这里的这种差别。

[7] 一般而言，一些哲学家是关于颜色的非实在论者。不过，我在这里只是把颜色当作候选项。如果你怀疑这些颜色，那么可以将其替换为某个其他的属性——例如，是一个三角形，或有质量。

[8] 这一建议呼吁人们注意功能主义的功能指示形式（参见注释5）。

[9] 如果意向状态的这一概念看起来还是很奇怪，那么思考一下油画中的一个部分——一张微笑的脸，设想这张脸在不同的画境中转移：在一幅油画中，这张脸似乎处于充满欢乐的婚礼场景中间；在另一幅油画中，这张脸属于集中营中的一名士兵。语境影响了这张脸表达的意义，尽管这些脸没有任何内在的差别。

[10] 这是我对贝克（1993）和伯奇（1993）的理解。

[11] 一些哲学家（例如，特勒［Teller］，1986）一直认为，某些类型的量子状态违反了这一原则。不过，即使这种观点是对的，人们也很难理解这一点如何有助于解决心灵的“宽”状态所引起的困惑。

[12] 这是一个大问题，不可能在几个段落之内解决。相关的有趣读物包括马丁和海尔在1998年合著的、海尔在1998年独立著述的书中之115～119页、148～158页，里面有更详细的讨论。这里值得注意的是，该问题的原因通常根据它们的外在特征予以描述：如果我打开灯的开关导致房间被照亮了，那么我的行动——打开开关——可以被描述为我照亮了房间；并且我们可以说，这个行动吓坏了小偷。

[13] 斯马特（1959）和普赖斯（1956）讨论了这种混淆，他们说这种混淆是“现象学的谬误”（关于进一步的讨论，参见海尔，1998：78-81，206-209）。

[14] 一些哲学家认为，倾向属性以范畴属性为基础，其他的哲学家则认为，所有的属性都是倾向性的（关于这两种可能性的讨论，参见芒福德［Mumford］，1998；海尔，1998：第6章）。

[15] 罗伯特·柯克（Robert Kirk）最早将怪人当作哲学例子使用。查默斯则在1996年详细地讨论了这些怪人。

[16] 这种观点的一个结论是，作为意识基础的各种规律在以下这种较强的意义上说一定是基本的，即它们无法从支配基本粒子和基本力的那些规律中推导出来。

参考文献

Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Baker, Lynne Rudder (1993). “Metaphysics and Mental Causation.” In Heil and Mele (1993): 75-95.

Block, Ned (1980). “What is Functionalism?” In *Readings in Philosophy of Psychology*, vol. I. Cambridge, MA: Harvard University Press: 171-184.

Boghossian, Paul (1990). “The Status of Content.” *Philosophical Review*, 99: 157-184.

Burge, Tyler (1993). “Mind-Body Causation and Explanatory Practice.” In Heil and Mele (1993): 97-120.

Capitan, W. H. and Merrill, D. D. (eds.) (1967). *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Davidson, D. (1970). “Mental Events.” In L. Foster and J. W. Swanson (eds.),

Experience and Theory. Amherst, MA: University of Massachusetts Press: 79－101. Reprinted in Davidson (1980).

—— (1974). "Psychology as Philosophy." In S. C. Brown (ed.), *Philosophy of Psychology*. New York: Barnes and Noble Books: 41－52. Reprinted in Davidson (1980): 231－239.

—— (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

Dretske, Fred (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dupré, John (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Harman, Gilbert (1990). "The Intrinsic Quality of Experience." *Philosophical Perspectives*, 4: 31－52.

Heil, John (1998). *Philosophy of Mind: A Contemporary Introduction*. London: Routledge.

Heil, John and Mele, Alfred (eds.) (1993). *Mental Causation*. Oxford: Clarendon Press.

Kenny, A., trans. and ed. (1970). *Descartes: Philosophical Letters*. Oxford: Clarendon Press.

Kirk, Robert (1974). "Zombies vs. Materialists." *Proceedings of the Aristotelian Society*, Supplementary, vol. 48: 135－152.

Levine, Joseph (1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly*, 64: 354－361.

Lewis, David (1966). "An Argument for the Identity Theory." *Journal of Philosophy*, 63: 17－25. Reprinted in *Philosophical Papers*, vol. 1. New York: Oxford University Press (1983): 99－107.

Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Martin, C. B. (1997). "On the Need for Properties: The Road to Pythagoreanism and Back." *Synthese*, 112: 193－231.

Martin, C. B. and Heil, John (1998). "Rules and Powers." *Philosophical Perspectives*, 12: 283－312.

—— (1999). "The Ontological Turn." *Midwest Studies in Philosophy*, 23: 34－60.

Mumford, Stephen (1998). *Dispositions*. Oxford: Clarendon Press.

Place, U. T. (1956). "Is Consciousness A Brain Process?" *The British Journal of Psychology*, 47: 44－50.

Poland, Jeffrey (1994). *Physicalism: The Philosophical Foundations*. Oxford:

Clarendon Press.

Post, John F. (1991). *Metaphysics: A Contemporary Introduction*. New York: Paragon House.

Putnam, Hilary (1967). "Psychological Predicates." In W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press: 37-48. Reprinted as "The Nature of Mental States," in Putnam's *Mind, Language, and Reality: Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press (1975): 429-440; and in Ned Block (ed.), *Readings in Philosophy of Psychology*, vol. I. Cambridge, MA: Harvard University Press (1980): 223-231.

Shoemaker, Sydney (1980). "Causality and Properties." In Peter van Inwagen (ed.), *Time and Cause*. Dordrecht: Reidel Publishing Co.: 109-135. Reprinted in *Identity, Cause, and Mind: Philosophical Essays*. Cambridge: Cambridge University Press (1984): 206-233.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review*, 68: 141-156.

Teller, Paul (1986). "Relational Holism and Quantum Mechanics." *British Journal for Philosophy of Science*, 37: 71-81.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge: MIT Press.

Winch, Peter (1958). *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge and Kegan Paul.

第 10 章　民间心理学

斯蒂芬·P·斯蒂克，肖恩·尼科尔斯
(Stephen P. Stich and Shaun Nichols)

235 有关常识心理学（或者“民间心理学”，如人们经常所称的那样）的争论已进入到了当代心灵哲学的中心舞台。研究者不仅对于什么是民间心理学，而且对于它与出现在心理学和神经科学中的关于心灵/大脑的科学理解是怎样联系的都有严重的分歧。在这一章，我们将首先解释为什么民间心理学在心灵哲学中起着如此重要的作用。而在此以前，我们又需要对哲学讨论心灵的历史作一点简要回顾。接着，我们将把注意力转向当代出现的一些热烈争论，这些争论的目的在于弄清人们期望民间心理学所能发挥的哲学作用，并运用认知科学的成果寻求关于民间心理学的实在本质的更清晰理解。

10.1　为什么民间心理学在心灵哲学中起着重要作用?

要想充分理解哲学家们对民间心理学的迷恋，首先应简要回顾一下心灵哲学中两个最为重要的问题，以及几个世纪以来由对其中一个问题最有影响的回答所导致的难题。这两个问题就是心身问题和他心问题（the problem of other minds），前者问的是心理现象和物理现象如何发生联系，后者问的是我们何以能知道别人的心理状态。根据笛卡尔所提出的对心身问题的回答，宇宙中有两种截然不同的实体：物质实体和心理实体，前者位于空间和时间之中，而后者位于时间之中但不位于空间之中。按照笛卡尔所说，心理现象是发生在心理实体中的事件或状态，而物理现象是发生
236 在物质实体中的事件或状态。笛卡尔坚持认为，心理与物理之间有双向的

因果相互作用，但是很多哲学家相信，如果这两者中一个处于空间之中，而另一个不处于空间之中，那么它们如何能发生相互作用就难以理解。笛卡尔主义观点的另一个问题是：它似乎使他心问题变得非常棘手。如笛卡尔所认为的，如果只有我能感知我的心理状态，那么你似乎就无法排除这种假说，即：我是一个无心的怪人——一个物质的躯体，只在行为举止上与心灵存在因果联系。

20 世纪中叶，关于意义的证实主义解释对哲学思想有重要影响。根据证实主义者的研究，一种经验性主张的意义与能证实该主张的观察是密切相关的。在证实主义的影响之下，哲学行为主义者认为，笛卡尔主义关于心灵是“机器中的幽灵”的解释（使用赖尔［1949］的著名比喻）是彻底错误的。他们认为，如果“信念”、“愿望”和“疼痛”之类的普通心理状态术语具有意义，它们就不能指称发生在人体内部（或者还要糟糕地说，它们根本不存在于空间之中）无法观察的事件。相反，使用了这些术语的句子的意义一定是根据这些条件句来分析的，这些条件句详细说明了某个人在各种环境下会如何行为。例如，一个哲学行为主义者也许会指出：

（1）约翰相信雪是白的。

该句子的意义可以由类似于下面的句子表达：

（2）如果你问约翰：“雪是白的吗?”他会作出肯定的回答。

哲学行为主义者的最大困难可能是：他们进行意义分析的最终结局一般要么是明显错误的，要么是循环论证——用一个心理术语来分析另一个。因此，与（2）相反，即使约翰相信雪是白的，他也可能不会作出肯定的回答，除非他在**注意**、他**想要**让你知道他的想法，他**相信**作出肯定的回答就能做到这一点，等等。

哲学行为主义者越来越相信他们没有办法解决循环论证问题，而想对科学术语的意义作出证实主义解释的哲学家也遇到了一个非常类似的难题。证实主义要求理论术语的意义一定是能根据可观察的东西来说明的。但是当哲学家们真的想提供这样的定义时，他们似乎总需要另外的理论术语（亨普尔，1964）。在科学哲学中，对这个问题的反应是对于理论术语如何获得其意义的问题的一种截然不同的解释。这种新解释指出，一组理论术语的意义不是仅仅根据可观察的东西来定义的，它们是通过嵌入一种经验理论而共同获得的。任何给定的理论术语的意义都在于它和其他术语

的理论规定的（theory-specified）相互关系——**既有观察的也有理论的**。
237 关于这种观点最有影响的陈述能在大卫·刘易斯（1970、1972）的著作中找到。根据刘易斯，理论术语的意义是由他所称的“功能定义”给出的。理论实体被“定义成**由理论所规定的**因果作用的占有者……被定义成**这些**实体：不管这些实体可能是什么，它们**相互之间**以及与O（观察）术语的指称有一些因果关系”（1972：211；第一个和最后一个强调为引者所加）。

基于由威尔弗雷德·塞拉斯（1956）首先提出的看法，刘易斯接着指出：表示心理的或心理学的状态的普通术语可能是以完全类似的方式获得其意义的。如果我们“把常识心理学看成是一种引进术语的（term-introducing）科学理论，尽管它是创立于像专业科学这样的建制出现之前的”，那么，关于科学中的理论术语之意义的“功能定义”解释就能被直接运用于常识心理学所使用的心理状态术语中（刘易斯，1972：212）。刘易斯指出，这就是思考常识心理学的正确方法：

> 设想在某个天才创立心理状态理论之前——这种理论用其新引进的T［理论］术语来解释刺激与反应之间的规律性，我们的祖先最初都只谈论外部的事物、刺激和反应，但这并没有发生。我们的常识心理学绝不是一种新创立的引进术语的科学理论——甚至不是史前的民间科学。这个故事，即心理术语是作为理论术语被引进的，明显是一个神话。
>
> 它其实就是塞拉斯的神话……尽管它是一个神话，但它能成为一个好神话，也能成为一个坏神话。如果我们关于心理状态的名称实际上指的是它们在这个神话正确时所指的东西，它就是一个好神话。我把它是一个好神话作为工作假说。（1972：212-213）

自刘易斯等人[1]提出这种解释以来的30年间，它已经成为研究者关于心理状态术语意义的最普遍看法。由于这种解释主张心理状态术语的意义是由功能定义提供的，因此这种观点常被称为**功能主义**。[2]我们现在就能明白心灵哲学家们为什么对理解常识（或民间）心理学的实在本质很关注的一个原因了。根据功能主义，**民间心理学提供了普通心理状态术语的意义**。

哲学家们关注民间心理学的第二个原因很快也能得到解释。关键之处在于：根据像刘易斯这样的解释，民间心理学是一种**经验**理论，它自认为解释了在人的（可能还有在动物的）行为中所发现的“刺激与反应之间的

规律性”。当然，如果常识心理学是一种经验理论，那么就像任何经验理论一样，它最终可能会被证明是**错误的**。我们或许发现，介于刺激和反应之间的状态和过程还未被民间理论作出恰当的描述，而后者确定了心理状态术语的意义。几乎每个重视功能主义的人都承认这种可能性：常识心理学**也许**最终是错误的。然而，最近几十年来，很多杰出的心灵哲学家都认为这不仅仅是一种**纯粹的**可能性，相反，他们主张：认知科学和神经科学中不断丰富的理论和经验研究结果都令人信服地证明常识心理学**是**错误的，而且不只是细枝末节的错误。如这种观点的热情支持者，保罗·丘奇兰德曾说：

> FP［民间心理学］遭受了大范围的解释失败……它已经至少停滞了2500年，而且……它的范畴（迄今为止）与作为其背景的物质科学似乎要么是不可通约的，要么是毫不相干的。从长期来看，依据物质科学来解释人类行为似乎是无可否认的。任何符合这种描述的理论都必须严肃地面对被彻底取消的可能。(1981：212)

丘奇兰德并未止步于抛弃（或“取消”）民间心理学理论。他和别的“取消主义者”还提出，由于民间心理学是一种有严重缺陷的理论，因此我们也应当得出这样的结论：内置于民间心理学的理论术语并不是真的有什么指称。他们认为，信念、愿望以及民间心理学的其他假设，和燃素、以太以及被证明严重错误的经验理论的其他假设是完全类似的；和燃素、以太等一样，**它们均不存在**。显然，这些是极富争议性的主张。从它们第一次被提出以来，对于它们的可信性的争论就在心灵哲学家的议事日程中占据重要地位。[3]由于取消主义者的核心论题是坚持民间心理学乃一种极其错误的理论，因此，想对这个论题作出评价的心灵哲学家对于什么是民间心理学以及它的主张是什么，显然需要一种清晰而准确的理解。

10.2 什么是民间心理学？两种可能的回答

如我们所看到的，功能主义者主张：普通心理状态术语的意义是由它们在常识心理学理论中所起的作用决定的。但是，这种理论究竟是什么？在哲学和认知科学的文献中，对于这个问题有两种截然不同的回答。[4]对于刘易斯和很多遵从他的指引的人来说，常识的或民间心理学和这些关于

心理状态的主张是密切相关的，它们几乎得到了每个人的同意，并且还被认为是显而易见的。

> 将所有你能想到的与心理状态、感觉刺激和运动反应的因果关系相关的老生常谈汇集起来……还要加上这些老生常谈：一个心理状态从属于另一个——“牙疼是一种疼痛”，等等，可能还有其他形式。只包括我们大家都知道的老生常谈——每个人都知道它们，每个人都知道其他的每个人也都知道它们，如此等等。由于这些词语的意义是大家都知道的，因此我将主张**心理状态的名称的意义来自于这些老生常谈**。（1972：212；强调为引者所加）

因此，根据这种回答，民间心理学只是老生常谈的集合，或者说由于这个集合一定是巨大而笨拙的，因此我们可能会认为民间心理学是一个概括的集合，它以简单易懂的方式使这些老生常谈变得系统。这种系统化可能也会使人们更加自然地将民间心理学描述成一种理论。我们将此称为民间心理学的**老生常谈解释**（platitude account）。

第二种回答关注的是一组哲学家和心理学家都很感兴趣的技能。在很多情况下，人们都很擅长**预言**别人的行为。在要求预言一个驾车者在接近红灯时会做什么时，几乎每个人都会说他将停车，而且幸运的是，我们的预言通常是正确的。我们经常也很擅长**归属**心理状态[5]——擅长说出他们知觉、认为、相信、渴求、害怕等等的东西，擅长**预言**未来的心理状态，擅长依据过去的心理状态来**解释**行为。[6]在最近的讨论中，**心灵阅读**（mindreading）这个离奇的称号经常被用于描述这组技能，而且自20世纪80年代中期以来，发展心理学家和认知心理学家已经创作了大量文献，其目的就是要考察心灵阅读的突现，解释支撑它的认知机制。

对于支撑心灵阅读的认知机制，最为普遍接受的观点（20世纪80年代中期以前**唯一**的观点）是：人们对心灵有丰富的心理表征信息，这些信息在引导这些生成了我们的归属、预言和解释的心理机制方面起着重要作用。有些支持这种观点的心理学家认为，心灵阅读中所涉及的信息具有和科学理论几乎完全相同的结构，获得、存储和使用它的方式与其他常识和科学理论几乎完全相同。这些心理学家经常将他们的观点称为**理论论**（the theory theory）（戈普尼克和韦尔曼［Gopnik and Wellman］，1994；戈普尼克和梅尔佐夫［Gopnik and Meltzoff］，1997）。还有的人认为，心灵阅读中所使用的很多信息是天赋的、存储于心理“模块”之中，在心理模块

中这些信息与心灵其他部分中所存储的信息只能发生非常有限的相互作用（肖勒和莱斯利［Scholl and Leslie］，1999）。由于模块性理论家和理论论者都同意心灵阅读依赖于关于心灵如何工作的丰富信息，因此我们就能用**信息丰富理论**（information-rich theories）这个术语来指称这两种理论。这两种理论提出用另一种方法来说明这种理论（如果功能主义者是正确的），它决定着心理状态术语的意义——它是支撑心灵阅读的理论（或全部信息）。我们将此称为民间心理学的**心灵阅读解释**（mindreading account）。

现在请允许我们问一个问题：民间心理学的老生常谈解释和心灵阅读解释是如何发生联系的？心灵阅读的信息丰富理论所假设的关于心灵的心理表征信息是如何与这个老生常谈的集合相联系的，（根据刘易斯）这个集合决定着心理状态术语的意义吗？一种可能是：这些老生常谈（或者它们的某种系统化）与引导心灵阅读的信息几乎是完全**同一的**——除了每个人都欣然接受的常识信息之外，心灵阅读很少（或者说根本没有）使用关于心灵的信息。如果这是真的，那么民间心理学的老生常谈解释和心灵阅读解释就趋向一致。但是，和大多数研究心灵阅读的认知科学家一样，我们也相信这种趋同是**非常**不可能的。我们持怀疑态度的出发点是与认知科学家所考察的其他复杂技能所作的比较。在从面部识别（扬［Young］，1998）到决策（吉仁泽［Gigerenzer］等，1999）再到常识物理学（麦克洛斯基［McCloskey］，1983；海斯［Hayes］，1985）的几乎每个案例中，都能发现心灵所使用的信息和原则根本不可能通过内省得到。在这些领域，我们的心灵使用了大量人们不能识别或者赞成的信息，这与刘易斯据认为能识别和赞成的老生常谈不同。其次，让我们持怀疑态度的理由是：在很多心灵阅读任务中，人们归属心理状态似乎是根据（他们没有察觉到他们在使用的）暗示。例如，埃克曼（Ekman）已经证明存在大范围的“欺骗暗示”（deception cues），它们使我们相信一个对象（target）并不相信他所说的话。这些暗示包括“面部表情的变化、身体的运动、噪音的抑扬变化、咽喉的吞咽、很深或很浅的呼吸、词语之间长时间的停顿、口误、面部表情微小变化、手势错误”（1985：43）。在大多数情况下，人们根本不知道这一事实，即他们使用了这些暗示。因此，尽管对于支撑心灵阅读的心理机制还有很多内容需要了解，但我们认为，这些机制所使用的关于心灵的信息很可能比刘易斯式老生常谈中所包含的信息要丰富得多。

如果我们在这一点上是对的，那么某些研究者——他们认为关于普通心理状态术语的意义的功能主义解释是正确的——就必须面对一个非常关键的问题：哪种民间心理学解释经过择取能实际决定心理状态术语的意义的理论？这些术语的意义是由我们能够通过汇集并对老生常谈进行系统化而表达的理论决定的，还是由这种丰富得多的理论所决定的——我们只有研究支撑心灵阅读的机制所使用的这类信息才能发现它？

我们不认为对这个问题真有什么决定性的回答。如果我们能借助一种具有良好动机并被普遍接受的一般意义理论，那当然是非常有用的。但是，众所周知，目前并没有这样的理论。意义是一个甚至在最基本的问题
241 上也充满了分歧的话题，很多哲学家都认为，用以说明心理状态术语之意义的整个功能主义方案都是完全错误的。[7] 然而，在说了所有这些之后，我们倾向于认为那些对功能主义方案持同情态度的人应当选择民间心理学的心灵阅读解释而不是老生常谈解释。因为根据心灵阅读解释，民间心理学是人们在识别和归属心理状态，作出关于心理状态的推理，以及基于心理状态归属而进行预言和解释时实际使用的理论。很难理解为什么有人——他和功能主义者一样，认为心理状态术语获得意义是通过内置于一种理论——想把注意力集中在人们很容易就能承认其原则的以老生常谈为基础的理论，而不是另一种更丰富的理论——它实际指引着人们关于心灵的思考和谈论。

10.3 来自模仿论的挑战

让我们花点时间来看看我们的讨论进行到哪里了。在 10.1 节，我们解释了为什么民间心理学在最近的心灵哲学中起着如此重要的作用：功能主义者认为，民间心理学是暗中对普通心理状态术语作出定义的理论；而取消主义者（他们在心理状态术语的意义方面通常和功能主义者的意见一致）认为，民间心理学是一种严重错误的理论，不仅这种理论本身，而且它所假设的心理状态都应当被抛弃。在 10.2 节，我们区分了两种不同的民间心理学解释。我们认为——尽管是试探性地——功能主义者应当选择心灵阅读解释，根据这种解释，民间心理学是大量丰富的信息或者理念，它构成了人们在归属心理状态、预言和解释行为时的技能的基础。在这一节，我们将把注意力转向一种重要的新挑战，它是针对所有这些内容而提

出来的。从20世纪80年代中期以来，很多哲学家和心理学家就一直争辩说，认为心灵阅读利用了大量关于心灵的丰富信息的观点是错误的。相反，他们主张，心灵阅读可以被解释成一种**心理模仿**，它很少甚至根本不需要关于心灵如何工作的信息（戈登［Gordon］，1986；希尔［Heal］，1986；戈德曼，1989；哈里斯［Harris］，1992）。如果这些模仿论者是正确的，如果我们同意民间心理学的心灵阅读解释，那么也就**不存在民间心理学这样的理论**。这对功能主义者来说是坏消息，它对取消主义者来说也是坏消息，因为如果不存在民间心理学，那么他们的核心论证——主张民间心理学是一种严重错误的理论——就同样犯了严重的错误。

支撑心灵阅读的心理机制何以能不需要大量丰富的信息呢？模仿论者在回答该问题时通常都是先使用一个类比。假如你想预言一架特定的飞机在某种气流条件下是如何工作的。一种办法是从航空学理论以及对这架飞机的详细描述来进行预言。另一种迥然不同的策略则是建造一个飞机模型，把它放入一个复制了那些气流条件的风洞中，接下来只要观察这个模型是怎样工作的就行了。第二种策略不同于第一种，它不需要丰富的理论。模仿论者认为，与第二种策略相类似的策略可以用于解释人们的心灵阅读技能。因为如果你打算预言另一个人的心灵将有何举动，再如果那个人的心灵与你的相似，那么你或许就能把你自己心灵的成分视为关于另一个人（我们称之为“对象”）的类似心灵成分的模型。

至于这个过程是如何进行的，可以简要地概括如下：假如你想预言对象在某个重要问题上将决定做什么。我们以为，对象的心灵将利用这样一个决策或者“实践推理”系统进行决策，这个系统把他的相关信念和愿望当成输入，并（设法）提出关于要做什么的决定。图10—1中的细线勾画的是这种可能构成了正常决策过程之基础的认知结构。假如你的心灵能暂时使你的决策系统“脱机”，从而使你没有按照它所产生的决策作出实际的行动。再假如在这种脱机模式下，你的心灵能为你的决策系统提供一些假设的或者说“假装的”（pretend）信念和愿望——你可能不实际具有而对象却实际具有的信念和愿望。于是，你的心灵就可以只是袖手旁观，让你的决策系统来作决策。如果你的决策系统与对象相似，或者你输入脱机系统的这些假设的信念和愿望与对象所具有的相近，那么你的决策系统所作出的决策就会与对象的决策系统将作出的相类似或者近乎相同。如果现在将这个脱机决策发送到你心灵的某个部位，它产生了关于其他人将做什么的预言，那么你就会预言：这就是对象将会作出的决策，而且你的预言

很可能是正确的。根据模仿论者，所有这些过程的发生你很少甚至根本不能有意识地察觉。此外，关键的论点在于：这个过程没有使用任何关于决策系统怎样工作的理论或者丰富的信息。相反，你只是用自己的决策系统来**模仿**对象将实际作出的决策。图 10—1 中的粗线勾勒了这种认知结构，它可能构成了这种基于模仿的预言的基础。

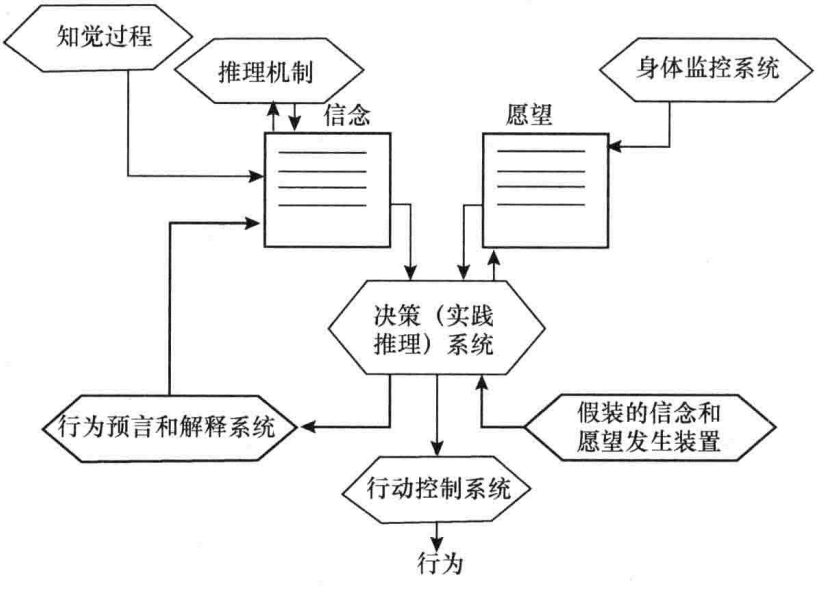


图 10—1

我们刚才描述的过程是让决策系统脱机并通过模仿来预言决策。但是几乎完全相同的过程或许也能用于使推理机制或者心灵的其他成分脱机，并因此用于作出关于其他种类的心理过程的决策。一些更热心的模仿论支持者指出：**所有**心灵阅读技能都可以由类似于这个模仿过程的其他过程来完成，因此我们不必认为民间心理学理论在心灵阅读中有**任何**重要作用。
243 如果这是对的，那么功能主义和取消主义就都有麻烦了。[8]

10.4 对心灵阅读的三种解释：信息丰富解释、基于模仿的解释和混合解释

模仿论者和心灵阅读的信息丰富解释的支持者对于支撑心灵阅读的心理过程提出了竞争的经验理论[9]，有关这个话题的很多文献都被看成是这

两个群体之间孰为胜者的争论。[10]然而，近年来，人们越来越意识到心灵阅读是一种复杂的、有多个面向的现象，心灵阅读的某些方面可能是由类似于模仿且信息贫乏的过程推动的，而有些方面则是由信息丰富的过程推动的。这种混合方案就是我们多年来所提倡的方案（斯蒂克和尼科尔斯，1995；尼科尔斯等，1996；尼科尔斯和斯蒂克，即出），在这一节我们将简要概括一下支持这种混合方案的论据。[11]我们首先集中于心灵阅读的一个重要方面，在这个方面信息丰富的解释特别不可信，而模仿式的解释很有可能是正确的。之后我们将处理心灵阅读的另外两个方面，我们认为，对于这两个方面，信息丰富的解释显然比基于模仿的解释更可取。

10.4.1 推理预言：一种由模仿推动的心灵阅读技能

对于正常成年人的心灵阅读技能，一个突出的事实是：我们都非常擅长预言对象的推理，甚至是他们明显的**非论证性的**推理。例如，假如弗雷德在听到广播的一条简讯后开始相信美国总统辞职了，此时弗雷德认为谁会成为总统呢？我们很快会产生这样的预言：弗雷德认为副总统将成为总统。我们完全知道而且我们以为弗雷德也完全知道：有很多突发情况都会使他的推理出错。副总统可能被暗杀了，副总统也许在宣誓就任总统前辞职了，一桩丑闻可能使副总统被罢免了，可能发生政变了，等等。很容易产生很多副总统不会成为新总统的可能。然而，我们仍会毫不犹豫地作出类似弗雷德的非论证性推理的预言，而且在大多数与此类似的情况下，我们的预言都是正确的。任何合适的心灵阅读理论都必须承认这些事实。

心灵阅读的信息丰富方案的支持者显然对推理预言保持了沉默。的确，就我们能够决定的事实而言，还没有一位这种方案的支持者**打算**对这一事实作出解释，即：我们都非常擅长预言别人作出的推理。我们都倾向于认为这种遗漏的原因是非常清楚的。对于信息丰富方案的坚定支持者来说，我们的推理预言技能之所以能够存在，唯一可行的解释是**更多的信息**。如果我们擅长于预言其他人是如何推理的，那么这一定是因为我们以某种方法获得了一种关于人们怎样推理的非常合适的理论。但这种解释似乎太过分（profligate）。要理解其中的原因，可以考虑预言推理与预言某个和我们说相同语言的人的语法直觉之间的类比。要解释我们在后一项任务中的成功，信息丰富方案的支持者就必须说：对于在别人身上推动语法直觉产生的过程，我们有一种**理论**。但是如哈里斯（Harris，1992）所指出的，这似乎是很牵强的。更简单的假设是：我们依赖于**我们自己**的生成

语言学直觉并决定我们自己关于特定句子的直觉的机制，我们把这些机制归入到对象之上。

使我们接受综合的心灵阅读理论的过程一定有不同表现，模仿过程就是其中之一方面，哈里斯所谓的**从简单性进行的论证**（argument from simplicity）起着重要作用。然而，我们认为，对于觉得以模仿为基础的推理预言解释比信息丰富解释更可取的看法来说，从简单性进行的论证并不是唯一的理据。当然，如果从简单性进行的论证是唯一可取的论证，那么
245 信息丰富方案的坚定支持者可能就只是坚定自己的立场，并辩称大自然所产生的系统经常决不是简单的。对于生物系统中多余的和明显不必要的复杂性，确实有很多的例证。因此，信息丰富理论者可能会认为，单独这一事实，即对推理预言的基于理论的解释比模仿式的解释更简单，几乎不是反对它的有力论证。然而，还有另外一种论证，它可以用于反对关于推理预言的信息丰富方案。我们认为这是一种特别重要的论证，因为它可以推广到很多别的心灵阅读技能，因此，在帮助我们决定心灵阅读的哪些方面能被合理地看成是以模仿为基础的内容时，它是一种宝贵的启发法（heuristic）。

第二个论证，我称之为**从准确性进行的论证**（argument from accuracy），它由以下观察开始：在很多情况下（包括和大多数心灵阅读者以前所遇到的都迥然不同的情况）推理预言都非常准确。例如，在认知社会心理学的“启发法和偏见”（heuristics and biases）传统中有大量文献记载着这样的情况：人们在很多需要演绎和归纳推理的问题上都作出了似乎很糟糕的推理。[12]然而，在所有这些文献中都不存在这样的主张：人们不擅长**预言**其他人的推理，不管这些推理是好还是坏。这与我们下面要讨论的愿望归属方面的文献形成了鲜明对比，在那里，人们经常会谈到，人们的愿望和决策有多么出乎意料和**难以**预测。虽然还没有进行系统的研究，但我们认为：一般来说，即使是在与他们以前遇到的都完全不同的任何问题上，预言其他人行动的人，所作的也很可能就是他们对自己所作的这些糟糕推理。如果真是这样，那么它就对信息丰富解释提出了一个问题：普通的心灵阅读者最后想如何提出这样一种关于人们如何作出推理的**准确的**理论——这种理论甚至支持关于非常不熟悉的推理的正确预言？这个问题会被这一事实弄得更加尖锐：人们在其他一些心灵阅读任务上的表现非常糟糕。为什么人们在推理方面得到了正确理论，而在其他心理过程方面却得到了**错误**理论呢？与此相反，基于模仿的推理预言解释对我们的准确性有

一种现成的解释。根据模仿论解释，我们对于**作出**推理和**预言**推理使用的是相同的推理机制，因此，可以预料，我们能预言其他人会作出与我们相同的推理。

显然，从准确性进行的论证是一把双刃剑。在我们尤其擅长预言或归属不熟悉的情形中的心理状态下，这个论证表明：心灵阅读过程不可能是由信息丰富的过程推动的。但在我们**不擅长**预言或归属心理状态的情形下，这个论证表明：这个过程是不可能由**模仿**过程推动的。我们认识到，回应从准确性进行的论证可以采取各种办法，因此，我们不认为这个论证是决定性的。然而，我们确实相信，这个论证辩明了一个强有力的初始假设，即：准确的心灵阅读过程是由类似于模仿的过程推动的，而不准确的心灵阅读过程则不是。如果这是对的，那么就有强有力的根据来支持这个假设：推理预言是以模仿为基础的。

10.4.2 愿望归属：一种不能用模仿来解释的心灵阅读技能

心灵阅读另一个非常重要的方面是向别人归属愿望的能力。如果没有这种能力，我们就不会知道别人想要什么，我们在预言或解释他们的行动时就会有难以克服的障碍。有很多过程都能产生对一个对象的愿望的信念。在某些情况下，我们用有关对象的言语和非言语行为（包括他们的面部表情）的信息来确定他们想要什么。在其他情况下，我们是根据别人对于对象说了什么来归属愿望的。而且，十有八九很多别的线索和资料源也都被用于愿望归属过程。我们的主张是：这些愿望归属技能不依赖于模仿，而是由信息丰富的过程推动的。关于这一主张我们有两个迥然不同的理由。

首先，愿望归属表现出了一种系统误差（systematic inaccuracy），这至少会支持一个初始假定：这个过程不是以模仿为基础的。一个例子来自于社会心理学中一组非常著名的实验。米尔格拉姆（Milgram，1963）让一个被试“老师”控制一些开关，他可以向另一个被试即“学习者”（实际上是一个同谋）进行电击。学习者每出一次错，就要求老师进行逐渐增强的电击，其中一个标有“危险：严重电击”，达到峰值的开关上标有“450伏，XXX”。如果老师向试验助手表示了保留态度，他就被平静地告知要继续实验。实验的结果是令人吃惊的，绝大多数被试都执行了**所有**电击。人们常常会觉得这些结果难以置信。当然，米尔格拉姆的研究结果是如此违背直觉，以至于在言语上重演实验时，人们仍不能预言这些结果

(比尔布劳尔 [Bierbrauer]，1973；尼斯比特和罗斯，1980：121 页的讨论)。对这些研究结果的一种可信解释是：在米尔格拉姆的实验中，实验者发出的指令产生了服从的愿望，它在多数情况下压倒了被试不伤害他们相信处于电击装置接收端的人的愿望。人们觉得这些结果令人吃惊，觉得比尔布劳尔的被试并不能预言这些结果，这一事实指出，我们确定别人愿望的能力有重要的缺陷。

在认知社会心理学中，有大量文献详细描述了很多别的案例，在这些案例中，愿望和偏好在一些显著的、出乎意料的方面受到了被试所面对的环境的影响。就现在的目的来说，重要的一点是：人们一般都觉得这些结果令人惊讶，有时候是非常令人不安的，它们令人吃惊的这个事实（甚至在看到或者获得关于实验情况的详细描述时）表明：他们用于预言被试的信念和偏好的心理机制是有系统偏差的。尽管这里不适合对文献中的很多例子作深入探讨，但我们不禁想提一下我们乐意提出的一个。[13]

洛温斯坦和阿德勒（Loewenstein and Adler，1995）对被试的这种能力进行了研究，即在**他们自己**的偏好受到一种惊人的、知之不多的效应影响时对这些偏好作出预言。洛温斯坦和阿德勒所使用的理论是**馈赠效应**（endowment effect），即人们坚定而迅速地表现出来的一种倾向：当他们实际拥有一种东西时，他们会给这种东西设定比他们不拥有它时高得多的价格（塞勒 [Thaler]，1980）。洛温斯坦和阿德勒是这样描述这种现象的：

> 在对馈赠效应的典型演示中……一组被试（卖者）被赠予了一种东西，并被给予这种选择：用它来换一定数量的现金；另一组被试（选择者）没有被赠予这种东西，但被赠予了得到这种东西还是得到一定数量的现金等一系列选择。虽然这两组的客观的财富情况是相同的，就像他们所面对的选择一样，但是受馈赠的被试坚持认为他们比未受馈赠的被试得到的钱多得多。(1995：929-930)

在这样一个实验中，它是用来检验“未受馈赠的”被试在实际拥有所讨论的东西时，是否能预言他们设定的价格。实验人员首先让被试（他们是同属一个班的大学生）检查一个印有学校校徽的大杯子。然后给随机挑选的大约一半的被试发一张表，要求他们在上面“想象他们拥有所展示的杯子，并预言他们是否愿意用这个杯子换取一定数量的钱”(1995：931)。当收到表格的被试填写完时，**所有**被试都被赠予了一个杯子，并收到第二

张表，上面的填写说明和预言表上的相似，但在第二张表上清楚地写着他们真能用杯子换钱，而且他们在第二张表上所作的选择决定着他们能得到多少钱。“被试被告知：他们在其中一条线上画圈就表示接受了这种选择——至于具体哪条线则是由实验者事先确定的。”（1995：931）结果表明：填完第一张表的被试对于他们愿意用杯子交易的钱预言过少（under-predicted）。在一组被试中，预言的交易价平均为 3.73 美元，而被试（**同样**作出预言的被试）的实际交易价平均为 5.40 美元。此外，实验中似乎有一种“参照效应”（anchoring effect），它压低了实际交易价，因为对于没有就他们自己的卖出价作出预言的被试来说，平均的实际交易价要更高，达到 6.46 美元。这里我们再次发现，人们在预言环境对愿望的影响方面有系统的误差，在这种情况下，他们不能预言的愿望是他们自己的愿望。如果这些愿望预言是由模仿过程推动的，那么，为什么这些预言有系统的误差就难以作出解释。但是，如我们所相信的，如果它们是由信息丰富的过程推动的，这种误差就很容易理解。引导着预言的理论或者丰富的信息对于产生这些愿望的相当惊人的心理过程根本没有准确的信息。

对于这种看法，即推动愿望归属的心理机制使用了信息丰富的过程而非模仿，我们的第二个理由是：要理解这些机制所完成的工作如何**能**由模仿来完成，是很困难的。当然，就我们所知，对于这样一些愿望探测任务是怎样被完成的，模仿论者只是提出了一种建议，而这一建议明显是不可信的。这一得到戈登（1986）和戈德曼（1989）认可的建议首先提到了这样的事实，即像图 10—1 中所描述的这类模仿过程能用于作出行为预言，又指出这些过程也能用于产生关于这些愿望和信念——它们利用像综合分析战略这样的战略（最初是霍尔和斯蒂文斯［Halle and Stevens］在 1962 年的著作中针对意志识别而提出来的）能产生所观察到的行为——的信念。在用图 10—1 中的过程来预言行为时，假设的或“假装的”信念和愿望被输入心灵阅读者的决策系统中（当然，是“脱机”使用的），心灵阅读者预言：对象将做心灵阅读者根据这些信念和愿望决定要做的事情。在对关于愿望和信念的信念的产生所作的一种综合分析解释中，这个过程实际上是逆向进行的。它开始于一个行为情景，这个情景通过寻找假设的信念和愿望已经发生并继续进行，当这些信念和愿望被输入心灵阅读者的决策机制时，就会产生执行我们想要解释的行为的决策。

这种策略显而易见的问题是：产生它的候选者太多。一般来说，可能会有无穷多个信念和愿望的组合使心灵阅读者决定执行其所讨论的行为。

戈登完全意识到了这个问题，而他似乎认为他有一种解决办法：

> 不管我对这些假说作多长时间的检验，我都不会把关于［对象的］行为的所有候选解释都尝试一遍。如果我要作出决定的话，有些未经检查的候选者可能至少会和我接受的这种做得同样好，也可能其中有无限多个会这样。但是从直觉上我会说，这些“太牵强了”。在这里，我表现出了我惯性的偏见。对于追踪别人的行为，我必须做得越不“动人”（或者像演员们所说，“夸张”）就越好。我只在必要时，只在别人行为中的某种东西不适合时……才伪装，这种惯性偏见可能被看成是一种“努力最少”（least effort）原则：“假装最少原则”（principle of least pretending）。它解释了在其他情况相同时，我为什么偏爱更少地违背“真实的”世界——即违背我自认为的世界。（戈登，1986：164）

不幸的是，我们还完全不清楚戈登用一种反“动人”的惯性偏见所想
249 到的是什么。最明显的解释是：根据某种直觉的标准，归属越“牵强”，它们就离自己的心理状态越远。但如果这就是戈登的意图，那么似乎就可以断定这种看法是没有效果的。因为在很多情况下，我们解释行为是借助于和我们自己**完全**不同的愿望或者信念（或者两者兼有）。例如，我解释猫追老鼠，可能是借助于猫想吃老鼠的愿望。但是，有无限多的愿望会使我去追老鼠，这些愿望从直觉上说比吃老鼠的愿望更接近于我的真实愿望。对于缩小综合分析策略所产生的这个候选信念和愿望的无穷集合，模仿论者没有提出别的主张。如果对这个问题没有某种可信的解答，这种策略看起来就相当没有希望了。因此，在过去的十年间，这种解释在很大程度上会从模仿论的文献中消失，就不足为奇了。而这可能反映出模仿论者至少在私底下会承认：愿望归属只能借助于信息丰富的过程来解释。

10.4.3 矛盾的信念归属：另一种不能被模仿解释的心灵阅读技能

然而，心灵阅读的另一个重要方面是这种能力：归属我们自己并不具有的信念——**矛盾的信念**（discrepant beliefs），如人们有时称呼它们的那样。有很多过程都推动着矛盾的信念归属，有些依赖于关于对象的知觉状态的信念，有些则利用了关于对象的言语行为的信息，还有些依赖于关于对象的非言语行为的信息。我们猜测，所有这些都是由信息丰富的机制而非模仿的机制推动的。我们的理由大体上和我们为愿望归属所提出的理由

相似。首先，有大量证据表明，矛盾的信念归属系统表现出了我们预料这样一种信息丰富的系统会产生的系统误差，这个系统还不够丰富，它不含有与产生这些矛盾信念的范畴的过程相关的信息。其次，典型的模仿机制没有可信的办法来做矛盾的信念归属系统所做的事情。

对于矛盾的信念归属方面的系统性失败，一个让人忧虑的例子来自于对坚定信念（belief-perseverance）的研究。在心理学实验室和日常生活中，有时会发生这样的事情：一些很有说服力的证据（例如，试验结果）证明它们有某种迄今仍料想不到的特质。根据这些证据，人们一般会形成这样的信念：他们确实有这种特质。此后不久，假如提供给人们这样一种令人信服的情况，它使第一组证据受到了怀疑，那么这一信念会发生什么变化？例如，假定他们相信他们所依赖的试验结果实际上是其他人的伪造，或者说根本没有做真正的试验。大多数人的预测都将是：受到怀疑的信念将被彻底抛弃。这种观点是这样的一代社会心理学家共同具有的，他们骗被试相信关于他们自己的各种事情，这经常是通过做一些受操纵的心理学实验，观察被试的反应，然后通过解释骗术来向被试“说明情况”（debriefed）。这里的假设是：造成持久的伤害是不可能的，因为一旦骗术得到解释，所诱导的信念就会被抛弃。但在一系列引起广泛讨论的实验中，罗斯（Ross）及其合作者都证明了事实绝非如此。一旦某个被试相信他有一种特质，那么向他说明让他产生确信的证据完全是虚假的也并不能成功地消除那个信念（尼斯比特和罗斯，1980：175－179）。如果所讨论的这种特质是自杀倾向或者是“潜在的同性恋者”，那么坚定信念就会导致严重的问题。矛盾的信念归属系统的这个部分会使心理学家和其他所有的人认为，这些矛盾的信念在说明情况后就会被抛弃，这个部分对于坚定信念的过程明显有不准确的信息，因此它会导致完全错误的信念归属。

另一个对国家政策具有重要意义的例子是由洛夫特斯（Loftus，1979）等人的研究提供的，他们研究“事后干预”（post-event interventions）对人们关于他们所目击事件的信念的影响。在一个实验中，给被试看一段汽车事故的影片，很短的时间之后，再问他们一系列与事故有关的问题。对于有些被试，其中的一个问题是：“这辆白色跑车沿着乡间小路行驶，在经过谷仓时速度有多快?”问另一些被试的是：“这辆白色跑车沿着乡间小路行驶时的速度有多快?”一周后，所有被试都被问到这个问题：“他们是否看见了谷仓?”尽管在影片中并没有谷仓，但是被试如果被问到了提及谷仓的问题，他们就更可能相信他们看见了谷仓。在另一个实验中，这个

实验是在火车站和其他自然场景下做的，洛夫特斯和她的学生表演了一幕“盗窃案”：一个男同伙从两个女学生临时留下的包里拿走了一种东西，并把它塞到外套下面。片刻之后，其中一个女士注意到她的包被动过并开始大喊：“哦，我的天哪，我的录音机不见了!”她接着哭诉说录音机是老板借给她的，非常昂贵。此时，旁观者中的大多数人都非常合作，他们被要求留下电话号码，以便在保险索赔时能为这个意外事件作证。一周后，一位“保险代理人”致电目击者并询问偷盗的细节。其中一个询问的问题是：“你看见录音机了吗?”一半以上的目击者都记得看见了录音机，几乎所有人还能对它作出详细的描述——尽管事实是：**并没有录音机**。根据类似的实验，洛夫特斯得出的结论是：即使随意地提及并不存在的东西或者并未发生的事件（例如，在警察询问过程中）也能显著地增加这种可能性：这些东西或者事件将被混入人们关于他们所观察到的东西的信念。洛夫特斯研究的一个核心主题是：司法系统对依赖于目击者的证词应当非常
251 谨慎。至于司法系统为什么并**不**像它应当的那样谨慎，一个主要原因是：我们的信息驱动的矛盾信息归属系统缺乏像洛夫特斯所证明的事后信念形成过程的信息。

正如在愿望归属的情况里那样，我们看到，在任何可信的意义上，推动矛盾的信念归属的心理机制所做的工作都不**可能**由模仿来做。再者，模仿论者在这里所提出的唯一主张是综合分析解释，而这种策略对于信念归属而言并不比愿望归属更有效。

10.5 结　论

在上一节，我们概述了接受关于心灵阅读的混合解释的一些理由，根据这种解释，这种技能的某些方面是借助于信息丰富的过程解释的，另一些方面是由模仿解释的。尽管我们只考察了少数心灵阅读技能，但我们在别的地方（尼科尔斯和斯蒂克，即出）论证了完全相同的模式能够普遍存在。心灵阅读是一种复杂而有多个面向的现象，它的很多方面最好是由信息丰富的方案来解释，而其他方面最好是由模仿的方案来解释。如果这是正确的，那么它就不仅给功能主义者，也给取消主义者提供了一些相当棘手的选择。如我们看到的，功能主义者认为，普通心理状态术语的意义是由民间心理学决定的，而取消主义者一般也同意这一观点。在 10.2 节，我

们主张：如果民间心理学被看成是推动心灵阅读的信息丰富的理论，那么功能主义就是最可信的。但现在似乎只有**部分**心灵阅读依赖于信息丰富的理论。功能主义者是应当主张支撑心灵阅读这些方面的理论决定着心理状态术语的意义，还是应当退回到民间心理学的老生常谈解释？我们倾向于认为：不管功能主义者接受哪种选择，他们的理论的吸引力都会比其之前的要小。显而易见，老生常谈方案和心灵阅读方案会分道扬镳，只有部分的心灵阅读依赖于民间心理学。

【注释】

[1] 尽管我们将集中于刘易斯那些有影响的解释，但很多别的哲学家也提出了相似的观点，包括普特南（1960）、福多和齐哈拉（Fodor and Chihara，1965）和阿姆斯特朗（1968）。

[2] 但是要当心，在心灵哲学中，“功能主义”这个术语被用于指称多种观点。其中有些与我们刚才概括的这种观点有明显的家族相似性，有些则没有。相关研究的较好综述，参见莱肯（1994）和布洛克（1994）。

[3] 对这些争论的一种综述，参见斯蒂克（1996：第1章）和本书第2章。 252

[4] 这种区别是在斯蒂克和雷文斯克罗夫特（Stich and Ravenscroft，1994）的描述中第一次提到的。

[5] 但并非总是如此，如我们在10.4节会看到的。

[6] 当然，取消主义者不会同意我们在归属和预言心理状态或者依据过去的心理状态来解释行为等方面做得不错，因为他们认为我们所归属的心理状态并不存在。但他们并不否认：在研究者就其他人的心理状态所作的说明中存在令人印象深刻的**共识**，而这种共识是需要解释的。

[7] 例如，参见福多和莱泼尔（1992）。对于在意义理论方面的很多争论，相关有分量的综述参见德维特（Devitt，1996）。

[8] 罗伯特·戈登是这个观点最热心的支持者，即所有心灵阅读技能都能由模仿来解释。其典型的表述是：

> 民间心理学在一千年间没有发生大的变化，这是……不可思议的……丘奇兰德认为，这是证明民间心理学是一种错误理论的依据；事实上，这也可能证明它根本不是理论，至少不是这种公认意义上的理论，即（大致来说）一个未明确定义的一组术语的规律系统。相反，它可能只是实践推理的能力，并且得到了儿童时期和原始人的假装游戏能力的一种特殊用法的补充。

当然，取消主义者或许会提出反驳，说模仿论者是以尚待证明的假定为论据，因为对决策预言的模仿解释通常预设信念、愿望和民间心理学的其他假定是存在的。而取消主义者认为，这些常识的心理状态并不存在。为这种反驳建立一种可信的答复暂

时留给读者作练习吧。

[9] 然而，希尔（1998）认为，有一种关于模仿理论的解释认为，模仿论是先验地正确的。相关批评参见尼科尔斯和斯蒂克（1998）。

[10] 这个文献中的很多重要论文都已收录至戴维斯和斯通（1995a、1995b）的著作中。

[11] 我们也认为，心灵阅读的一些重要方面是由这样一些过程推动的，它们不能恰当地归入信息丰富的或模仿一类。但由于篇幅所限，我们在这里就不考察这种情况了，参见尼科尔斯和斯蒂克（即出）。

[12] 在这种实验中，最著名的是那些对所谓的**合取谬误**（conjunction fallacy）作出说明的实验。在一个著名的实验中，卡尼曼和特维斯基（Kahneman and Tversky，1982）给被试提出了下列任务：

> 琳达31岁、单身、坦率并且非常机灵。她主修哲学。当学生时，她非常关注歧视和社会正义问题，也参与了反核武器示威游行。
>
> 请根据其概率对下列陈述排序，用1表示可能性最大，用8表示可能性最小。
>
> (a) 琳达是一位小学教师。
>
> (b) 琳达在书店工作并教瑜伽课。
>
> (c) 琳达积极参与女权主义运动。
>
> (d) 琳达是一位精神科社会工作者。
>
> (e) 琳达是妇女选民联盟的一名成员。
>
> (f) 琳达是一位银行支票员。
>
> (g) 琳达是一位保险推销员。
>
> (h) 琳达是一位银行支票员并且积极参与女权主义运动。

253

在一群天真的被试中——他们在概率和统计学方面没有任何背景知识，尽管明显的事实是：除非一个人是一位**银行支票员**，否则她就不可能成为一位**女权主义的**银行支票员。但有89%的人判断陈述（h）比陈述（f）更有可能！当向精通统计学的被试——斯坦福商学院决策科学专业的研究生——提出相同的问题时，有85%的人会给出相同的答案！这种结果——在其中被试判断，复合事件或事态比它其中的某一个成分更有可能——自卡尼曼和特维斯基的开创性研究以来多次被发现，而且它们是非常可靠的。对于启发法和传统的偏见研究所作的有用评论，参见卡尼曼等（1982）、尼斯比特和罗斯（1980）、巴伦（Baron，2001）和萨缪尔斯等（Samuels et al.，2003）。

[13] 对有关文献所作的精彩评论，参见罗斯和尼斯比特（1991）。

参考文献

Armstrong, D. (1968). *A Materialist Theory of the Mind*. New York: Humani-

ties Press.

Baron, J. (2001). *Thinking and Deciding*, 3rd edn. Cambridge: Cambridge University Press.

Bierbrauer, G. (1973). *Effect of Set, Perspective, and Temporal Factors in Attribution*, unpublished doctoral dissertation, Stanford University.

Block, N. (1994). "Functionalism." In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell: 323–332.

Churchland, P. (1981). "Eliminative Materialism and Propositional Attitudes." *Journal of Philosophy*, 78: 67–90. Reprinted in W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell (1990): 206–223. Page reference is to the Lycan volume.

Davies, M. and Stone, T. (1995a). *Folk Psychology*. Oxford: Blackwell.

—— (1995b). *Mental Simulation*. Oxford: Blackwell.

Devitt, M. (1996). *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. Cambridge: Cambridge University Press.

Ekman, P. (1985). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York: W. W. Norton and Co.

Fodor, J. and C. Chihara (1965). "Operationalism and Ordinary Language." *American Philosophical Quarterly*, 2 (4). Reprinted in J. Fodor, *Representations*. Cambridge, MA: MIT Press (1981): 35–62.

Fodor, J. and LePore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Gigerenzer, G., Todd, P. and the ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.

Goldman, A. (1989). "Interpretation Psychologized." *Mind and Language*, 4: 161–185.

Gopnik, A. and Meltzoff, A. (1997). *Words, Thoughts and Theories*. Cambridge, MA: MIT Press.

Gopnik, A. and Wellman, H. (1994). "The Theory-Theory." In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press: 257–293.

Gordon, R. (1986). "Folk Psychology as Simulation." *Mind and Language*, 1: 158–170. Reprinted in Davies and Stone (1995a). Page reference is to the Davies and Stone volume.

Halle, M. and Stevens, K. (1962). "Speech Recognition: A Model and a Program for Research." In J. Fodor and J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.

Harris, P. (1992). "From Simulation to Folk Psychology: The Case for Development." *Mind and Language*, 7: 120-144.

Hayes, P. (1985). "The Second Naive Physics Manifesto." In J. Hobbs and R. Moore (eds.), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex: 1-36.

Heal, J. (1986). "Replication and Functionalism." In J. Butterfield (ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press: 135-150.

—— (1998). "Co-cognition and Off-line Simulation: Two Ways of Understanding the Simulation Approach." *Mind and Language*, 13: 477-498.

Hempel, C. (1964). "The Theoretician's Dilemma: A Study in the Logic of Theory Construction." In C. Hempel, *Aspects of Scientific Explanation*. New York: The Free Press: 173-226.

Kahneman, D. and Tversky, A. (1982). "The Psychology of Preferences." *Scientific American*, 246 (1): 160-173.

Kahneman, D., Slovic, P., and Tversky, A. (eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Lewis, D. (1970). "How to Define Theoretical Terms." *Journal of Philosophy*, 67: 17-25.

—— (1972). "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, 50: 249-258. Reprinted in N. Block (ed.), *Readings in the Philosophy of Psychology*, vol. I. Cambridge, MA: Harvard University Press: 207-215. Page references are to the Block volume.

Loewenstein, G. and Adler, D. (1995). "A Bias in the Prediction of Tastes." *The Economic Journal: The Quarterly Journal of the Royal Economic Society*, 105: 929-937.

Loftus, E. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Lycan, W. (1994). "Functionalism." In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell: 317-323.

McCloskey, M. (1983). "Intuitive Physics." *Scientific American*, 248 (4): 122-129.

Milgram, S. (1963). "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*, 67: 371-378.

Nichols, S. and Stich, S. (1998). "Rethinking Co-cognition: A Reply to Heal." *Mind and Language*, 13: 499-512.

—— (forthcoming). *Mindreading*. Oxford: Oxford University Press.

Nichols, S., Stich, S., Leslie, A., and Klein, D. (1996). "Varieties of Off-line Simulation." In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press: 39–74.

Nisbett, R. and Ross, L. (1980). *Human Inference*. Englewood Cliffs, NJ: Prentice-Hall.

Putnam, H. (1960). "Minds and Machines." In S. Hook (ed.), *Dimensions of Mind*. New York: New York University Press: 138–164.

Ross, L. and Nisbett, R. (1991). *The Person and the Situation: Perspectives of Social Psychology*. Philadelphia: Temple University Press.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Samuels, R., Stich, S., and Faucher, L. (2003). "Reasoning and Rationality." In I. Niiniluoto, M. Sintonen, and J. Wolenski (eds.), *Handbook of Epistemology*. Dordrecht: Kluwer: 1–50.

Scholl, B. and Leslie, A. (1999). "Modularity, Development, and 'Theory of Mind'." *Mind and Language*, 14: 131–153.

Sellars, W. (1956). "Empiricism and the Philosophy of Mind." In H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science*, vol. 1. Minneapolis: University of Minnesota Press: 253–329.

Stich, S. (1996). *Deconstructing the Mind*. Oxford: Oxford University Press.

Stich, S. and Nichols, S. (1995). "Second Thoughts on Simulation." In Davies and Stone (1995b): 86–108.

Stich, S. and Ravenscroft, I. (1994). "What is Folk Psychology?" *Cognition*, 50: 447–468. Reprinted in Stich (1996).

Thaler, R. (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1: 39–60.

Young, A. (1998). *Face and Mind*. Oxford University Press.

第11章　个体主义

罗伯特·A·威尔逊（Robert A. Wilson）

11.1 导　言

256 在过去的大约25年间，围绕个体心灵的本质与个体以及其心灵所处世界的特征之间的关系这一总问题，心灵哲学中产生了大量的讨论。有时候这里的基本问题是根据心理学或者心理的状态是否“在头脑中”来解释的，但对于不识此道者来说，可能这听起来像是争论一个莫名其妙的问题：心理状态当然在头脑中！（参见罗兰兹［Rowlands］，1999；威尔逊，2000a、2001）因此，我们最初的任务是对这个问题的一种版本作出清晰的阐述，这一版本更清晰地说明了为什么这个问题成为一个长期争执不休的问题，它首先阐述了这个问题对于很多其他不同问题的重要意义，例如认知科学的方法论、自我知识的可能性以及意向表征的本质等。

现在考虑这个问题：自主体的环境特征是否在决定或者确定该自主体心灵的本质方面起着某种关键作用。一种自然的想法——它是不同意上述对该问题的回答的人所共有的——是：自主体与他们的世界有因果的相互作用，他们通过感官收集世界方面的信息，因此他们心灵的本质，尤其是他们的思想所关于的东西，部分地**是**由他们世界的特征决定的。因此，世界是人的思想，从而是人心灵的一种**因果**决定因素。对于人心灵的**内容**，对于人所知觉和思考的东西，世界是一种促成性的原因（contributing cause）。这就是说：一个人心灵的内容和他的环境从因果上说并不是相互孤立的。在心灵哲学中，区别个体主义者和反个体主义者的是对这个问题的回答：是否在某种**更深的**意义上，心灵的本质是由个体世界的特征决定的。

257 我们可以这样思考这个问题：扩展关于这一思想（即心灵的内容部分

是由自主体的环境因果地决定）的简短讨论，以此来考察世界方面的不同在什么条件下暗示了心灵方面的不同。个体主义者认为，之所以会如此，只是由于世界方面的不同使个体内部发生的事情出现了某种相应的变化；反个体主义者否认这一点，因此他们承认有另一种可能性，即在其所有内在特征方面都等同的个体所具有的心理学或心理的状态可能有不同的内容。而且，假如内容不同的心理状态**本身**就是不同类型或种类的状态，那么这就意味着一个个体的内在属性并不决定或确定该个体的心理状态。

这为我们提供了另一种理解问题的方式，而且是一种更准确的方式，以说明关于心灵的个体主义（或**内在主义**）与对它的否定即反个体主义（或**外在主义**）之间的区别。个体主义者主张而外在主义者否认的是：某一个体身体内发生的事情在**形而上学**上决定着该个体心理状态的本质。该个体主义决定论题与因果决定论题不同，它表达了一种关于心灵状态之本质或实质的观点，并点明了这样一层含义：心理状态尽管是由世界的状态因果地决定的，但它们是自主的，或者说是不依赖于个体之外的世界的特征的。个体主义要说的意思不外乎是：所有内在方面都完全相同的两个个体，**一定会有**相同的心理学状态。由于存在关于**孪生人**（Doppelganger，即分子层次同一的个体）的想象，以及有关于孪生地球的相应想象，个体主义的这种含义以及围绕个体主义的争论常常会变得更为明朗起来。下面，我将通过概述关于个体主义争论的历史而转向这两种想象。

11.2 到达孪生地球：什么在头脑中？

希拉里·普特南的《意义的“意义”》（The Meaning of “Meaning”，1975）一文在讨论自然语言术语之意义的语境下引进了两种想象。他是想证明“意义”不是并且不能共同满足两个论题——那时关于自然语言指称的流行观点一般认为它能满足以下两个论题：一个论题主张术语的意义决定其指称，另一个主张意义就“在头脑中”。这里的这个短语应该这样理解：它提出了一个关于上述意义的形而上学决定的类型同一的主张。这些论题代表着**描述的**指称理论，它们自弗雷格和罗素首次提出以来就占据着突出位置，根据这些论题，一个术语的指称是由说话者加到这个术语上的描述确定或者说从形而上学上决定的。举个经典的例子，假如我认为亚里

士多德是一位已故的伟大哲学家，他写了很多重要的哲学著作，例如《尼各马可伦理学》（*Nicomachean Ethics*），而且他是柏拉图的学生，是亚历
258 山大大帝的老师。那么，根据描述主义的指称观，我的术语“亚里士多德”指称的就是世界上任何满足我加给这个术语的各种描述的东西：它是世界上的这种东西——一位伟大的哲学家，已经去世了，写了很多重要的哲学著作（例如，《尼各马可伦理学》），是柏拉图的学生，是亚历山大大帝的老师。

这些关于**专名**指称的描述主义观点是索尔·克里普克在其有影响力的《命名与必然性》（*Naming and Necessity*，1980）一书中批判的焦点，他那时批判的是上述这组观点及其预设，而普特南关注的是**自然种类术语**，例如“水”和“老虎”。克里普克和普特南都想让他们的批判及随后的自然语言指称理论的替代者（即因果指称理论）变得非常一般且提供替代性方法，从而思考语言与世界的关系。但是，请允许我们继续深入考察普特南的论证，并揭示它与个体主义的关系。

考虑一个平凡的居民奥斯卡，他生活在地球上并和我们大多数人一样和水发生相互作用：他喝它，用它洗漱，看见它作为雨从空中落下。奥斯卡对于水的本质并没有专门的知识，因此他将很多描述与他的术语“水”联系在一起：它是人可以喝的一种液体，它被用于洗漱，它作为雨滴从空中落下。根据描述的指称观，这些描述决定着奥斯卡的术语“水”的指称。也就是说，奥斯卡的术语“水”指称世界上任何满足了他加给这个术语的这组描述的东西。由于这些描述是“在头脑中”的，因此，根据这种观点，自然语言的指称就是个体主义的。

让我们继续看普特南的论证。现在想象奥斯卡有一个分子层次的孪生人奥斯卡*，他生活在这样一个行星上，除了某一方面之外，这个行星的所有方面都和地球完全相似：人们喝的、用以洗漱的和看见从空中落下的物质**不**是水（即 H_2O），而是一种具有不同化学结构的物质 XYZ。我们姑且把这个行星称作“孪生地球”，这种物质的化学成分我们可以用“XYZ”来表示，它在孪生地球上也被称为“水”，而且奥斯卡*作为奥斯卡的孪生人或者孪生兄弟，他对这种物质具有和奥斯卡对地球上的水相同的信念。（回想一下，奥斯卡以及他的孪生人奥斯卡*对于水的化学结构都没有专门知识。）孪生地球上有我们所称的“**孪生水**”（twin-water）或“**孪水**”（twater），而没有水；奥斯卡*与之发生相互作用的是孪水，而不是水——毕竟在孪生地球上是没有水的。假如奥斯卡的术语“水”指称或者是关于

水，那么奥斯卡*的术语“水”就指称或是关于**孪水**。也就是说，假如指称至少是意义的一个方面，那么他们所具有的自然语言术语就有**不同**的意义。但是根据假设，奥斯卡和奥斯卡*是孪生人，因此他们的所有内在属性都是同一的，因此他们“头脑中”的东西也应是同一的。这样，普特南就推断：奥斯卡所用的自然语言术语的意义不是由奥斯卡头脑中的东西从形而上学上决定的。

普特南批判的对象是一个思考语言的传统，用普特南从鲁道夫·卡尔纳普的《世界的逻辑结构》（*The Logical Construction of the World*，1928）中借用的术语来说，这个传统就是**方法论的唯我论**（methodologically solipsistic）：它在处理自然语言术语以及（更一般地说）语言的意义时，并不假设个体的语言使用者之外的世界是存在的。既然普特南的主要论点涉及自然语言术语以及它们的语义学与头脑中的东西的关系，那么一个人要得出一种关于心灵本身的个体主义的否定立场，就至少需要将他的推理从语言扩展到思想。但是考虑到他所反对的传统，这种扩展也许会被认为微不足道，因为其实方法论唯我论传统中的哲学家——从布伦塔诺到罗素、胡塞尔，再到卡尔纳普——都是用心理学术语来想象自然语言及其用法的。

对“个体主义”这个术语本身的介绍可以在泰勒·伯奇的《个体主义与心理》（Individualism and the Mental，1979）一文中找到。伯奇在这篇论文中提出了一系列思想实验，它们在很多方面都与普特南的孪生地球思想实验相似。伯奇把个体主义看作至少从17世纪的笛卡尔以来就在现代哲学思想中流行的一种总体的心灵概念，他认为我们解释行为的常识心理学框架（即我们的**民间心理学**）不是个体主义的。重要的是，伯奇明确提出了一个反个体主义的案例，这个案例并不依赖于可能有争议的关于自然种类术语之语义学的主张——他提出其反个体主义案例，用的是具有关于关节炎、沙发和合同的思想的自主体——因此，他的论证并不预设任何一种关于自然种类的科学本质主义。然而，就像普特南的论证一样，伯奇的论证也确实预设了一些关于自然语言理解的观点。

最核心的论题是：对于很多东西——我们有关于它们的思想，也有表达它们的自然语言术语，我们能够而且确实有**不完备的**理解。假如一个人有可能具有一些依赖于这种不完备的理解的思想，如一个人的大腿肌肉组织得了关节炎（关节炎是只出现在关节上的一种病）。或者如我们所说，“关节炎”在我们的语言共同体中只适用于关节疾病。考虑一个人伯

特（Bert），他具有由他说“我大腿上得了关节炎”来表达的信念。在现实世界上，这是一个关于关节炎的信念；只是伯特对关节炎的本质有一种不完备的或者说部分错误的观点，因此他用相应的句子表达了一个错误的信念。

但是，现在设想伯特生活在一个不同的语言共同体中，在其中“关节炎”这个术语**确实**既适用于关节的疾病也适用于身体其他部位（包括大腿）的疾病。在**这个**语言共同体中，伯特的思想就不会存在它在现实世界上会涉及的不完备理解；实际上，他在这个世界上的思想是**正确的**。假如这两个语言共同体有这些不同，那么一个具有涉及他所称的“关节炎”的思想的人在这两个共同体中似乎就会有不同的思想：在现实世界上，伯特有关于关节炎的思想；而在反事实的世界中，他有关于**另外的病**的思想。**我们**可以称之为“大腿关节炎”（tharthritis）[①]，以便将它与我们在现实世界上所患的病区别开。原则上，我们可以认为伯特本人在这两个语境下是同一个人——也就是说，他在所有内在方面都是同一的。然而，我们向伯
260 特归属了具有不同内容的思想，而且这样做似乎**只是**由于他所处的语言共同体不同。因此，人的思想内容就不是由个体的内在属性从形而上学上决定的。再次，如果两种思想在内容方面的不同蕴含思想本身的不同，那么这就蕴含思想不是个体主义地个体化的。

有时在普特南的反个体主义观点和伯奇的反个体主义观点之间所作出的一种对比（例如，西格尔［Segal］，2000：第 2～3 章），是将普特南的观点描述成一种**物理的**外在主义，而把伯奇的观点描述成一种**社会的**外在主义：根据普特南，物理世界的特征（例如，水自身的本质）部分地从形而上学上决定着人类心灵的内容；而根据伯奇，是社会世界的特征（例如，一个人的语言共同体的本质）决定着心灵的内容。虽然这种区别有助于人们想到这两个观点相区别的一个方面，但我们也应当谨记普特南自然语言观也有“社会的”方面：语言的劳动分工。就这两种观点来说，重要的是这一观念：语言使用者和心理存在物相互之间依靠和信赖的情况能够反映在我们关于语言和思想的日常的常识思维方式之中。因此，根据这两种观点，意义的本质和思想的本质都有着社会性的一面，这就部分地证明反**个体主义**的名称适合其中的每一种观点。

① 这是作者杜撰的一个词，是取 thigh（大腿）一词的前两个字母与 arthritis（关节炎）一词组合而成的，姑且译为“大腿关节炎”。——译者注

11.3 认知科学的姿态

自认为对认知科学有贡献的哲学家在个体主义者与外在主义者的争论中占据着最活跃的舞台。大约在个体主义受到普特南和伯奇的抨击之时，它也得到了辩护，它被看成是一种特别适合于对心灵作出真正科学理解的心灵观，尤其是在认知科学这个新兴的跨学科领域中所陈述的那些观点。对于那些提出辩护的人来说，普特南—伯奇论证有些可疑的非自然主义的成分，他们的结论也似乎有一些反科学的东西，对个体主义的一部分辩护和对外在主义的相应抨击都取决于我所称的**认知科学的姿态**（the cognitive science gesture）——如当代经验的认知研究所指出的，关于心灵的任何真正科学的理解都一定是个体主义的。

在熟悉了普特南的“方法论的唯我论”的用法之后，杰里·福多就将方法论的唯我论辩护成了这种学说：心理学本身应当只考虑**窄**（narrow）心理状态。在这里，这些状态并不预设“任何个体的存在，除被归属了那种状态的主体之外”（福多，1980：244）。福多把方法论的唯我论看成是思考心理状态的首选方法，尤其是考虑到语言学中的乔姆斯基革命和心理学中相伴随的计算革命。如果心理状态由计算规则控制的转换，那么认知科学的任务就是说明这些规则；就心理状态是计算的这一判断来说，对个体所处的物理的或社会的世界所作的更宽广的考虑似乎与个体的心理学本质是不相干的。

斯蒂芬·斯蒂克（1978）的自主性原则为阐述个体主义的认知科学观提供了另一种方法，它的一些变化形式已经成为阐述个体主义的标准方式。这个原则说：“心理学家应该关注的状态和过程是那些随附于有机体当下的、内在的物理状态上的状态和过程。”（斯蒂克，1983：164－165）**随附性**概念为说明我们前面介绍的那种形而上学决定提供了一种更准确的方法。一组属性 S（随附的属性）随附于另一组属性 B（基础属性），仅当有某种东西，它在 B 属性方面相同一，一定也会在 S 属性方面相同一。这不仅部分地是由于随附性在阐述物理主义的诸版本中的重要地位，以及被知觉到的物理主义与个体主义之间的联系（有时候还不止这些），而且，还部分地是由于普特南和伯奇的论证中对孪生人的强调。依据这一种或那一种随附性阐述来表达个体主义及其否定，就变得最有代表性了。

福多的认知科学观和斯蒂克的认知科学观都包含有这一思想：一个个体的心理学状态应该与个体发现他们自己所处的纯粹的头脑之外的环境**隔开**。与前文所述的普特南和伯奇不同，福多和斯蒂克关注个体主义与心理学方面的解释实践的相关性，他们用各自的原则支持关于心理学和认知科学的范围和方法论的实质性结论。福多将唯我论的心理学与他所说的自然主义心理学进行了比较，认为由于后者（其中包括吉布森［J. J. Gibson］关于知觉的方法、学习理论和威廉·詹姆斯的自然主义）不可能证明心理学中一种可靠的研究策略，因此方法论的唯我论就为理解认知提供了唯一富有成效的研究策略（也参见福多，1987）。斯蒂克支持一种句法或者计算的心灵理论，它在本质上根本并未使用意向性或心理内容概念，因此它在辩护关于内容的**取消主义**观点时使用了自主性原则（也参见斯蒂克，1983）。

尽管我认为认知科学姿态**是**一种姿态（而不是一种借助于经验实践的可靠论证），但它不是一种**空洞的**（empty）姿态。虽然福多和斯蒂克的论证在哲学界或认知科学界还未得到普遍认可，但他们已经触动了认知科学研究者的心弦，毫不奇怪，这可能是由于在认知科学中占支配地位的研究
262 传统是个体主义的。对于认知科学诸方面的个体主义观点的承诺，比较明确的陈述包括乔姆斯基（1986、1995、2000）所使用的两个语言概念（“I”语言和“E”语言，分别表示“内部的”和“外部的”）之间的区别，杰肯道夫（Jeckendoff，1991）在“心理学的”和“哲学的”心灵概念之间所作的区别，以及科斯米德斯和图比（Cosmides and Tooby，1994）对我们内部的、进化—特化认知模块的构成本质之强调。

对于实践的认知科学家来说，个体主义的一部分吸引力是其被知觉到的与心灵的表征理论的联系。心灵的表征理论认为，通过关于世界是怎样的（作为知觉的结果）或者世界应当怎样（作为行动的指令）的内部心理表征，我们与世界发生知觉的和行为的相互作用。杰肯道夫在表达这种观点时说：

> 不管真正实在的本质是什么，在我们看来，实在都是由我们的内部心理表征的本质所决定和限制的……物理刺激（光子、声波、皮肤上的压力、空气中的化学成分，等等）对感觉神经元发生机械作用。感觉神经元充当了派利申（1984）意义上的传感器，它们建立了外围层次的表征，如视网膜阵列和耳朵所产生的任何听觉分析。反过来，

外围的表征刺激着更中心层次的表征的构成，最终导致了像三维层次模型这样的中心格式中的表征形成。(1991：159-161)

只要有机体这些合适的内部表征状态保持不变，有机体与其环境的更外围的因果关系就和认知是不相干的，因为这种因果关系要想对认知有重要影响，只能通过改变表征该环境的内部心理状态。

11.4 功能主义、物理主义和个体主义

就很多对认知科学感兴趣的哲学家来说，个体主义吸引人是由于这种观点与心灵哲学中的物理主义和功能主义之间有一种关系已被发现，后两者自20世纪80年代以来就被普遍接受了。物理主义（或者说唯物主义）是一种有多种表达方式的观点，其中最常见的一种可能是根据随附性概念：一切事实、属性、过程、事件和事物都随附于物理的事实、属性、过程、事件和事物，因为它们是基础物理学所假定的。对物理主义的这种本体论阐述（涉及何物**存在**）常常伴随着一个**解释**论题，其称：物理解释在某种意义上是对任何现象的终极解释。

个体主义被认为是和物理主义联系在一起的，因为它凭借随附性阐述 263
不外乎是声称：如果没有个体在内在物理状态方面的相关不同，就不会有心理学的不同。那些反对个体主义的人有时所受的指责（例如，由布洛克和福多所提出的指责），是说他们认可了某种形式的心灵二元论，或者说由于他们忽视或误解了因果力在心理学分类中的作用而使心理因果性成了一个谜。如果把这与对认知科学自身有影响的方法论构想联系起来，我们就可以认为，个体主义是对获得这些心理学解释——它们将心灵恰当地定位在物理世界之中——的一种最小约束，是一种依据因果力对实体进行分类的心理学。（然而，我们已经看到，个体主义者在其对心理学意味着什么的方面意见并不一致。）

功能主义者相信，心理学状态和过程应当根据它们的因果作用或功能作用以实现个体化——也就是说，根据它们在有机体的整个因果组织中的位置——而且这种看法也是很常见的：这些功能作用或因果作用是个体主义的。当然，这些因果作用可以用不同的方法来理解，但有两种（互补的）方法在认知科学中最流行——一种方法是根据计算概念（例如，福

多，1980；派利申，1984）；另一种方法是根据分析性的分解（analytical decomposition）的观念（例如，丹尼特，1978；卡明斯［Cummins］，1983），这两种方法同意一种个体主义的解释。计算过程只对心理状态的**句法**属性起作用，它们已经被合理地看成是个体主义的；人们自然会认为分析性的分解是始于一种心理能力（例如，记忆、深度知觉、推理）并寻找有机体借以例示这种能力的内在属性。

然而，尽管它们初看起来是可信的，但是，物理主义与个体主义之间以及功能主义与个体主义之间的关系都不是没有问题的，我认为，实际上这两个所谓的推理在仔细考察之后都是无效的。通过考察明确的个体主义论证——这些论证更准确地说明了这些关系，我们就能对这些主张作出更充分的考察。

11.5 诉诸因果力

一个被广泛讨论的个体主义论证来自于福多《心理语义学》（*Psychosemantics*，1987）的第二章。虽然在我看来，一系列相关的批评（范久利克，1989；伊根［Egan］，1991；威尔逊，1992、1995：第 2 章）在证明这个论证有致命缺陷方面是决定性的，但这个论证本身却也利用了一个或者也许是一系列在哲学界影响深远的直觉。这个论证本身容易叙述，一般科学中的分类方法或者个体化能满足一般化的个体主义版本：科学对其假定和发现的实体进行分类是依据它们的因果力。心理学和认知科学在这方
264 面也应当毫不例外。但是，任何东西的因果力都随附于其内在的物理属性之上。因此，科学的分类方法（从而心理学的分类方法）一定是个体主义的。

为了引出这个论证的问题，一种方法是问：是什么使得第一个前提（关于一般的科学分类方法）为真。假如人们猜想的自然主义转向被当代哲学心理学的研究者接受了，那么就可以认为，这里的支持来自对各个科学领域的实际分类实践所作的考察。然而，一旦一个人真的转向考察这些实践，他就能轻松地发现：各种科学分类并**不是**“根据因果力”；相反，它们是**从关系上**（relationally）对其自身的种类进行分类的，在这里，决定种类成员身份的**实际**关系就够了。这里经常援引的例子包括进化生物学中的物种——它们是从种系发生学上（因而从历史上）个体化的——和地

质学中的大陆，它们的因果力与它们作为大陆的角色是完全不相干的（参见伯奇，1986a）。这个问题在这一支持个体主义论证的语境下尤其严重，因为论证中的另一个前提说：一种东西的因果力随附于其内在属性，因此一个人不能只是规定：在“因果力”的某种引申意义上，这些科学中的个体化是“根据因果力”。（如果一个人这样做，那么“因果力”就不再保持这样的随附。）

尽管人们承认这一论证本身在上述确认的方面是有缺陷的，但这种直觉仍然存在：个体主义**确实**清楚地表达了认知解释的一个约束条件，它是更一般的科学能满足的，这个直觉会导致一种从物理主义上说值得尊重的心理学（例如，参见沃尔什［Walsh］，1999）。我的观点是：这个直觉本身严重低估了各种科学中的分类和解释实践的多样性（参见威尔逊，2000b），它必须被彻底放弃。要想复活这种个体主义论证，就必须对各种关于科学的分类和解释之本质作出先验假设，这些假设会使人想起对控制逻辑实证主义的科学观的科学实践所作的普遍合理重构，而这在任何自称当代自然主义心灵哲学家们听起来都应当是警钟。

11.6 外在主义与形而上学

那么，物理主义与个体主义之间公认更一般的关系是什么？如果能证明对个体主义的否定要求否定一个可信的、一般的物理主义版本，那么我认为外在主义本身就真的有麻烦了。但是，就像个体主义借助于因果力和科学的分类法一样，我怀疑从促成这个论证的一般直觉到这个论证本身的这种举动本身都将被证明是成问题的。例如，外在主义者能够尊重这句物理主义口号：“没有物理的不同就没有心理学的不同。”因为相关的物理不同处于个体之外；对这句口号进行改进的尝试（例如，没有此时此刻的物 *265*
理的不同就没有心理学的不同）可能要么会以待证明的假定为论据来反对外在主义者，要么会利用这种物理主义解释，它至少和个体主义本身一样存在争议。

实际情况是：外在主义者自身并不像他们原本可以的那样注意到处于当代唯物主义核心的形而上学概念，而当他们这样表示关注时，他们听起来就是反物理主义的。这里最著名的案例是伯奇（1979）对个体主义关于相关心灵观点的意义所作的开创性讨论，他主张：反对个体主义不外乎是

说既反对关于心灵的类型同一论，也反对关于心灵的个例同一论，它们是两种主要的唯物主义形式。

在我看来，这些概念中讨论最不深入的是**实现**概念。尽管它通常这样表达唯物主义，即唯物主义要求所有心理状态都是作为物理状态而实现的，而且它把相关的物理状态看成是大脑状态，但对这种实现关系的属性或者说对实现者状态的属性还鲜有一般性的讨论（不过，参见休梅克，2000；吉勒特［Gillett］，2002）。这就向外在主义者提出了一个问题，因为关于实现的标准观点偷偷掺杂了一种个体主义偏见。根据这种标准观点，实现者被认为不仅对于它们所实现的状态**在形而上学上是充分的**，而且还**物理地构成了**具有所实现的这些属性的个体。如果否定第二个合取支（如我认为个体主义者应该做的那样），就会为这种观念创造空间：心理状态有一种宽实现，这是我在其他地方和更一般地讨论实现的语境下试图辩护的一种选择。

11.7 关于马尔视觉理论的争论

我说过：乍看起来，个体主义得到了关于心灵的计算理论和表征理论的支持，从而得到了受到那些理论影响的认知科学界的支持。但是我也指出：真正解释性的认知科学是个体主义的这种主张有一种认识性的基础，它更像一种姿态而非证明。如果想根据第一种观点来证明第二种，有种方法是去考察这场持续的哲学争论，它争论的是：大卫·马尔（David Marr）著名的初级视觉理论是否是个体主义的？除了争论本身的内在兴趣之外，我们这里的考察也有助于引出与心灵有关的一些更广泛的问题，个体主义问题处于其中的核心位置，这些问题包括计算和表征的本质问题。

在《个体主义与心理》一文的最后一节，伯奇指出：他的思想实验和他所得出的结论——心理内容从而有内容的心理状态不是个体主义的——
266 对于认知的计算解释是有影响的，而且这些影响是双重的。首先，对心灵的纯计算解释——从个体主义理解的——是不合适的；其次，就这些解释的确诉诸心理内容概念来说，它们也不是个体主义的。伯奇在《个体主义与心理学》（Individualism and Psychology，1986a）一文中所探讨的就是后一种思想，他在文中惊人地主张：马尔的视觉理论**不**是个体主义的。这是考察一种认知科学内部与个体主义相关且受到普遍尊重的观点的第一

次尝试，而且在超越认知科学姿态，马尔走向一种真正利用认知科学自身的经验实践的论证方面，因此这也是一个关键的转折点。

如人们通常所指出的，所谓的“马尔视觉理论”是对初级或者说“低层次”视觉中一系列过程的解释，它是从20世纪70年代中期开始由马尔及其同事，如埃伦·希尔德雷思（Ellen Hildreth）和托马斯·波吉欧（Tomas Poggio）在麻省理工学院（Massachussetts Institute of Technology）发展而来的。这些过程包括立体视觉、运动知觉、形状和表面知觉以及明显的计算方法。马尔的《视觉：一种对人类的视觉信息的表征和加工的计算研究》（*Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*，1982）一书是在1980年马尔不幸早逝之后才出版的，它成为了这种方法的代表性意见，尤其是对哲学家来说，由于马尔将计算的细节与他对于理解视觉的前景和影响的粗线条的、提纲挈领的陈述很好地结合了起来，因而很容易理解。自马尔的书出版以来，对他的视觉理论的研究就在持续进行，并延伸到了构成更广阔的低层次视觉的领域（例如，参见希尔德雷思和厄尔曼[Hildreth and Ullman]，1989）。有意思的是，借助马尔理论的个体主义方面的哲学文献在解释这种理论时大都几乎无一例外地愿意依靠他的《视觉》一书。

马尔所倡导的计算理论的关键，是认识到人们能够——其实对马尔来说，是一定——在不同的层次上对视觉进行研究。根据马尔的看法，在研究信息加工装置时，可以从三个层次上来分析。首先是计算理论的层次（下面称**计算层次**），它说明了计算的目标，在这个层次上这种装置本身被抽象的形式术语描述成“从一种信息向另一种信息的映射”（1982：24）。其次是表征和算法的层次（下面称**算法层次**），它选择“输入和输出的表征和用来将一种表征转换成另一种表征的算法”（1982：24-25）。第三是硬件执行的层次（下面称**执行层次**），它向我们说明了表征和算法在实际装置中是如何物理地实现的。

哲学讨论和马尔本人的讨论一样，都集中在视觉的计算层次和算法层次，马尔本人（1982：23）把它们分别描述为关于视觉“是什么和为什么”（what and why）的问题和“怎么样”（how）的问题。如我们将看到的，对于计算层次涉及什么尤其有争议。除了经常引用的进行信息加工分析的三层次结构之外，对于马尔的视觉方法还有另外两个有意思的维度，它们在哲学文献中有些被忽视了。这些维度不仅为马尔的理论，而且为 267

如何在这个理论中理解“计算”和“表征”的问题增加了某种复杂因素。

第一个维度是这种观念：视觉计算是在**计算推理的舞台**上依次进行的。马尔说：视觉理论的总目标是“要理解如何从世界的图像中有效而可靠地获得关于世界的描述”（1982：99）。马尔认为，从视网膜图像的亮度变化到成熟的三维描述的推理是这样进行的，即通过建立一系列初级表征：原始要素图、完全要素图以及 2.5 维图。我们把这个维度称为视觉计算的**时间性维度**（temporal dimension）。第二个维度是视觉加工受制于模块性设计，由此三维图像结构的特定方面——立体视觉、深度、运动，等等——原则上能进行独立地研究，我们把这个维度称为视觉计算的**模块性维度**（modular dimension）。

对视觉计算的时间性维度和模块性维度的承认，会使关于“这个”视觉过程的“这些”计算层次和算法层次是什么的讨论变得复杂。从最低限度上说，在辨认马尔的三个层次时，为了分析一个给定的视觉**过程**，我们至少必须首先确定模块性维度，而为了分析一个给定的视觉**计算**，我们又必须至少确定时间性维度。

伯奇认为马尔的理论不是个体主义，他在如下段落中对这一观点作出了明确而充分的表达：

> （1）这种理论是意向的。（2）这种理论的意向基元（primitives）和它们所携带的信息是参照偶然存在的物理项目或者在正常情况下它们借以被引起的条件和它们所适用的条件进行个体化的。（3）因此，如果这些物理条件以及（可能）伴随的物理规律是不同的，那么传送给主体的信息和他的视觉表征的内容就不同。（4）说相关的不同（比如说，光学的）规律在主体的眼睛和神经系统中……通常会引起同一些从非意向的、个体主义的角度进行个体化的物理规律，这种看法并非不融贯的。（5）在这种情况下［根据（3）］，这个人的视觉表征会携带不同的信息，有不同的表征内容，但这个人的整个非意向的物理历史……也许会保持不变。（6）假设有些知觉状态在这个理论中是依据它们的信息或意向的内容来辨认的，那么由此就能推出：个体主义对视觉理论是不成立的。（1986a：34）

第二个和第三个前提对于马尔的视觉理论提出了具体主张，而第一个前提连同第四个和第五个表明：这个论证和我们前面所讨论的伯奇的孪生

地球式论证具有相似性。

伯奇本人的注意力集中于为（2）—（4）作出辩护，马尔在叙述视觉的计算层次和算法层次时，对他所诉诸的方面似乎依赖于“现实世界的结构”。在《视觉》的各个部分，马尔无疑都大量地借助于这种结构。例如，他说：

> 这些表征的目的，在于有效地描述现实世界的各个方面。因此，在确定所使用的表征的本质以及推导并保持这些表征的过程的本质时，现实世界的结构就起着重要的作用。理论分析的一个重要部分就是搞清楚在设计表征和过程时所使用的物理约束和假设。（1982：43；也参见 68，103-105，265-266）

马尔的确主张：他假定的初级视觉中的表征基元——例如“斑点、线条、边缘、组合群等”——“对应于所看到的表面上的真正物理变化”（1982：44）。总之，这些评论被认为是尤其支持（2）和（3）的。

围绕如何解释马尔理论的很多争论都聚焦于这是否是理解他对“现实世界的结构”的诉诸的正确方法。对于下述看法至少有两种代替者：它认为这些评论主张头脑之外的世界对于视觉状态的计算分类法很重要。

第一种代替者认为，这些评论**只**给予现实世界一种在建立马尔所说的计算理论时的作用。由于视觉是为了保证有机体在世界上有效活动而从世界中提取信息的一个过程，因此，我们在解释视觉**为了**什么、视觉**做**什么、视觉**被设计来执行**什么功能时，显然必须对世界的结构有所了解。如果这是正确的，那么似乎就能相信：一个人在建立表征和算法的理论时，**不**必考虑头脑之外的东西。由于视觉状态就是在这个层次上被作为计算机制的对象而进行分类的，因此，马尔提到“现实世界”这一事实并未使他对视觉状态和过程的分类承诺一种外在主义观点。

第二种代替者认为，这些评论只是指出了现实世界的结构在提出一种计算的分类法和在关于视觉的更一般的计算理论中的**启发性**作用。也就是说，借助头脑之外的世界，对于理解视觉是如何工作的和理解视觉状态和计算的本质来说只是一条有用的捷径。它要么能提供重要的背景信息，这些信息使我们能理解表征基元并因而理解最早的视觉计算阶段；要么是充当解释性的透镜，使我们能用富有意义的术语来建立一个计算过程的**模型**。此外，就像前一种选择一样，头脑之外的世界在计算视觉中只是起辅助性作用，即使马尔有时候在概述他的理论时主要是参照它。

个体主义者主要从两个方面对伯奇的论证作出了反驳。首先，西格尔（1989）和马修斯（1988）实际上都否定了（2），西格尔还认为，这些意向基元（如边缘和广义椎［generalized cones］）在马尔理论的语境下最好被解释成是由它们的**窄内容**进行个体化的。其次，伊根（1991、1992、
269 1995、1999）更惊人地否定了（1），他认为：作为计算理论，马尔的理论根本不是意向的。这两种反驳都值得详细考察，尤其是就它们都强调在当代讨论中仍有争议的问题而言。事实上，对于心灵的本质以及我们应该如何研究心灵，有关马尔理论所提出的基本问题比已解决的要多。

西格尔指出：在一个人想把意向内容归属给两个个体的视觉状态时，有两种一般的解释可用。首先，一个人可以仿效伯奇，依据在正常情况下引起某种视觉状态的东西来解释这种状态的内容。因此，如果起这种作用的是表面上的一条裂纹，那么对应的视觉状态的内容就是“裂纹”；如果起这种作用的是环境中的阴影，那么这种视觉状态的内容就是“阴影”。甚至在孪生人的案例中也可能如此，因此，被这样个体化的视觉状态不是个体主义的。其次，一个人对这两种情况下的视觉状态的内容可以作出一种更自由的解释，这种解释在这种状态的原因方面保持中立，为了表示这种中立性，我们可以称之为“裂影”（crackdow）①。这种内容是孪生人所共有的，因此它**将**是个体主义的。

西格尔论证的关键部分是他更喜欢第二种解释的论据，人们就是期望在这里发现他是如何模仿马尔视觉理论的细节的。虽然西格尔的一些论证在这里确实是模仿的，但他也引入了大量的一般性考虑，这些考虑与马尔理论没有多少关系。例如，他指出第二个解释有“它本身的经济性”（1989：206），从而模仿了简单性的考虑，他说：

> 最好的理论描述**总是**这样的描述——在这种描述中，在一个对这两种远因进行区别的层次上，表征不能说明它们的外延，而且认为外延包括的这两种东西**总是**更好的。（1989：207；强调为我所加）

为什么“总是”？西格尔在支持他反外在主义的论据时，一般是说“好解释的基本标准”（1989：207），但就像诉诸科学解释的本质一样——

① 这是作者杜撰的一个词，由 crack（裂纹）这个词和 shadow（阴影）这个词的后三个字母组合而成，姑且译为“裂影”。——译者注

科学解释依赖于科学的分类法，而科学的解释是由“因果力”个体化的，我们在这里也应该对描述科学实践的一般性水平（以及对应的实质性细节的缺失）表示怀疑。就像伯奇本人在阐述一般的外在主义论证时借助了知觉表征的客观性一样（1986a：第3节；1986b），在我看来，这些对表征姿态过失的先验要求是和对心理学中的个体主义的那些更有意思、更实质的经验论证缠绕在一起的。

在西格尔更明确地借用马尔理论的特征时，他提出了三个一般性观点，后者与他论证这种理论是个体主义的有内在关联。这三个观点是：每一个表征的归属都需要一种“自下而上的解释”（1989：194），都需要一种“自上而下的动机”（1989：195），而且都要“根据行为证据来检验”（1989：197）。总之，这三个观点意味着，在马尔的理论中，假设表征是不容易实现的，它其实是受整个任务的要求和方法严格限制的。第一个观点表明：这个理论所假定的任何高层次表征都一定派生于低层次的输入表征；第二个观点表明：所有被假定的表征的动机都导源于它们在整个知觉过程中的作用；第三个观点表明：“意向内容是从已辨别的行为中推导出来的。”（1989：197）

西格尔使用第一个假设作了这样的论证：既然孪生人身上最早期的表征内容——“等于并且包含零交叉（zero-crossings）”（1989：199）——是相同的，那么就有一种直接的证据证明：下游的高层次表征一定是相同的，除非假定某一种不同能有一种自上而下的动机。但是，既然我们是在考虑孪生人，那么就没有什么行为的证据能用于诊断这两者之间的表征是不同的（西格尔的第三个观点），从而就没有可用的自上而下的动机。如他所说：“如果只要一种内容，那么使用（这两个孪生人）的这两种内容就没有任何理论意义，因为没有区别这些内容所服务的理论目的。”（1989：206）

外在主义者如何抵制这个挑战性的论证？有三种不同的策略，其中每一种都比它前面的一种更少赞同西格尔的观点。

首先，有人可能会承认西格尔从其对马尔的解释中所抽出的这三个观点和他关于低层次表征是个体主义的主张，但怀疑这些观点和主张的意义。有人可能会同意马尔理论由之起步的灰色阵列在某种意义上表征了光强度值，而零交叉在相同的意义上表征了光强度的突然变化。但这两者都只是某个视网膜状态的表征，而不是世界的表征，而且这些有机体内部的表征存在窄内容应当是不足为奇的。此外，这些表征的意向性或“关于

性”的深度可能会受到质疑，这正好也是因为它们并不涉及任何能延伸到头脑之外的因果关系；它们可能被认为是表征，就像我的胃咕咕叫表征我当下的饥饿状态一样。然而，一旦我们转到下游过程，即后来出现在视觉加工的时间维度上的过程，**真正**强健的表征基元就开始起作用了，例如“边缘”和“广义椎”等基元。有的人可能会主张，使用了**这些**基元的状态的内容就是某种世界状态的表征，它们从形而上学上说依赖于它们在世界上的对应物，因此不是个体主义的。西格尔这种反驳的可信性既取决于马尔理论中的“表征”的弱意义和强意义之间的区别强度，也取决于这种主张，即我们要有这样的状态时就需要这种强意义，这些状态从哲学上说就是表征状态。

其次也是更激进的策略是：有人可能承认这个理论中假定的所有表征基元都是在相同意义上进行表征的，但他怀疑这一主张，即**任何**一个对应
271 状态的内容都是窄的：如果你愿意，不妨说它始终都是宽内容。使用灰色阵列和零交叉状态的表征内容实际上是宽的，这种观念本身的线索可能来自于西格尔的第二个观点，即表征需要一种自上而下的动机，因为正是通过反思可靠的三维视觉世界的三维图像之整个建构过程的要点，我们才能看到（甚至初级视网膜）表征也一定是关于世界上的状态和条件的表征。这一观点必然会超出马尔理论本身，它显然只涉及关于我们如何从贫乏的视网膜信息推出三维图像的计算问题，但我认为，它在精神上非常类似于我们说的马尔理论的吉布森主义方面（比较夏皮罗［Shapiro］，1993）。

最后也是最不折中的策略是：有人可能会反对西格尔关于马尔理论中的三个观点之一个或者几个，或者更确切地说，反对西格尔赋予这些观点的意义。时间上靠后的表征**是**派生于之前的表征，但它对于如何对**任意一个**的内容进行个体化并未向我们作任何说明。同样，马尔本人是从视网膜图像的低层次表征开始的，这对于这些表征是窄的还是宽的也未向我们作多少说明。自上而下的动机**是**为假定表征作辩护所必需的，但是由于在马尔理论内部有一系列的目标都与三维视觉过程的总目标有关，因此，对于这些表征的内容是窄的还是宽的也并没有给我们多少引导。行为证据在诊断特定表征的内容方面确实起着一种作用，但马尔不是行为主义者，因此行为的辨别没有为分析表征提供试金石（夏皮罗，1993：498-503）。

第三种反驳似乎最可信，可以深入发展，但在我看来，它也意味着这

个问题可能并没有定论，即马尔理论使用的是窄内容概念还是宽内容概念，或者都运用了，还是一个也没用？尽管马尔本人根本未涉及这种理论的基元的意向本质问题，但是，就对他的理论可能作出的解释来说，他的方法论评论等的深入程度给我们留下了多得无所适从的选择。这不仅是关于马尔的意思和意图方面的一种不确定性，也是关于视觉本身的计算方法内部——而且我认为，也是更一般的计算心理学内部——的一种不确定性。记住这一点后，我现在就转向伊根的这一主张：这种理论**根本不是意向的**。这是关于马尔理论的一种少数派观点，我相信它还未得到应有的认可（比较巴特勒［Butler］在1996年和1998年以及夏皮罗在1997年对伊根的批评；也参见乔姆斯基在1995年著作之55页的脚注25）。

伊根关于马尔的看法的核心，是一种关于马尔的**计算的**描述层次的本质的特殊观点。马尔的评论者们几乎普遍认为这相当于其他人所说的描述的“知识层次”（纽厄尔［Newell］，1980）或者“语义层次”（派利申，1984），即它是为控制视觉和其他认知过程的计算机制提供**意向的**描述的。伊根并不像有些人（如夏皮罗，1997）所认为的那样（据说是为了只关注马尔的算法描述层次）忽视了马尔的计算层次，相反，她反对这种关于计算层次占支配地位的理解方式，她认为，使之成为计算层次的，是它用精确的数学术语说明了由给定的算法所计算的功能。也就是说，虽然这个描述层次是功能性的，但是使之在建立一种**计算**理论时处于第一阶段的，是它为计算提供了一种功能—理论的描述，从而从所有别的功能描述中抽象出来。因此，虽然视觉或许具有各种能用非常接近于常识语言的语言来说明的功能（例如，它是为了从世界中提取信息，是为了知觉客观世界，是为了引导行为），但在伊根看来，其中没有一种功能是马尔的计算的描述层次的一部分。鉴于这种观点，我们可以主张马尔理论是个体主义的，因为计算理论能轻松地推导出来：

> 计算理论不考虑现实环境，因为它的目标是为这样一种机制提供一种抽象的并因而是完全一般的描述，这种机制为预言和解释其在任何一种环境中（甚至是在这种装置的所作所为不能被恰当地描述成**认知**的环境中）的行为提供了基础。当计算描述伴随有一种适当的意向解释时，我们就能明白，计算特定数学函数的机制在特定环境中是如何能促进像视觉这样一种认知功能的。（1995：191）

根据伊根的观点，虽然意向解释把计算理论与我们关于认知功能的（以常识为基础）理解联系了起来，但它并未成为计算理论本身的一部分。自然，伊根的观点提出了一些问题，这些问题不仅与马尔用计算的描述层次指涉什么有关，而且更一般地说，也与关于认知的**计算**方法的本质有关。

当然，马尔在有些地方的确说到，计算层次就是对视觉的目的所作的一种高层次的功能描述，因此它主要是引导研究者提出一些一般性问题。例如，他的一张表提供了这一理论在计算的层次上所回答的下列简要问题："计算的目的是什么？为什么这一计算是合适的？执行这一计算的策略逻辑是什么样的？"（1982：25，图 1—4）支持马尔理论是外在主义的研究者，通常都以对这一理论的计算层次所作的宽泛理解为理据（参见：例如伯奇，1986a：28；夏皮罗，1993：499-500、1997：134）。

对计算层次以及认知的计算方法所作的这种宽泛的理解，其问题在于：虽然它在计算心理学和关于认知的更具民间色彩的思考方式之间架设了桥梁，但是，它在计算方法内部的计算层次和算法层次之间造就了一条鸿沟。例如，如果我们假定：计算层次只是说明了某些视觉状态有表征边缘的功能，有些有表征形状的功能，等等，那么这些描述中就没有什么能在建立这样的**算法**方面引导我们——这些算法产生了处于视觉的计算方法之核心的状态到状态的（state-to-state）转换。关于视觉的目的或者它进化的目的更通俗阐述不太能弥合这条鸿沟。

这里的观点是：计算的说明本身是一种非常特殊的功能描述，至少是当它们在自动的算法层次上被完成或者执行的时候。就这种关于认知的计算方法的宽泛理解来说，其支持者要么必须认为计算层次包括且超越了伊根所确认的关于认知能力的功能—理论描述，要么必须把这些描述分配给算法层次。后一种选择只会使上面确认的鸿沟进一步扩大。但在我看来，前一种选择是把很多截然不同的东西都置于"计算层次"的标题之下，因而认识不到计算的假设会使之被意识到的这些问题所限制。马尔理论的时间性维度和模块性维度会使这个问题进一步恶化。

这里隐藏了一个重大问题，它涉及的是功能主义在认知的计算方法内部应当如何理解，以及相应地这些方法实际的包容性如何。人们通常认为功能主义提供了一种方法，能使我们的民间心理学和关于心灵的明显图像（塞拉斯，1962）与发展中的心灵科学相一致，即便这种一致涉及到要沿着个体主义路线对民间心理学作出修改（例如，借助窄内容概念将它分解

成一种**窄**民间心理学）。而计算主义被看成是一种说明相关的功能作用是什么的方法：它们就是“计算的作用”。但是，如果对于马尔关于计算的概念是功能—理论概念的理解，那么伊根就是对的。如果我们接受这种关于这一理解一般为认知的诸计算方法中所共享的观点，那么关于心灵的功能主义版本就一定相应地会是功能—理论的：它不仅一定“不考虑现实环境”，就像她认为计算层次一定会做的那样，而且也不考虑功能主义经常借助的这种**内部的**因果作用。照此看来，认知机制是用能从数学上描述的输入来传送能从数学上描述的输出，而作为计算装置，这就是一切。协调我们的“两幅图画”的希望一定在其他地方。

在为马尔视觉理论的非意向特征进行辩护时，伊根就计算心理学的核心提供了一幅严格的图画，这幅图画在过去的发展中与计算的认知科学的个体主义取向是一致的（比较乔姆斯基，1995），即使是在计算心理学家（例如派利申，1984）想把他们的理论置于更具包容性的语境之中的时候。如夏皮罗（1997：149）所指出的：这种计算观的一个问题是，关于 X 的计算理论对于 X 的本质很少向我们作出说明，甚至包括足以将 X 个体化为（比如说）一个**视觉**过程的信息也是阙如的。虽然伊根（1999）似乎愿意接受这个结论并将这种担心恰当地置于计算的理论之外，但这种答复强调得到严格解释的计算理论与无数这样的理论之间是存在鸿沟的，这些理论是表征的、功能的或者本质上是生态学的，计算理论要对任何给定的认知过程作出彻底的机械论解释，就必须与这些理论相结合。对计算的解释越严格，这个鸿沟就越大，计算理论对我们认知的理解贡献就越少。有人可能认为，对于心理学中的计算理论的观点，伊根的失误就在于这方面的限制**过于**严格了。

11.8 利用性的（exploitative）表征与宽计算主义

在考察另一种思考计算和表征的方法时，让我们首先考虑一个关于马尔理论的个体主义解释与外在主义解释之间有意思的区别，它涉及马尔式的计算系统是由什么构成的。关于计算的个体主义者，如伊根和西格尔，认为它们包含了各种关于世界是什么样子的**天赋假设**。这是因为视觉过程涉及从两维视网膜图像中恢复三维的信息，如果没有其他输入，这个过程就会决定不足（underdetermined）。要解决决定不足的问题，唯一的办法

是作出一些关于世界的天赋假设。其中最著名的是厄尔曼的**刚性假设**（rigidity assumption）："任何一个经历了两维转换的要素集合，都有一种唯一的解释，即它是在空间中运动的一个刚性体（rigid body），因此它应当被解释成一个运动着的物体。"（1979：146）个体主义者提出的主张是：与此类似的假设是驱动认知加工的计算系统的一部分。这是理解马尔的视觉方法的标准方法。

像夏皮罗这样的外在主义者已经对这个问题给出了不同的回答。尽管为了使我们的计算机制能解决决定不足问题，某些关于世界的假设一定是正确的，但这些假设只是我们的计算机制所**利用的**（exploited）假设（夏皮罗，1997：135，143；比较罗兰兹，1999），而不是我们认知结构中的天赋假设。也就是说，这些假设涉及外部世界的特征之间的关系或者说内部视觉阵列的属性与外部世界的属性之间的关系，但这些假设本身并未被编码在有机体之中。为了说明这两个观点之间的区别，让我们看一些简单的例子。

里程表是记录汽车行驶里程的，它是通过计算车轮旋转的圈数来达到这一目的，而且它的制造原理就是为了显示与圈数成比例关系的一个数字。其方法是使"车轮转 1 圈＝x 米"的假设成为它计算机制的一部分；
275 实现这个目的的另一种方法是：只记录车轮每转 1 圈是 x 米，从而利用了这一事实：车轮转 1 圈＝x 米。在第一种情况下，它编码了一个表征假设，并用这个假设来计算它的输出；在第二种情况下，它没有包含这种编码，但**使用**了其结构与世界结构之间的一种现有关系。在任何一种情况下，如果它发现在自身所处的环境中车轮旋转与行驶距离之间的关系被调整了（例如，车轮更大了，或者说在跑步机上行驶），那么它就不会发挥它应发挥的功能，而将错误地表征汽车已行驶的距离。

考虑关于学习如何击打一个正垂直下落的棒球的两种不同策略。由于棒球的加速度是 $9.8ms^{-2}$，因此在挥拍与击打之间就有一段时滞。一个人要么可能假设球在下落（比如说，以特定的加速度），然后用这个假设来计算他应当何时挥拍；或者他在挥拍时只是瞄准距离他所知觉到的球所在位置一定距离的地方（比如说，两英尺）。在后一种情况下，这个人会利用加速度、时间与距离之间的关系，但不一定会把这种关系编码到他在这项任务所使用的假设里。

实现同一个目的会有这两种不同的策略，这一事实至少应当使我们在接受下述主张时保持警惕：天赋假设是计算系统解决决定不足问题的**唯一**

方法。但我也想提出这种想法，即我们特殊的知觉系统和更一般的认知系统通常都利用而非编码了关于世界以及我们和它的关系的信息，对于马尔本人在这个问题上的立场似乎也可以持类似看法（也参见威尔逊，即出）。

伊根提出了一个假设，并得到了有关个体主义和计算的哲学研究者的普遍接受。这个假设就是：至少计算心理学中的**算法**的描述层次是个体主义的。我认为，这里的这种观念似乎过于明显，以至于它很少被详细地表达出来：算法是对符号的句法或形式属性进行操作的，而这些算法是内含于例示这些符号的有机体的。不过这一点值得怀疑，它的目的可能既不能通过争论有多少算法被当成了马尔的计算层次的一部分，也不能通过争论马尔的计算层次与算法层次之间的界限来实现，而是要证明计算本身能延伸到有机体的头脑之外，它们涉及个体与其环境之间的关系。这种立场认为，至少有些驱动认知（尤其是人类认知的计算系统）延伸到了有机体之外了。我在其他地方（1994；1995：第 3 章）将这种立场称为**宽计算主义**（wide computationalism），它在马尔视觉理论上的运用标志着对这些参数的背离，这些参数控制着关于该理论的个体主义—外在主义争论。对于这个方面，即认知——甚至从计算上考虑——被“嵌入”或“置于”其本质之中，宽计算主义构成了一种思考的方式（也比较哈钦斯［Hutchins］，1995；麦克拉姆罗克［McClamrock］，1995），它还提供了一个框架，利用性的表征概念可以在其中得到探讨。

宽计算主义的基本思想很简单。过去，控制认知的这种计算被认为是 276
既起始于也终止于头颅。但是，为什么会认为头颅构成了这样一条奇妙的界限，在它之外真正的计算止步而纯粹的因果关系却起步？假如我们是被安置于具有丰富信息的复杂环境中的生物，发生在头脑中的计算是对应的计算系统的一个重要部分但非其全部。这种观点引出了这种可能性：探索计算的单元，它们既包括大脑，也包括头脑之外的环境。因此，宽计算系统所涉及的心灵，其实超出了头颅的界限而进入了世界。

说明由宽计算主义在个体主义争论中所造成的这种背离的本质，一种方法是利用**定位的**（locational）心理学状态概念和**分类的**（taxonomic）心理学状态概念之间的区别（也参见威尔逊，2000a；比较罗兰兹，1999：第 2～3 章）。个体主义者和外在主义者一般在怎样对心理状态进行分类或个体化方面意见不一致，但两者一般（尽管并非总是）认为相关的状态是我们所说的定位上的个体主义（locationally individualistic）：它们位于有

机体的界限之内。个体主义者和外在主义者通常的分歧在于：除了定位上的个体主义之外，心理状态从分类上说是否也一定是个体主义的。然而，宽计算主义反对这种定位的个体主义假设，它主张：有些“相关的状态”——其中一些构成了相关的计算系统——不是位于个体的头脑之中，而是位于他的环境之中。

宽计算主义背后的直觉观念很容易理解，但是对于主张宽计算主义是一种思考和研究认知加工的切实可行的模型来说，有两个至关重要的主张是有争议的。第一个主张是：对有机体的环境作出一种形式的或计算的描述，并把有机体大脑的各部分——从计算上描述的——和被这样描述的环境看成一种统一的计算系统，这一做法有时候是合适的。如果这一点不正确，就很难把宽计算主义看成是一种融贯的观点。第二个主张是：由此产生的这个心灵—世界计算系统本身而不只是它处于头脑内部的部分，是真正的认知系统。若没有这第二个主张，宽计算主义至多只是提出了一种划分计算世界的荒唐可笑的方法，这种方法对于我们应该如何思考真实头脑中的真实认知没有任何明显的影响。由于篇幅有限，我不打算对每个问题都作出答复，而是集中研究这个问题：这个关于表征和计算的一般观点是如何被马尔的视觉理论接受的。

如我们所看到的，马尔本人认为视觉理论的任务是要证明：我们如何从“由视网膜上的光感受器所检测到的图像强度值阵列”（1982：31）中提取视觉信息。因此，如我们前面所指出的，视觉问题**开始**于视网膜图
277 像，而不是开始于那些图像之外的世界的属性，而且“视知觉的真正核心是要**从**图像的结构中推导出外部现实世界的结构”（1982：68；强调为我所加）。马尔接着描述了一系列适用于世界的约束，它们“使这种推导成为可能”（1982：68），但他解释说：“只要把这些约束变成一种能或者不能从内部证实的假设，它们就能得到运用。”（1982：104）马尔关于头脑之外的世界事实对建立视觉理论中的计算层次的重要性所说的一切，代表了他对相关情况的看法（例如，1982：43，68，99，103－105，265－266）。用我在这一节前面所介绍的术语来说，马尔本人对计算和表征采取的是一种编码的观点，而不是一种利用性的观点。根据马尔的看法，视觉系统从定位上说是一个个体主义系统。

这里，不管马尔本人的观点是什么，要对关于其理论的宽计算解释作出辩护，一种直接的方法是抵制他从“x是在世界上有效的一种物理约束”到“x是一个被编码于大脑中的假设”所作的推导。这实质上就是我以前

提出的一个人在坎贝尔和罗布森（Campbell and Robson，1968）所开创的关于形式知觉的多空间通道理论中应该做的事情。与马尔的视觉理论相似——马尔的理论部分地是以这项研究为基础的（特别参见马尔，1982：61-64），人们通常认为这个理论假设了一个定位上的个体主义计算系统，这个系统起步于视觉通路上的初级通道，这些通道对四个参数有不同的敏感程度：朝向、空间频率、反差和空间相位。我建议要认真对待这一主张（威尔逊，1994；1995：第3章），即任何（世界上的）视觉景象都能分解成这四种属性，因此要认识到，计算系统本身延伸到了世界之中而刺激与视觉通道之间的因果关系本身是由转换规则模拟的。这不只是要把这些属性编码在神经系统里的不同视觉通道中，而是要认识到形式知觉系统在头脑里的部分利用了头脑之外的世界上的形式属性。就马尔的理论来说，有一个方面是这种宽计算解释能**轻松**辩护的，另一个方面是它**难以**辩护的。

第一个方面是马尔的“假设”，如空间一致假设（1982：70）和“立体影像基本假设”（1982：114）通常都把反映世界结构的物理约束作为起点；在上述例子中，它们都把空间定位约束（1982：68-69）和三个相匹配的约束（1982：112-114）作为起点。因此，这个策略就是要证明：这些约束**本身**——而不是它们派生的编码——在定义计算系统方面起着一种作用，而不只是使我们能为一个定位上的个体主义认知系统在进行计算描述时充当一种启发性作用。

关于马尔理论的宽计算解释难以辩护的对应面是：这些约束本身没有说明什么是计算的基元。一种可能性是简单地将马尔归属给**图像**的基元归属给被知觉到的景象特征本身，但这过于草率了。例如，马尔认为零交叉 *278*
是这种计算中的步骤，它们表征图像强度上的剧烈变化。虽然我们可以认为它们是表征世界上的刺激强度的变化，但零交叉本身在视觉系统中位于脑中某个初级的部位，它们可能靠近视网膜。我想，更好的策略是紧缩这种关于视网膜图像的解释，从它出发“上溯”，进而设法辨认世界上更丰富的外部结构，即满足马尔所假设的物理约束的结构。也就是说，一个人应当把时间性维度延伸到马尔理论，从而使基本视觉过程中最早的阶段就**开始于世界**，而不是头脑。由于视觉研究主要是在全面的个体主义框架中进行的，因此这种策略就要求对视觉理论本身进行改写，以至于它能涵盖一个因果地延伸到视网膜图像之外的过程（也参见罗兰兹，1999：第5章）。

11.9 窄内容与马尔的理论

让我们考虑西格尔在论证马尔的视觉理论是个体主义时所采取的第一个步骤，即主张一个人在希望将意向内容归属给两个个体的视觉状态时，有两种一般性的解释可资利用：一种是“受限的”（伯奇的解释），一种是“自由的”（西格尔的解释）。与这两种一般性解释相似的解释就隐含在我们一开始所考察的孪生地球案例和关于个体主义的争论中；孪生人一定分享了某种关于似水物质（或者在伯奇的标准案例中，是关于似关节炎的疾病）的意向状态，这种观念是阐述**窄**内容（即随附于个体的内在物理属性的意向内容）概念的基础。我在别的地方（威尔逊，1995：第 9 章）表达了我对这些尝试的怀疑，这里我想将这种怀疑与在西格尔的解释中看起来无伤大雅的第一个步骤联系起来。

在西格尔的解释中，第一个步骤（即关于能对马尔理论作自由解释的预设，以及一般对孪生地球案例所引出的观点）本身是有问题的。首先，注意这些表征，为了使它们的析取内容清楚易懂，我们可称之为“裂影”或者“水或孪水”，它们**确实**表征了它们可靠的环境原因：“裂影”是由裂纹或者阴影可靠地引起的，它有裂纹或者阴影的内容；“水或孪水”同样如此。但由此一来，正如伊根（1995：195）所指出的那样，这种析取内容就是一种宽内容，而不是窄内容。简言之，尽管被孪生人分享是必要的，但心理内容是窄的，因而它是不充分的。

再强调一下，如果视觉状态的内容是个体主义的，那么**不管孪生人的环境如何不同**，它都一定是他们共同具有的。因此，“孪生人”的案例只
279 是思考潜在无限多个体的一种启发法。但是这样一来，就必需的内容（不仅仅是“裂影”，而且还是更难把握的析取的东西）而言，关注两个个体共同具有的内容，从而关注在两个环境原因之间保持中立的内容，就代表着一种误导的简化，因为潜在地有无限多种环境，它们能在个体的视觉系统中产生相同的内在物理状态，就像裂纹在现实世界上所展现的那样（也参见伊根，1991：200，脚注 35）。这不是说我们不能为这种状态的内容拼凑一个名称（我们能，称之为“X”），而是说我们很难认定一个被这样个体化的状态是**关于什么东西**的。如果关于某种东西是意向性的核心，那么这就会使这些被窄的个体化的状态作为意向状态的地位受到怀疑。

西格尔（1991：490）主张：“裂影”——也暗指“水或孪水”——的窄内容不一定是析取的，它们只是比裂纹或水有更大的包容性（也参见西格尔，2000）。但是根据析取内容来思考上述观点时，就会使这些一般性问题变得非常清晰：（1）根据它们的内容来对状态进行个体化仍然要参照那些会使它们被标记的（tokened）东西；（2）在思考如何对认知系统的状态进行分类时，一旦一个人不考虑这种观念，即认知系统被设定并与现实世界有相互作用，那么就很难清楚地把这些状态描述成是具有某种明确内容的意向状态。如人们有时所说的，窄内容会变得**难以言表**。对于这第二个反驳，学界已作出两种回应。

首先，有的人可能承认，严格说来，窄内容**的确**难以言表，但“悄悄接近它”的途径也有很多（福多，1987：52）。有的人指出，这种途径是通过谈论一个人何以能将窄内容“锚定”在宽内容上（福多，1987：50-53）；或者谈论如何说明一个命题的**实现条件**（洛尔，1988）。然而，尽管这些意见很流行，但在我看来不过是借吹口哨壮胆罢了，而且它们所依靠的这种让步是致命的。所有“悄悄接近”窄内容的途径都涉及以某种方式使用宽内容。然而，如果宽内容是这样一个成问题的概念（因为它不是个体主义的），那么这个问题确实会波及任何概念，例如悄悄接近窄内容，因为宽内容概念对于它的可理解性是很关键的。

此外，如果窄内容真的难以言表，那么下述观念就不能得到合理地支持，即这个概念对于现实实践中所作的心理学解释是至关重要的，而且在自然种类和认知科学的规律中确实会起重要作用。除了道格拉斯·亚当西斯克（Douglas Adamsesque）的科学欺骗之外，还没有一种科学的核心解释结构是难以言表的。此外，这个观点也会使下述主张变得极不可信：人是借助考察认知科学中的实际解释实践而获得窄内容概念的，因为如果窄内容难以言表，那么就不能发现它被表达在任何现存心理学理论之中的事实。总之，用我之前介绍的术语来说，悄悄接近窄内容是认知科学所必需的，这种观念代表了一种向认知科学姿态的回归。

其次，人们主张：虽然普通人提出意向内容的名称确实**是**很困难的，但在相关认知科学中的研究者却始终都能并且确实在这样做，我们应该听从他们。例如，有的人主张，在马尔的理论中，很多——即使不是所有——表征基元，例如斑点、边缘和线条，都有窄内容。这些概念和很多科学术语一样，也是技术概念，因此和常识的概念及术语可能没有明显的关系，但它们仍能使我们了解窄内容是如何被表达的。有人认为这种回应

感觉是在以有待证明的假定为论据，就像这种主张一样：民间心理学状态本身是窄的。然而，哲学观点不能被它们所解释的科学理论——例如马尔的理论——的材料充分决定，这对于个体主义者和外在主义者来说都是个未解决的问题。正如我对利用性表征和宽计算的讨论时所指出的，我自己的观点是：我们必须复兴这些方式。在认知科学中，关于心灵的计算理论和表征理论就是用它们来解释的。如果超出姿态的方式也能做到这一点，那么窄内容的（难以）表达性问题基本上就没有实际意义了。

11.10 个体主义与自我知识问题

到目前为止，我一直在讨论当代心灵哲学中和认知科学最接近的个体主义和外在主义。各种讨论都在探讨这些立场与心灵哲学以及更一般的哲学中的传统问题的关系，这也证明了个体主义和外在主义在更一般的哲学中的核心地位——完全不同于它们与经验的认知科学的相关性。在我看来，其中最有意思的讨论主要围绕三个相关的认识论问题展开：自我知识、先验知识和怀疑论。

自我知识的基础是关于自己心灵的知识，传统上这种知识被认为涉及通向个人自己的心灵状态的某种优越通道（privileged access）。这种第一人称的优越通道常常是根据一种或更多种由此产生的二阶心理状态所拥有的独特属性来理解的。这些状态，例如我相信地球绕着太阳转的信念，据说是**不可错的**（也就是说，不会为假或者出错），它意味着只需具有二阶信念就能保证一个人具有一阶信念，它是前者的对象；或者说这些状态是**不可纠正的**（也就是说，即便出错了，也不能被除拥有它们的人之外的人来纠正），这至少意味着它们有一种其他类型的心理状态所缺乏的认识的可靠性。不论是哪种情况，关于自己心灵的知识与关于他人心灵的知识以及关于世界上其他东西的知识之间，都有一种不对称性。在现代哲学中，表现出这些深层次不对称性的是以下事实：认识论导论，尤其是带有历史色彩的、反思怀疑论的认识论导论，可能会介绍**他心问题**和**我们关于外部世界的知识问题**，但不会介绍对应的**自我知识问题**。关于人自己心灵的怀疑论似乎被自我知识的本质排除了。

281

在新近的心灵哲学中，虽然关于心理状态的第一人称知识和第三人称知识之间的反差有所弱化，但我们关于心灵的部分常识概念仍然是：我认

识自己心理生活的方式与我认识别人心理生活的方式不同（比较西沃特，1998）。因此，第一人称具有认识的优越性观念仍然存在。关于一个人的**自我**、关于一个人的心灵或者躯体的状况或状态的知识，看起来不过是个内省的问题，不过是内指的（inward-directed）反思或注意的问题，它不需要通过观察或者实验收集证据。我只是**感觉**我的皮肤痒，或者因留意而注意到我的鞋子夹脚；但要想了解**你的**皮肤是否痒或者你的鞋子是否夹脚，我就要观察你的身体及行为（包括你所**说**的话），然后从这种观察中推导出关于你的身体状态的结论。自我知识是**直接的**，而关于他人的知识是以某种方式推导出来的或者说是间接的，它以观察和其他形式的证据为基础。由于自己的心理状态通常是第一人称思想的对象，因此我们熟悉自己的心灵，但不熟悉他人的心灵。

对于理解第一人称的优越通道以及与之伴随的自我知识与关于他人心理状态的知识之间的不对称性，个体主义的心灵概念似乎是很合适的。如果心理状态是通过抽象作用从个体之外的环境进行个体化的，那么，在理解内省过程，即将我们的心灵之眼朝内转向时（用一个常见的隐喻来说），如何透露这些状态的内容方面，似乎就没有什么问题了。以笛卡尔的想象为例——它在某种意义上产生了自我知识与其他形式的知识之间的不对称性。即使有一恶魔诱骗我相信：外部世界是存在的——其他具有像我的心理状态一样的心理状态的人是存在的，但我仍然坚信的一件事情是：我有种种经验，它们有某种内容。正如人们有时所说的那样，对于我面前是否真的有一棵树，进而对于我是否实际看到了一棵树，我可能会上当受骗；但对于**我是否认为我正在看一棵树**，我是不会被骗的。因此，个体主义似乎有利于这样一种认识的可靠性，它是关于人自己的心理状态的第一人称知识所具有的，而相应的第三人称知识所缺乏的。

相反，即使是对更谦逊的第一人称优越通道形式，外在主义也提出了一个明显的问题，人们甚至认为它对任何一种自我知识形式的可能性也提出了质疑。因为外在主义主张：什么**是**心理状态，部分地是由这些状态的主体之外的世界之本质形而上学地决定的。因此，要知道一个人在想**什么**，即要知道一个人的心理状态的内容，似乎就必须对他自我之外的世界有所了解。但这会使我们关于自己心灵的第一人称知识与我们关于其他东西的知识相类似，从而否定自我知识所具有的**优越**通道。这就意味着：要 *282*
知道我自己的心灵，我必须知道关于我所生活的物理或社会世界中可能难以发现的事实。因此它也就表明：在很多普通案例中，我可能是不假思索

地归属自我知识的，在这里我其实根本没有自我知识。

我们可以换种方式来表达这个问题，这种方式是从以下两种关于优越通道的具体解释与具体的外在主义版本之间的不同中抽取出来的。不管它是否是不可错的、不可纠正的、私密的、内省的或者先验的，但是关于自己心理状态的知识有一种特殊的特征。除其他东西之外，知道自己的心理状态也就意味着知道它们的内容。根据外在主义，一个主体的心理状态的内容部分地是由关于她的物理或社会环境的事实从形而上学上决定的。然而，关于这些事实的知识却没这种特殊特征。这样一来，假如自我知识的这种特殊特征依赖于世界方面的知识的非特殊特征，那么前者是如何与后者相容的呢（也参见勒德洛和马丁［Ludlow and Martin］，1998：1）？有些人更生动地陈述了这个问题。例如，戴维森认为它是“关于笛卡尔怀疑论的一幅颠倒的图像”（1987：94），根据它，“我们关于外部景象的信念是……指向世界的，但我们不知道我们相信的东西”（1987：94）。戴维森因而主张：外在主义似乎暗示了我们根本没有自我知识；海尔指出：“如果外在主义是正确的，那么一个人只检测一种状态是不可能发现它的意向属性的。”（1988：137）他接着把这一点与戴维森的研究联系了起来，戴维森关注的是一个“难对付的怀疑论者，他怀疑这个假设：我们思考我们思考我们思考的东西”（1988：137）。

这个问题可以简要地表述成三个不一致的命题（也比较麦金西［McKinsey］，1991，他的三个命题是不同的）。假定P＝我们心理状态的内容，E＝关于环境的事实，再假定“通过内省”代表自我知识的独特特征：

1. 我们通过内省知道P。（自我知识）
2. P部分地是由E形而上学地决定的。（外在主义）
3. E不是通过内省被知道的。（常识）

这里的主张是：这三个命题之一必须被放弃。如果我们反对自我知识，那么我们会放弃这一观念：我们有通向自己心灵的优越通道；如果我们反对外在主义，那么我们会放弃一种独立可信的心灵观；而如果我们反对常识，那么我们对于我们关于物理或社会世界的知识就会作出一种奇怪且不可信的主张。

在这样不加掩饰地陈述时，我认为对“自我知识问题”的正确回答是主张：所有这三个论题都是正确且一致的，并不存在需要外在主义者去解

决的自我知识问题。它们的一致性取决于这一事实：提出了认识论主张的（1）和（3）只是通过（2）——它提出了一种形而上学主张——联系起来的。作为对形式不一致性指责的一个反例，考虑这个论证的一个例子，这里 P＝感觉疼痛的状态，E＝中枢神经系统的一个特殊的复杂状态。在 283
（1）—（3）的这些实例中并没有不一致性：我们通过内省确实知道我们感觉疼痛；那个状态是由我们中枢神经系统中的某个特殊状态形而上学地决定的，我们不是通过内省知道**那个**状态的。（或者更谨慎地说，我通过内省所认识到的，并不是作为我们中枢神经系统的状态。）这同样适用于我们最初的三元组，也适合于这些命题的变化形式——它们用自我知识的其他独特特征替换了“通过内省”。

对于描述所谓的外在主义者的问题以及恰当回答这个问题的基础，如果这是正确的方法，那么这个问题的两个特征就值得注意。

第一个特征是：这个问题的核心不是外在主义的心灵观本身，而是关于形而上学决定的任何论题。在这里，决定状态（determining state）不是通过认识心理状态所用的那种特殊方式来认识的。由于并非有机体所有内部的、个体主义地个体化的状态都是这样认识的，因此，如果自我知识问题真是一个问题，个体主义者就一定会遇到它的一种变化形式。因此，即使一个人反对上述那种解决问题的办法，外在主义者和个体主义者仍要解决自我知识问题的另一个版本，这不外是说，尽管最初有这样的表象，但外在主义者并**没有**遇到一个关于自我知识的**特殊**问题。

第二个特征是：这样描述的问答与一个问答对于家族有密切关系——这个家族的“问题”包括摩尔的未决问题论证和分析悖论，它们的近亲可能是对心脑同一论的标准反驳。人们主张，疼痛不可能同一于 c-纤维激活，因为一个人能知道他感觉疼痛，却不一定知道他的 c-纤维正在激活。现在标准的回答是：这种反驳在试图从认识论前提推出本体论结论时出现了一个谬误。那么，由于是作为一个据说有矛盾的三元组，而不是一个得出这种结论的论证，因此，自我知识问题本身并不会遇到这个特殊问题，但反对把外在主义看成对这个问题的一种回答的反驳也**将**遇到这种反驳。然而，此处这种宽广的相似性需要谨记于心。对一个人来说，对自我知识问题作出何种回答，可能和这类答复与更一般的反驳的合适程度联系在一起。

自我知识问题的支持者应当反对这种主张：（1）—（3）就足以表达这个困境。特别是他们应当（并且实际地）反对（2）是这个三元组的成员

时。相反，自我知识问题是由下面这个三元组构成的（比较麦金西，1991）：

A. 我们有关于P的先验知识。（自我知识*）
B. 我们有P蕴含E的先验知识。（关于外在主义的知识）
C. 我们不能先验地知道E。（常识*）

284 （A）—（C）**是**不一致的。但和（1）—（3）相比，这种对自我知识问题的解释在**每一**点上都会遇到质疑。

首先，外在主义者承诺了（B）吗？就肯定的回答来说，必须对之前的两个问题作出肯定回答：根据外在主义，（i）我们知道P蕴含E吗？（ii）P衍推出了E吗？以（ii）为例：外在主义主张（例如）有一个具有“大腿上患了关节炎”这种内容的心理状态能够意指关节炎真的患在大腿上吗？如我们所看到的，否定此推论的一个理由是：外在主义体现了这种观念：在思想和语言方面都有一种社会的劳动分工，它甚至承认在“空案例”（vacuous cases）中也有意向性：我们能够思考P，不是因为P而是因为其他人思考P。

然而，假如外在主义主张：个体心理状态的本质与个体之外世界的情况之间有很深的个性化关系，那么P与E之间的这种关系似乎就是可信的。这表明：E需要用更细微的方式来解释，它可能包含了各种析取支，如果外在主义的心灵观是正确的，那么这些析取支合在一起就一定是正确的。例如，人们可能主张：具有关节炎发生在大腿上这种思想，要么意指关节炎确实发生在大腿上，要么意指一个人生活在某个人的语言共同体之中；可能这里需要增加更多的（或者更复杂的）析取支（比较布朗[Brown]，1995）。但是这样一来似乎不太可信的是：“我们”（即我们每一个普通人）会知道被这样解释的（2），更别说先验地知道这一点了。毕竟，我们当中很少有人会系统地思考我们思想的内容对于世界的意蕴；的确，很多作过这样思考的人——个体主义者——都得出了这样的结论：对于世界的特征，它们**什么也没**告诉我们。

这反过来又引起了以下回应：要与（A）和（C）组成一个有矛盾的三元组，（B）只需主张我们**能够**有这样的知识就行了。如果外在主义是对的，如果至少有些人会相信它及其衍推，那么这就足以能产生这种矛盾。

在我看来，这是提出自我知识问题的一种奇怪方法，因为它现在听起来像是这样一个问题：它主要出现在那些精通外在主义文献的人的自我知

识上，而不是出现在自我知识本身上。但是，这里的真正问题以及这种关于三元组的解释的第二个问题是：这个三元组含有一种对（C）的可疑解释。因为即使（C）**是**一种常识的原则（并且是模态化的，而（3）不是，这似乎是可疑的），它似乎也是错误的。因为，虽然普通人通过经验一般都知道 E 中所提到的东西，因此他们一般不是先验地知道 E，但根据对（B）的这种解释，有的人能以这种方式知道 E，这至少看起来是可能的。如果加上这种提示（即通常这不是我们认识关于经验世界的事实的方法）那么这种让步似乎完全没有害处，它还能维护这一直觉：自我知识从认识上说是优越的。

借助于（A），我们就能看到对三元组的这种解释是如何破坏它作为外在主义的问题的地位的：我们先验地知道我们思想的内容吗？麦金西认为，先验知识是“不依赖于经验研究而得到的”（1991：175）知识，而且他相信，内省和推理就是我们获得这种知识所凭借的典型过程。如果这种主张被看成是声称：自我知识能够**完全不依赖于**关于世界的经验研究而得到，那么外在主义者就应当对它保持警惕；他们愿意承认以及支持（A）可能需要的是：在无需对特定时刻的世界进行经验研究的情况下，我们能知道我们那些特定时刻的心理状态的内容。

根据这种解释，（A）为真是因为内省的存在，而（B）的真则依赖于我们能够理解关于内容和意向心理状态的外在主义本质的论证。虽然只要我们认为内省或推理单独就能成为保证先验知识的手段，（C）似乎就是正确的（在上述意义上），但一旦我们把内省和推理**放在一起**考虑，它就会变得可疑。由于我们很少既内省我们自己的心理状态又把这种内省的结果作为前提来进行复杂的哲学推理，因此，（C）会被证伪的情况本身是很罕见的；尽管如此，严格说来，（C）仍是错误的。

【注释】

我应该感谢加布里埃尔·西格尔（Gabriel Segal）、弗朗西斯·伊根（Frances Egan）和劳伦斯·夏皮罗（Lawrence Shapiro）阅读了这篇述评的初稿。

参考文献

Block, N. (1986). “Advertisement for a Semantics for Psychology.” In P. French, T. Uehling Jr., and H. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 10 (Phi-

losophy of Mind). Minneapolis: University of Minnesota Press.

Brown, J. (1995). "The Incompatibility of Anti-Individualism and Privileged Access." *Analysis*, 55: 149–56. Reprinted in Ludlow and Martin (1998).

Burge, T. (1979). "Individualism and the Mental." In P. French, T. Uehling Jr., and H. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 4 (Metaphysics). Minneapolis: University of Minnesota Press.

—— (1986a). "Individualism and Psychology." *Philosophical Review*, 95: 3–45.

—— (1986b). "Cartesian Error and the Objectivity of Perception." In P. Pettit and J. McDowell (eds.), *Subject, Thought, and Context*. Oxford: Oxford University Press. Also in Grimm and Merrill (eds.), *Contents of Thought*. Tucson, AZ: University of Arizona Press (1988).

Butler, K. (1996). "Individualism and Marr's Computational Theory of Vision." *Mind and Language*, 11: 313–337.

—— (1998). "Content, Computation, and Individuation." *Synthese*, 114: 277–292.

Campbell, F. W. and Robson, J. G. (1968). "Application of Fourier Analysis to the Visibility of Gratings." *Journal of Physiology*, 197: 151–166.

286 Carnap, R. (1928). *The Logical Construction of the World*, trans. by R. George, 1967. Berkeley, CA: University of California Press.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

—— (1995). "Language and Nature." *Mind*, 104: 1–61.

—— (2000). *New Horizons in the Study of Language and Mind*. New York: Cambridge University Press.

Cosmides, L. and Tooby, J. (1994). "Foreword" In S. Baron-Cohen, *Mindblindness*. Cambridge, MA: MIT Press.

Cummins, R. C. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Davidson, D. (1987). "Knowing One's Own Mind." *Proceedings of the American Philosophical Association*. Reprinted in Ludlow and Martin (1998).

Dennett, D. C. (1978). "Artificial Intelligence as Philosophy and as Psychology." In M. Ringle (ed.), *Philosophical Perspectives on Artificial Intelligence*. New York: Humanities Press and Harvester Press. Reprinted in D. C. Dennett, *Brainstorms*. Cambridge, MA: MIT Press.

Egan, F. (1991). "Must Psychology be Individualistic?" *Philosophical Review*, 100: 179–203.

—— (1992). "Individualism, Computation, and Perceptual Content." *Mind*, 101:

443-459.

—— (1995). "Computation and Content." *Philosophical Review*, 104: 181-203.

—— (1999). "In Defense of Narrow Mindedness." *Mind and Language*, 14: 177-194.

Fodor, J. A. (1980). "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioral and Brain Sciences*, 3: 63 - 73. Reprinted in J. A. Fodor, *Representations*. Sussex: Harvester Press (1981).

—— (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Gillett, C. (2002). "The Dimensions of Realization: A Critique of the Standard View." *Analysis* (October).

Heil, J. (1988). "Privileged Access." *Mind*, 97: 238-251. Reprinted in Ludlow and Martin (1998).

Hildreth, E. and Ullman, S. (1989). "The Computational Study of Vision." In M. Posner (ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jackendoff, R. (1991). "The Problem of Reality." Reprinted in his *Languages of the Mind*. Cambridge, MA: MIT Press (1992).

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Loar, B. (1988). "Social Content and Psychological Content." In R. Grimm and D. Merrill (eds.), *Contents of Thought*. Tucson, AZ: University of Arizona Press.

Ludlow, P. and Martin, N. (eds.) (1998). *Externalism and Self-Knowledge*. Palo Alto, CA: CSLI Publications.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W. H. Freeman.

Matthews, R. (1988). "Comments on Burge." In R. Grimm and D. Merrill (eds.), *Contents of Thought*. Tuscon, AZ: University of Arizona Press.

McClamrock, R. (1995). *Existential Cognition: Computational Minds in the World*. Chicago: University of Chicago Press.

McKinsey, M. (1991). "Anti-Individualism and Privileged Access." *Analysis*, 51: 9-16.

Newell, A. (1980). "Physical Symbol System." *Cognitive Science*, 4: 135-183.

Putnam, H. (1975). "The Meaning of 'Meaning'." In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis, MN: University of Minnesota Press. Reprinted in H. Putnam, *Mind, Language, and Reality*. New York: Cambridge University Press (1975).

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Rowlands, M. (1999). *The Body in Mind*. New York: Cambridge University Press.

Segal, G. (1989). "Seeing What is Not There." *Philosophical Review*, 98: 189-214.

—— (1991). "Defence of a Reasonable Individualism." *Mind*, 100: 485-494.

—— (2000). *A Slim Book About Narrow Content*. Cambridge, MA: MIT Press.

Sellars, W. (1962). "Philosophy and the Scientific Image of Man." In R. Colodny (ed.), *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press. Reprinted in W. Sellars, *Science, Perception, and Reality*. Atascadero, CA: Ridgeview Publishing Company (1963).

Shapiro, L. (1993). "Content, Kinds, and Individualism in Marr's Theory of Vision." *Philosophical Review*, 102: 489-513.

—— (1997). "A Clearer Vision." *Philosophy of Science*, 64: 131-153.

Shoemaker, S. (2000). "Realization and Mental Causation." Reprinted in C. Gillett and B. Loewer (eds.), *Physicalism and it Discontents*. New York: Cambridge University Press.

Siewert, C. (1998). *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.

Stich, S. (1978). "Autonomous Psychology and the Belief-Desire Thesis." *Monist*, 61: 573-591.

—— (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

Van Gulick, R. (1989). "Metaphysical Arguments for Internalism and Why They Don't Work." In S. Silvers (ed.), *Rerepresentation*. Dordrecht, The Netherlands: Kluwer.

Walsh, D. M. (1999). "Alternative Individualism." *Philosophy of Science*, 66: 628-648.

Wilson, R. A. (1992). "Individualism, Causal Powers, and Explanation." *Philosophical Studies*, 68: 103-139.

—— (1994). "Wide Computationalism." *Mind*, 103: 351-372.

—— (1995). *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. New York: Cambridge University Press.

—— (2000a). "The Mind Beyond Itself." In D. Sperber (ed.), *Metarepresenta-*

tion. New York: Oxford University Press.

—— (2000b). "Some Problems for 'Alternative Individualism'." *Philosophy of Science*, 67: 671-679.

—— (2001). "Two Views of Realization." *Philosophical Studies*, 104: 1-31.

—— (forthcoming). *The Individual in the Fragile Sciences I: Cognition.*

第12章 情 绪

保罗·E·格里菲思（Paul E. Griffiths）

12.1 无理性的感受还是理性的判断？

288

12.1.1 感受的情绪理论

直到20世纪初，人们都理所当然地认为情绪就是感受（feelings）[①]，即主观的经验状态。达尔文对情绪的生理和行为成分进行了广泛的经验研究，但他没有认识到，这些因素只是“表达”了人和动物所经验的感受。达尔文仿效赫伯特·斯宾塞（Herbert Spencer），把情绪定义成由身体之外的事态引起的感觉，他之所以这样做，目的是想把情绪与饥饿和疼痛等感觉区别开。感受也是威廉·詹姆斯的情绪理论的核心。尽管詹姆斯提出了更极端的意见，即情绪感受是由与情绪相联系的生理变化引起的，而不是引起了那些生理变化，但他仍然将情绪等同于感受，而不是等同于引起它们的生理变化或早期的神经过程。很显然，行为主义是最早站出来质疑这种感受理论的人。约翰·B·华生（John B. Watson）认为，成年人的情绪性行为是以婴儿身上的三种无条件反应——他称之为害怕、愤怒和愉快——为基础的条件反应。后来，在行为主义和意义的证实理论的影响下，像吉尔伯特·赖尔这样的哲学行为主义者主张，对情绪词语之意义的正确分析不必参照主观的经验状态。

20世纪60年代，哲学家欣然接受了心理学、语言学和新兴的人工智能领域中的“认知革命”。因此，对行为主义的否定并未伴随着感受理论

① feeling 和 sensation 通常都译为“感觉”，但这里所述的情绪理论明确指出，情绪虽然是一种主观的经验状态，但它不同于疼痛、饥饿等感觉。为了与 sensation 一词相区分，这里暂将 feeling 译为“感受”。——译者注

的复兴。20 世纪 60 年代初，情绪哲学研究中反而达成了一项共识，即情绪是根据它们所涉及的认知来定义的。这一共识一直延续至现在。有些哲学家已经承认了感受在情绪中的作用，但这不是决定情绪同一性的作用。情绪感受只是给“热认知”（hot cognition）增加了“热情”（heat）。例如，帕特里夏·格林斯潘（Patricia Greenspan，1988）认为，情绪是指向关于外部（或虚构的）刺激的评价性思想的舒适或不舒适的感觉。正是这种评价性思想定义着情绪。不同的负面情绪——例如愤怒和害怕——的区别仅仅在于它们涉及不同的评价性思想。哲学家们一般认为，情绪的同一性来自与它们相联系的思想，这是一个概念的真理，但对觉醒状态的“认知定名”（cognitive labeling）的心理学研究已经被引证为依据，以证明经验的发现找到了与概念分析相同的结论。最常被引用的研究表明：在实验人员提供语境暗示的影响下，可以诱导被试将注射肾上腺素所产生的感觉描述成愉悦或者愤怒（沙克特和辛格［Schachter and Singer］，1962）。

12.1.2 命题态度理论

从 20 世纪 60 年代初以来，认知主义者或者命题态度学派在情绪哲学中一直居于主导地位（格里菲思，1989；戴格［Deigh］，1994）。这个学派的基本承诺包括两个方面：首先，情绪彼此之间的区别在于它们所涉及的认知状态；其次，情绪中所涉及的认知状态能根据心理内容的命题态度理论来理解。心理状态是对命题的态度，例如信念、愿望、希望和意向。最简单的命题态度理论将情绪等同于评价性的判断（所罗门［Solomon］，1976）。如果一个人对这个命题——即他被冤枉了——持信念的态度，那么他就会愤怒。其他种类的著名命题态度理论还有信念/愿望理论、混合的感觉理论以及“看作”（seeing as）理论。信念/愿望理论把情绪分析成信念和愿望的组合（马克斯［Marks］，1982）。例如，希望被分析成某个事态是可能的信念和想让它成为现实的愿望。混合的感受理论——例如，上面讨论的格林斯潘的理论——将情绪分析成命题态度和感受的组合。感受的成分常用来区分冷认知（cold cognition）和热（情绪性）认识，而且在有些理论中常用来区别积极情绪和消极情绪。情绪的特定同一性是由命题态度成分界定的。最后，日益流行的“看作”方法认为，除非主体以正确的方法“看”对象，否则主体对于对象的信念和愿望还不足以构成情绪。一件有代表性的奇闻讲的是一位登山者，据说当她在将爬山看作可怕的与愉快的之间摇摆不定时，她持有相同的信念和愿望。这种方案的早期

版本沿着维特根斯坦后期著作所开创的某些路径，倾向于把“看作”当成原始概念（莱昂斯［Lyons］，1980）。而当代版本则依据认知中的注意现象来分析“看作”。情绪是认知中的这样一些倾向，它们使注意力指向某些信息源而非另一些信息源，或者它们引导对一种考虑而非另一种考虑给予更高的权重，从而导致了在缺乏情绪时不会发生的行动（卡尔霍恩［Calhoun］，1984；德索萨［De Sousa］，1987）。

在情绪哲学中，命题态度学派主要关注情绪是否是“合理的”，这指的是一种情绪性反应相对于引发它的刺激而能被判断为正确或是错误。感受的情绪理论因将情绪置于理性评价的领域之外而受到指责。这被看成是一个（更宽广的、易招致不满的）将道德领域与理性领域分开的倾向之一部分。最简单的判断主义理论通过将情绪等同于信念而把它们带回了理性领域。如果构成它的评价性信念是由主体能得到的证据证成的，那么，一种情绪就是合理的。更复杂的命题态度理论对情绪的合理性作出了更复杂的解释。信念/愿望理论遇到的困难是：关于合理性的形式解释——例如决策理论——仅限于给手段对于结果的适用性作出评价，而且它们把结果（愿望）看成是给定的。因此，这些理论必须对什么愿望是合理的作出一种说明。混合的感受理论能够用与刚才提到的理论相同的方法对具有某种特定情绪的合理性作出评价，因为情绪的同一性只是由其命题态度成分决定的。然而，状态是否首先是一种情绪，这取决于感受成分，因此，混合的感受理论必须对什么时候认为一个人的认知是热认知而不是冷认知给出某种说明。“看作”理论有其自身的困难，例如对处于愤怒或者爱之下会知觉到什么作出非循环的说明，但它们有一些有希望的资源能被有效地运用于合理性问题。根据这种理论，我们可以对构成情绪的认知偏见在产生真信念、成功的行动等方面的启发价值作出评价，并且断定，如果它们在这些方面是成功的，那么它们就是合理的。

12.2 进化的情绪理论

12.2.1 达尔文与情绪

几个世纪以来，面部表达一直是解剖学精心研究的主题，其一般的目的是为绘画和雕塑提供帮助。这一传统为查尔斯·达尔文的《人类和动物的表情》（*The Expression of the Emotions in Man and Animals*，1872）

提供了大量解剖学资料。达尔文从19世纪30年代末的M笔记本和N笔记本开始，就一直收集有关情绪的资料，而且他最初的设想是把这些资料放到《人类的由来》（*The Descent of Man*，1871）中。因此，这两本书是密切相关的。在《人类的由来》中，达尔文旨在证明动物的社会行为与人类的道德以及动物的美感与人类的美感之间进化的连续性。在《人类和动物的表情》中，他旨在证明人类的面部表情与动物的面部表情的进化的连续性，从而含蓄地证明支撑这些表情的进化的连续性。两本书的直接目的都是要证明：人类与动物在每个方面的区别都只是程度不同，因此人类可能是从更简单的前辈进化来的。在《人类和动物的表情》的序言中，达尔文明确地把查尔斯·贝尔爵士（Sir Charles Bell）的主张——人类面部的肌肉是上帝创造来表达人类情绪的——作为批判的对象。达尔文认为，现在表达情绪的很多动作都是以往生活方式的遗迹："人类的某些表达，例如在极度恐惧时头发直竖，或者在狂怒的影响下露出牙齿，除了相信人类曾经在很低等且类似动物的条件下生活过之外，我们难以得到合理的理解。"（1872：12）

达尔文认为，表情可以通过三个互补的进化原则来理解。其中最重要的是"有用的联想性习惯原则"（principle of serviceable associated habits），它是达尔文的本能理论在情绪研究中的直接运用。达尔文相信，本能行为来自通过心理强化而获得的习惯。很多世代稳定地获得相同的习惯会变成一个通过获得性特征遗传的可遗传行为或本能行为，达尔文对此坚信不疑。多数与特定情绪相关的不同行为，例如害怕时头发直竖等都反映了如今已消失的生活方式，在这些生活方式中，这些行为在每一代都会受到奖赏或强化，最终它们作为本能成为了遗传物的一部分。达尔文还为这个原则补充了另外两个原则，即"对立原则"（principle of antithesis）和"直接作用原则"（principle of direct action）。他的对立原则假定了一种对立的感受状态会导致对立的行为的本能倾向。达尔文对一条温顺的狗作了这样的评论：

> 在这些如此明显地表达了情感的动作当中，没有一种对这个动物有丝毫的直接用处。在我看来，它们只能这样来解释，即它们是同这种表达愤怒的态度完全相反或对立的。（1872：51）

达尔文把应用这两个原则之后所未涉及的行为解释为神经系统的"直接作用"的结果。在一个情绪过程中聚集起来的过剩能量会在诸如流汗和发抖

之类的行为中释放出来，而且这种能量释放只服从这样的理由，即它必须到达某个地方，在生理上这些孔道对于它的释放来说是有利的。

12.2.2 古典生态学中的情绪

本能行为的概念在20世纪20年代和30年代不太流行，那时，行为主
292 义是比较心理学中占主导地位的流派。本能行为的概念是被古典生态学的创始人复活的，他们认为自己是达尔文的心理进化研究的直接传人（洛伦兹，1965）。他们对表情的进化的解释保留了达尔文的原则，但对它们作了重新解释，以便与20世纪30年代产生于达尔文主义和孟德尔遗传学的“现代综合”中的进化理论相适应。“有用的联想性习惯原则”变成了“仪式化”（ritualization）和“派生活动”（derived activity）这些生态学概念（尼科·廷伯根［Niko Tinbergen］，1952）。派生活动是指最初进化是为了某一个特定目的，但后来被选择却是由于另外一个目的的行为。仪式化行为建基于某些派生活动，它们最初进化是为了执行某项实践的功能，但后来被选择却是为了执行信号功能。因此，尽管恐惧和愤怒时的毛发竖立并不会使一个人看起来比对手更庞大，但它确实传达了人类的情绪状态。派生活动需要一种特殊的进化论解释模式，不能仅仅根据它们当前执行的功能和保持其种群的选择压力来理解他们。就信号来说，这一点尤其明显。毛骨悚然作为恐惧的信号本质上并不比微笑或者大笑更好，这种特殊的行为作为一种信号被选择，仅仅是因为它在遥远的过去已经和某些情绪状态联系在了一起。它与那些状态联系在一起，不是因为它是一个信号，而是因为它使动物看起来更庞大。仪式化的概念可以使生态学重构达尔文的有用的联想性习惯原则，同时避开他关于获得性特征遗传的承诺。达尔文关于有机体获得的好处——它们使某些情绪行为变成了习惯——的大多数描述，和关于**选择优势**——它导致了那些行为通过自然选择的进化——的描述同样是合理的。达尔文的其他两个原则同样可以作出重新解释。对立原则是根据无歧义的信号的选择价值来解释的。对一条狗来说，发出它想避免冲突的信号与发出要攻击的信号同样重要。因此，可能有些行为之被选择只是由于它们看起来不同于发出攻击信号的行为。直接作用原则变成了替换活动（displacement activity）这个生态学概念。早期的生态学研究者分享了达尔文的这一看法，即本能的动机引起了一种积聚的心理能量，它必定会在某种行为中得到释放。通常研究者给出的例子是一只愤怒的猫，它不愿意攻击并开始洗澡。尼科·廷伯根评论道：“我认为，替换——通

过一个安全阀——确实充当了作为危险的过剩冲动的出口的功能，这是可能的。”（1952：23）这种对达尔文三原则的大规模重新解释进行得如此顺利，而且还允许保留原始研究如此多的细节，以至于早期的生态学研究者似乎完全没有意识到达尔文的理论与他们自己理论之间的差别。

古典生态学既对情绪本身提供了一种说明——这种说明用奥斯卡·海因罗斯（Oskar Heinroth）的隽语概括就是：“我把动物看成是智慧很少且非常情绪化的人类。”（洛伦兹，1966：180）也改进了达尔文对表情的解释。洛伦兹及其早期的追随者相信，动物行为是围绕一定数量的天赋行为序列——它们在合适的触发刺激出现时作为一个单元而被执行——组织起来的。和早期的本能理论家相反，洛伦兹否认动物具有寻求动物行为的实际进化目标——营养、庇护、生殖等——的动机。相反，动物的动机是执行特有的天赋行为，例如收集筑巢的原料或者织造一个巢，这些不为它们所知的行为将使它们获得庇护所和其他增强适应性的目标（洛伦兹，1957[1937]）。情绪是执行这些天赋的行为模式的心理伴随物。因此，鸟用一种刻板的、该物种特有的颈部动作向巢上插小树枝时会感受到一种令人满足的情绪（洛伦兹，1957：138）。之前使它能执行这种令人满足的动作行为会受这种动作的强化，并在将来更频繁地做这个动作。相反，野生火鸡对其空中捕食者所作出的反应会伴随一种消极的情绪，它使火鸡在将来会主动避开和执行那种行为模式相关的环境。洛伦兹情绪理论一个最与众不同的原则是：动物所具有的情绪的种类比人类的多得多（洛伦兹，1957：163）。洛伦兹认为，动物在执行一个愉快的天赋行为（比如捕捉猎物或宣示领土）之前，经常会有“欲望行为”（appetitive behavior），动物以此来主动地找出能释放这种天赋行为模式的“释放器”。在人类身上，天赋的行为序列越来越萎缩，而欲望行为则被精心设计成了智能的、目的指向的行为。尽管一只鸟筑巢是由于它在某种激素状态下发现收集小树枝是有益的，而且——毫不相关地——踩踏收集就绪的小树枝是有益的，而人建造栖身之所则是一种目的指向的行为，由此他能够获得单一的、有益的“在家”的感觉。人类丧失了极其多如此特殊的天赋行为同时也意味着很多极其特殊的情绪的丧失。我们并非对空中的猎食威胁有一种情绪反应，同时对陆地上的猎食威胁有另一种不同的情绪反应，事实上只存在一种恐惧情绪（洛伦兹，1957：163）。同样，当另一种灵长类动物面对威胁其统治权或挑战其权威时可能有不同的情绪，人类却只有一种愤怒情绪。

洛伦兹及其早期追随者的情绪理论并未幸免于 20 世纪 60 年代整个古

典生态学的动机理论（即所谓的“水力模型”[hydraulic model]）的反驳。然而，这种思想，即情绪感受在某种内在条件作用过程中起着关键性的作用，却是很多当代理论的重要组成部分，例如下面讨论的安东尼奥·达马西奥（Antonio Damasio）的理论。自洛伦兹以来，另一种一直流行的思想是：情绪在种系发生学上是一种古老的行为控制形式，尽管后来有智能行为的进化，但其中的某些部分仍然在人身上保留了下来。最后，如下面所讨论的，生态学家罗伯特·海因德（Robert Hinde）用于反对洛伦兹理论的一些论证也指出了一种看待情绪的全新方式。

12.2.3 埃克曼与“基本情绪”

在20世纪70年代以前，心理学和人类学领域有一个非常一致的共识，即各文化之间的人类情绪存在普遍的差异。但是，动物行为界的很多科学家却持截然相反的观点，他们相信，文化上特有的情绪状态是用一种文化上特有的面部表情和手势的符号（code）——它是个体在抚养期间获得的——表示的。这种文化主义传统在20世纪60年代末被达尔文方法在心理学自身内部的强有力复兴取代了。如今，保罗·埃克曼（Paul Ekman，1972）及其合作者的研究重新发掘了一项前人业已提及的共识：有些“基本情绪”能在所有人类文化中找到。

一个著名的实验使用的被试来自于新几内亚的沸尔（Fore）语族，他们在此之前几乎从未接触过西方人及其文化产品。这些被试都被分发了三张照片，其中每一张都显示一张面孔，研究者向这些被试讲述了一则包括一种情绪的故事。实验人员要求他们挑选出显示故事中这个人情绪的照片。这种设计的好处是不需要翻译情绪的名称。被试们在挑选合适表情的照片方面非常成功。实验人员还要求这些新几内亚被试表演故事中所描述的人的面部行为，然后，有关他们面部反应的录像带向美国大学生放映，学生们一般能准确地判断新几内亚岛人想表达的情绪。大约在同一时间，人类生态学家证明这些表情中的一部分很早就出现在人类婴儿身上（艾布尔-艾贝斯费尔特［Eibl-Eibesfeldt］，1973），而灵长类动物学家重申了人类的面部表情与非人的灵长类的面部表情之间的同源性（谢瓦利埃-斯科尔尼科夫［Chevalier-Skolnicoff］，1973）。其中普遍认可的“基本情绪”有：恐惧、愤怒、惊奇、悲伤、高兴和厌恶，这个名单中的每个词语都指称一种短暂而特殊的面部表情。

埃克曼（1984）把面部表情看成是**情感程序**（affect programs）的成

分。每种基本情绪都对应着一个储存于大脑某个部位的情感程序。这个程序被激活时就能协调一大群行为，包括面部表情、自主神经系统的变化、富于表情的声音的变化以及像畏缩或定向这样的肌肉—骨骼反应。情感程序这个概念则继承了天赋的行为序列这一早期生态学概念的很多特征。这两个概念都表明，有些表面复杂的行为实际上是行为的原子单位，当它们被合适的引发刺激触发时，会按同一个刻板的序列展开。埃克曼把这种引发情感程序的机制称为“自动评价机制”（automatic appraisal mechanism）。这是一个专门的神经系统，它将自己独特的刺激评价规则应用于一个有界的数据集——它来自于知觉信息加工的最初阶段。如果放在一起考虑，在最近的进化心理学家所支持的意义上，评价机制和情感程序就形成了一种认知模块（巴可［Barkow］等，1992）。

情感程序理论是与一种关于情绪系统的进化理论相伴生的。这个系统被解释成一种古老的认知形式，它最初是自主运转，后来被增加了高级认知功能。这种观点得到了神经科学家保罗·D·麦克莱恩（Paul D. MacLean，1952）的“三位一体脑”（triune brain）理论的支持。根据这种理论，情绪位于大脑的“古哺乳类脑”（paleomammalian）的部分，而高级认知功能是在新近进化出来的“新哺乳类脑”（neomammalian）结构中实现的。这些古老的行为控制形式在灵长类动物身上幸存下来，可以根据它们设置故障安全反应——它确保至关重要的行为在必要时会被执行，即使这意味着它们被执行得过于频繁——的价值来解释。情绪系统是一个原始但可靠的故障安全机制集合的观点在当代神经科学中仍有影响力（潘克沙普［Panksepp］，1998）。

埃克曼关于基本情绪的说明是对常规的一种彻底背离，这不仅是由于他早期对情绪中的文化差异的强调，而且是因为他重新采用了关于情绪本身的类型学说明。在情绪行为之下，有一定数量的离散情绪。在这个方面，埃克曼的研究显示出了西尔万·S·汤姆金斯（Silvan S. Tomkins）的影响，汤姆金斯将情绪当作“首要激励因素”（primary motivator）的再次论证，有助于使情绪重新变成主流心理学的一个课题（汤姆金斯，1962）。关于基本情绪理论的一条最持久的批评来自于这些理论家，他们认为情绪状态不属于离散的类型，而是沿着很多像愉快和觉醒这样的坐标轴或多或少连续地分布（罗素和费尔［Russell and Fehr］，1987；罗素，1997）。

12.2.4 社会生物学与情绪

20世纪70年代到80年代，生物社会学提出了一种与情绪进化有关的新观点。它将研究的焦点从基本情绪转移到了人类社会相互作用中涉及道德和准道德的情绪。像信任、忠诚、内疚和羞耻等情绪在调解竞争性的社会交往——它们是多数人类社会生物学研究的焦点——中发挥着一种明显的作用，很多社会生物学家都对此作了简要评论，大意是说：道德的情绪必定是作为执行稳定进化的社会相互作用策略的心理学机制而进化的（温里克［Weinrich］，1980）。罗伯特·A·弗兰克（Robert A. Frank，1988）指出，道德的情绪是作为“承诺问题”（commitment problem）的解决办法而进化的。当进化的相互作用中的制胜策略（winning strategy）涉及作出一个有约束力但有条件做某件事的承诺——如果这个条件被满足，它就会违背自身的利益——时，就会出现承诺问题。如果这种承诺是可信的，那么就需要某种特殊的机制，它会使有机体去违反自身的利益。弗兰克指
296 出，像愤怒和报复心这类情绪的进化，使得有机体能作出有效的威吓，如表示要用毁灭自身的进攻来阻止更强大的入侵者。相反，像爱和内疚这类情绪的进化，则能让有机体在某些条件中作出互惠的利他行为，例如即使一个没有给予回报，另一个也不可能有报复行为发生。关于情绪的博弈论解释说明，弗兰克的贡献对近来的道德心理学有相当大的影响（吉伯德［Gibbard］，1990）。

12.2.5 狭义进化心理学与情绪

“进化心理学”这个术语通常是在狭义上使用的，用以指称约翰·图比（John Tooby）和勒达·考斯麦茨（Leda Cosmides）所拥护的一种特殊方法（巴可等，1992）。心灵是由高度专门化的、领域特殊的认知装置或模块组成的一个集合，其中每一个都与我们进化史上的一个特殊的生态学问题相适应。与情绪系统一样，这些模块对特殊种类的数据进行操作，它们使用的算法与其他模块所使用的不同。因此，进化心理学认可关于基本情绪的情感程序理论，但是想走得更远，这既是通过增强情感程序的复杂性，又是通过寻找支撑其他情绪行为的模块机制来实现的（图比和考斯麦茨，1990；考斯麦茨和图比，2000）。大卫·巴斯（David Buss）认为存在一种性忌妒模块——是图比和考斯麦茨预测的其他模块之一。巴斯认为，性忌妒具有简单的知觉激发因素，例如，奇特的气味、变异的性行为、过多的目光接触以及违反控制个人空间的规则（2000：45）。忌妒模块使用

了专用的算法而且和基本情绪一样起着故障安全机制的作用，即“被用来不仅在发现了背叛的时候，而且在环境使背叛更可能发生的时候发出警报”（巴斯，2000：224）。这个模块产生了各种形式的反对雌性性伴侣的暴行，包括那种在祖先存在的环境中适当的谋杀。这些新近的理论与更传统的情绪进化解释之间一个值得注意的差异是缺乏这一思想，即情绪代表了一种更原始的行为控制形式，它能够与理性的、有计划的行动形成对比。情绪仅仅被看成是另一种认知模块，它反映了进化环境的适应性细节。这一思想，即存在一些基本情绪而且这些基本情绪是更复杂的情绪反应的成分，已经受到了一些进化心理学家的批评，因此他们认为每种情绪都是一个用来解决某个独特进化问题的特殊模块，所有情绪同样都是基本的。据说，甚至保罗·埃克曼也成为了“标准的社会科学模型”的受害者，他没有意识到人类情绪生活的一切方面都同样能得到进化的解释（高林和麦克伯尼［Gaulin and Mcburney］，2002：265-267）。

12.2.6 相互影响的情绪理论

进化论也被用于为相互影响的情绪观（transaction view of emotion）作辩护，根据这种观点，情绪是在有机体之间的社会相互作用中所作出的“举动”（moves）。这种思想，即情绪性行为是离散的、潜在的情绪的表达，受到了相互影响论者——根据进化的理由和动物的情绪模型——的质疑。相互影响的情绪观可以追溯到罗伯特·A·海因德的研究，他在动物行为研究的生态学传统的发展中是一个重要人物。从20世纪50年代中期开始，海因德（1956）就认为洛伦兹和廷伯根依据特定行动的内驱力而提出的古典的动物动机模型已经过时了。到20世纪60年代末，依据假定的潜在机制对动物行为的分析已经被行为本身在动物之间以及动物与其环境之间相互作用的适应模型取代了。以前被看成是本能的内驱力的表达行为，现在被看成是动物未来的可能行为或其动机状态的信号。但是进化的博弈理论向情绪化行为的应用预示着：它将是被设计来操作其他有机体的期望，而不是明显“表达”有机体的真实动机状态。情绪性行为不是表达动物的潜在动机，而是发送动物动机的信号，这种信号是可信的，而且被其他有机体接收后是有利的：“［恐吓］信号只有在发出恐吓的个体试图虚张声势、欺骗或者操作对手时才有意义……否则对于接下来做什么就是不确定的，因为它应当做什么部分地依赖于对手的行为。”（海因德，1985b：989）这些关于动物交往的思想到20世纪80年代已经是老生常谈了，但是

海因德用它们来设问：关于人类情绪的民间心理学对于研究动物行为——它们看起来与人类的情绪性行为是同源的——是否是一个好的出发点？民间心理学使我们以为，作出宣示领土的攻击性行为的动物正在“感受愤怒”。它也试图表明，正是这个基本的刺激情景——领土上的一次入侵——引起了愤怒，而且这种愤怒并未涉及一种控制或抑制它的心理努力——这是很困难而且只能部分成功的事情。最后，民间心理学表明，正是这同一种愤怒状态促使这个动物在宣示之后立即对入侵者发动了攻击。海因德指出，尽管动物的有些情绪性行为符合这些预期，但大部分却并不符合。他认为，宣示领土的行为是矛盾动机的一种征兆，与其说它是攻击的表达，不如说它是决定着动物是否具有攻击性过程的一部分。最重要的是，社会背景和行为的可能效应不仅决定动物是表达还是抑制它的“真实感觉”，而且实际地决定着动物具有什么情绪。

298 尽管海因德的讨论（1985a、1985b）主要是针对非人类的动物进行的，但他显然认为这些思想也可以应用于人类的情绪，而且人们可以用动物行为研究来解除对一种作为民间心理学话语之组成部分的情绪模型的掌控，从而考虑替代的选择。海因德的思想引起了对社会认知在情绪的产生和调节中的作用感兴趣的心理学家的关注。其中最著名的可能就是阿兰·弗里德伦德（Alan Fridlund，1994、1997），他的思想主要是对埃克曼基本情绪模型的批判。弗里德伦德认为，面部表情不可能像埃克曼所指的那样是对简单刺激情景的强制性反应，因为这种强制性的信息交流经常不是为了有机体的利益。弗里德伦德认为，如果人类依据面部的信息能决定彼此的动机，那么这必定是一场“军备竞赛”的结果，在竞赛中，发信号的有机体极力隐藏他们的动机而接收者则竭尽全力发现它。弗里德伦德的论证当然和关于动物交往的博弈论文献的基本方向是一致的。然而，进化论因不能预测进化演变的过程而声名狼藉，相对于经验的可靠性研究，研究者或者对这个理论论证给予太多重视，或者对之缺乏重视，这些态度都是错误的，人们就是用它来识别彼此的情绪。相互影响论者试图用对背景在面部表情解释中起重要作用的经验研究来化解这个挑战（罗素和费尔南德斯-多尔斯［Russell and Fernández-Dols］，1997）。他们认为，观察者是根据他们对出现于社会相互作用的面容的理解来解读其情绪意义的。虽然背景很重要而且人们经常觉察不到它的作用（这是很明显的），但似乎也不容否认的是，人类和其他灵长类动物一样，确实从面部行为本身得到了与其他有机体的动机和行为倾向有关的一些信息。这可能是“军事竞赛”的结

果，在这场竞赛中，信号接收者战胜了信号发送者，但这可能在很大程度上是受另一个事实的影响，即如海因德所认识到的，交流性的相互作用不纯粹是竞争性的。进化“博弈”的范围从零和博弈延伸到几乎纯合作的博弈，而塑造面部表情的进化博弈分布于这个连续统一体的各个位置。

弗里德伦德自己的经验研究关注的是社会背景在情绪性行为产生中的作用。他和其他的相互影响论者都记录了基本情绪产生方面的观众效应，他们认为，观众效应与情感程序理论是不一致的。例如，在某个时刻发生的社会交往比正在微笑的人所感觉到的主观满意程度更有力地预言了微笑。在一系列巧妙的实验中，弗里德伦德也试图证明，孤独的面部行为表现是由一个“头脑中的观众”——潜在的社会交往者，他们是孤独者思考的焦点——之存在而得到预言的（弗里德伦德等，1990）。弗里德伦德认为，这些结果是对基本情绪理论的证伪，但是他还不能确定这些结果支持这种解释。的确，埃克曼已经论证了依据社会背景调节情绪性行为的“展示规则”（display rules）是后天习得的、文化上特殊的，而且不会干扰自动评价机制和情感程序内部的实际运转（见下文）。但是，没有其他内容阻止情感程序论者把观众效应视为进化的“情绪模块”本身的组成部分。情绪性行为在很多有机体（例如家养的鸡）身上都表现出了观众效应，在这些有机体身上，观众效应作为进化的情绪系统本身的一部分，比作为习得的行为似乎更有可能（马勒［Marler］，1997）。

在其对动物交往文献的权威性回顾中，马克·豪泽（Mark Hauser）也认为弗里德伦德和埃克曼的看法是内在一致的。他指出，弗里德伦德的论证涉及的是有关情绪性行为的生物学功能问题，而埃克曼的情感程序模型则涉及产生那种行为的机制（豪泽，1996：495-496）。这种说法无疑解释了他们某个部分的分歧。然而，在某些地方，弗里德伦德似乎确实在讨论潜在的情绪过程的本质，而不只是在讨论它们的生物学功能。当然，关于情绪更宽广的相互影响观点包含着一种对关于支撑情绪性行为的心理过程之标准思想的质疑。愤怒的人感觉到有人冒犯了他们，他们就有纠正那个错误或者获得错误赔偿的动机。**由于**行为的社会效应而作出愤怒的行为就是假愤怒。然而，相互影响的情绪理论主张：情绪是“同一性改组和关系重构的非言语策略”（帕金森［Parkinson］，1995：295）。虽然表面听起来这像是一种更时髦的思想，即情绪是“社会的建构”（习得的社会作用），但这种情绪观进化的理由和非人动物中所存在的观众效应都告诫我们：不要轻易地把情绪是社会相互影响的观点同它们是习得的或者说在各

种文化间千差万别的观点同一起来。当然，一旦情绪是策略性的社会行为这种观念与它们是习得的行为或者意向行动这种观念区别开了，那么，在很多人看来，可能相互影响观就不太显得荒谬了。

12.3 情绪的普遍性

12.3.1 它为什么重要

人们普遍认为，情绪是道德动因和审美反应的一个关键因素。有一种观点认为，所有健康的人都能表现出并认识到相同的情绪，并能对之作出反应。这常被当作是支持道德和审美判断可能具有普遍有效性观点的根据。相反，如果人类的情绪像包含在不同语言中的情绪概念那样五花八门，如果人类只能理解他们自己文化群体中的表达技能，那么这似乎会倒向关于伦理学和美学的文化相对主义。

12.3.2 埃克曼的“神经文化理论”

埃克曼及其合作者用**展示规则**这个概念——是在他们另一个著名的实验中所例示的一个概念——处理基本情绪表达中的文化差异（埃克曼，1971、1972）。在25个美国大学生和25个日本大学生独处时，实验者将让人紧张的黑白影片放映给他们看。在这种持久的紧张状态中，这两组被试所表现出来的全部面部行为都非常相似。然而，当一名实验人员被带到房间，并允许学生询问与再次放映使人紧张的影片的情绪有关的问题时，日本大学生的面部行为就与美国大学生的截然不同了，录像带显示，日本大学生的消极表情瞬间出现之后就被礼貌性的微笑代替了。这例示了文化差异的展示规则概念化的一个重要特征：进化的表情保持不变，而是与文化上特定的行为相互作用来决定可观察的面部动作模式。假扮情绪的企图会“遗漏掉”无意识的情绪反应。要想压抑情绪行为，只能同时使用服务于其他目的的表情所关涉的肌肉才能进行，它们不能干扰情绪系统的实际运转。我前面讨论了这种可能性，即社会背景在这个系统的实际运转中可能起着一种作用。当然，这和刚才概述的实验中所例示的展示规则的进一步操作是完全一致的。

12.3.3 情绪的社会建构论

关于情绪的文化相对主义在20世纪80年代之复兴，是作为对心理现象的社会建构的更广泛兴趣的一部分。这次复兴使分析哲学家第一次真正参与到了有关普遍性的争论之中，因为关于社会建构论的新论证既是概念的也是经验的（所罗门［Solomon］，1984；哈勒［Harré］，1986）。一个有影响的论证是从这一被普遍接受的观念开始的：情绪涉及到对刺激的认知评价。人们认为，在这种情况下，刺激如何被表征的文化差异将导致情绪方面的文化差异。如果两种文化对危险的看法不同，那么，由于恐惧涉及到对危险的刺激的评价，因此恐惧在这两种文化中就是不同的情绪。埃克曼的基本情绪理论的拥护者未被这个论证说服，他们仍是根据其行为和生理的特征来定义情绪，而且承认在不同文化中触动相同情绪的东西是有很大差异的。社会建构论者定义情绪范围的方式也使基本情绪研究不太相干。有六七种基本情绪似乎需要对刺激作出最低限度的认知评价。社会建构论者通常反对把这些生理反应看成是情绪本身，他们保留了这个表示参与到这一社会情景——这些生理反应在其中可能被描述成（例如）愤怒或忌妒——中的人的更广泛认知状态的术语。因此，还不能确定建构论者与其普遍论对手之间的争论是否仅限于语义的争论。一方偏爱容易处理的还原解释，即使这些解释涉及的范围有限；而另一方关注科学可能忽略了人类情绪的社会和文化的那一面。

12.3.4 普遍性争论中的概念混淆

埃克曼的研究及随后的讨论有助于澄清与情绪的普遍性有关的一些问题。情感程序在各种文化中有相同的输出，但没有相同的输入。情感程序在孩童时期有一些普遍的引发因素，例如突然的巨大声响会引起恐惧。情感程序反应的条件作用中也存在系统的偏见，它们可能导致引发成年人的反应条件方面的趋同（奥曼［Öhman］，1993）。然而，一般的图景是：情感程序是与任何局部地执行一种明显的功能作用的刺激联系在一起的，因而恐惧的情感程序也是与任何局部地构成一种威胁的东西联系在一起的，厌恶反应是与任何局部看起来是有害的或不干净的东西联系在一起的，等等。因此，基本情绪的普遍性并不意味着导致情绪的东西就没有文化的差异。我们通过对下述两个问题作出区别就可以进一步作出说明：一个问题是情绪是否是跨文化的（能在一切文化中找到）；另一个问题是情绪是否是单态的（monomorphic）（能在所有健康的个体身上找到）。这类通常由

普遍论者收集到的证据可用以证明：情绪是跨文化的，而且与单态性问题几乎没有关系。情绪可能是跨文化的，但仍然像血型或者眼睛的颜色一样，每个人群中都有几类不同的个体。关于社会情绪进化的模型通常都预测：竞争的类型通过竞争将被保存在相同的人群中。令人吃惊的是，情绪是否已经进化的问题仍和另一问题，即是否有唯一的、普遍的人类情绪的本质紧密地联系在一起。

人们也能通过抛弃最后残存的习得行为与天赋行为之间的传统二分法来澄清普遍性的争论。情感程序理论的一些批评者认为，关于情绪的生物学观点是不合适的，这只是因为情绪性反应的出现和保持依赖于环境因素（拉特纳［Ratner］，1989）。但支持者一般认为，情绪是跨文化的，因而可能是进化的产物。这一证据意味着这些情绪是遗传决定的而且抵制随着环境的变化而改变。这些推理忽视了这些事实，即环境甚至在最老套的生物特性——例如躯体形态学或性行为——的发展中也起着大量建设性作用。和进化心理学的其他部分一样，进化的情绪为了建构和保存它们自身可能使用了很多发展中的儿童环境的可靠特征。作为这些特征的变化结果，它们会有文化的和个体的变化，也会经历遗传变异。狭义进化心理学家已经接受了这一思想并且指出，文化之间的心理学差异可能代表着能在关于由进化设计的一种发展程序内部得到不同选择。然而，这一思想并不承认环境变化会产生这样的情绪表现类型，即它们没有特殊的进化历史，因此没有构成进化的发展程序的一部分。为了避开这个困难，我已经提出的普遍性问题通常可以根据达尔文的**同源性**（homology）概念得到有效的重构（格里菲思，1997：135）。如果两种情绪性反应在那些个体或文化的共同祖先那里是一种反应的改进形式，那么它们就是同源的。使用同源性概念可以避免关于两种反应必须相似到什么程度才能算是“同一种”反应的毫无结果的争论。如果两种反应是同源的，那么它们就分享了同一个进化历史，而且不管它们从那时起有多大的差异，所分享的这个历史就能用来解释它们保留下来的共同特征。

12.4 认知科学中的情绪

12.4.1 感受理论的复兴

近年来，认知神经科学研究对情绪和认知之间的关系作出了新的解

释，并导致了感受的情绪理论的复兴。安东尼奥·达马西奥认为，实践推理依赖于经验情绪的能力。患有前额皮层双侧损伤的病人既表现出情绪性心理减弱，又表现出以解决现实世界问题的方式分配认知资源的能力减弱，然而，他们在抽象的推理能力方面并没有缺陷。达马西奥（1994）认为这些发现表明：情绪在对数据及其与手头任务的相关性的目标进行分类时起着至关重要的作用。这些看法引起了认知科学家的兴趣，他们在“情感计算”中看到了“框架问题”（frame problem）——选择所有相关的数据，却不对所有与可能的相关性有关的可用数据作出评价——一种可能的解决办法（皮卡德［Picard］，1997）。达马西奥的理论和一些哲学的“看作”理论存在相似性，即都将情绪等同于认知中的启发式偏见（heuristic biases）。然而，与那些理论相反，达马西奥是把情绪本身看成感受。这一点很重要，因为如果情绪发挥认知的作用，那么他的看法就是：认知的优先性是通过计算最相关、最重要的东西来指派的。这不是框架问题的解决办法，而是它的一个实例。达马西奥通过给予情绪感受以认知优先地位来避开这个陷阱。他描述了一类“首要情绪”（primary emotions），它们和埃克曼的基本情绪非常相似。达马西奥把情绪的发展想象成一个过程，在这个过程中，与基本情绪联系在一起的感受，与他称之为“次要情绪”（secondary emotions）的认知/感受合成物的特殊认知状态联系在一起。不过到目前为止，达马西奥对他的理论只给出了一种提示性的概括，这个概括是否能发展成一个可行的认知过程模型还有待研究。迄今，各种扩展达马西奥思想的尝试都与传统的行为条件作用相似——用思想来代替充当强化刺激的行为和情绪感受。诚然，条件作用模型作为复杂的认知表现的解释的局限性是大家所熟知的。

12.4.2 对双路径情绪模型的神经学支持

20世纪80年代情绪理论中最激烈的争论之一是围绕罗伯特·扎伊翁茨（Robert Zajonc，1980）的“情感首要性论题”（affective primacy thesis）展开的。扎伊翁茨指出，被试能够获得对于下意识刺激的偏好，而在向它们呈现更长时间的刺激时，它们却没能表现出识别那些刺激的能力。他认为，在正常情况下，两条不同的路径通向了情绪性反应和有意识的觉知以及回忆这样典型的认知反应。扎伊翁茨的主张是有争议的，这是由于另一种看法流传很广：情绪实质上涉及对刺激的“评价”，即人们所认为的典型认知过程的东西（拉扎勒斯［Lazarus］，1982；拉扎勒斯等，1984；

扎伊翁茨，1984）。扎伊翁茨关于通向认知和情绪的双路径概念显然类似于埃克曼的看法，即“自动评价机制”与基本情绪联系在一起，而且它的运行并不依赖于形成关于刺激情景的有意识的或可报告的判断。近年来，约瑟夫·勒杜（Joseph LeDoux，1996）对恐惧条件作用中所涉及的神经通路所绘制的详细地图已经证实了与扎伊翁茨关于恐惧的双路径模型相类似的东西：关于刺激的信息通过一条穿过皮层下结构——其中扁桃腺（amlygdala）尤其重要——的快速的“低层通道”（low road）激活了很多方面的情绪性反应；另一条更缓慢的“高层通道”（high road）激活了皮层结构，而且它对于相同刺激的更长期的、有计划的，而且经常是有意识的反应是必不可少的。勒杜的发现表明，至少对某些基本情绪来说，情绪涉及对刺激的认知评价这一思想需要被另一种思想取而代之，即情绪涉及两种评价，它们是冲突的而且具有互补但独立的认知功能。双路径模型也为支持很多进化的解释（即它们把基本情绪看成是普通生存问题的“匆忙而草率的”解决办法）提供了一些根据。

12.5 情绪是一个自然种类吗？

达马西奥将情绪定义为“由特殊原因引发的一种有机体状态转换”（1999：282）。弗里德伦德在遇到类似的定义时谈到：“这里的逻辑问题是什么**不是**情绪。情绪实际上已经代替柏格森的**生命冲动**（élan vital）和弗洛伊德的**力比多**（libido）而成为了所有人类生活的能量基础。”（1994：185）对很多理论家来说，情绪总体上确实是和动机同义的。达马西奥完全了解这一事实，而且为了在他所理解的概念变革时期便于交流，他还自觉地使用了一个熟悉的术语（达马西奥，1999：341）。例如，假如将心境（mood）与情绪区别开或者除了依据实际的情绪理论之外确定（有些?）愿望是不是情绪是极其困难的，那么将达马西奥的宽泛定义作为研究起点就是值得推荐的。不过，我在其他地方已经证明，对情感现象领域的科学研究受到了这种持续信念的阻碍，即“情绪”是一种统一的心理状态（格里菲思，1997）。科学的目的是将现象归类于“自然种类”：即可以发现很多可靠概括的范畴。民间心理学的情绪领域是如此五花八门，以至于该领域中所有心理状态不可能形成一个自然种类。故而，关于情绪的可靠概括即使有也将非常少，或者换句话说，没有一般的情绪理论。科学的进步是通

过对这个领域进行划分并研究可能构成了自然种类的现象类型而进行的，就像在记忆研究中所发生的那样。现在，代替情绪概念将需要更特殊的概念，而情绪哲学家的核心作用是推动这种概念的不断修正。

多数情绪哲学家都没有看到情绪范畴的严重问题，但他们都承认它是含糊的，涵盖了多种多样的现象。他们关注的是日常语言中“情绪”这个词以及它背后的概念。情绪概念的哲学分析与研究“情绪”这个术语在西方文化中的用法的心理学家所开展的分析具有合理的一致性（费尔和罗素，1984）。有一些明显的情绪范例，例如爱、幸福、愤怒、恐惧和悲伤，而且多数哲学家的情绪定义都涵盖了这些。他们的定义在相同的案例上是有分歧的，它们在经验研究中导致了被试之间的分歧，这些案例有骄傲、希望、性欲、疼痛和饥饿等。哲学的定义包括了这些特征——哲学家认为它们是情绪概念原型的一部分。情绪指向外部事态、相对短暂，并且有一个评价的方面，可以对其对象作出有吸引力还是令人厌恶的判断。多数定义也提供了情绪感受的作用。因此，哲学家像普通的说话者一样，对于什么算是一种情绪而非一种心境、一个愿望或一个意向等也可以达成合理层次的共识。心理状态是否是从单一的、多产的科学研究对象集中在一起的，其他文化是否是以相同的方式形成情绪的概念，这些问题还都有待研究。

12.6 结　论

关于情绪的哲学心理学是一个繁荣的领域，每年都有大量的书籍和论文问世。如今也存在着一种与心灵哲学更密切整合的趋势，而这种整合在关于认知、知觉和行动的哲学心理学中很常见。近年来关于情绪的进化心理学已经受到了哲学家们的关注（格里菲思，1997；霍斯特［Horst］，1998；埃文斯［Evans］，2001），尽管情绪的潜能对派生于认知系统中的认知科学中的观点提出了挑战（德兰西［Delancey］，2001；埃文斯，出版中），而神经科学家根据情感神经科学的最新发展所提出的情绪理论也受到了哲学家们的详细审查（普伦兹［Prinz］，即出），指向伦理学和美学中更传统的问题的哲学研究也开始利用情感神经科学的主张，这可能是由于达马西奥的主张——即情绪和理性是分不开的——和更古老的哲学观点产生了强烈的共鸣（布莱克本［Blackburn］，1998；努斯鲍姆［Nussbaum］，2001）。

参考文献

Barkow, J. H., Cosmides, L., and Tooby, J. (eds.) (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford and New York: Oxford University Press.

Buss, D. M. (2000). *The Dangerous Passion: Why Jealousy is as Essential as Love and Sex*. New York: Simon and Schuster.

Calhoun, C. (1984). "Cognitive Emotions?" In C. Calhoun and R. C. Solomon (eds.), *What is an Emotion: Classic Readings in Philosophical Psychology*. New York: Oxford University Press.

Chevalier-Skolnikoff, S. (1973). "Facial Expression of Emotion in Non-human Primates." In P. Ekman (ed.), *Darwin and Facial Expression: A Century of Research in Review*. New York and London: Academic Press: 11–89.

Cosmides, L. and Tooby, J. (2000). "Evolutionary Psychology and the Emotions." In M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of the Emotions*, 2nd edn. New York and London: Guildford Press: 91–115.

Damasio, A. R. (1994). *Descartes Error: Emotion, Reason and the Human Brain*. New York: Grosset/Putnam.

—— (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

306 Darwin, C. (1872). *The Expressions of Emotions in Man and Animals*, New York: Philosophical Library.

—— (1981/1871). *The Descent of Man and Selection in Relation to Sex*, Facsimile of the 1st edn. Princeton: Princeton University Press.

De Sousa, R. (1987). *The Rationality of Emotions*. Cambridge, MA: MIT Press.

Deigh, J. (1994). "Cognitivism in the Theory of Emotions." *Ethics*, 104: 824–854.

Delancey, C. (2001). *Passionate Engines: What Emotions Reveal About Mind and Artificial Intelligence*. New York and Oxford: Oxford University Press.

Eibl-Eibesfeldt, I. (1973). "Expressive Behaviour of the Deaf and Blind Born." In M. von Cranach and I. Vine (eds.), *Social Communication and Movement*. London and New York: Academic Press: 163–194.

Ekman, P. (1971). "Universals and Cultural Differences in Facial Expressions of E-

motion." In J. K. Cole (ed.), *Nebraska Symposium on Motivation 4*. Lincoln, Nebraska: University of Nebraska Press: 207-283.

—— (1972). *Emotions in the Human Face*. New York: Pergamon Press.

—— (1984). "Expressions and the Nature of Emotions." In K. Scherer and P. Ekman (eds.), *Approaches to Emotions*. Hillsdale, NJ: Erlbaum.

Evans, D. (2001). *Emotion: The Science of Sentiment*. Oxford: Oxford University Press.

—— (In Press). *Rethinking Emotion: A Study in the Foundations of Mind*. Cambridge, MA: MIT Press.

Fehr, B., and Russell, J. A. (1984). "Concept of Emotion Viewed from a Prototype Perspective." *Journal of Experimental Psychology: General*, 113: 464-486.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. New York: Norton.

Fridlund, A. (1994). *Human Facial Expression: An Evolutionary View*. San Diego: Academic Press.

—— (1997). "The New Ethology of Human Facial Expressions." In J. A. Russell and J. M. Fernández-Dols (eds.), *The Psychology of Facial Expressions*. Cambridge: Cambridge University Press: 103-129.

Fridlund, A. J., Schaut, J. A., Sabini, J. P., Shenker, J. I., Hedlund, L. E., and Knauer, M. J. (1990). "Audience Effects on Solitary Faces During Imagery: Displaying to the People in Your Head." *Journal of Nonverbal Behaviour*, 14 (2): 113-137.

Gaulin, S. J. C. and McBurney, D. H. (2001). *Psychology: An Evolutionary Approach*. Upper Saddle River, NJ: Prentice Hall.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.

Greenspan, P. (1988). *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge.

Griffiths, P. E. (1989). "The Degeneration of the Cognitive Theory of Emotion." *Philosophical Psychology*, 2 (3): 297-313.

—— (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Harré, R. (1986). "An Outline of the Social Constructionist Viewpoint." In Harré (ed.), *The Social Construction of Emotion*. Oxford: Oxford University Press: 2-14.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hinde, R. A. (1956). "Ethological Models and the Concept of 'Drive'." *British*

Journal for the Philosophy of Science, 6: 321-331.

—— (1985a). "Expression and Negotiation." In G. Zivin (ed.), *The Development of Expressive Behavior*. New York: Academic Press: 103-116.

—— (1985b). "Was 'The Expression of Emotions' a Misleading Phrase?" *Animal Behaviour*, 33: 985-992.

Horst, S. (1998). "Our Animal Bodies." In P. A. French and H. K. Wettstein (eds.), *The Philosophy of Emotion*. Notre Dame, IN: University of Notre Dame Press.

Lazarus, R. S. (1982). "Thoughts on the Relations Between Emotion and Cognition." *American Psychologist*, 37: 1019-1024.

Lazarus, R. S., Coyne, J. C., and Folkman, S. (1984). "Cognition, Emotion and Motivation: Doctoring Humpty Dumpty." In K. Scherer and P. Ekman (eds.), *Approaches to Emotions*. Hillsdale, NJ: Erlbaum: 221-237.

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

Lorenz, K. (1957 [1937]). "The Nature of Instinct." In C. H. Schiller (ed.), *Instinctive Behavior: The Development of a Modern Concept*. New York: International Universities Press: 129-175.

—— (1965). *Preface to "The Expression of the Emotions in Man and Animals" by Charles Darwin*. Chicago: University of Chicago.

—— (1966). *On Aggression*, trans. by M. K. Wilson. New York: Harcourt, Brace and World.

Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.

MacLean, P. D. (1952). "Some Psychiatric Implications of Physiological Studies on Frontotemporal Portions of the Limbic System (Visceral Brain). *Electroencephlography and Clinical Neurophysiology*, 4: 407-418.

Marks, J. (1982). "A Theory of Emotions." *Philosophical Studies*, 42: 227-242.

Marler, P. (1997). "Animal Sounds and Human Faces: Do they have Anything in Common?" In J. A. Russell and J. M. Fernández-Dols (eds.), *The Psychology of Facial Expression*. Cambridge: Cambridge University Press: 133-226.

Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge and New York: Cambridge University Press.

Öman, A. (1993). "Stimulus Prepotency and Fear: Data and Theory." In N. Birbaumer and A. Öman (eds.), *The Organization of Emotion: Cognitive, Clinical and Psychological Perspectives*. Toronto: Hogrefe.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and*

Animal Emotions. Oxford and New York: Oxford University Press.

Parkinson, B. (1995). *Ideas and Realities of Emotion*. London and New York: Routledge.

Picard, R. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Prinz, J. (forthcoming). *Emotional Perception*. Oxford: Oxford University Press.

Ratner, C. (1989). "A Social Constructionist Critique of the Naturalistic Theory of Emotion." *Journal of Mind and Behaviour*, 10 (3): 211-230.

Russell, J. A. (1997). "Reading Emotion from and into Faces: Resurrecting a Dimensional-Contextual Perspective." In J. A. Russell and J. M. Fernández-Dols (eds.), *The Psychology of Facial Expression*. Cambridge: Cambridge University Press: 295-320.

Russell, J. A. and Fehr, B. (1987). "Relativity in the Perception of Emotion in Facial Expressions." *Journal of Experimental Psychology (General)*, 116: 223-237.

Russell, J. A. and Fernández-Dols, J. M. (1997). *The Psychology of Facial Expression*. Cambridge: Cambridge University Press.

Schachter, S. and Singer, J. E. (1962). "Cognitive, Social and Physiological Determinants of Emotional State." *Psychological Review*, 69: 379-399.

Solomon, R. (1976). *The Passions*. New York: Doubleday.

—— (1984). "Getting Angry: The Jamesian Theory of Emotion in Anthropology." In R. A. Schweder and R. A. LeVine (eds.), *Culture Theory: Essays on Mind, Self and Emotion*. Cambridge: Cambridge University Press: 238-254.

Tinbergen, N. (1952). "Derived Activities: Their Causation, Biological Significance, Origin and Emancipation During Evolution." *Quarterly Review of Biology*, 27 (1): 1-32.

Tomkins, S. S. (1962). *Affect, Imagery and Consciousness*. New York: Springer.

Tooby, J. and Cosmides, L. (1990). "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments." *Ethology and Sociobiology*, 11: 375-424.

Weinrich, J. D. (1980). "Towards a Sociobiological Theory of Emotions." In R. Plutchik and H. Kellerman (eds.), *Emotion: Theory, Research and Experience, vol. 1: Theories of Emotion*. New York: Academic Press: 113-140.

Zajonc, R. B. (1980). "Feeling and Thinking: Preferences Need no Inference." *American Psychologist*, 35: 151-175.

—— (1984). "On the Primacy of Affect." In K. Scherer and P. Ekman (eds.), *Approaches to Emotion*. Hillsdale, NJ: Lawrence Erlbaum Associates: 259-270.

第 13 章　人工智能与理性的多面性

安迪·克拉克（Andy Clark）

13.1　理出头绪

309 在本章的讨论中，我将集中关注人工智能（AI）和心灵哲学之间日益复杂的交织联系中的一个细小方面，即尝试解释理性思维如何在机器中成为可能的问题。在历史上，AI 和心灵哲学正是在这一关键的地方相遇的。但是，我将证明这个位置是不断变化的。因为我们关于理性的思想和理性（reason）**是**什么以及哪种机制能解释它们的构想，就处在过渡的状态之中。为了弄清这场巨变，我将对几种见解和方法作出比较，首先是被称作关于机器理性的图灵—福多构想，接着是联结主义及其基于技能的理性模型，最后转向这样一些问题，即由机器人学、关于情绪和理性的神经科学研究以及关于“生态学的合理性”的工作所提出的问题。与人们最初所看到的相比，人类理性中所包含的东西可能既多又少。

好吧，先来看基本的（而且我确实指基本的）故事。

13.2　经典形态的核心观念

我将在本章考虑的**所有**方案之共有的一个核心观念是：形式有时可以当成意义使用。这当然是一切对理性给出一种机械论解释的企图所依赖的核心见解。如果宽泛地理解，在逻辑学、图灵机、符号主义的 AI、联结主义的 AI，以及甚至是“反表征主义的”机器人学中起作用的正是这同一种把戏。这种把戏就是把某一组能通过非语义说明的属性或特征组织和协调
310 起来，从而使这样建造的一种装置在合适的环境中能表现出“语义良好行

为”（semantic good behavior）。“语义良好行为”这个术语有意涵盖了各种各样的东西。它涵盖了进行演绎推理、作出合适猜测、根据收到的输入或刺激作出适当行为等能力。任何这样的东西——（大致来说）它**看起来好像知道它正在做的事情**——都表现出语义良好行为，这类案例包括从（－A∨B，－B）推出－A的逻辑学家，包括这个人——他选择带上雨伞，因为他相信要下雨并且不希望被淋湿，包括这条狗——它选择食物而不是毒素，还包括这个机器人——它在一条腿受损后恢复平衡并继续行走。周围还有**很多**语义良好行为，我们对其中一些的理解要比其他的好得多。可是，**理性**是在哪里引起人们的注意呢？

至少可以证明，理性支配的行为是我所称的语义良好行为的一个特殊子集。例如，杰里·福多的看法是：一直到图灵的研究发表后，我们才开始对**合理性**（rationality，我将假定它是指理性支配的行为）在机器中何以可能有所了解（一个出色而简洁的陈述，参见福多，1998：204－205）。形式逻辑向我们表明，保持真值只有通过关注形式而非意义才能有保证。**不管**A和B意指什么，B都能从A并且B推出来，如果你遵循依据符号的形状和联结词定义的规则，你就不可能从正确的前提推出谬误，即使你不知道前提或结论关涉的是什么。正如福多所指出的，图灵证明：就所有这些可形式地（“根据形状”）说明的程序来说，编程合理的机器就可能代替人。

大致就是在这一点上，最初纯粹的物理主义信念的断言（即语义良好行为以种种方法始终并且处处都有一个充分解释的物质基础）变成了一种以理性支配的行为作目标的真正研究纲领。这种观念——它很快就被铭刻在经典的符号主义AI的研究纲领之中——是：理性在机器中能被解释成适当的计算过程对符号的操作，这里的符号是不可非语义地分开的项目（这些项目以形式、形状、电压等为代表），而计算过程是依据规则（它们本身在某种程度上也能被表达成符号）识别、书写和修改符号的自动的机械过程。在这些系统中，如豪格兰的著名评论：“如果你照顾句法（非语义的特征和属性），那么语义将照顾自己。”（1981：23）因此，就像透过图灵的非凡成就和后来经典AI进一步发展这副透镜所看到的，这种核心观念不但开始显得更具体，而且更特殊了。用福多的话说，它变成了这种观念：“使心灵具有理性的东西至少一部分是由于心灵对思想进行计算的能力；这时思想……被认为是有句法结构的，而这里的‘计算’指像图灵那样的形式操作。”（1998：205）

于是，用形式（宽泛地理解）代替意义这种一般观念就逐渐变成了关于阅读、书写和替换符号等受图灵机控制的看法：这种看法能在早期的 AI 研究找到完整的表述。在这里我们遇到了纽厄尔和西蒙（Newell and Simon，1976）的描述，即智能是奠基于所谓的**物理符号系统**的操作之上的：在这些系统中，可非语义识别的实体充当了特定内容的载体（因而变成了“符号”），并且承担着各种熟悉的操作（通常是根据指令复制、组合、产生和破坏符号）。例如，香克（Schank，1975）的理解故事（story-understanding）程序使用一种特殊的事件描述语言来对这种背景知识——它们是对关于简单故事的问题作出明智反应所必需的——进行编码，从而发展了一个符号数据库来帮助它“填充”遗漏的细节。

由于这种经典方法被看成是对合理的、理性引导的思维在机器中何以可能的解释，因此它就表现出了一种令人满意的直接性。通过想象每个参与分享的思想都有一种内部的符号回响，这些内部的回响都分享了这种思想结构的相关方面，它对可语义地感觉到的思想转换作出了解释（“他们享用了美餐，因此他们可能留了小费”；“天在下雨，我讨厌雨，因此我将带一把伞”）。因此，对句法敏感的过程可以凭借尊重思想之间语义关系的方式调节推理过程（从思想到思想的转换）。

13.3 非经典形态的核心观念

这种观念（理性引导的思想转换是以对内部符号串的句法驱动的操作为基础的）有一个著名的竞争对手。这个对立的观点得到了许多研究人工神经网络的科学家的支持，他们认为理性引导的思想转换基于由简单的加工元件构成的一个平行网络支撑的矢量到矢量（vector-to-vector）的转换。完整表述这一差异的全部细节是本章力不能及的（就我个人最好的尝试来说，参见克拉克，1989、1993）。但是我们至少能注意到（我认为）一个特别相关的真正差异之处。它涉及我所称的这两种方法的“最佳目标”（best target）。就经典的（类似于图灵机的）AI 来说，最佳的目标是能在**句子空间**（sentential space）中展示和模拟的合理推理。我用“句子的空间”指由携带意义的结构（得到解释的句法项目）——它们共享了句子的逻辑形式：关于有意义的元素的序列串——所居住的抽象空间，在这个空间中，不同种类的句法项目代表着不同的东西，而且在这个空间中，整体

的意义是这些项目（标记）及其序列顺序的一种功能，这包括其他标记的修改效应（例如，“天未下雨”中的“未”）。能在句子空间中重构起来的合理推论包括：福多中意的所有例子（关于选择带上雨伞，等等），定义在句子表达式之上的所有演绎推理的情形，以及所有诱导推理（基本上是有效的猜测）的情形——在这些情形下前提与结论之间的联系可以由其他句子的创造性恢复或运用来建立（就像在之前提到的香克的理解故事程序中那样）。

相反，人工神经网络方法的最佳目标是各种各样的合理“推论”，在 312
这些推理中，输入一般来说是知觉的，而输出（通常）一般来说是肌肉运动。例如，这类合理推理隐含在猫快速地对树枝的承载力作出评估，快速而优雅地跳到更安全的休息点，或者隐含在笔迹专家关于某个签名是赝品的快速直觉判定，这种判定通常是在有意识地分离出透露真相的特定迹象之前作出的。

然而，这不是说联结主义方法局限于知觉—运动领域。相反，这种观点是：它采用合理的推理（而且更宽泛地说，采用合理的选择）与它采用知觉—运动技术在结构上是连续的。（大体而言）推断和推理**在所有层次上**都被重构成模式完成（pattern-completion）和模式进化（pattern-evolution）的过程，而这些过程是由简单的加工单元群之间串联的从矢量到矢量的转换执行的。例如，接触到一个输入——它描述了一个面部长有红点的年轻人的视觉特征——的网络可能学会了产生一种相当于麻疹诊断的活动模式作为输出。这个诊断通过一个类似的机制可能开出了一张青霉素的处方。所涉及的这种从矢量到矢量的转换（在这个模型中）与我们借以执行更基本的识别和控制行为的转换是完全连续的，就像我们在识别一张熟悉的面孔或者走路时协调视觉的本体感受输入时那样。这些模式完成过程是在由简单加工单元——由数值加权链接联系起来的——组成的网络中进行的，它们初看起来完全不同于句子的 AI 模型，在后者中，医学诊断（例如）可能依赖于查阅存储的一组规则和原理。这种区别的一个重要的来源就在于联结主义系统通常**获得**联结权重——它们既充当知识储备，又充当加工引擎——的方式。这些权重是通过让系统接触一系列样本（训练的实例）而获取模型：这种体制凭借运用特殊的学习规则，导致一个**受原型支配的知识数据库的**发展（参见丘奇兰德，1989）。这实际上意味着系统学会了依据一批样本案例最突出的特征来“思考”一个领域，而且它的反应、判断和行动是由所知觉到的当前案例与样本的最典型特征和反应模

式的相似性引导的。反过来，这意味着这个系统所知道的东西几乎不能恰当地表达成关于这个领域的一组句子、规则或者命题。根据这个模型，与查阅符号数据库中的一组规则相比，作出内行的医学诊断与知道如何骑自行车更为相似。因此，一个调整好的联结主义网络可能产生这样的判断：它们是合理、合适的，但却抵制将准演绎的句子重构成一个将符号表达式作为其前提的论证的结论。根据这种观点，这些合适的反应和判断是理性
313 与合理性的基础。语形论证和推理只是被描述成这种以原型为基础的一般推理能力的特例，区别仅仅在于这里的目标与训练领域涉及的是公开的言语和文本的符号串。

因此，联结主义和经典主义的区别（至少就我所考虑的这些典型的化身中）在于它们对理性本身的看法。后者实际上把理性描述成了在准语言空间中的受符号引导的状态转换。前者实际上把理性描述成引导着同一类（通常是）非句子空间中从矢量到矢量的转换——它们也支撑着知觉—运动反应——的原型式知识的发展。然而，在这种区别之下却是一项重要的共识。两个阵营都同意合理的思想和行动涉及用内部的资源来表征突出的事态，用所设计的转换操作（它们对于那些内部表征的非语义特征是很关键的）来产生进一步的表征（在联结主义的情形下是一连串从矢量到矢量的转换），最终产生行动。

13.4 机器人学：超越核心？

完全不借助于内部基于形式的意义载体来解释合理的行动，这有可能吗？内部表征是我们生活中不可或缺的工具吗？

考虑低级的家蝇。马尔（1982：32－33，根据麦克拉姆罗克，1995：85 的报告）指出，苍蝇在维持生活时无论如何都没有编码这种知识：飞的行动需要拍打翅膀的命令。相反，苍蝇在未与地面接触时会自动地活动翅膀。因此，跳的决定就自动地导致了（通过取消足部接触）翅膀的拍打。

现在设想这种回路增加了。比如，假定“跳的决定”**本身**通过将一个“逼近的阴影”探测器直接连线到跳的神经命令上而被绕过去了。再设想这个“逼近的阴影”探测器本身不过是一个无智能的（dumb）程序，它用视觉细胞的原始输出来计算某种简单不变的知觉。最后，你设想一个非常

简单的生物，它是由很多这种基本的自动程序组成的，但通过完全相同的方法，这些程序本身都协调一致，因而它们在（一般来说）生态上合适的时候能够彼此开和关。例如，一个“消灭食物”程序可能被“某种东西逼近就跳”程序压倒了，而这个程序反过来又引起“拍打翅膀”程序，等等。大体而言，你所设想的就是罗德尼·布鲁克斯（Rodney Brooks，1991）等机器人学家所支持的“包含结构”（subsumption architecture），并且是造成像“无表征的智能”这样引发争议的文章题目和像“快速、廉价并失去控制”这样的口号（现在被用作了影片名!）的原因。

然而，我们还**完全**不清楚的是，这个故事能否（甚至原则上）简单地按比例放大以便给予我们“无表征的合理性”。首先，还不清楚我们何时应当承认某种复杂的内部状态至少构成了**某种**事件或事态的表征。苍蝇拍打翅膀程序看起来像一种简单的反射，然而，甚至在这里有的人也可能指出：鉴于这种反射回路的进化史，该回路的某些状态（由打破足和地面的接触所激活的状态）表征了这一事实，即双脚已经离开地面。通常看起来，布鲁克斯等人实际上指出的反而是某种**类型**的内部表征——即经典的AI所支持的大致的语言形式表征——的缺失。

然而，一个更根本的困难（它完全超出了“内部表征”这个术语的模糊性）涉及的是这**类**行为——它能由任何类似于反射的复杂机制作出合理解释。这里成问题的情形显然是深思熟虑的理性和抽象的思想。可能涉及的行为包括：计划明年的家庭旅行、思考美国的枪支控制问题（例如，“枪支制造商应该对生产了比已知的**合法**市场需求更多的枪支负责吗”）、在你放松地躺在泰晤士河边时靠心理意象来计算你的西班牙公寓的窗户的数量，等等。这些情形绝对不是完全一致的。但它们至少都有一个共同的特征：它们在一种很强的意义上都是“表征饥渴”（用克拉克和托里维奥[Clark and Toribio] 1994年著作中的一个术语来说）。表面上看，所有这些情形都要求大脑使用外部事态的内部**替身**，这里，一个“替代物”在这种强的意义上（参见克拉克与格拉什[Clark and Grush]，1999）是这样一个项目，它不仅是设计来携带关于某个事态的信息（例如，就像内部的回路可能携带关于苍蝇身上的打破足与地面接触的信息），而且是设计来使系统**甚至在缺乏直接物理联系的情况下**使它的行为与特定事态的特征相适应。因此，这样的系统——它必须将其活动与远端的东西（我的西班牙公寓的窗户）和非存在的东西（工具房里的妖怪）协调起来——就是一种合适的、注定要使用（强的）内部表征的东西，这种内部表征就是这些内

部状态，它们打算充当真正的替身，而不只是周围的信息携带者（关于联系和不联系问题的一些出色的讨论，参见史密斯，1996）。相反，通常几乎所有（除了斯坦［Stein］，1994；比尔［Beer］，2000）这类案例都被用于证明无表征的适应性反应，在这些案例中，相关行为不断受到来自与这一行为相适应的事态周围的输入的推动和调整。

在某种意义上，合理的行为就是由理性引导或者对理性敏感的行为。从直觉上看，这似乎涉及某种反思和评价这些选择的能力；预知后果并采取相应行动的能力。但是**这种**关于合理性的看法（“深思熟虑的合理性”）直接将合理的行动放到了“表征饥渴”箱中。显然，将来的结果不能直接引导目前的行动（比如说，就像背景光源可能直接引导光敏机器人那样）。仅仅在如下意义上后果才是有效的：系统可以用其他某种东西替换推理过程中所产生的后果。至少从表面上看，这就要求在某种相当有力的意义上使用内在表征。

13.5 情绪与理性

看起来，关于我们作出受理性引导的行为能力的机械论解释，不能抛弃关于内部替身的最基本观点：这些替身在缺乏相关的事件和事态的情况下也能引导行为和推理（我认为最好），把联结主义和真实世界中的机器人学所做工作看作是我们关于这些替身的可能本质的概念的扩展，看作是对这样一些方面的说明，正是在这里，我们的身体和环境构造、运动和主动干预都有助于改变需要大脑解决的那些问题。看起来，关于我们作出受理性引导的行为能力的机械论解释，也不能离开最基本的内部替身的观念，这些替身能指引行为和推理。例如，笔和纸的使用，可能极大地改变了需要大脑在遇到复杂的算术任务时、在规划长期战略时以及甚至在就枪支控制进行推理时所面对的问题。但是这些转换并非不需要对结构敏感的内部操作——这些操作定义于内部具有内容的载体之上；相反，它们改造了需要这种内部结构解释的问题。

然而，强调对理性敏感的思想和推理会使我们对人类认知的另一个维度的极端重要性视而不见，因为人类的理性与人类的情绪紧密地——可能也难以摆脱地——交织在一起。公正地对待这种重要的相互关系是下一代AI模型要面对的两个主要挑战之一。

情绪长期以来（至少在宽泛的康德的传统中）都被看成是理性的敌人。我们当然会说，某些判断笼罩着忌妒的阴影，是由狂怒和激情的短暂爆发而非合理的反思等推动的。然而，越来越清楚的是，情绪对合理反应的正常影响远非有害的。它们实际上最好被看成是理性自身机制的一部分。举一个著名的例子。考虑菲尼亚斯·盖奇（Phineas Gage）的情形。盖奇是一名19世纪的铁路工人，他的大脑在一次爆炸中因铁棒刺穿颅骨而受损。尽管前额皮层大面积受损，但是损伤并未影响盖奇的语言、运动技能和基本的推理能力。看起来他似乎避免了所有认知的损害。然而在随后的几年中，此前的诊断令人遗憾地被证实是不正确的。盖奇的个人和职业生活都明显地恶化。他丢了工作，陷入打架斗殴，不能规划未来和遵守正常的社会行为规范，他变成了另一个明显不成功的人。达马西奥等人（1994）对此的解释是：前额皮层的损伤妨碍了（他们所称的）一个“躯体标记”（somatic markers）系统——将一个事件的影像/轨迹与一种本能反应（根据结果，或厌恶或吸引）连接起来的大脑状态。这个标记系统自动运转（在正常主体身上），它既影响当下的反应，也影响我们最初基于进一步考虑和反思而作出的一系列选择。它也是积极的，而且关键在于当我们想象一个事件或可能的行动时，它会产生一种肯定或否定的情感信号，这种信号是在（特别是）皮肤电反应中表现自身的。人们推测，盖奇将没有这些反应，而且他不会使他的推理和思虑受躯体标记系统中自动的删减选项和影响选择的操作——它们是他在经验社会和职业的行动过程中逐渐获得的——的约束。当代研究似乎确证并澄清了这幅宽广的图画。E. V. R.（一位出现类似的前额内侧损伤的病人）和盖奇有相同的外观。尽管在标准的智商和推理测试中分数不错，但E. V. R.同样控制不了他的职业和社会生活。在一系列有趣的实验中（比彻拉［Bechera］等，1997），正常控制组和前额损伤的病人玩一种纸牌游戏，它涉及（不为被试所知）两副赢牌和两副输牌。被试可以选择从任何一副（A、B、C或D）中挑选牌。玩了一会儿之后，正常控制组确定了更好的纸牌（直接的报酬更小，但长期来看保证金更少而且更为可靠）并且很快就表现出在拿到“坏”牌时增强的皮肤电反应。有趣的是，这种皮肤反应在被试能清楚地说出偏爱更好的牌的任何理由之前就出现了。相反，E. V. R.没有出现这种皮肤反应。躯体线索的这种缺失似乎妨碍了他选择更好的牌的能力，**甚至在其有意识的心灵已经估计到所有这些牌的时候**——他将知道A和B是输牌——仍在游戏期间继续支持它们。

这里显然有很多东西需要讨论。如丘奇兰德所指出的，这些情形最好被理解成是起因于“情绪不能影响（病人的）理性和决定”（1998：241）吗？或者它是**不适当的**涉及情绪的情形——短期的回报胜于长期（但更多的）报酬吗？可能这些并非真的不相容：无论如何，正是这些现场无意识的否定性反应（由一致的皮肤电反应证明的）的缺失为认知错误敞开了大门。

似乎可以直接得出这个结论，即人类理性最好不被看成是一个无情绪的逻辑引擎——它有时与情绪爆发的战斗交织在一起——的操作。相反，真正的合理行为（人类身上）是一系列复杂而叠加的相互作用的结果，在这些相互作用中，审慎的理性和（经常是非常无意识的）带有情感的巧妙反应共同引导着行动和选择。事实上，情绪因素（至少如躯体标记假说所指出的）的功能就是帮助理性选择在各个时间点起作用。因此，躯体标记所起的作用非常类似于内部表征（宽泛地理解）；它们使我们能根据过去的经验进行投射推理。相对于使我们能从想象上探测未来的东西——它使用了关于一生中的选择和经验的知识，这种来之不易的知识都巧妙地浓缩成了一个自动情感反射网——什么更适合于被看成是理性本身机制的一部分呢？

317

13.6 整体推理

复杂情况的另一个来源涉及福多（1983：111）所说的“信念系统的整体属性”。福多认为，AI 在这里遇到了一个特殊的问题。因为人们认为合理推理的图灵机模型（回想一下前面的 13.2 节）不可避免地是**局部的**，它擅长于解释下雨的思想（句法地标记的）是怎样让位于需要一把雨伞的思想的。它也擅长于解释（已知有一些传统的假设——参见福多和派利申，1988）为什么可能思想的空间（对一个个体来说）在重组之下表现出了某种封闭性——“系统性”的属性，正是在这里能够思考 aRb 的人通常也能思考 bRa，等等。但是，福多坚持认为，当前基于 AI 的模型在遇到各种对整体更敏感的形式推理时就会彻底失败。例如，在诱导推理情形中，对某个事件的最佳解释可能隐藏在整个系统知识库的某个地方：这个知识库太大了，因而不适合任何穷举搜索的过程。福多反对通过启发法和简化假设来回避对这个问题的传统尝试（例如使用“框架”——参见明斯

基［Minsky］，1975；福多，1983：116），他认为，这只是把这个问题重新定位成了“执行控制”问题——即怎样在正确的时间找到**正确的**框架（或者无论什么）。由于带上雨伞以防雨的这个决定也潜在地对来自于知识库的某个地方的对抗性信息敏感，因此福多实际上留下了一个机械合理性的模型，它（就我所能看到的）可能并未述及任何真实但非演绎的推理情形。如福多似乎经常承认的，合理机制的福多—图灵模型仅仅在“信息封装系统”领域才最有效——通常是加工有限范围内的输入信号的知觉系统，它们据信对各种形式的自上而下且受知识驱动的推理都不敏感。人们情不自禁地会感到，这里很难是理性的居所。

考虑到这种悲观主义情景——它铭刻于福多“关于认知科学不存在性的第一定律：一个认知过程……越是整体的，人们就越不可能理解它。非常整体的过程……根本无法理解”（1983：107）——因此，发现某些理论家（丘奇兰德，1989：178；克拉克，1993：111）主张把联结主义方案作为对这个“整体上敏感的理性”问题的一种解答，就不足为奇了。福多专门对这些方法作出了反驳，因为它们不能解释系统性和对句法敏感的局部推理。但是，现在在我看来（尽管这是一个冗长的故事——参见克拉克，2002），整体的诱导推理问题实际上也确实影响着联结主义方法。大致来说，它在这里是作为路径选择和搜索问题出现的，即以下问题：如何用可能来自于知识库的某个地方的信息来塑造和改变加工流程本身的方向，以保证正确的输入探测器在正确的时间被正确的神经亚群加工。

丘奇兰德（1989）和克拉克（1993）都认为这个问题已经解决了（在联结主义的背景下），因为“生物相关方面的全部信息都是由编码的刺激本身自动存取的”（丘奇兰德，1989：187）。当然，输入探测器（回想一下 13.3 节）将沿着被表征的任何刺激维度自动地激活与探测器最符合的原型。但是这在合理响应过程中至多只是第一步。由于已经发现了这些最好的句法匹配（因为这最终仍是一个受形式驱动的过程），因此有必要看一看极其重要的信息是否存储在某个地方，它们无法存取是由于缺乏与这个探测器相匹配的表面。我认为，在福多（专门）考虑的这类情形中起最大作用的就是这一步。 *318*

好消息——我在克拉克（2002）中作了特别强调，这里就不再赘述了——是，现在看来第二步在计算上可能是可行的，这是由于神经联结主义研究与一种创新性的“二阶”搜索程序——是为了在万维网上使用而开发的——的奇特组合（克莱因伯格［Kleinberg］，1997）。这里的思想就是

要把第一遍（无智能的、模式匹配的、基于句法的）搜索与后续的搜索——它以与第一遍搜索所确认的因素相联系和相分离的模式为基础——结合起来。但是，就现在的目的来说，关键是要承认，真正在整体上易受影响的搜索过程对现有的为人类理性提供基础的计算模型所提出的特殊问题。

13.7 快速而节俭的启发法

然而，人们可能合理地反驳说：这种关于人类理性的看法是极其言过其实的。我们经常**不**想存取相关的知识；我们经常**不**选择使我们最幸福或最成功的东西；我们甚至会（接下来会承认它）犯简单的逻辑错误。然而，令人吃惊的是，我们经常做得和实际的一样好。最近关于“生态的合理性”的理论对此的解释是：我们（的大脑）使用了简单的捷径策略，它们是用以根据特定的约束条件和机会——它们描述了人类学习和人类进化的典型情景——来产生好的结果。一个生动的例子就是所谓的“识别启发法”（recognition heuristic）。如果你问我：圣地亚哥和圣安东尼奥哪座城市人口多？我可能会认为是圣地亚哥，这只是因为我**听说过**圣地亚哥。一旦我知道这两个名字，我可能就会使用一种不同的快速而节俭的启发法来检查其他的线索。可能我认为一个好线索是“我听说过它们的交响乐团吗”等等。重要的是，我没有作出比此**更艰苦的**努力。可能有很多我打算“包括在内”的小线索和指示。但是，根据一些令人印象深刻的研究（例如，参见蔡斯［Chase］等，1998），这样做可能既耗时间又（这是最重要
319 的）毫无结果。试图用更缓慢的和（明显）更聪明的东西来代替这种快速而节俭的启发法，可能是我作出的**更坏的**选择。

尚不清楚的是，就你为了解释合理性何以从机械上说是可能的就必须解释的**东西**来说，这些重要研究（究竟）应该对我们的看法有怎样的影响。我们有可能看到机器人学的爱好者与基于人体工程学的方法（13.4节）可能共同使用相对简单的神经网络控制器（13.3节）来学习快速而节俭的启发法，启发法最大限度地使用了局部的时机和结构。在某种意义上，躯体标记机制（13.5节）可能被认为是在执行另一种快速而节俭的启发法，它能使当前的决策轻松地从过去的经验中受益。在这种冲击之下，很多有关整体的诱导推理的担心（13.6节）可能会完全消失。如上所述，

我自己的看法是，有些难题仍然存在。但我所支持的解决办法（参见克拉克，2002）**本身**可以被看成是快速而节俭的启发法的一个特例：即一种容易得到的程序，它用别的东西（之前提到的第二遍的基于连通性模式的搜索）代替了基于内容的整体搜索。

13.8 结论：活动目标与多种技术

我们现在已经看到，合理性涉及的东西比最初看到的既多得多又少得多。它涉及的东西比基于句法的局部推理——定义在易处理的准句子编码之上——要多得多。甚至福多也承认这一点，或者他至少承认，他还不清楚如何用这些资源来解释整体的诱导推理。它涉及的东西比（可以说）冷静地使用服务于目标的信息要多得多。因为人类理性似乎依赖于一种微妙的相互作用，在这种作用中，情绪性反应（经常是无意识的反应）通过**增强**我们作出流畅、合逻辑的合理反应的能力而帮助我们筛选选项，并使我们的选择出现倾向性。我已经证明，这些情绪系统可以被有效地看成是一种得到很好提炼的、来之不易的知识储备，它涉及我们一生中的选择和行为的经验。

但合理性涉及的东西也可能比我们通常所认为的要**少**得多。可能人类理性（我把这当成我们永恒的目标）实质上是在我们生态环境的火炉中锻造的一种蹩脚的折中办法。快速而节俭的启发法适合于充分利用使我们维持生活的最廉价资源，它通常可能和自然一样近地接触到了理性空间。机器人学和联结主义的研究进一步增强了这种关于既更多又更少的看法，因为身体和世界的特征被用于从基于原型的基本机载推理能力榨取最大的利益。有迹象表明，甚至整体的诱导推理的怪物也可能完全屈从于快速而节俭的启发法与简单句法搜索之间的某种巧妙组合。

那么，这会将“合理性在机器中何以可能?”这个据说是根本的问题置于何处呢？我认为，它将这个问题留在了在新旧之间摇摆不定的一个重要十字路口上。如果（如我所认为的）13.4 到 13.8 节中所描述的每种研究纲领都处理了这个问题的重要方面，那么合理性问题就恰好变成了这一问题，即在社会的、环境的和情绪的背景下解释大致合适的适应性反应结果。合理性（或者就像我们人类通常所享有的其中的那些）就是你在这全部因素的混合以某种方式协调一致和相互促进时所得到的东西。弄清楚了 *320*

这种复杂的生态平衡行为，就是弄清楚了合理性在机器中何以可能。

参考文献

Bechera, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science*, 275: 1293-1295.

Beer, R. D. (2000). "Dynamical Approaches to Cognitive Science." *Trends in Cognitive Sciences*, 4 (3): 91-99.

Brooks, R. (1991). "Intelligence Without Representation." *Artificial Intelligence*, 47: 139-159.

Chase, V., Hertwig, R., and Gigerenzer, G. (1998). "Visions of Rationality." *Trends in Cognitive Sciences*, 2 (6): 206-214.

Churchland, P. M. (1989). *The Neurocomputational Perspective*. Cambridge: MIT/Bradford Books. Churchland, P. S. (1998). "Feeling Reasons." In P. M. Churchland and P. S. Churchland (eds.), *On The Contrary*. Cambridge, MA: MIT Press: 231-254.

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*. Cambridge, MA: MIT Press.

—— (1993). *Associative Engines: Connectionism, Concepts and Representational Change*. Cambridge, MA: MIT Press.

—— (1996). "Connectionism, Moral Cognition and Collaborative Problem Solving." In L. May, M. Friedman, and A. Clark (eds.), *Minds and Morals*. Cambridge, MA: MIT Press: 109-128.

—— (2002). "Local Associations and Global Reason: Fodor's Frame Problem and Second-Order Search." *Cognitive Science Quarterly*.

Clark, A. and Grush, R. (1999). "Towards a Cognitive Robotics." *Adaptive Behavior*, 7 (1): 5-16.

Clark, A. and Toribio, J. (1994). "Doing Without Representing?" *Synthese*, 101: 401-431.

Clark, A. and Thornton, C. (1997). "Trading Spaces: Connectionism and the Limits of Uninformed Learning." *Behavioral and Brain Sciences*, 20 (1): 57-67.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., and Damasio, A. R. (1994). "The Return of Phineas Gage: Clues about the Brain from the Skull of a Famous Patient." *Science*, 264: 1102-1105.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

—— (1998). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. and LePore, E. (1993). "Reply to Churchland." *Philosophy and Phenomenological Research*, 53: 679-682.

Fodor, J. and Pylyshyn, Z. (1988). "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition*, 28: 3-71.

Haugeland, J. (1981). "Semantic Engines: An Introduction to Mind Design." In J. Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press: 1-34.

Kleinberg, J. (1997). "Authoritative Sources in a Hyperlinked Environment." IBM Research Report (RJ 10076). A version also appears in H. Karloff (ed.), *Proceedings of the 9th ACM-SIAM Symposium on Discreet Algorithms* (1998), and an extended version in *Journal of the ACM* 46 (1999).

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

McClamrock, R. (1995). *Existential Cognition*. Chicago, IL: Chicago University Press.

Minsky, M. (1975). "A Framework For Representing Knowledge." In P. Winston (ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill.

Newell, A. and Simon, H. (1976). "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the Association for Computing Machinery*, 19: 113-126.

Schank, R. (1975). "Using Knowledge to Understand." *TINLAP*: 75.

Smith, B. C. (1996). *On the Origin of Objects*. Cambridge, MA: MIT Press.

Stein, L. A. (1994). "Imagination and Situated Cognition." *Journal of Experimental and Theoretical Artificial Intelligence*, 6: 393-407.

第 14 章　心灵哲学与神经科学

约翰·比克尔（John Bickle）

322 大约 20 年前，帕特丽夏·丘奇兰德（Patricia Churchland）以其特有的激情大声疾呼："再明显不过的是，心灵哲学家至少会从对大脑如何工作的认识中受益。"（1986：4）自那时以来，神经科学获得了呈指数级的迅猛发展。我们如今已经经历了"脑的十年"，我们对产生认知、知觉和行动的神经机制有很多认识，我们的知识跨越了从膜通道的生物物理学到大规模平行神经网络的大范围动力学的所有层次。有的人或许认为，"神经科学哲学"现在支配着心灵哲学，有的人可能认为，在对相关且可用的神经科学信息一无所知时，哲学家将耻于谈论如意识、认知表征、知觉的认识论甚至略带规范性的问题，但其实他们是错误的。在很大程度上，主流心灵哲学仍然对这些方面的问题漠不关心。（你在这个文集中发现了多少神经科学呢?）

另外，为什么理性而聪明的思想家会忽视这种"显而易见"的东西呢？答案并不复杂。就历史上尤其是其当下的形式来说，神经科学是一项**还原的**事业。而"还原论"在当代哲学中已不流行了。在同一本书中，丘奇兰德宣称："在关于还原论的争论中，对立双方的争论经常不能正常进行，因为他们对他们的分歧之处意见不一致。"（1986：278）这种评价现在仍然成立。还原仍然被哲学家们深深地误解，包括它对"特殊的"、潜在地还原的科学的方法论意蕴。本章的一个主要目标就是澄清为当代神经科学提供灵感的这种"还原论"的含义和方法论的重要性。

当代哲学家讨厌"还原论"事业还有别的原因。例如，职业的安全感。只有具有乡村无神论气质的哲学家才乐于看到"心灵"被科学篡位。
323 这个概念长久以来就一直处于哲学的核心位置。如果"心灵"被**还原的**科学驱除了，归入到了"神的意志"、"自然界"和"生物"的等级，那么世界还有什么留给哲学家们来沉思默想呢？

显然，这些评论并未处理这些**论证**，它们推动着关于还原论的模糊观点继续发展，而且最近越来越完善了。它们也没有为反还原论者提供任何停顿下来的经验理由，而这些是我随后的任务。在接下来的四节中我将对下述主张作出辩护。

- 关于心理神经还原论“要么行动要么闭嘴挑战”的争执已经被化解了，而对来自于最近科学中的例子的残余隐忧表明：关于还原的普遍误解仍在哲学（和认知心理学）中蔓延。
- 当下单细胞神经生理学层次上的研究正在产生一些结果，它们与哲学关注的事情直接相关，甚至和意识有关。
- 哲学家不是唯一想处理意识的“质的”和“主观的”方面的理论家；坚定的神经科学家提出了与这些特征相关的问题，并用巧妙而直接的经验方法来解决它们。感受性质和主观性再也不仅仅为哲学家所独占。

我最后将论述一个有点离题的问题，即我认为大肆宣扬哲学家、心理学家和神经科学家之间的“跨学科性”在实践中大都还是一种神话。在同一领域，很少有人会相信其他人的贡献，而人们也仅是在这一点上达成了共识，诚然，经过适当训练的哲学家可以对改变这一现状作出独特的贡献。

14.1 现实神经科学中的真正还原

对现有的科学还原理论作出评价并提出一种替代选择是科学哲学的一项艰巨任务，它也远远超出了本章的范围。[1]但是，有两个特征需要明确地提出，以避免出现丘奇兰德所警告的言语争论。首先，科学的还原是**理论间的还原**。它是科学**理论**而非实体、属性或事件之间的一种关系。科学的还原可能产生跨理论的本体论**结果**，但这些结果**从属并依赖于**这种首要的**理论**间关系。其次，**理论间的统一**这个概念处于科学还原的中心。当还原实现时，还原的理论会完整地解释被还原理论的材料，这些材料通常仍是用后者的术语和框架表达的。（不用说，这个条件在原则上是成立的，但在实践中并不总成立。）

追求还原心理学的当代神经科学在以下引自著名教科书的两句话中得到了很好的表述。戈登·谢泼德（Gordon Shepherd）写道：

很多认知心理学家……都相信关于学习和记忆的**理论**应该是前后一致的和自足的，无需借助于神经机制……

对大多数神经生物学家来说，这种观点过时了，**现代研究的目标之一就是要将这两个层次结合成一种一致的框架**。(1994：619；强调为我所加)

以上所强调的短语反映了前文所述的科学还原的两个特征，即其首要的理论间特征和统一主义目标。坎德尔（Kandel）等把还原论的位置又降了一个层次：

神经科学的目标是理解心灵，即我们是怎样感知、运动、思维和记忆的。在本书之前的版本中，我们强调行为的重要方面可以在个别的神经细胞的层次上考察……这种方法……多半是用细胞生物学的术语讲述的。现在也可能**直接在分子的层次上**处理这些问题。(1991：xii；强调为我所加)

他们在20世纪90年代初期推动了这种重新定位，那时“分子革命”刚开始横扫神经科学。只要花五分钟读一读从20世纪90年代初期到现在的《神经科学学会摘录》(*Society for Neuroscience Abstracts*)，就能看出分子理论和实验方法论**更加**流行的程度。还原论在当代主流的神经科学中是活跃而繁荣的。

然而，研究目标是一码事，最终的结果则是另一码事。当代神经生物学实际提出了**真正的认知**心理学理论向之还原的理论吗？我将这个问题称作关于心理神经还原的“要么行动要么闭嘴挑战”，我支持一种肯定的回答（比克尔，1995；1998：第5章）。我的论证包含两点理由：

1. 当代关于联想学习的心理学理论借助于被普遍接受的“真正认知的标志”的合标准资源（在它们内容之上的表征和计算）。

2. 这些心理学理论还原成了这些神经生理学理论，这些理论是关于适当脑区中的神经元回路以及某些形式的突触可塑性的细胞和分子机制（神经元之间的电化学传递的效率就是借助于这些机制而随时间提高或降低的）。

真正的认知心理学理论的神经生物学还原已经是一个既成事实。

第一个主张很有趣而且普遍未被哲学家和认知心理学家认识到（雷斯科拉 [Rescorla]，1988)；但是我在已出版的著作中已有两次讲到它（引

证在前面的段落中），这里就不赘述细节了。只要这样说就够了，20 世纪 70 年代以来，由于在实验技术、巧妙的实验设计以及产生反直觉预测——得到了经验的证实——的数量模式方面的进步，联想学习理论

> 强调一种刺激导致另一种刺激的**信息**……这些理论强调世界的实际状态与有机体关于该状态的**表征**之间差异的重要性。它们把学习看成**这两者借以协调一致的一个过程**……一种简写形式就是：有机体只有在他们“感到惊讶”时才调整他们的巴甫洛夫式联想。（雷斯科拉，1988：152-153）

这种方法是完全通用的。学习理论家既把它运用于异乎寻常的联想现象，例如封闭效应（blocking effect）和行为上无声的学习，而且把它运用于经典的条件作用。他们在这篇论文中第一次用准确的量化方式讲述了这样一则故事，雷斯科拉和瓦格纳（Wagner）用“明确的认知主义术语”陈述了它：

> 有机体只有在事件违背他们的预期时才学习。有些建立起来的预期跟一个刺激综合体之后的事件相关：这样一来，由这个综合体及其成分刺激所引起的预期只有在随之发生的事件与合成的预期不一致时才会被修改。（1972：75）

很快，他们的模型有了进一步的发展并且出现了实际测试，到 20 世纪 70 年代后期它就统治了这个领域（迪金森 [Dickinson]，1980）。

我的第二点主张又是什么呢？神经科学家主张，学习和记忆涉及到中枢神经系统（CNS）回路中的变化，这可以追溯到雷蒙·凯尔（Ramony Cajal），它是由赫布（Hebb，1949）第一次明确地提出来。自 20 世纪中叶以来，最受关注的位置就是突触，即神经元之间的狭小间隙，电化学活动的传递就发生在那里。在 CNS 中，这种传递主要是通过突触前神经元向突触间隙释放化学神经递质，这些递质随后与突触后神经元上受膜限制的蛋白质（受体）捆绑在一起。这种捆绑引起了一系列生物化学事件，它们打开离子选择膜通道，在那片突触后膜上要么导致了去极化（兴奋性的突触后电位），要么导致了超极化（抑制性的突触后电位）。大量突触前的、突触后的和突触间隙内的生物化学因素都影响着突触传递效能。这些因素是**可塑的**：对内生的和外部的各种生物化学事件的要求是可以改变的。[2]

大量差异很大的实验证据都支撑突触可塑性是主要的学习和记忆机制。[3]研究者运用各种实验方法论，如动物标本（脊椎动物和非脊椎动物）、解剖的区域和行为任务等已经产生了一种关于学习和长时记忆的突触基础的一般模型（谢泼德，1994：648-649）。基本的细胞生物学概念是**长时程增强**（Long-term potentiation，LTP）（参见图 14—1）。一个动作电位——从突触前轴突膜向下传播到它的终球——打开了电压门控钙离子（Ca^{2+}）通道。Ca^{2+}流入突轴前终末（沿着它的浓度和电的梯度），然后产
326 生了一个生物化学梯级（cascade）——它导致囊泡捆绑的增加，它包含传递到突触前膜上的活性带的神经递质谷氨酸盐（GLU），以及随后谷氨酸盐向突触间隙的释放。[4]这种谷氨酸盐与两类突触后的受体捆绑在一起。一类是离子型 AMPA（α-氨基-3-羟基-5-甲基-4-异恶唑丙酸）受体，它打开了直接的钠离子（Na^+）通道，导致了 Na^+的流入（沿着它的浓度梯度和电梯度）和随后的 EPSP。另一类是 NMDA（N-甲基-D-天冬氨酸盐）受体。当处于或者接近静息膜电位时，谷氨酸盐捆绑没有产生离子的流入，因为 NMDA 受体受到了镁的阻挡。然而，当膜被充分地去极化时（即被处于 AMPA 受体附近的谷氨酸盐捆绑），镁的阻挡就会突然消失，与 NMDA 受体捆绑在一起的谷氨酸盐就打开了突触后的 Ca^{2+}通道。

327 从突触后来看，Ca^{2+}充当了第二信使。它激活了：

- 一串分级的捆绑蛋白质和蛋白激酶的 Ca^{2+}，它分解突触后神经元的细胞骨架，之后又将它重建为一种不同的构型，引起受体数量和位置方面的变化；
- 磷蛋白和（可能）其他转录因子，它们反过来激活了突触后神经元核中直接的早期基因，在受体和其他蛋白质合成方面产生了长期变化；
- 氧化一氮合成充当一种逆行性的传感器，它依靠突触前神经元来增强随后的谷氨酸盐的释放。

另外，NMDA 受体突触后的激活引起了花生四烯酸（AA）的产生，AA 似乎也充当了一种逆行性的传感器。从突触前来看，AA 启动了一串分级的蛋白激酶，它们最终与基因转录因子相互作用，在蛋白质产生、细胞结构和功能方面产生了长期变化。[5]

关于 LTP 诱导的突触可塑性的理论如何能被整合进认知主义学习理论（例如现代联想学习理论）的还原中呢？关键就是霍金斯（Kawkins）和坎

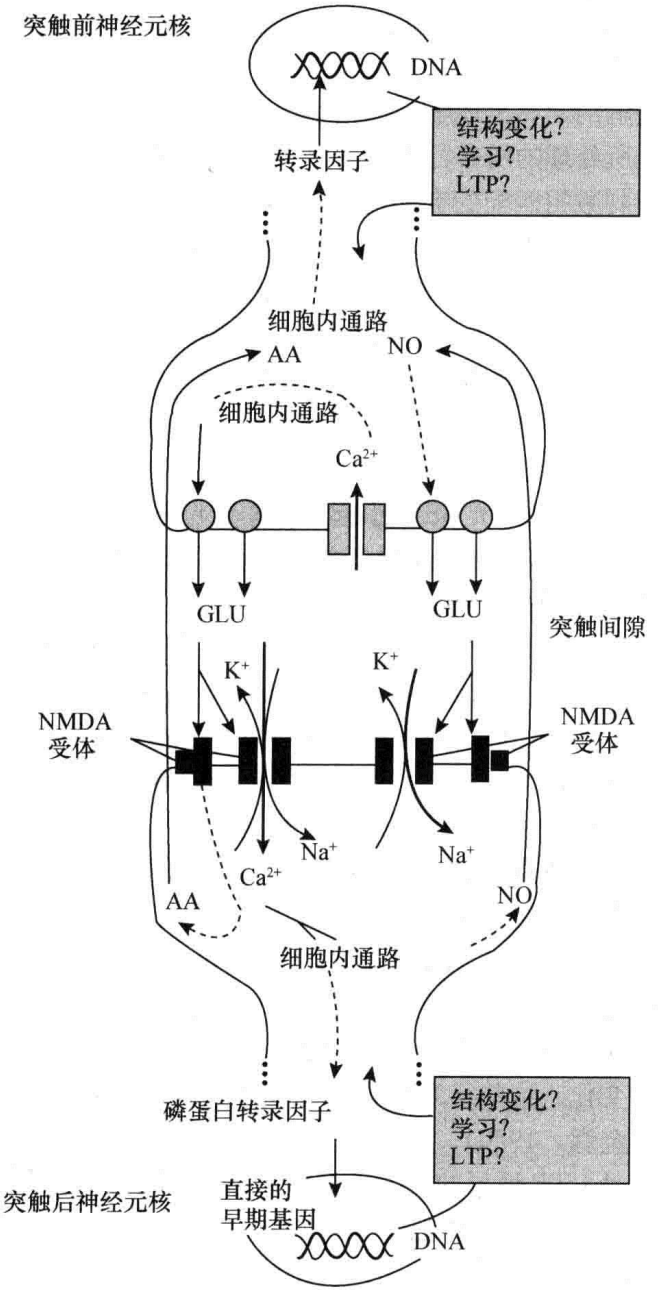


图 14—1 当前关于 LTP 诱导的突触可塑性理论简化图

有关解释和缩写请参见正文。(改编自谢泼德，1994：648，图 29.18)

德尔（1984）所说的“学习的细胞生物学字母表”（cell-biological alphabet of learning）和我所说的“组合还原”（combinatorial reduction）（比克尔，1995；1998：第5章）。细胞生物学的和分子的机制提供了“字母”，为日趋复杂的神经解剖学和生理学利用的字母序列和组合（“词语”）解释了认知心理学理论所处理的所有行为数据。例如，由经典条件作用所引起的行为变化是由与刺激相匹配的突触前神经递质释放的增加解释的。（这是霍金斯和坎德尔的“字母”之一。）最初，中性的条件刺激（CS）引起了从受刺激的感觉受体传导而来的中枢通路中的弱神经递质释放（基准率之上）。在行为方面非常重要的非条件刺激（US）引起了强释放。US通路中的活动激活了在CS通路中的突触前终球上形成突触的中间神经元。这些中间神经元释放出神经递质5-羟色胺——它与突触前CS通路终末上的受体的捆绑在一起。这在这些终末中引起一个生物化学梯级——它抑制钾
328 离子（K^+）流出，增加了CS所引起的动作电位并使更多Ca^{2+}流入。额外的Ca^{2+}使囊泡与端膜、神经递质释放和突触后反应的捆绑得到增强。这样，CS-US的配对就能使更弱的CS到达同样由突出的US引起的第二信使系统。[6]之后，CS诱导的活动在激活非条件反应运动通路方面就取代了US。最终的结果就是运动神经元中的活动，运动神经元凭借骨架产生了合适的肌肉收缩，而这些肌肉收缩随着时间的推移就形成了行为动力学。

霍金斯和坎德尔（1984）只借助突触前的机制就对联想学习——它是通过细胞生物学的“字母”的序列和组合进行的——的一些高阶认知特征作出了解释。他们所有的回路假设都是以已知的解剖学和生理学知识为基础的。他们说明了封闭效应的行为动力学、刺激的特异性和一般性、灭绝与自然恢复、二阶（S-S）条件作用以及US的预接触（pre-exposure）如何能直接由细胞生物学“字母”在生物学上的合理序列和组合来解释。久而久之，这些序列和组合所产生的最终结果就是驱动行为反应的运动神经元活动方面的变化。这些行为数据就是引起了“认知主义的”联想学习模型的数据（迪金森，1980）。另外的分子资源是由新近的发现提供的，这些发现涉及的是LTP诱导的突触可塑性将神经生物学“组合还原”的范围扩展到的多种类型的学习和记忆（比克尔，1998：第5章）。最近，关于“陈述性长时记忆”和从短时记忆到长时记忆的“合并转换”（consolidation switch）的认知心理学处理办法已经加入了这个群体（斯夸尔和坎德尔，1999：第7章）。

最后产生的这种细胞和分子生物学解释不仅获得了定性的行为数据。

例如，霍金斯（1989）对他和坎德尔在以前的还原中所使用的突触前特征提出了一种数量模型。这种模型使他能在一个解剖学方面合理的计算机模拟中模拟这些细胞生物学“字母”。霍金斯表明，随着时间的变化，模拟的运动神经元的动作电位率的曲线与学习曲线、行为动力学以及由雷斯科拉—瓦格纳方程所预测的变化的强化模式是完全一致的。他的定量测量，即模拟的运动神经元随时间变化的激发率，是由横跨模拟的感官、易化因素（facilitator）以及运动神经元的参数和可变的突触权重值计算的。所有权重值都是通过选择来模拟已知的生物特征的。因此，甚至当这些神经生理学“字母”被限定于霍金斯和坎德尔早期所说的突触前细胞生物学机制时，模拟的运动神经元活动——是由这些字母在日益复杂的神经解剖中的序列和组合产生的——完全获得了认知心理学解释的行为动力学和预测。[7]

这种情况只是真正的认知心理学理论向细胞生物学/分子神经科学理论还原的一个例证。还有其他例子，它们使用了当代关于LTP诱导的突触可塑性的理论的最新材料。[8]因此，对于各种真正的认知心理学理论来说，我的第二个论点已经阐明了。真正的认知心理神经还原已经是一个既成的科学事实。

14.2 神经功能?

心理学家莫里斯·斯考滕和H·卢伦·德扬（Maurice Schouten and H. Looren de Jong，1999）对我的第二个论点提出了质疑。他们的批评值得在这里讨论，这至少有两个理由。首先，他们表达了流行的并且**具有科学根据的**反还原论主题；其次，他们直接处理我的经验案例研究。因此，对那些关于心理神经还原的经验论证负责任的反对者来说，可以把他们的论证当成很好的样板。然而，他们的论证也有重要的缺陷，把这些缺陷指出来有助于澄清隐含在当代主流神经科学中的还原论的一般主题。

斯考滕和德扬（1999）在他们的批评中一直强调两点：

- 需要详细说明综合的科学解释中的功能；
- 功能单从物理机制的理论上是难以理解的。

在论及具体运用于**心理神经的**理论之间的关系时，他们主张，大脑的功能不能通过“纯粹自下而上的”理论化来发现，甚至不能通过这种方法来发

现详细说明细胞生物学的和分子的过程的序列和组合。他们的第一个论证有两个前提。首先，高层次的倾向**多样随附于**（multiply supervenient）物理的基质和机制。换言之，很多高层次的倾向都随附于同一种物理基质。很多读者会把多样随附性看成是**可多样实现性**这个更熟悉的概念的翻版。多样随附性已经越来越多地出现在反还原论的论证之中（金凯德［Kincaid］，1988；恩迪科特［Endicott］，1994）。其次，倘若有多样随附性，那么高层次理论是在给定情况下将**因果相关的**低层次特性与**因果不相关的**低层次特性区别开所需要的。就给定事件来说，只有某些低层次特性是因果相关的（当时出现的很多低层次特性之一）。回避高层次理论会造成与物理基质的特殊倾向特性——它们对于给定的解释是相关的——相关的客观信息的丧失，因此这种“纯粹自下而上的”方法论——斯考滕和德扬认为它是组合还原和细胞生物学/分子“字母表”方法的典型特征——“将是无效的”。[9]

这个论证有很多问题。第一个问题就是完全误解了还原论的方法论承诺。当代神经科学所使用的（以及霍金斯、坎德尔和我分别分析的）方法论并不“回避”高层次理论，现在多数还原论者都明确地接受了**协同进化研究意识形态**（coevolutionary research ideology，它首先得到了胡克的支
330 持）。有些人甚至认为，**在方法论上**高层次理论化在完成还原之前及其之后**都是必不可少的**（比克尔，1996；1998：第4章）。协同进化本身是一个方法论建议，而不是对理论选择的一个限制或强迫接受。它不是用来排除某些高层次理论（即缺乏还原潜能的理论），而是要使新出现的理论建议得以传播，给它们机会展示其解释力和经验的准确性。历史上（甚至是现在）正是生理学理论面对着主流心理学、社会科学和哲学最有力的抵制。[10]此外，由于合适的心理神经还原论必须与各种科学中的跨层次理论关系和方法论相一致，因此，心理神经还原论者必须承认高层次的概括在获得理论间的还原前后能够具有“强有力的认识保证”（霍根，1993）。科学史提供了很多成功理论的案例，它们在长期的发展过程中只是被大致地确认为理论。甚至理论间还原关系的这种逻辑也需要承认高层次理论。还原是（发达的）理论与在认识上有保证且需要发展的高层次实例之间的一种二元（two-place）关系。如果要得到理论**间**的还原，特殊的科学就必须继续在它们的还原者发展时提供理论。最后须知，在斯考滕和德扬的第一个论点中，归属给高层次理论的作用（即将特殊解释对象之因果相关的和因果不相关的低层次倾向特性作出区别）与归属给它们一种实质的但却是

纯方法论的作用是一致的。对于引导低层次理论的发展但又不向我们承诺反还原论的结论来说，它们是不能取消的。

斯考滕和德扬还对我把学习和记忆—LTP联系作为一种**既成的**心理神经还原提出了批评。他们的错误理解很常见，有必要在这里讨论。他们首先指出高层次的神经心理学研究是怎样推动最初对哺乳动物的海马——LTP最先就是在这里发现的——的生理学研究的。他们的历史论述是正确的，但只是表明高层次的理论化在方法论上对于神经科学很重要，我们只是拒绝接受还原论者为什么不应该否认这一点的原因。历史的细节只不过证明了高层次理论化的一种方法论作用。斯考滕和德扬也主张，从20世纪80年代中叶以来，“对于‘LTP作为记忆基质’假说的经验支持**主要来自于**药理学药剂的应用……它们似乎对NMDA（受体）活动起反作用并削弱空间学习”，他们主张“在这类研究中，空间学习的操作定义是：**在水上迷津中的表现**”（1999：247；强调为我所加）。甚至在一篇以哲学家和认知心理学家为探讨对象的论文中，这种关于神经科学研究的“事实陈述”也是幼稚的，甚至放在20世纪90年代早期也是不正确的。检索提交给2000年神经科学学会年会的报告摘要的标题词或关键词，使用了“LTP”或“突触可塑性”的论文有200篇以上（见www.sfn.org）。其中只有4篇在题目中包含了“水上迷津”或者将它列为一个关键词。所有这200多篇摘要显示出大量种类繁多的分子操作和行为范例已被用于研究LTP，它们还显示出了斯考滕和德扬主张的幼稚。

这个问题远非是由非神经科学家所犯的一个关于神经科学的事实性错误。斯考滕和德扬反过来提出了NMDA受体拮抗肌和水上迷津任务所特有的一些方法论的和解释的问题，他们暗示这些问题对学习和记忆—LTP还原构成了一个一般的挑战。他们的问题没有提供这样的东西，是由于各种分子操作、行为任务以及最近的基因剔除和转基因操作（在哺乳动物身上）都提供了还原的证据（斯夸尔和卡德尔，1999：第7章）。事实上，最近证明LTP是关于学习和记忆的一种细胞/分子机制的最有说服力的实验研究，既未涉及药理学的操作，也未涉及莫里斯水上迷津。相反，它来自于转基因的成年老鼠，它们受到操控来过度表达（overexpress）一种基因，这种基因的蛋白质生成阻碍了海马体内蛋白激酶A这个催化剂子单元。它使用了一种双重恐惧条件作用行为测试，这种测试包括环境线索、一种中性的CS以及一种足部电击US（参见斯夸尔和坎德尔，1999：149-153）。对于心理神经的反还原论者来说，斯考滕和德扬的错误留下的教训

是：研究者可以为考虑经验研究欢呼，但首先要懂得在这个问题上所进行的科学研究的范围和种类。

斯考滕和德扬还对学习与记忆—LTP 的诱导联系的神经科学证据提出了一种更一般的担忧。他们指出，重要的是要把对学习和记忆的影响与对可能促成行为的其他系统的影响区别开。很多系统都对 NMDA 受体拮抗肌敏感，包括感觉、运动、动机系统和注意系统，而所有这些系统都与水上迷津任务有关。或许 LTP 是主要适合于其他系统之一的可塑性的一种机制？或许是如此，但这对于神经科学家来说并不陌生。事实上，这就是神经科学家为什么在他们的实验设计中如此小心的原因。（顺便提一句，关于**特殊的**药理学药剂的**特殊的**方法论担忧，已经促使研究学习和记忆的细胞和分子机制的神经科学家把他们的实验方案转到了基因剔除和转基因药剂；例如，参见斯夸尔和坎德尔，1999：119－124，151－153。）哲学家和认知心理学家不应该跳过神经科学论文的“方法”部分，因为就是在这里神经科学家们透露了他们为了消除可能会破坏解释的混合因素而对有分歧的实验变量的**控制**——学习、视觉、注意、运动等等。显然，无论是谁对他们所使用的**特殊**控制所作的建设性的批评，神经科学家都会留意。但是哲学家和心理学家实际上不需要告诉他们需要控制可能混淆的影响，这些影响和斯考滕和德扬所指出的那些一样显而易见。了解到神经科学家通常所使用的控制是多么巧妙，可能会使哲学家们大吃一惊。[11]

332

最后，斯考滕和德扬证明功能理论不只**是**在方法论上是必不可少的或重要的。因为我们只有用这些理论才能回答“‘为什么’和‘为了什么’”的问题，即关于“（这种机制）应当做什么”的问题，关于“必要的规范性维度”的问题（1999：255－256）。他们强调，这些问题需要“一种关于功能的更具本体论意味的解释”（1999：256）。这种强调和他们对**目的论功能**（teleofunctions）的研究是密切相关的。一个系统的目的论功能依赖于它的选择（进化）史，目的论功能理论详细说明了这些种类，它们参照不同物理系统在其选择史中共有的目标而将这些系统统一了起来。只有适当的功能理论才能解释“实在的这些客观属性”（1999：256）。

目的论功能目前是生物哲学、心理学哲学、心灵哲学和语言哲学中的一个流行概念（米利肯，1984；波斯特，1991），它们处于受到最强烈的科学启发的反还原论论证的核心。接下来我将要作的说明不应当被看成是

我“最终的答复”（类似的东西并不存在）。但是，就这个概念和论证来说，还有很多方面是有问题的。一旦我们理解了每个不同的物理系统各自是如何工作的——即是说，一旦这样一种功能解释已经发挥了其本质的（然而是专门的）方法论作用，为什么我们需要一种关于这些不同的物理系统的“统一说明”（unifying specification）呢？这种统一为我们的**本体论**添加了什么呢？为什么会认为回答“为什么”和“为了什么”问题就是在作出本体论的承诺——这种承诺已经超越了发挥作用的各种物理机制？还要注意，就在正面的目的论功能讨论中归属给物理机制的大多数目的来说，“高层次的理论化”是普通而明显的。考虑这个受欢迎的例子：心脏的目的论功能是泵血，但得到这个见解真的需要很多“高层次的理论化”吗？这个例子是直观的，可见大多数系统的“目的论功能”通常是很明显的，尤其是在我们理解它们的物理机制的时候。（请注意：这**不是**说揭示它们的选择史的任务通常是平凡的或明显的——事实并非如此，就像**真正的**进化生物学和生态学所遇到的困难所证明的那样。）

除了主流的进化生物学之外，还有一种科学也与解释**为什么**一种特性存在于某个系统中有关。这种科学就是**分子遗传学**，它的抱负就是无情地还原。有些分子遗传学家甚至认为进化论发挥着一种本质但却是专门的方法论作用。例如，分子生物学家詹姆士·夏皮罗（James Shapiro）曾说：

> 传统进化论中的大多数基本概念都出现于1953年之前，那时人们对DNA还一无所知。20世纪上半叶，人们使用像基因、等位基因、支配、外显率、突变、异位显性、适应性和选择等术语，精心设计了进化过程的数学处理方法……虽然分子遗传学家仍然大量使用旧式语言……但他们实际上是在一个新的概念领域中工作的。分子遗传学的概念领域与经典遗传学**和进化理论**是不同的，就像量子物理学不同于经典力学一样。（1999：23；强调为我所加）

由于建立在诺贝尔奖得主芭芭拉·麦克林托克（Barbara McClintock）的独创性见解的基础之上，夏皮罗的图画涉及由许多细胞生物化学事件所引起的遗传突变。“大多数进化论者都试图（不切实际地）模拟这些细胞功能的行动，以此类比传统进化论的随机突变事件。”（1999：28）相反，内部的（对基因组来说）“信号转导网络”调节着遗传变化的计时和位置，包括在多个地点同时发生的变化。“为什么”和“为了什么”的问题是可

以由这种信号转导网络在 DNA 生物化学调节的层次上解决的，人们也越来越多地用分子术语来理解这些网络本身。“环境”只不过（偶尔）启动（kick-start）了这些内部的网络。照此看来，“到 20 世纪末必须对**进化**作出重新审视”（1999：23；强调为我所加）。分子遗传学被这些生物学哲学家——他们在发展目的论功能概念方面是最积极的——忽视了，因为他们从一种科学理论中获取灵感。根据有些分子遗传学家的说法，这种理论现在必须作出彻底的反思。也许，对分子遗传学这缕新曙光的无知是目的论功能概念所依赖的重大错误。

最后，在生物学哲学和心灵哲学的语境下，（目的论）功能的“本体论解释”必须了解它是什么。最终的解释是关于古典的属性或事件变种的一种**二元论**。这种解释暗示了有些属性或事件是不能由物理机制解释的警示，目前还没有什么能单独地、自然而然地对这种观点提出反驳。我们最好的生物学和心理学可能最终都向我们承诺了非物理的属性或事件。但是，这些人——他们寻求为任何意义上的**物理主义**辩护——自身不得不如此地蔑视得到本体论解释的（目的论）功能。忽视还原论者对当代神经科学和分子生物学的同情似乎也是极其傲慢的。[12]如果有什么领域构成了当代主流生物学的“最高荣誉”，那就是这两个领域了。当然，这一现状会发生改变，但我仍必须指出的是，当代生物学哲学家和心灵哲学家对它们的漠不关心有悖常理。

14.3 意识与细胞神经科学

意识是一种心理现象，很多人认为它被远远地排除在还原主义的神经科学之外。几乎在整个 20 世纪，它都被主流的心灵科学忽略了。最近，它咆哮着重返科学和哲学。继意识的复苏之后是明显的二元论的复兴（杰克逊，1982；内格尔，1989），关于我们（人类）解决意识/大脑问题的能力的“新神秘主义”担忧（麦金，1989）以及“革命化”物理学的要求（查默斯，1996；彭罗斯［Penrose］，1994）。甚至对神经科学持同情态度的物理主义者也认为，解释意识需要“异乎寻常的”、“全脑的”（whole-brain）资源：成熟的脑成像技术、大规模并行神经网络以及关于其整体活动的数学分析。他们共同的看法是，即使神经科学最终能胜任这项任务，特定的神经生理学技术仍难以胜任。

一个值得注意的例外是知觉神经生理学家威廉·纽瑟姆（William Newsome）。在就主流神经科学的“单一单元方法”进行交流时，他惊呼：“我们尚未着手彻底研究它的效用……令我激动的是最近出现了一种趋势，即将单一单元方法运用于有行为表现的动物身上，它们是训练来执行简单的认知任务的。”（加扎尼加，1997：57）纽瑟姆提到的任务包括知觉、注意、学习、记忆和运动计划等。从他自己的实验室得出的结果能以这种方式来解释，这种方式使得它们与最近关于意识的**哲学**争论产生关联。

现象外在论（phenomenal externalism）认为，个体感受器表面之外的环境决定着（“个体化”）他的感觉经验的质的内容（“感受性质”）（德雷斯基，1996；莱肯，1996）。它在发展过程中的一部分推动力是：人们最近都纷纷涌向“表征的”感受性理论，以及表征内容外在论在一般哲学中占据着支配性地位。对后者的论证是借助于一位哲学家的流行想象：孪生地球。人们通常用这些思想实验为语言意义和认知内容的外在论进行辩护。弗雷德·德雷斯基强调，没有什么会阻止这个人——他接受关于内容外在论的论证——接受关于现象外在论的相同论证：“正如我们区别和确认信念是根据它们是关于什么的信念，区别和确认它们是关于什么的信念是根据它们与什么具有适当的关系；我们区别和确认经验也必须根据它们是关于什么的经验。”（1996：145）[13]这种观点的本质表现在德雷斯基的口号式短语中即是：“经验本身处于头脑之中……但头脑中没有什么……必须具有对这些经验作出区别的性质。”（1996：144－145）尽管弗雷德和孪生的弗雷德是物理的复制品，从而是神经生理的复制品，但他们有意识的感觉经验可能具有不同的感受性质，这是因为他们的外部环境不同。

然而，关于纽瑟姆的“单一单元”结果——它们使用了对猕猴的视觉区 MT 的微电流刺激（参见图 14—2）——的一种解释证明：关于实质内容的外在主义直觉在经验上是不可信的。MT 区（中间颞皮层）是通向“背侧”（“顶部”、“哪里”）视觉加工流的大门（参见图 14—3），损伤研究和电生理学记录都揭示了它在关于运动方向的视觉判断中的作用。多数 MT 神经元的方向是有选择的，它们最适合于激发在单一方向上运动的视觉刺激。就像其他的皮层区域一样，MT 也有一种柱状组织，一个给定的柱上的神经元都共享了相似的感受野和优选的运动选择性。这些特征在柱与柱之间是有变化的，因此，MT 作为一个整体表征了视野中每个点上的所有运动方向（奥尔布赖特［Albright］等，1984）。

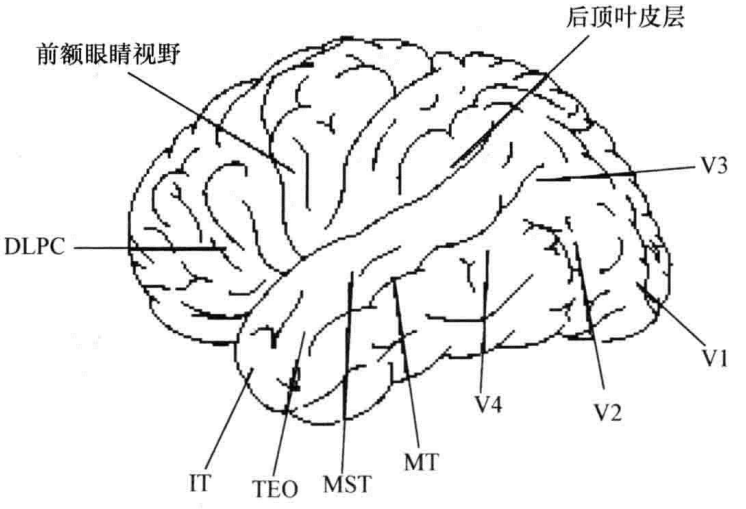


图 14—2　灵长类的动物（猕猴）视觉系统的解剖结构

缩写：MT（中间颞皮层）；MST（中间上颞部）；IT（下颞皮层）；DLPC（背侧前额皮层）。

纽瑟姆及其同事提出一种对运动刺激的强度进行量化的技术（萨尔兹曼［Salzman］等，1992；参见图 14—4）。点的序列是在计算机屏幕上绘
335 制的，其中用百分比的相关性表示的某种运动刺激的强度，以反映某个点被再次绘制在一个固定时空区间中的可能性。例如，在一个“有 50%的相关性的垂直刺激”下，有一半的点被重新绘制在一个固定向上的区间上（提供了垂直运动的错觉），而另一半的重新绘制则是随机的。纽瑟姆的团队还就确定运动方向的判断提出了一种行为范例。他们的控制是巧妙的，但基本思想很简单：猴子注视一个闪亮的中心点，而且在出现一个具有特殊强度的视觉运动刺激时保持这种注视（参见图 14—5）。关注点和运动刺激都消失后，靶灯（LED）出现在周围。一个 LED 定位于运动刺激的方向上。另一个定位于相反的区域。猴子显示其关于运动方向的判断是通过扫视（快速地运动眼睛）恰当的 LED。猴子只有在扫视正确时（即扫视处于运动刺激方向上的 LED）才会得到奖赏。在实验中，实验人员首先定位一个 MT 细胞的感受野（刺激引发一种反应所处的这部分视野）和优选的运
336 动选择性，这就能将运动刺激限制于该视野区域。然后，他们就能比较在刺激出现期间将微电流刺激运用于那个细胞，以及不运用于那个细胞时猴子关于各种刺激强度下运动方向的报告。在细胞优选的运动方向上的靶 LED 被称为优选 LED，相反方向上的靶 LED 被称为无效 LED。猴子的扫视由此构成了一种关于（知觉到的）似动方向的报告。

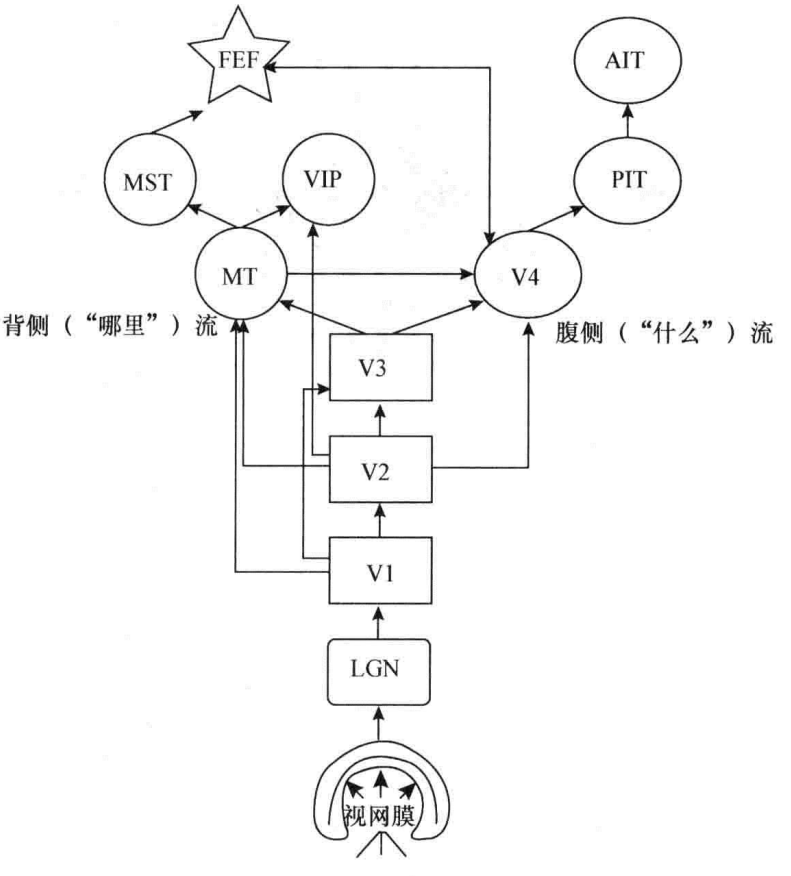


图 14—3　关于哺乳动物视觉系统的主结构、皮层分析器区、回路和加工流的流程图

缩写：大部分和图 14—2 中相同，除 LGN（背侧丘脑的外侧膝状体核）；PIT（后下颞皮层）；AIT（前下颞皮层）；VIP（腹侧内顶区）；FEF（前额眼睛视野）之外。

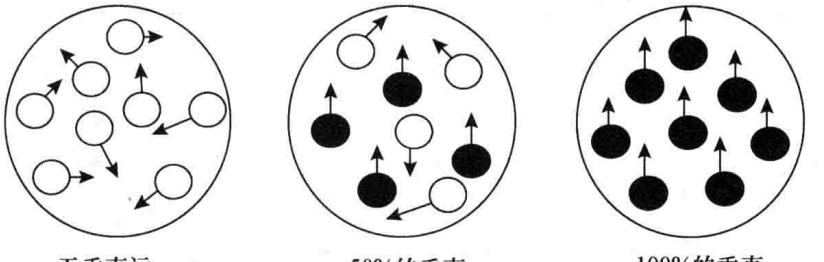


图 14—4　关于运动方向刺激强度的定量测量

实际展示包含的点比这里所例示的要多得多。（改编自萨尔兹曼等，1992：2333，图 1）

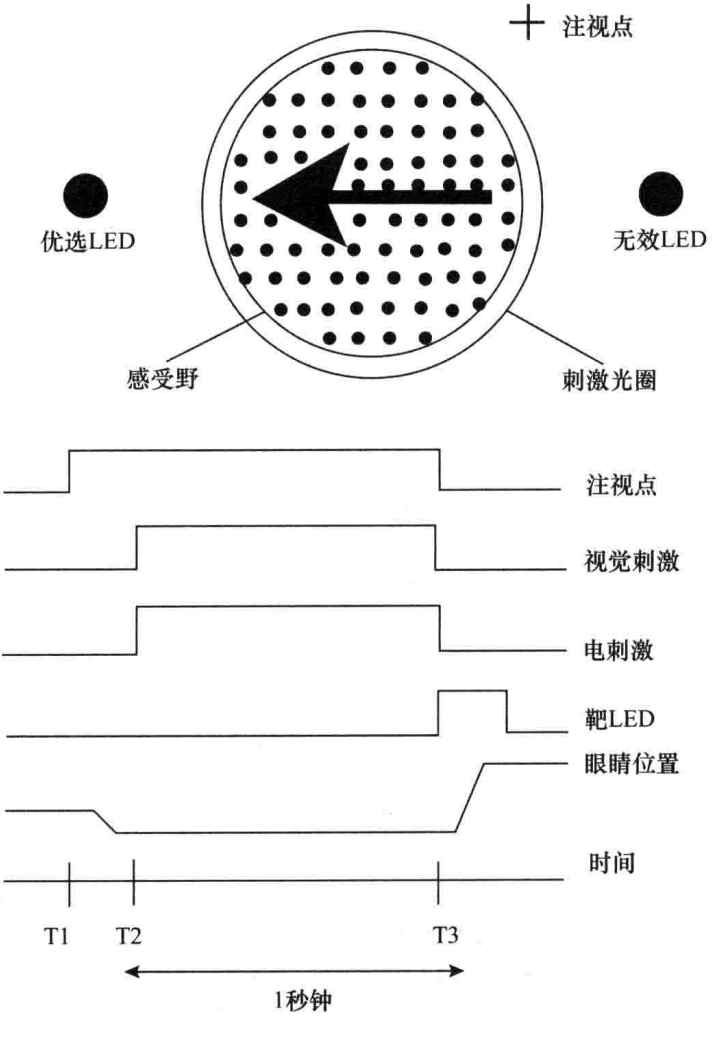


图 14—5　纽瑟姆的实验范式涉及对 MT 区个别神经元的微电流刺激

（改编自萨尔兹曼等，1992：2334，图 2）

这种运动强度测量方法和行为范式使纽瑟姆团队能将猴子关于一个 MT 神经元的优选方向上的似动报告的比例设计成运动刺激强度的函数（参见图 14—6）。如果对选择方向的 MT 神经元的微电流刺激向神经过程——它们支撑着关于运动方向的视觉判断——增加了“信号”，那么它就会使猴子关于神经元的优选方向的报告产生偏向。用图例表示，就将会造成心理测量函数向左侧偏移（参见图 14—6）。这些恰好就是纽瑟姆及其同事在各种刺激强度和微电流刺激频率下所观察到的结果（萨尔兹曼

[Salzman] 等，1992；穆拉苏吉 [Murasugi] 等，1993）。几乎在每个用百分比表示的相关性中，对选择方向的 MT 细胞的微电流刺激都使猴子对优选 LED 的扫视出现了严重的偏向。这种偏向甚至在相反方向上出现强运动刺激时（例如，＞－50%的相关性）也会出现。在此我们回想一下，猴子只有在正确地报告了刺激的运动方向时才会得到奖赏，它们从来不会因为这些接连不断的错误选择而受到奖赏。在神经元优选的方向上，不断增加的微电流刺激频率会提高似动报告的比例（甚至是在相反方向上有更强的运动刺激的情况之下）。

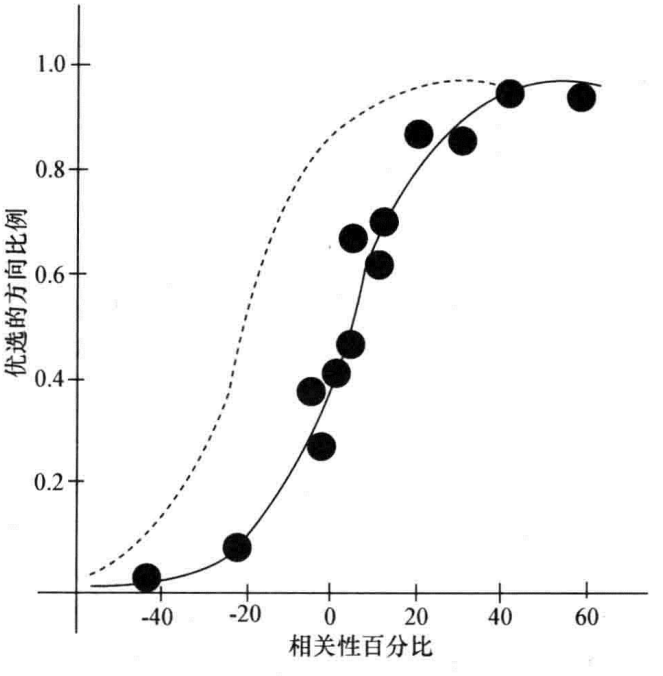


图 14—6　示意性的心理测量函数

它作为运动信号强度的一个函数，描绘了在运动选择 MT 神经元优选的方向上的决策比例（点和实线）。函数向左偏移是遵循微电流刺激而被预测的，当然，前提是微电流刺激为支撑运动方向的视觉判断的神经过程增加了信号（虚线）。（改编自萨尔兹曼等，1992：2335，图 3）

这些结果自然会引出一个问题：猴子在微电流刺激试验中**看见了**什么？甚至是在运动刺激出现在相反的方向上时，猴子也**有意识地觉察到了**出现在神经元优选的方向上的运动吗？纽瑟姆及其同事承认，他们的结果不能对这些问题作出决定性的回答。但是他们也敢于提出这样的建议：

> [一种] 可信的假说是：微刺激**引起了一种主观的运动感觉**，它类似于在出现运动后效或瀑布错觉期间所经验到的那种……因此，运动看起来是**能在大脑内部独立计算的一种性质，似乎“被归属”给了环境中的模式化对象**。（萨尔兹曼等，1992：2352；强调为我所加）

他们是在主张：运动感受性质是由神经活动从内部产生的，并且与外部对象的表征“相联系”。幸好我们的“内部归属”通常与外部事件是非常一致的，自然选择对于其“归属”更随意的生物来说要更加残酷。但是在合适的条件下，我们内部产生的感受性质和外部事件是可以分开的。这就是纽瑟姆的微电流刺激研究中所发生的事情。

在这里起作用的一般观念就是神经科学家鲁道夫·利纳斯（Rodolfo Llinás）和神经哲学家帕特丽夏·丘奇兰德所称的**内源**（endogenesis）。如他们所说：“这里的关键是：感觉经验不是由来自外部世界的输入信号而是由大脑内在的连续过程产生的。”（利纳斯和丘奇兰德，1996：x）来自于和外部参数相连的感受器的输入信号所起的作用是：“支撑、塑造以及用其他方式形塑内部活动以产生一种方便生存的‘我在世界中’（me-in-the-world）的表征图式。”（1996：x）自然选择——适合于利用一种环境条件，这并不是真理——决定着一种图式是否“成功”。[14]

338 纽瑟姆的实验证据和解释，连同一般的内源概念都对现象外在主义非常不利。首先须知，“微电流刺激＋（强）无效方向刺激”实验中的猴子，相对于“没有微电流刺激＋（强）优选方向刺激”实验中的猴子，是孪生地球实验的一个经验类似物。其二者的大脑状态在这两种情况下（接近于）等同，至少从 MT 和进一步达到的背侧流来看是这样（这些部位对于视觉运动探测和判断是很重要的）。然而，环境的刺激则不同。在第一种
339 情况下，无效方向上的运动是和大脑状态联系在一起的（由于微电流刺激）；在第二种情况下，（相反的）优选方向上的运动和大脑状态联系在一起。如果现象的外在论是正确的，那么运动感受性质就应当不同。然而，两只猴子在这两种情况下报告了相同的似动方向（借助于他们被训练出来的对优选 LED 的扫视）。按照上面引证的纽瑟姆及其同事的解释，这表明运动感受性质在这两种情况下是相似而不是不同的。还有其他证据证明这种效应并不是猕猴所特有的。如纽瑟姆及其同事所说：“最近有报道称，未加工的运动知觉可以用人类顶枕叶皮层的电刺激来引起。”（萨尔兹曼等，1992：2352）它也不是运动所特有的，刺激强度的测量方法、行为范

式以及微电流刺激技术都可以推广到其他类型的视觉刺激，包括方位、颜色和立体视觉差异。最近，纽瑟姆及其同事已发表了关于立体视觉深度的相似的微电流刺激的结果（迪安杰利斯［DeAngelis］等，1998）。

就有意识的视觉经验的质的内容来说，重要的是发生“在头脑中” *340*
（大脑中）的东西。推动**现象的**外在论的直觉似乎在经验上是不可信的。正是旧式单一单元神经生理学为这个关于有意识的质的内容的哲学结论提供了经验证据。

考虑关于“单个神经元”神经生理学的第二个例子，这种神经生理学产生的结果可运用于对意识的哲学考虑。麦克亚当斯和曼塞尔（McAdams and Maunsell，1999）研究了外显的有意识注意对猕猴（视觉）的 V4 和 V1 区的单个神经元活动的影响（再看图 14—2）。V1（初级视觉皮层）通过背侧丘脑的外侧膝状体核接收视网膜定位输入。V4——在纹外皮层的更上一层——是通向“腹侧的”（“颞部的”、“什么”）视觉加工流的大门（再看图 14—3）。V4 既包含方位选择神经元，又包含颜色选择神经元。多数神经元都有一种优选的方位或颜色，它引起了最大限度的活动。相似的刺激引起较少的活动，而不同的刺激不引起任何活动（基线反应率以上，参见图 14—7）。

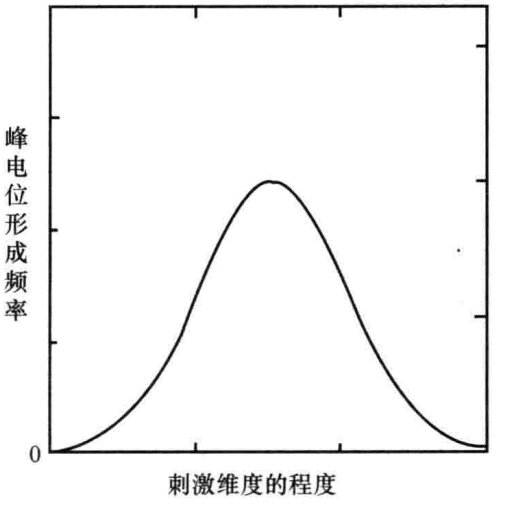


图 14—7　刺激维度选择神经元（如颜色、方位）的调谐曲线表现出一种标准的高斯反应

X 轴上处于曲线最高点之下的维度程度反映了神经元优选的刺激维度。

心理学家早就知道外显的有意识注意会产生得到改进的感觉表现，其

测量方法包括得到改进的探测阈值和更敏锐的辨别力。在个别的感觉神经元层面，外显的有意识注意能够以两种方式改变神经元反应，以此来解释这些行为的改进。[15] 首先，它能增加神经元活动的振幅（参见图 14—8A）。神经元的刺激选择性保持不变，就像两条调谐曲线的相似宽度所反映的那样。刺激所产生的动作电位的频率则增加了，就像实际上所有刺激维度上的调谐曲线的高度所反映的那样。

这种效应被称为“倍增缩放”。神经元反应更强，信号比通常的也更好，这可以对改进的行为探测阈值和速度作出解释。然而，外显的有意识注意的这种作用对于意识爱好者是紧缩的，他们坚持认为，意识至少在其神经实现模式上是“特殊的”或“独特的”。这个结果会使外显的有意识注意的作用与——比如——简单地强化视觉刺激的特征相类似。有意识的注意充当了一种内部的、内源性的机制，它只适合于个别神经元“获取收益”。另一方面，有意识注意可能有一种更强健、更独特的作用，也许它改变了个别神经元的刺激选择性，使这些神经元的活动更准确地显示所注意的刺激特征。在外显的有意识注意的条件下，神经元调谐曲线的峰化就反映了这种作用（参见图 14—8B）。更陡峭的调谐曲线将提供刺激维度更细密的表征，这能提升探测阈值和速度。意识爱好者可能会对这一结果感到振奋，因为增加神经元的刺激选择性不是一种常见的神经生理动力学。

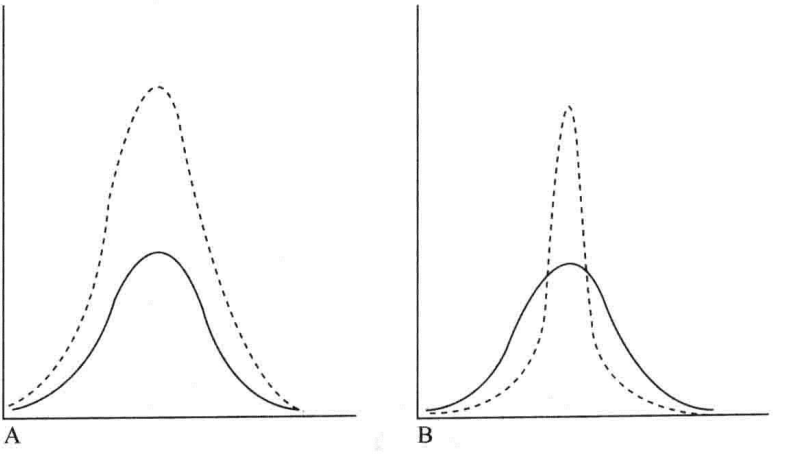


图 14—8　外显的有意识注意视觉神经元的感受野的位置对其活动轮廓可能的两种影响

A：神经元反应的倍增缩放（活动率方面相对稳定地增加，达到各种刺激维度程度），却不增加刺激选择性。B：增加的刺激选择性，这是由神经元的谐调曲线的峰化反映的。

为了检验这些竞争的解释，麦克亚当斯和曼塞尔（1999）提出了一种延时取样匹配任务与 V4 和 V1 中的单细胞记录相结合的方法（参见图 14—9）。他们首先确定 V4 和 V1 神经元的感受野和刺激选择性在试验期间得到记录，图 14—9 所有方框中的椭圆形虚线代表所记录的神经元感受野的位置。在检验试验之前，猴子要注意哪个位置已经受到了提示：神经元 *342*
感受野内的位置或者与它完全相反的位置。试验开始时，猴子注视一个中心点并按下按钮。样本刺激——方向杆或者一小块彩色胶片——在 500 毫秒后出现在屏幕上。一个刺激占据了神经元的整个感受野，另一个占据着相反的位置。这些样本出现在屏幕 500 毫秒后消失。延迟时间持续 500 毫秒，之后检验刺激出现。猴子必须指出出现在提示位置上的检验刺激是否与样本相匹配，它所用的方法是：如果刺激匹配，就在 500 毫秒内放松按钮；如果不匹配，就继续再按住按钮至少 750 毫秒。例如，在图 14—9 所示的情形中，如果猴子得到的提示是注意方向位置，它就必须继续按住按钮，因为样本和检验的方向杆不匹配。但是如果它得到的提示是注意颜色的位置，那么它就必须放松按钮，因为样本和检验的彩色胶片相匹配（尽管这在黑白示意图中不明显）。猴子只有在正确地报告了提示位置是否匹配时才会得到奖赏。两个位置上的匹配和不匹配没有关联，因此猴子注意错误的位置得不到什么好处。

猴子在试验的所有阶段始终都注视着中心点。这保证了细胞的视觉输 *343*
入保持不变。当向抽样检验阶段所提示的位置提供相同的视觉刺激时，已记录的神经活动方面的任何不同都可以归因于猴子注意状态的不同。由于所有 V4 的记录都来自于方位选择神经元，因此在这种动物执行方位匹配任务时，“注意的”（Attended）模式就会出现（参见图 14—9）。这样一来，已记录的神经元就对与匹配任务相关的刺激作出反应。当这种动物执行颜色匹配任务时，“未注意的”（Unattended）模式就会出现，因为已记录的神经元仍对方位刺激作出反应，但这种刺激与我们手头的匹配任务是不相干的。与未注意的模式相比，注意模式中神经元激发率的任何变化都反映了对神经元的感受野位置的外显的有意识注意的神经元影响。

用 200 多个方位选择的 V4 神经元和 124 个 V1 神经元所做的实验均明显地支持倍增缩放假说（参见上面的图 14—8A；参见麦克亚当斯和曼塞尔，1999：图 2、图 4、图 5、图 6、图 7、图 10）。就个别的细胞和群体内的平均值来说，与对同一个方位刺激的未注意的反应相比，注意的反应的

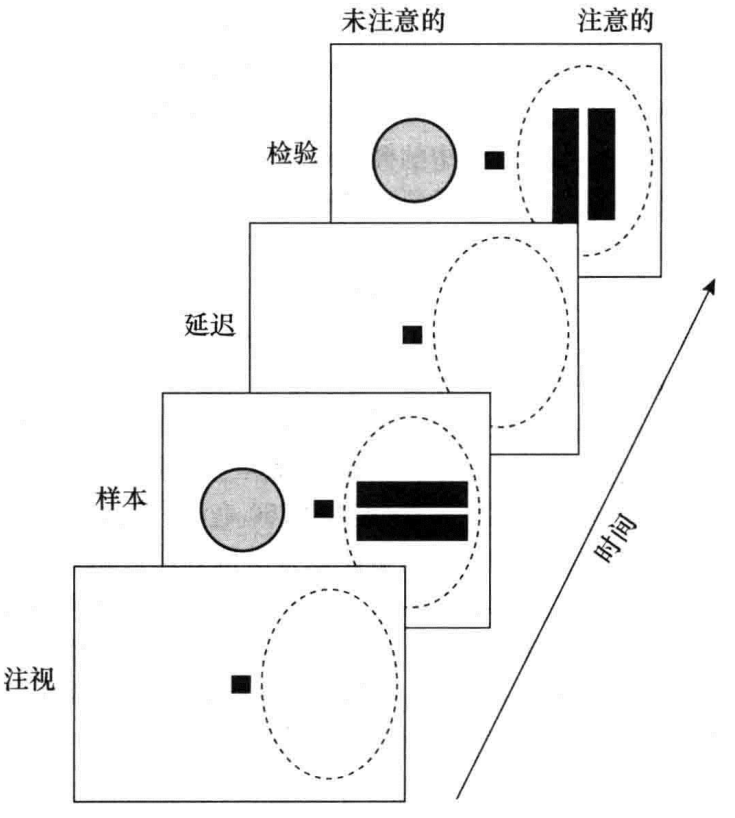


图 14—9　麦克亚当斯和曼塞尔的延迟取样匹配任务的示意图

有关解释参见正文。（改编自氏著，1999：433，图 1）

振幅（动作电位的频率）在统计上要大得多。对感觉神经元的感受野的位置的外显的有意识注意，将其动作电位频率提升到了它喜欢的相关刺激维度程度。然而，对全部刺激维度程度的标准偏离在注意的和未注意的模式之间保持不变。这意味着这两条调谐曲线几乎是等宽的。因此，外显的有意识注意并不影响神经元的刺激选择性。最后，注意的和未注意的调谐曲线几乎有相等的渐进线值。这意味着外显的有意识注意并不影响神经元对“非优选的”刺激维度程度的反应。综合所有这些结果就可以得出一个清晰的结论。引导对感觉神经元的感受野位置的外显的有意识注意，只是增加了这个神经元对优选的和相似的刺激的反应。它只是“获取了收益”，却没有增强这个神经元的刺激选择性。

因此，麦克亚当斯和曼塞尔指出，外显的有意识注意对单个神经元活动的影响，与像处理刺激的特征和反差这样寻常的程序是相同的：

注意的影响与刺激处理的影响之间的现象学类似性提高了这种可能性，即注意涉及的神经机制与在处理来自于视网膜的上行信号时所使用的神经机制相似，皮层神经元会同等地对待视网膜输入和注意输入。(1999：439)

它们的结果支持上面提到的关于意识“紧缩”的观点。就其对单个神经元的影响来说，外显的有意识注意只是另一种“收益增加因素”。

在单个神经元活动的层次上将因果作用归属给外显的有意识注意，听起来可能不可思议。单一的神经元在生物化学上是复杂的离子通道和离子泵，外显的有意识注意会改变通道蛋白质的形状和导电膜的梯度吗？细胞神经科学使二元论复兴了吗？当然没有。来自视觉流中的高级神经区域广泛的兴奋性投射和皮层区内的跨柱状体投射，对内源地产生的单个神经元动力学提供了一种简单易懂的物理解释（吉尔伯特等，2000）。尽管如此，甚至物理主义的意识爱好者也应当对麦克亚当斯和曼塞尔的结论保持谨慎。尽管他们对最终的意识的物理解释已经感到欣慰，但他们仍然坚持要求意识的神经实现和作用的特殊的、独特的本质。不知何故，意识在大脑中所做的必定不仅仅是增加刺激的特征和实现反差。麦克亚当斯和曼塞尔的结论甚至否定了这种意识爱好者。这个结果本身就有**哲学的**意义。它是由“单细胞”神经生理学点缀的，这进一步证明了还原论在神经科学中的潜力，甚至是对关于意识的**哲学**考虑的潜力。

14.4 还原论的神经科学与“困难问题”

有些神经科学家认为，大脑“只是另一种器官”。然而，很多人寻求用神经科学来“认识你自己”，而且直言不讳地表达这种态度。例如，在其很有影响的教科书导论中，神经生物学家戈登·谢泼德描述了研究神经生物学的一些理由。其中有两点特别有启迪意义：

随着我们越来越老，我们会完整地经历人类行为的丰富性——思维和感觉的能力、记忆和创造的能力——而且我们想知道，如果我们还有些惊奇的话，大脑是如何使这成为可能的。(1994：3)

种族主义——对不同的人的恐惧和憎恨——的神经生物学基础是什么？恐怖主义和犯罪是内置于我们大脑回路中的吗？为什么人类似

> 乎一心想通过环境污染和发展毁灭性武器而自我毁灭？为什么我们的大脑中有这种东西，我们怎么能控制它？在一切科学和医学中，神经生物学是能最终解决这些关键问题的唯一领域。(1994：3)

这些并不是某个左派奇想的激昂演说，它们来自于《神经科学杂志》(*Journal of Neuroscience*) 的编辑，它们也不是谢泼德的特有风格。类似的引文可以扩充很多倍。多数神经科学家都不是哲学的门外汉。

这仍不会让有些哲学家满意。很多人仍对心灵的“质的”和“主观的”方面的守护者心存忌妒。他们似乎认为，只有他们（可能还有少数心
345 理学家）在认真解决成为一个有意识的、有心灵的人“是什么感觉”的问题。他们以此暗示心灵的这些特征超出了神经科学家的专业兴趣和范围。但是他们甚至在这一点上也是错误的。考虑下述引自威廉·纽瑟姆的话。他提到的任务就是上一节所讨论的运动方向任务。

> 我相信，**内部经验的本质**对于我们理解神经系统的功能**很重要**……即使我能（用神经的术语）对猴子在我们整个任务中的**行为**作出解释，我也不会满意，除非我知道 MT 中的微电流刺激是否**实际地使猴子看见运动**。如果我们在回答这一问题并理解其含义之前就止步不前，那么我们就是捡了芝麻而丢了西瓜。（加扎尼加，1997：65-66；强调为我所加）

然而，纽瑟姆处理关于意识的“困难问题”并不需要特殊的学科或方法论。在宽阔的经验还原论道路周围并没有捷径：“目前……我猜想我们必须自下而上地向这些野心勃勃的目标摸索前进，让从每个探究层次得来的新发现都为下一层次所问的问题提供线索。”（加扎尼加，1997：67）[16]

在心灵哲学中，“困难问题”的热心守护者应该放轻松。他们不是唯一尊重或追求完整的心灵光环的人。如果神经科学家自身要受到信任，这些问题就不应该处于神经科学的职业兴趣或范围之外。例如，纽瑟姆推断说：“尽管我感觉到这个‘困难’问题对我们的理解力的限制，但整个认知神经科学的努力是很重要的。它是值得为之献身的一项科学事业，它当然优于克隆另一个基因！”（加扎尼加，1997：68）它也优于再编造泛滥了的哲学家幻想的变种，就像孪生地球和神经科学家玛丽那样（这里只说两个例子）。

14.5 走向真正跨学科的哲学和神经科学

20世纪末，“分析”哲学的著名主题之一就是科学与哲学的连续性。有蒯因的论述为证：

> 本体论的问题是和自然科学问题同等的……这个差别只是程度上的差别……它……取决于我们宁可调整科学织造物的这一股绳而非另一股以适应某些特殊的顽强经验这个模糊的实用倾向。(1949：45)

也能以威尔弗里德·塞拉斯（Wilfrid Sellars）为证：“正是这种‘整体眼光’显示出了哲学事业的特征。否则，就很少有东西能将哲学家与坚持反思的专家区别开。”（1962：39）还可以汉斯·赖辛巴赫（Hans Reichenbach）为证：“简言之：写作本书的目的是要指出，哲学已经从思辨进展而为科学了。”（1957：vii）后一代的心灵哲学家对这些主张谨记于心。希拉里·普特南（1960）在可计算性理论中发现了他早期的功能主义灵感。杰里·福多（1975）在认知心理学和乔姆斯基语言学中发现了有利于他的证据。丹尼尔·丹尼特（1978）发现了“意向立场”潜伏在人工智能之中。保罗·丘奇兰德和帕特丽夏·丘奇兰德发现另一种关于认知结构和认知运动学的解释则是从神经科学中产生的。

哲学中这种跨学科转向是整个理智潮流的先锋。跨学科计划开始在所有科学中蓬勃兴起。心灵哲学受这种影响程度之深并不是偶然的事。认知科学尤其是最近的“认知神经科学”是自称的“跨多科性”的最明显（和得到充分资助）的例子。心理学家斯蒂芬·科斯林（Stephen Kosslyn）的描述很有代表性：

> 认知神经科学是对大脑研究、行为和认知研究以及计算系统——它们具有大脑的属性，能产生行为和认知——研究的一种跨学科结合。我认为，认知神经科学不是这些领域的交集、重叠点，而是它们的结合：这不只是说每种方法都限制着其他方法，而是说每种方法都为同一种现象的不同方面提供了真知灼见。(加扎尼加，1997：158-159)

但有，在放弃小心谨慎和无需顾忌礼貌的时候，人们会在认知神经科学家

中间发现一种不同的态度。很少有人会抵制原则上的跨学科的理想。但在实践中，几乎所有人都相信其他学科的人仍对他自己的贡献一无所知。

找到这种态度的公开证据是不容易的。科学的著作往往将这些态度隐藏起来，而有关的哲学家由于希望得到科学家的重视，因此他们只在罕见的情况下才表达它。不过，由神经科学家迈克尔·加扎尼加（1997）编辑的一本书提供了必要的格式，很多读者都将惊讶地看到这种态度的影响多么深远。这本书收入了十位杰出的认知神经科学家的“访谈”，他们来自于构成这项努力的各个学科。在最初发表在《认知神经科学杂志》上时，这些访谈原是电子邮件，只进行了最低限度的编辑。这里的想法是模仿工作之余的攀谈，它们让人兴奋、令人鼓舞，有的甚至催人奋进。（上述引用科斯林的话就来自他的访谈。）其中出现的一个主题是：“跨科性”虽然原则上值得表扬，但在实践中仍是一个神话。

放射线学者马库斯·雷切尔（Marcus Raichle）的工作在发展正电子发射X射线断层照相术（PET）和功能性磁共振成像（fMRI）技术和分析方面很有帮助，他称这些“简单化的行为方法”和“不加区别地用软件包分析数据”是很多功能成像实验的“阿基里斯之踵”（加扎尼加，1997：33）。心理学家兰迪·加利斯特尔（Randy Gallistel）声称，记忆的细胞机制的神经生物学方法甚至受到了关于联想学习的范围和关键参数的过时观念的阻碍：“当前对记忆（的神经基础）的研究，是建立在关于记忆形成的因素是什么的完全错误的概念基础之上的。”（1997：75）计算上复杂而现实的记忆模型是在人类认知心理学和生态学中发展起来的，但这些模型却被神经科学家们忽视了。心理学家恩德尔·塔尔文（Endel Tulving）将这种抱怨又向前推进了一步。研究记忆机制的神经科学家彻底忽视了一半现象：检索（retrieval）。认知心理学家“发现了检索，并在20世纪60年代研究了如何从分析和实验上将它与存储区别开”。这些发现在20世纪70年代使认知心理学中的记忆研究发生了一场革命，但是“那场革命仍未波及脑科学家”（1997：95-96）。语言学家史蒂文·平克（Steven Pinker）坚持认为，进化视角对语言的重要性，以及对“主流进化生物学”的熟悉都没有影响乔姆斯基派的心理语言学家。更直接地说：“绝大多数认知科学家和神经科学都没有真正地思考大脑的进化。”（1997：113-114）神经心理学家阿方索·卡拉马佐（Alfonso Caramazza）认为，来自于一般科学哲学的关于可预测性的过时观点，已经阻碍了研究者对关于语言的新“认知神经心理学”方法的接受，尽管这种方法不断地显示出各种新的缺陷

347

(1997：142-143)。

即使这些只是少数研究者的态度。但是他们都是杰出的研究者。有人对加扎尼加的访谈留下了这样的感觉：低层次上的研究者仍固守于行为方法论以及认知的理论和概念，而这些理论和概念在其原本提出的学科中已落后了30年。关于高层次从业者从低层次研究者得到的细胞和分子机制的知识的相似判断也很常见。例如，回想一下在前面14.1节刚开始时引用谢泼德的话。[17]这些几乎不是人们能指望在这种努力中发现的态度，而这种努力认为自身是最先进的跨学科科学。

这里的问题是：包括认知神经科学在内的每种学科都很困难。这项努力需要一个**共同体**愿意教授**并学习**个别学科所收集到的大量细节的一部分。愿意同在其他层次上工作的人交流是研究者必要的训练过程，但最终认知神经科学要求研究者在各个层次的方法和事实细节方面都接受训练。它需要**跨越**学科的研究者。这是一种令人畏惧的工作描述。但是，它确实给了这些哲学家以希望，他们想对真正的神经科学有所贡献，而不只是对这个学科进行反思。在哲学的“综合视野”与神经科学的事实和实验的细节方面受到研究生学业训练的思想家，就为这项任务作了理想的准备。哲学专业认识到这种潜在的境况是不容易的，但是有些研究生课程、出版社和基金机构正在设法实现某些希望。[18]

【注释】

要特别感谢马里卡·伯恩斯坦（Marica Bernstein）和罗伯特·理查森（Robert Richardson），前者创作和改编了这些图表，并对本文的初稿提了意见，后者对再改稿所提的意见使我作了大量修改。

[1] 在比克尔（1998：第2～3章）中，我提出了这种评价和替代方案。在之后的各章，我将这个一般性解释推广到了心理神经还原的特殊特征。对这种一般性解释的评价，参见理查森（1999）。对于我区别“新潮”还原论和“古典”还原论尝试的一种批判性回应，参见恩迪科特（Endicott，1998）。至于反对我将这个一般理论推广到心理神经情形的一些经验和概念论证，参见斯考滕和德扬（1999）以及我在14.2节中的讨论。

[2] 关于突触传递的优秀研究综述可以在任何不错的神经生物学或生理心理学课本中找到。谢泼德的著作（1994：第6～8章）格外好。对于20年前学习了初等神经科学并且没有继续学习的人需要给予告诫：这个故事已经变了！代谢型受体、第二信使、逆行传递的重要作用以及突触前神经元和突触后神经元中的基因表达上的生物化学效应，已经产生了一幅关于突触传递和突触可塑性的不同图画。我将在随后的讨论

中对这种复杂情况作一些介绍。

[3] 例如，参见谢泼德（1994）著作中的表 29.1 和 29.3，那是一系列历史上的实验支持。这些表只包括 20 世纪 70 年代中期以前的结果。自那时起这两张表有了很大发展。

[4] 我的讨论中将省去大量已知的生物化学。对此方面的一个很好的介绍，参见谢泼德（1994：第 6 章）。

[5] 最近，我们对分子遗传学和 LTP 诱导的生物化学的理解力已经显著提高。对有些新细节的优秀介绍，参见斯夸尔和坎德尔（1999：第 6～7 章）。（顺便提一句，这包括埃里克·坎德尔分享了 2000 年诺贝尔生理医学奖的研究。）

[6] 关于这种细胞生物学“字母”的完整的分子细节，参见坎德尔等（1991：第 65 章）。斯夸尔和坎德尔（1999：第 3 章）包含有最近的发现。

[7] 在比克尔（1998：第 5 章，第 2 节）中，我说明了根据我在那本书第 3 章所提出的关于理论间还原的一般解释，以及这个案例的特征是如何满足所有这些条件的。

[8] 在比克尔（1998：第 5 章，第 2 节）中，我简单描述了另一个例子：关于有等级结构的记忆储存的一个认知理论向哺乳动物感觉皮层中的 LTP 机制的还原。关键的神经科学证据是由神经生物学家加里·林奇（Gary Lynch）、计算机科学家理查德·格兰杰（Richard Granger）及其同事所得到的电生理学的和计算机模拟的结果（参见林奇等，1988；格兰杰等，1989）。

[9] 例如，斯考滕和德扬认为：“比克尔的思想是：在**不提及高层次的功能**的意义上，还原的方法**必须**以纯粹自下而上的方式**进行**。”（1999：253；强调为我所加）

[10] 如果一种高层次理论所假定的实体或过程与低层次理论可以利用的有冲突或者完全不一致，那么这就是否定前者的充分理由。但是，这个“约束条件”是一般的科学方法论的一部分。要排除这类高层次理论，我们不需要特殊的“协同进化”原则
349 （参见比克尔，1996；1998：第 4 章）。

[11] 这里一段自传式的注解是合适的。我记得我还是一名哲学研究生时，在神经科学结业和实验室课堂讨论中被主导着讨论的问题弄得心灰意冷。我们已经读过当时流行的神经科学文献中的论文，里面充斥着丰富的理论思想和含义——并且把课堂讨论会用在了谈论例如摄像机的胶片速度和电极端的直径上。我那时还没有意识到，科学方面的研究生训练有多少是实验设计的艺术。“抽象的批判性推理”的禀赋不能代替给优秀实验人员作学徒的亲身经历。

[12] 这两种科学本身在发展生物学之下逐渐统一起来了。根据突触可塑性和神经元发展共享的分子机制，埃里克·坎德尔邀请我们“将学习想象成……神经元分化的……后期……阶段”（1979：76），这是在 25 年前，自那时起，我们对学习和神经发展共享的分子基础的知识已经增加了（例如，参见谢泼德，1994：第 9 章）。学习是神经元分化的后期阶段，这个论断得到了主要的主流神经科学家的支持。一个学科还有

可能成为“疯狗式”(mad dog)还原论吗?

[13] 这句引语最后一个从句中的模态是骗人的。在这篇文章里,德雷斯基小心地指出,他是在强调现象的外在论的**可用性**而非真实性。莱肯(1996)则要更大胆一些。

[14] 参见利纳斯和丘奇兰德(1996)的文章,特别是利纳斯和佩尔(Paré)合作的著作(第1章),它涉及的是内源的神经生物学证据。

[15] 就这些人来说,他们担心这种关于外显的有意识注意对单个神经元活动的因果作用的谈论近似于故弄玄虚。我们应当感到安慰的是,从神经科学方面对这些作用的解释正处于积极的发展之中。参见我在下面五个段落的简要讨论。

[16] 注意:纽瑟姆“自下而上”的方法论也没有像斯考滕和德扬(1999)所批评的那样“回避了高层次理论”(参见14.2节)。

[17] 也参见坎德尔等(1991)的前言和导论。虽然作者们没有挑选出不了解细胞和分子神经科学的进步的高层次理论家,但从内容看他们明显是主要的批评对象。

[18] 例子包括华盛顿大学的“哲学·神经科学·心理学”课程、牛津大学的“哲学·心理学·生理学”课程、帕特丽夏·丘奇兰德的麦克阿瑟基金会“天才”补助资金、麦克唐纳哲学计划和最近颁给凯思琳·阿肯丝(Kathleen Akins)的神经科学奖金,以及克卢沃学术出版社(Kluwer Academic Publisher)的新杂志《大脑与心灵:跨学科的神经科学和神经哲学杂志》(*Brain and Mind: A Transdisciplinary Journal of Neuroscience and Neurophilosophy*)。

参考文献

Albright, T. D., Desimone, R., and Gtoss, C. G. (1984). “Columnar Organization of Directionally Selective Cells in Visual Area MT of Macaques.” *Journal of Neurophysiology*, 51: 15–31.

Bickle, J. (1995). “Psychoneural Reduction for the Genuinely Cognitive: Some Accomplished Results.” *Philosophical Psychology*, 8 (3): 265–285. 350

—— (1996). “New Wave Psychoneural Reduction and the Methodological Caveats.” *Philosophy and Phenomenological Research*, 56 (1): 57–78.

—— (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P. M. (1985). “Some Reductive Strategies in Cognitive Neurobiology.” Reprinted in *Neurocomputational Perspective*. Cambridge, MA: MIT Press (1989): 77–110.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

DeAngelis, G. C., Cumming, B. G., and Newsome, W. T. (1998). “Cortical

Area MT and the Perception of Stereoscopic Depth." *Nature*, 394: 677-680.

Dennett, D. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.

Dickinson, A. (1980). *Contemporary Animal Learning Theory*. Cambridge: Cambridge University Press.

Dretske, F. (1996). "Phenomenal Externalism." In E. Villaneuva (ed.), *Perception*. Atascadero, CA: Ridgeview: 143-157.

Endicott, R. (1994). "Constructival Plasticity." *Philosophical Studies*, 74 (1): 51-75.

—— (1998). "Collapse of the New Wave." *Journal of Philosophy*, 95 (2): 53-72.

Fodor, J. A. (1975). *The Language of Thought*. New York: Thomas Crowell.

Gazzaniga, M. (ed.) (1997). *Conversations in the Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Gilbert, C., Ito, M., Kupadia, M., and Westheimer, G. (2000). "Interactions Between Attention, Context and Learning in Primary Visual Cortex." *Vision Research*, 40 (10-20): 1217-1226.

Granger, R., Ambros-Ingerson, J., and Lynch, G. (1989). "Derivation of Encoding Characteristics of Layer II Cerebral Cortex." *Journal of Cognitive Neurosciences*, 1: 61-87.

Hawkins, R. D. (1989). "A Simple Circuit Model for Higher-order Features of Classical Conditioning." In J. H. Byrne and W. O. Berry (eds.), *Neural Models of Plasticity: Experimental and Theoretical Approaches*. San Diego: Academic Press: 74-93.

Hawkins, R. D. and Kandel, E. R. (1984). "Is There a Cell-biological Alphabet for Simple Forms of Learning?" *Psychological Review*, 91: 375-391.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hooker, C. A. (1981). "Toward a General Theory of Reduction." Part I: "Historical and Scientific Setting." Part II: "Identity in Reduction." Part III: "Cross-categorial Reduction." *Dialogue*, 21: 38-59, 201-236, 496-529.

Horgan, T. (1993). "Nonreductive Materialism and the Explanatory Autonomy of Psychology." In S. Wagner and R. Werner (eds.), *Naturalism: A Critical Appraisal*. Notre Dame: University of Notre Dame Press.

Jackson, F. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127-136.

Kandel, E. R. (1979). "Cellular Insights into Behavior and Learning." *Harvey Lectures*, 73: 19-92.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (eds.) (1991). *Principles of Neural Science*, 3rd edn. New York: Elsevier.

Kincaid, H. (1988). "Supervenience and Explanation." *Synthese*, 77: 251-281.

Llinás, R. and Churchland, P. S. (eds.) (1996). *The Mind-Brain Continuum*. Cambridge, MA: MIT Press.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Lynch, G., Granger, R., Larson, J., and Baudry, M. (1988). "Cortical Encoding of Memory: Hypotheses Derived from Analysis and Simulation of Physiological Learning Rules in Anatomical Structures." In L. Nadel, L. Cooper, P. Culicover, and R. M. Harnish (eds.), *Neural Connections, Mental Computations*. Cambridge, MA: MIT Press: 180-224.

McAdams, C. J. and Maunsell, J. H. R. (1999). "Effects of Attention on Orientationtuning Functions of Single Neurons in Macaque Cortical Area V4." *Journal of Neuroscience*, 19 (1): 431-441.

McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind*, 98: 349-366.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Murasugi, C. M., Salzman, C. D., and Newsome, W. T. (1993). "Microstimulation in Visual Area MT: Effects Ovarying Pulse Amplitude and Frequency." *Journal of Neuroscience*, 13 (4): 1719-1729.

Nagel, T. (1989). *The View from Nowhere*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Post, J. F. (1991). *Metaphysics: A Contemporary Introduction*. New York: Paragon House.

Putnam, H. (1960). "Minds and Machines." In S. Hook (ed.), *Dimensions of Mind*. New York: Collier.

Quine, W. V. O. (1949). "Two Dogmas of Empiricism." Reprinted in *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Reichenbach, H. (1957). *The Rise of Scientific Philosophy*. Berkeley, CA: University of California Press.

Rescorla, R. A. (1988). "Pavlovian Conditioning: It's Not What You Think It Is." *American Psychologist*, 43: 151-160.

Rescorla, R. A. and Wagner, A. R. (1972). "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement." In

A. H. Black and W. F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts: 64-99.

Richardson, R. C. (1999). "Cognitive Science and Neuroscience: New Wave Reductionism." *Philosophical Psychology*, 12 (3): 297-308.

Salzman, C. D., Murasugi, C. M., Britten, K. R., and Newsome, W. T. (1992). "Microstimulation in Visual Area MT: Effects on Direction Discrimination Performance." *Journal of Neuroscience*, 12 (6): 2331-2355.

Schouten, M. and de Jong, H. L. (1999). "Reduction, Elimination, and Levels: The Case of the LTP-learning Link." *Philosophical Psychology*, 12 (3): 237-262.

Sellars, W. (1962). "Philosophy and the Scientific Image of Man." Reprinted in *Science, Perception, and Reality*. London: Routledge and Kegan Paul.

Shapiro, J. A. (1999). "Genome System Architecture and Natural Genetic Engineering in Evolution." In L. H. Caporale (ed.), *Molecular Strategies in Biological Evolution*. New York: New York Academy of Sciences: 23-35.

Shepherd, G. (1994). *Neurobiology*, 3rd edn. Oxford: Oxford University Press.

Squire, L. and Kandel, E. (1999). *Memory: From Mind to Molecules*. New York: Scientific American Library.

第 15 章　人格同一性

埃里克·T·奥尔森（Eric T. Olson）

15.1　人格同一性问题

人格同一性究竟是什么？这是难以说清的。在这个标题下进行的讨论通常都围绕着以下问题展开。 *352*

我是谁？对多数人来说，“人格同一性”（personal identity）这个短语首先让人想到的是我们称作人的个人的同一性，或者说它就是你在自己身上所看到或所规定的样子，这与你实际的样子可能是不同的。

持存（persistence）。心理学家在谈论人格同一性时，他们通常是在“我是谁?”的意义上理解它的意思。而哲学家一般用它指的是截然不同的东西。通常他们指的是使一个人从某个时刻持续生存到另一时刻的东西，即存在于不同时间的同一个人。他们寻求的是我们的持存条件。什么样的冒险是你有可能幸免的？什么样的东西必然会使你的存在终止？是什么决定着哪个未来的存在者或哪个过去的存在者是你？你指着一张老照片中的女孩说她就是你，是什么使你成为那个人——而不是其他人中的一个？她与现在的你所具有的哪个相关联的方面使得她成为你？在历史上，这个问题经常起因于这一希望，即我们死后可能继续存在。这是否在任何意义上都可能取决于生物学死亡是否是人**有可能**幸免的事情。设想在你死后的阴间或这个世界上真的会有某个人，他在某些方面与你相关。如果有的话，那么，使那个人成为你——而不是（比如说）我，或者成为之前不存在的一个新人的是什么？他要成为你必须怎样与现在的你相联系？

证据。我们如何查明谁是谁？在判定此时此刻的这个人是不是昨天在这里的那个人时，我们借助于什么证据？当各种证据支持相反的判定时，

353 我们应当做什么？证据的一个来源是记忆：如果你能记得做了某件事，或者你至少记得似乎做了某件事，那么可能就是你做了它。另一个来源是物理的连续性：如果做这件事的人看起来和你完全一样，或者甚至更好，如果她在某种意义上是和你在物理或时空上连续的，那么这就是认为她是你的理由。20 世纪 50—60 年代，哲学家就这些标准中的哪一个更基本展开了争论。例如，记忆是否单独就能被当成是同一性的证据，或者是否是仅就它能对照第三人称的“身体方面的”证据来检验而言它才算是证据。这和持存问题不同，但这两者有时是搅和在一起的。什么使你在时间中持存是一回事，我们如何查明你是否持存则是另一回事。如果罪犯的指纹和你的完全一样，法庭可能会判定他就是你。但即便它是真凭实据，使某个过去的存在者或未来的存在者成为你，也并不是因为他具有你的指纹。

人口数量（population）。如果我们认为，持存问题研究的是故事开始时所采用的角色中哪些会存活下来成为故事结束时的角色，那么，我们也能问任意一个时刻舞台上有多少角色。是什么决定着现在有多少个我们？或者一个人在何处离开？下一个人在何处出现？你可能认为：人（people，或个人［person］，我认为这些术语是同义词）的数量就是人类动物的数量——灵长类物种**智人**（Homo sapiens）的成员，也许扣除了有缺陷状态不算作人的成员。但这是有争议的。外科医生有时会将连接一个人大脑两半球的神经束切断（连合部切开术），这会导致奇特的行为，例如一只手在向上提裤子的同时另一只手却在向下拉裤子。这会给我们创造两个人——两个能思维、有意识的存在物吗（参见内格尔，1971。普切蒂［Puccetti，1973］认为，每个正常人的外表之下都有两个人）？患人格分裂的人实际上是两个、三个或七个不同的能思维的存在物的居所吗（威尔克斯［Wilkes］，1988：127f；奥尔森，2003）？

这有时被称为持存问题的“共时的同一性”（synchronic identity）问题，而不是“历时的同一性”（以及下面的“我本可能如何存在”问题的“反事实的同一性”）。我回避这些措词，是因为它们表明同一性有两种：共时的和历时的，它们还导致了这个荒谬的问题：形形色色的东西是否可能是共时地同一但历时地不同，或者反之亦然，只有一种量的同一性关系。只有在两种情况下才会出现人和其他具体事物的同一性和多样性的问题：只涉及一个时刻的共时情况和涉及几个时刻的历时情况。

人格（personhood）。什么成为一个人？什么特征使某种东西成为了一个人而不是非人？在你从一个受精卵开始的发展中，在哪一点你才终于成

为了一个人？什么会使一只黑猩猩或一个火星人或一台电子计算机成为一个人，如果它们能行的话？

有些哲学家似乎认为，所有关于人格同一性的问题都可以归结为这一个。当我们问什么使一个人在时间中持存，或者什么决定我们在任一时间是一个人还是两个人时，他们说我们是在探究我们关于人的概念（例如，佩里［Perry］，1975：7ff.；威克尔斯［Wilkes］，1988：viif.）。我认为这是错误的。倘若我的大脑被移植了，我是否应当和这个器官一起走，普通的“人”的定义对此并没有向我们作任何说明。如洛克所认为的，假定一个人是“有思想、有智慧的一种东西，它有理性、能反思，并且能在异时异地认自己是自己，是同一个能思维的东西”（1975：335），那么，我是一个人，并因而会得到我已移植的大脑的东西。但是这并未告诉我们他和我是两个人还是一个人。 354

我们是什么？从形而上学上说，你、我以及其他人是什么样的东西？我们是物质的还是非物质的？我们是实体、属性、事件还是别的东西？我们是由物质构成的，还是由思想和经验构成的，还是根本不是由什么东西构成的？对这个公认非常模糊的问题有一些可能的回答。我们是人类动物。令人惊讶的是，多数哲学家，既有过去的也有现在的，都反对这种回答。我在后面将对它作更多说明。在历史上，最常见的回答是：我们是没有部分的、非物质的灵魂（或者换句话说，是由一个非物质的灵魂和一个物质的身体构成的复合物，参见斯温伯恩［Swinburne］，1984）。休谟说，我们每个人似乎都是“不同知觉的集合体或一束知觉，它们以难以置信的速度彼此接续着，并处于永远的流动和运动之中”（1888：252。也参见昆顿［Quinton］，1926；罗文［Rovane］，1998：212）。这种观点的一个现代后裔是：你是一种计算机程序，是一种原则上能存储在磁盘上的完全抽象的东西（这是科幻小说中常见的一种思想）。如今最流行的观点可能是：我们是由人类动物“构成”的物质的东西：你是由和某种动物一样的物质构成的，但你和这种动物是不同的东西，因为使你持存的东西不同（威金斯，1967：48；休梅克，1984：112-114；贝克，2000）。甚至还有这种荒谬的观点：我们实际上根本不存在。人类的存在是一种形而上学的错觉。巴门尼德、斯宾诺莎、休谟和黑格尔（如我对他们的解读）以及更近一些的罗素（1985：50）和安格（Unger，1979）都否认他们自己的存在，而且我们在印度佛教中也发现了这种观点。

有什么重要意义？关于同一性和持存的事实有什么实践上的重要意

义？设想外科医生将要把你的大脑放进我的脑袋。由此产生的这个人（他将认为他就是你）对我的行动负责还是对你的负责？还是对两者都负责？还是对两者都不负责？他有权使用谁的银行存款？假如他在手术后会感觉到剧烈的疼痛，除非我们中有一个人提前支付了一大笔钱。如果我们俩都极其自私，那么我们当中哪位应当付钱？

你或许认为对这些问题的回答完全取决于由此产生的这个人**是**你还是我。只有**你**才能对你的行动负责。你不能忽视其未来幸福的唯有你自己。你对自己的未来而不是任何其他人的未来有一种特殊的、自私的兴趣。但是很多哲学家都否认了这一点。他们说，其他人可能对你的行动负责。你可能有一种自私的理由去关心其他人的福祉。我关心或者应该理性地关心明天发生在奥尔森身上的事情，这不是因为他是我，而是因为他和我“在心理上是连续的”，或者他是以别的某种方式——它并不意味着量的同一
355 性——和我相联系的。如果明天其他人和我在心理上是连续的，那么我就应当把我自私的关心传递给他（参见休梅克，1970：284；帕菲特，1971、1984：215；马丁［Martin］，1998）。

我本可能如何存在（how could I have been）？我本可能的存在与我实际存在的样子有多不同？我的哪些属性是我本质上具有的，哪些只是意外地或偶然地具有的？例如，我本可能有不同的父母吗？也就是说，别的父母生的某个人本来可能是我，还是他必定会成为其他人？我——这位哲学家——在获得任何心理特征之前就可能在母亲的子宫中停止存在吗？存在这样的可能世界吗：它们除了难以区别谁是谁之外和现实世界完全一样——那里的人们“位置不断变化”，因此你的经历实际上是我的经历，而且反之亦然？这些问题是否最好描述成关于人格同一性的问题，这是有争议的。（它们当然不是关于其他世界的存在物是否等同于现实世界的人的问题，参见范·英瓦根［van Inwagen］，1985。）但是，它们经常是和其他问题结合在一起讨论的。

我们的概述就这样完成了。这些问题都不相同，应当加以区别。我希望我能说出使这些问题以及它们中的某一个问题成为人格同一性问题的共同特征。但是就我所能看到的来说，除了名字以外没有一种这样的特征。不存在一个人格同一性问题，而是只有很多松散地联系在一起的问题。

我在这一章将聚焦于持存问题，这不是因为它最重要（倘若有的话，也是“我们是什么”的问题），而是因为它支配着自洛克以来关于人格同一性的哲学争论。但是我也将简要地提及其他一些问题。

15.2 理解持存问题

同一性和变化是众所周知的困难话题，即使专家也经常误解持存问题。我们已经提到了将它与证据问题合二为一的趋势，这里再作两点附加说明。

首先，它涉及的是量的同一性（numerical identity）。说各种各样的东西在量上是同一的，就是说它们是**一种**东西而不是两种。如果我们现在指着你之后又指着或描述存在于另一时刻的某个人或某种东西——比如说，某个年龄的男子——那么问题就是：我们是两次指着一个人，还是每次指着两个人中的一个。从量上说你同一于未来的某个，因为他那时拍的照片和你现在拍的照片是同一个人的两张照片。

量的同一性和质的同一性（qualitative identity）不同。当事物完全相同时，它们就是质的同一的。一个过去或将来的人要成为你，也就是说，在量上同一于你，不必和现在的你完全一样。你从生到死不会在质的方面一点不变：你的体形、外表以及其他很多方面都会变化。某个人和现在的 *356*
你完全一样，这也不能保证她就是你。在宇宙的某个地方有另外一个人可能和现在的你完全一样，可能最后一个原子和人格的怪癖都完全一样。然而，你和她不会是同一个人。（你们不能同时出现在两个地方。）**两个**人、两只猫或者两个烤箱可能在质上是同一的。

任何东西都不能**改变**其量的同一性。我们有时会说这样的话——“如果我失去了一切记忆，我就不再是我了”，或者“我不会是同一个人”，或者甚至“我会成为另一个人”。如果这些主张涉及的是量的同一性，它们就是自相矛盾的。任何东西都不能真实地在某个时刻成为一种东西，而后来又成为另一种在量上不同的东西。如果我说在某次冒险经历之后我将成为一个不同的人，或者说我不是此前的我了，那么我的意思必定是说：这个未来或过去的人和我在量上相同一，但在质上的某个重要方面是不同的。否则，它就不是我而是在彼时彼处存在的另一个人了。说这些话的人通常是在“我是谁”的意义上谈论某个人的个人身份。也许通过放弃我现在的身份并获得另一个身份——也就是说，通过改变我的特性或我看待自己的方式——我就能继续存在，而无需和现在的我是同一个人。我应该像一个参议员，她在当选总统后，所当选的官职就不再和她以前当选的一样

了，她已经把第一次当选的官职换成了另一个。在这两种情况下，我们始终都具有在量上相同的东西。

不幸的是，“同一性”和“相同性”（same）经常被用以表示很多不同的东西：量的同一性、质的同一性、个体的心理同一性，等等。更糟糕的是，有些哲学家谈论某个人“幸存下来”（surviving），以及“幸存下来成为”或“变成了”某个人的方式并不暗示量的同一性，因此我能在经历某次冒险之后“幸存下来”，即使我在那之后不存在了，由此混淆在所难免。当**我**问你是否会在经历某件事后幸存下来时，我的意思是说你在这件事的前后是否会存在。

这里有一种不同的误解。持存问题几乎都是这样陈述的：

> （1）在何种可能的情况下，在某个时刻存在的一个人会和存在于另一时刻的一个人相同一（或者是同一个人）？

我们有一个存在于某个时刻的人和一个存在于另一时刻的人，问题就是：“他们”成为一个人而不是两个人的充分必要条件是什么？

这问的是一个错误问题。我们可能想知道你是否曾经是一个胚胎或胎儿，或者你是否能在你的心理特征完全消失后作为一个植物人而存活下来。这些显然是有关什么使我们持存的问题，任何对我们在时间过程中的同一性的解释都应当对它们作出回答。（它们的解答可能有重要的伦理意蕴。）然而，多数对人格问题的回答——例如，前面引用的洛克的回答——都同意：你若没有某些心理的特征，就不可能成为一个人。而专家们说，早期的胎儿和持续处于植物人状态的人没有心理的特征。倘若如
357 此，那么他们严格说来就不是人。因此，如果持存问题就是使一个过去的或未来的人成为你的是什么，那么某个人——她问我们是否曾是胎儿或是否可能变成植物人——就不是在探询我们在时间过程中的同一性，但她显然的确是在探询。

一种对问题（1）的有代表性的回答对这个麻烦作了说明：“存在于某一时刻的一个人同一于存在于另一时刻的一个人，必然是当且仅当前一个人在前一个时刻能想起后一个人在后一时刻所具有的经验，或者反之亦然。”我们可以称之为**洛克主义观点**（Lockean View）。但它可能不完全是洛克的想法。它认为，过去或将来的一个人是你，仅当你现在能想起她那时具有的经验，或者她那时能想起你现在所具有的经验。它不是很可信，但不必太计较。重要的是：洛克主义观点似乎有可能排除你成为一个植物

人的可能，因为植物人不能想起任何事情。也就是说，它似乎可能意味着：如果你陷入了持续的植物人状态，那么由此产生的这个植物人就不会是你的判断。你要么停止了存在，要么转到了阴间。但是，实际上洛克主义观点并没有暗示这些。这是因为我们这里并没有存在于某一时刻的一个人和存在于另一时刻的一个人（假定植物人不是一个人）。洛克主义观点告诉我们你是哪个过去的或将来的**人**，而不是哪个过去的或将来的东西。它向我们说明了使一个人之为一个人持存的是什么，而不是使一个人无条件地持存的是什么。因此它在这里完全不适用。基于同样的理由，它对于你是否曾是一个胚胎也没有作任何说明（奥尔森，1997：22－26；麦凯[Mackie]，1999：224－228）。

因此，问题（1）太狭隘了。我们反而应当问：

> （2）在何种可能的情况下，在某个时刻存在的一个人会和存在于另一时刻的**某种东西**（不管它那时是不是一个人）相同一？

那么，为什么这么多哲学家会问（1）而不是（2）呢？因为他们认为每个人本质上都是一个人：任何实际上是一个人的东西，不可能不是作为一个人存在的。（相反，某种实际上是一名学生的东西，有可能不是作为一名学生存在的：没有一名学生在本质上就是一名学生。）如果真是这样的话，那么任何在某个时刻是一个人的东西，在她存在的所有其他时刻都必定同样是一个人。这一假设使问题（1）和（2）成了等价问题。但是，它是否正确却是另一个重大的问题（“我本可能如何存在”问题的一个实例）。如果你本质上是一个人，你就本不可能是一个胚胎或者一个植物人（如果这些东西不是人）。产生你的胚胎不是在量上和你相同一的，你只有在它发展出了某些心理能力时才会存在。这一假设也排除了我们是动物的可能，因为没有一个动物本质上是一个人：每个人类动物开始时都是一个无思想的胚胎，最终可能都是一个无思想的植物人。

我们是否是动物或者是否曾经是胚胎，这是人格同一性的解释应当回答的问题，而不是我们能根据我们表达这些问题的方式事先解决的问题。 *358*
因此，我们最好不要在一开始就假定我们本质上是人。思考问题（1）会对这个问题作出过早判断，因为它支持关于我们是什么以及什么使我们的持存超过其他人的一些解释（特别是疑问（1）会有效地排除下一节所描述的身体方法）。这就像在排除了它可能是个女子这种可能性之前就问哪个男子犯了罪一样。

15.3 对我们在时间过程中的同一性的解释

对持存问题主要有三种回答。第一种是说，某种心理关系对于一个人的持存或者是必要的或者是充分的（或者既是必要的也是充分的）。你是这个未来的存在物：它在某种意义上继承了你的心理特征——个性、信念、记忆等。你是这个过去的存在物：你已经继承了它的心理特征。我将其称为**心理的方法**（Psychological Approach）。自洛克以来，多数述及人格同一性的哲学家都赞同它的某个版本。洛克的观点就是一个典型例证。

另一种回答是：我们在时间过程中的同一性是某种严酷的物理关系。你是那个过去或将来的存在物，它拥有你的身体或者和你是相同的动物，等等。你是生是死和心理事实毫无关系。我将其称为**身体的方法**（Somatic Approach）。它相当不得人心，但我将在后面为它辩护。

你可能认为，事实处于两者之间：我们的生存既需要心理的连续性也需要物理的连续性；或者也许任何一个都是充分条件而无需另一个。这种看法通常是心理方法的版本。一个检测的实例是：你的大脑——据认为主要负责你的心理特征的上脑（upper brain）——被移植到我的头脑中。（这在物理上是可能的，虽然实践上是一件棘手的事情。）这就产生了两个人：最终拥有你的大脑和你的心理特征的人，和剩下空脑壳的人，他可能还活着但将没有心理特征。如果心理事实与我们的持存是完全相关的，那么你就是得到你的大脑的人。如果你是这个无脑的植物人，那么你的同一性就是某种非心理的东西。

心理方法和身体方法都同意，**有**某种东西使我们能够持存——我们在时间过程中的同一性就是或者必然来自于某种不同于它自身的东西。第三种观点否认这一点。心理的连续性和物理的连续性是同一性的证据，但并不能保证同一性，而且也不是必需的。没有一种连续性对于你的生存是绝对必要或绝对充分的。对持存问题唯一正确的回答是：此时此地的一个人与一个过去的或将来的人相同一，当且仅当他们是同一的。对人们来说，并没有能增进信息的、不同寻常的持存条件。这有时被称为**简单观点**（Simple View，奇泽姆，1976：108ff.；斯温伯恩，1984；洛，1996：
359 41ff.；梅里克斯［Merricks］，1998）。它通常和这种观点结合在一起：我们是非物质的或者是没有部分的，但也不是必须如此。（混合的观点也是

可能的：心理的连续性和物理的连续性对于生存可能是必要或充分的，即使没有什么东西是既必要又充分的。）

简单观点还未得到充分理解，应当给予它更多的注意。然而，我必须省略它。我提到之后将忽略的另一个观点是：我们根本没有持存。没有一个过去的或将来的人是和你在量上相同一的。严格地说，你不是这个刚才开始读这个句子的人（休谟，1888：253；塞德尔［Sider］，1996）。这可能是因为任何东西或者至少任何变化的东西都不能存在于两个不同的时刻。

15.4 心理的方法

心理方法通常似乎是从这种关于人的观念——从这种对人格问题的回答——发展而来（贝克，2000：124）。几乎所有人都同意：成为一个人至少部分地要有某些心理特征。根据定义，人都是心理的存在物。因此，难道他们不是必须有心理的持存条件吗？对一个完全失去其**所有**心理特征的人来说，我们一点也不能排除她的生存吗？一个失去了所有心理特征的人不是一定不再是一个人而且根本就不再存在了吗？这就使某种心理关系对于一个人的持存而言是必不可少的。

但问题并不如此简单。考虑一个类似的论证。成为一个少年，根据定义就是要达到一定年龄。因此，少年不是必须有与年龄相关的持存条件吗？在一个少年失去其少年时代时，我们一点也不能排除她的生存吗？显然不能。我用我自己来鲜活地证明：一个人能够活过他的 20 岁生日。这个类似的论证依赖于这个错误的假设：每个少年本质上就是一个少年，或者至少一旦你是一个少年，那么只要不死，你就不能不是一个少年。最初的论证依赖于类似的假设：每个人本质上就是一个人，或者至少不再是一个人就意味着不再存在。如我们之前所看到的，这个假设绝非毫无瑕疵的。

因此，心理方法并不是明显正确的，而是必须作出论证。最常见的论证是以这个观念为基础的：如果你的大脑被移植到另一个头脑，你会和它相伴，之所以如此，是因为这个器官带着你的记忆和其他心理的特征。但是，要从这个直觉的信念转到一种对持存问题的可信回答，则是出了名的困难。

我们首先必须说明我们在时间过程中的同一性是什么心理关系。15.2

节的洛克观点借助于记忆：一个过去的或将来的存在物是你，仅当你现在能想起她那时所具有的经历，或者反之亦然。这里我们面对着里德（Reid）和巴特勒（Butler）在18世纪所发现的两个著名问题（参见佩里1975年著作中的引文）。

360 第一，假如一位年轻学生因在图书馆借的书超过归还期限而被罚款。在她是一名中年律师时，她记得交过罚款。再后来，在她衰老的时候，她还记得她的律师经历，但把交过罚款和她年轻时所做的一切别的事情忘得一干二净。洛克观点意味着：年轻学生就是中年律师，而律师就是老妇，但老妇却不是年轻学生——这是一个不可能的结果。如果x和y是一个人并且y和z是一个人，那么x和z就不可能是**两个人**。如逻辑学家所说，同一性是传递的，而洛克的记忆连续性却不是传递的。

第二，它似乎属于这种关于记得一段经历的观念：你只能记得你自己的经历。记得交过罚款（或者换句话说，交罚款的经历）就是要记得**你自己**交过罚款。这会使这一主张——即你就是你能记得其经历的那个人——变成平淡无奇，并未增进任何信息（但是它并不影响这个主张：记忆的关系对于同一性是必要的）。如果你不是已经知道某个人是不是具有某段过去经历的人，你就不可能知道他是否真的记得这一过去的经历。在运用这种据认为告诉了我们谁是谁的理论以前，我们还得知道谁是谁。

对第一个问题的一种回答是从直接的记忆关系转到间接的记忆关系：老妇就是年轻学生，因为她能回想起这位律师在某个时候——那时她（律师）记得学生的生活——所具有的经历。传统上解决第二个问题的方法是要杜撰一个新概念："倒摄认知"（retrocognition）或"准记忆"（quasi-memory），它几乎和记忆一样但没有同一性的要求（佩内鲁姆［Penelhum］，1970：85ff.；休梅克，1970）。这种杜撰已经受到了批评，尽管我认为在这里并不重要（麦克道威尔［McDowell］，1997）。但是这两种解答对我们的帮助不是很大，因此洛克观点面临着这个明显的问题：我过去的很多时刻是我根本不能记住的，即使是间接地记住。我现在回想不起昨晚睡着时发生在我身上的事情。但如果说我们知道什么的话，那就是知道我们在入睡时并未停止存在。

今后的最佳办法是根据因果的依赖性来解释心理的连续性（休梅克，1984：89ff.）。存在于后一个时刻的一个人与存在于前一时刻的某个人**具有心理的联系**，仅当后者具有她在后一时刻所具有的心理特征，而这在很大程度上是由于前者具有她在前一时刻所具有的心理特征。我是从一个名

叫奥尔森的年轻人——他很多年前就开始喜欢哲学——那里继承了我现在对哲学的爱好：这是一种典型的心理联系。你在心理上与某个过去的或将来的人相连续，如果你现在的心理特征通过一系列心理联系与她那时所具有的心理特征联系在一起的话，那么我们就能说：在某个时刻存在的一个人同一于存在于另一时刻的某种东西，仅当前者在前一时刻与存在于后一时刻的后者在心理上是连续的。

这仍然留下了一些重要问题没有作出回答。例如，假如我们能将你大脑中的心理内容用电子的方法复制到我的大脑上，从而清除两个大脑之前的内容，由此产生的这个人和刚才的你在心理上是非常相似的。这是否是一种心理连续性的情况？这个问题的回答取决于相关的是哪一种因果依赖性。所产生的这个人在某种意义上继承了你的心理属性，但这不是通常的继承意义。就你能实际地通过“大脑状态转换”从一个人类动物转到另一个来说，这是合适的意义吗？心理方法的拥护者对此存在意见分歧（安格，1990：67-71；休梅克，1997）。

15.5 分体问题（the fission problem）

不管心理的连续性最后归结为什么，我们对心理方法一个极其深切的担忧是：你显然是和**两个**过去的或将来的人在心理上连续的。如果你的大脑被移植，由此产生的新人就是和你在心理上连续的，从而根据心理方法他就是你。而大脑有两个半球，如果其中之一受损，那么由此产生的新人在心理上也是和最初的人相连续的。这里，心理方法赞成现实生活的判断：大脑半球切除术（甚至是切除控制言语的左半球）被看成是针对不宜作其他手术的脑肿瘤的一种极端但可以接受的疗法（里格特林克［Rigterink］，1980）。实际遇到这种情况的人都不会怀疑手术后的这个人就是最初的那个人。因此，心理方法内在地暗示：如果我们破坏你的一个大脑半球并移植了另一个，那么你就是得到移植的大脑半球的这个人。

但是，现在假定外科医生移植了两个半球，每个都移植到一个不同的空脑壳里。把由此产生的人称作左派（Lefty）和右派（Righty），两者都和你在心理上相连续。如果你与**任何一个**和你在心理上相连续的将来的人是同一的，那么由此可以推出：你既是左派又是右派。这就意味着左派就是右派：两种东西不可能在量上同一于一种东西，但是左派和右派显然是

两种东西，因此，你不可能和两者都同一。我们可以用另一种方法得出同样的观点。假如左派在某个时刻感到饥饿而右派没有，那么，如果你是左派，你就感到饥饿；如果你是右派，你就不感到饥饿；如果你既是左派又是右派，那么你就同时既感到饥饿又不感到饥饿，但这是不可能的。

除了彻底放弃心理方法以外，避免这种矛盾似乎只有两种办法。一种是说：尽管有这样的表象，但“你”实际上始终都是两个人——这种立场被异想天开地称为双重占有观（double-occupancy view）（刘易斯，1976；努南［Noonan］，1989：122-148；佩里给出了一个更复杂的变种）。在同一地点有两个不同的但完全相似的人，而且他们同时是由相同的物质构成的，做着相同的事情并有相同的思想，外科医生只是把他们分离开了。根据很多理由，这都是难以置信的。这个例子很重要是因为它意味着：在我们知道后来发生的事情之前，我们不可能知道现在有多少人。（这种看法通常和“四维论”［four-dimensionalism］——这种有争议的形而上学论
362 题：所有持存的东西都在时间上延伸并且是由时间的部分构成的——结合在一起。）

另一种办法是放弃这一主张：心理的连续性本身对于你的持存是充分的。你和一个过去或将来的人——他和你在心理上相连续——是同一的，仅当那时**没有其他人**和你在心理上相连续，这就是一种“无分支观点”（non-branching view）（威金斯［Wiggins］，1967：55；休梅克，1984：85；安格，1990：265；加勒特，1998）。左派和右派都不是你。如果你大脑的两个半球都被移植了，你的生命就结束了——尽管只要一个大脑半球被移植而另一个被破坏，你仍会存活。这太难以置信了，如果你有一半大脑就能存活，那么保留另一半怎么能意味着你**不能**存活呢（参见努南，1989：14-18，149-168）？就此而言，如果你的一个大脑半球被移植而另一个留在原地，你就会死亡（尽管诺齐克1981年著作中的变种避免了这一点）。如果“脑状态转移”给了我们心理连续性，那么，假如在不删除你的脑状态的情况下把你的全部脑状态复制到另一个大脑上，那你就不复存在了。

这里有另一种考虑。在面对移植你的一个大脑半球这一前景时，似乎没有理由提出另一个半球被损坏了。相反，即使它们进入了不同的头脑，难道你不会把两者都保留下来吗？然而根据无分支观点，你宁选死亡也不选继续存在。这就是导致帕菲特等人说你不是真想要继续存在的理由。总之，就你是有理性的而言，你只是想要将来有某个人和你在心理上相连

续，不管他是否在严格意义上就是你。更简单地说，关于谁和谁相同一的事实并没有实践的意义。不过这样一来，我们首先就得弄清楚我们是否有理由接受心理的方法。假如你关心你的两个分体的分支的福祉，就像你通常关心你自己的福祉一样，尽管他们都不是你，那么，你关心这个人——他在最初的移植中得到了你的大脑——身上发生的事情这一事实，并不表明他就是你。

人们有时说，分体不是心理的方法的问题，而是困扰着（除简单观点之外的）所有持存问题的子问题。我不相信它是对身体方法中认为我们是动物[①]的版本而提出的（参见 15.7 节）。我怀疑发生在一个人类动物身上的一切未必会产生两个人类动物，若非另一个存在的话，我们应该乐意将其中的任意之一与最初的人类动物相同一。但是，这里我不能对这一主张进行论证。

15.6 思维动物问题

心理方法遇到的第二个问题和分体没有任何关系。它之所以产生，是因为那种观点暗示了我们不是人类动物。没有一种心理的连续性对于一个 *363*
人类动物的持存是必要的或充分的（卡特［Carter］，1989；艾尔斯［Ayers］，1990：278-292；斯诺登［Snowdon］，1990；奥尔森，1997：80f.；100-109；麦克道威尔，1997：237；威金斯，1980：160，180。后两者明显不赞成这一观点）。不必要是因为每个人类动物都以一个胚胎开始，而且可能都以植物人的状态结束。胚胎或植物人根本没有心理特征，从而也不会和什么东西在心理上相连续。因此，一个人类动物没有任何一种心理的连续性也能持存。如果**你**的持存需要心理的连续性，你就不是一个人类动物。不充分是因为：如果你的大脑被移植到另一个头脑，那么，得到这个器官的人（而不是其他任何人）是和手术前的你在心理上连续的。但是外科医生并不会因此而把某个人类动物从一个头脑转换到了另一个头脑。他们只是把一个器官从一个动物身上转移到了另一个动物身上。（留下来的这个无脑的东西仍然是一个动物，但一个移植走的大脑和一条刚切断的

① 这是笛卡尔的观点，即动物没有心灵，人的身体也没有心灵，因此也是动物。——译者注

手臂一样都不是动物。）任何一种心理的连续性对于一个人类动物的持存都不是充分的。如果它对于你的持存是充分的，那么你也不是一个人类动物了。

心理方法的拥护者都不否认你和某个人类动物——你在照镜子时所看到的人类动物——具有一种亲密的关系。人类动物能思考、有经验。未发育完全和大脑受损的人可能是例外，但是，这些具有成熟神经系统且拥有良好工作状态的人当然能思维。因此，有一个思维的人类动物现在就处于你所在位置。但是**你**的确就是处于你所在位置的能思维的东西。由此可以推出你就是那个动物。而由于这个动物具有非心理的持存条件，因此这就与心理方法相矛盾了。我们将此称作思维动物问题。

如果与你相关的这个人类动物不能思维，这个问题就不会出现，但这是难以置信的。它有一个处于良好工作状态的健康人脑，它甚至和你具有相同的环境和进化历史，是什么使它不能思维呢？如果“你的”动物不能思维，那必定是由于没有任何一种动物能思维。严格地说，动物和树一样没有智能。这表明思维的东西必定是非物质的：即使有物质的东西能思维，它也是一个动物。但是，心理方法的支持者很少会说我们是非物质的。任何人，如果他否认动物能思维但坚持认为我们（能思维）是物质的，那么最好让他对这个惊人主张作出解释。休梅克提出：动物不能思维，是因此它们有错误的持存条件（1984：92-97；1997；1999）。心理属性本质衍推的结果是：心理的连续性对于它们的载体在时间过程中持存是充分的。然而，具有正确持存条件物质的东西是能够思维的，虽然他发现追随者很少（《奴斯》[*Noûs*]，2002）。

另一方面，如果人类动物**能**思维，但你和我却不是动物，那么，在任何我们认为只有一种思维东西的地方，都至少有两种。这一章是一个动物
364 和一个非动物的哲学家共同撰写的，我应当想知道我是哪一个。我可能认为我是非动物，但这个动物有同样的理由认为**它**是这个非动物，就像我认为我是的一样，然而它是错误的。因此，我是如何知道我不是犯错误的那个呢？如果我是动物，我可能仍认为我是非动物。因此，即使我是不同于动物的某种东西，也很难明白我究竟是如何知道这一点的。

就此而言，如果“我的”动物能思维，它可能就和我有相同的心理特征。（否则我们就应该要求对这种不同作出解释。）这应当使它成为一个人。这样一来，人就表现为两种：动物的人和非动物的人。动物的人有非心理的持存条件，但是心理方法声称所有人的持存都是由于心理的连续

性。换句话说，如果人类动物不是人，那么在地球上漫步的有理性、有智能、有自我意识、有道德责任的存在物中至多只有一半是人。成为一个人本身没有什么实践的意义，更何况我们根本不可能知道我们是否是人，这和关于什么是成为一个人的大多数解释是相冲突的。

努南提出用一种语言学假说来解决类似的问题（1989：75f.；1998：316）。首先，不是任何有理性、有自我意识的东西就是一个人，而是只有具有心理持存条件的才是。因此，人类动物并不算是人。其次，像“我”这样的人称代词（以及像苏格拉底这样的名字）总是指称人。因此，当与你相关的动物说“我”时，它并不指称它自己，相反，它指称的是你，即“它的”人。当说“我是一个人”时，你并不是在错误地说**它**是一个人，而是在正确地说你是一个人。因此，这个动物并没有在它是哪个东西上出错，你也没有。你可以从这些语言学事实——即你就是你在说“我”时所指称的东西，“我”总指称一个人——推断你是一个人。你能够知道你不是一个动物，因为根据定义，人具有的持存条件与动物的不同。这个主张所遇到的困难我在这里不能详细阐述。总之，它仍给我们留下了多余的、让人不安的能思维之物，并使人格成了一种微不足道的属性。

当然，解决思维动物问题的另一种办法就是承认我们是动物，并放弃心理的方法。

15.7 身体的方法

心理方法有吸引力，是因为我们在设想心理的连续性与物理的连续性相分离的情况时，很容易相信我们是和前者相伴在一起的。但是一个同样有吸引力的观念是：我们就是动物。当然我们看起来就是这样。在你看见你自己或者另一个人时，你就看见了一个人类动物。如我们已经看到的，这个明显的事实——人类动物能思维——为支持我们是动物的结论提供了强有力的理由。但是，如果我们是动物，那么我们就有动物的持存条件。而动物在时间过程中的持存似乎是由于某种严酷的物理连续性。因此，关于我们是什么的最自然的解释就导致了身体的方法。

少数哲学家认可身体方法，但却不认为我们是动物。他们认为，我们 *365*
就是我们的身体（汤姆森［Thomson］，1997），或者我们在时间过程中的同一性就是我们身体的同一性（艾耶尔［Ayer］，1936：194）。这些就是

所谓的人格同一性的身体标准的诸版本。目前还不清楚它们与我们是动物这种诊断是怎样联系的。人们经常说，某人可能有一部分或完全无机的身体。但是任何动物都不可能部分地或完全地是无机的。如果你截去一个动物的一条肢并代之以无机的假肢，那么动物只是变得更小并装有一条假肢（奥尔森，1997：135）。如果这是正确的，那么你就能同一于你的身体却无需是一个动物。有些哲学家说：动物的身体和动物自身始终是不同的东西：一个动物死后就不再存在了，除非它的死是由特别的暴力所致，否则它的身体会继续作为尸体存在；或者一个动物能在不同的时间有不同的身体（坎贝尔，1994：166）。倘若如此，那么就没有一个人能既是一个动物又与他的身体相同一，但是我不想参与这些争论。我发现身体标准是很难理解的，因为对我来说还不清楚是什么使某种东西成为了某个人的身体（范·英瓦根，1980；奥尔森，1997：142-153）。我相信“人的身体”或“一个人的身体”这个短语是造成很多哲学混淆的原因，最好不使用它们。总之，我们是动物这种观点是身体方法的最清晰、最可信的版本，我将用这一章剩下的篇幅专门讨论它。

我们是动物，这并不意味着所有人都是动物。这与主张存在完全无机的人——上帝、天使或机器人——是一致的。这里的主张是：我们**人类**的人是动物。（人类的人就是与一个像你和我一样的人类动物相联系的某个人：如果你坚持，也可以说是某个有人类身体的人。）它也不意味着这个观点：所有动物甚或所有的人类动物都是人。人的胚胎和处于持续植物人状态的人都是人类机体，但我们可能不想称之为人。事实上，这种观点对于什么使人成其为一个人没有任何暗示。

因此，身体方法只为一部分人提供了持存条件：为我们但未为上帝或天使，如果存在这样的人的话。它还向有些非人（not-people）指派了与他指派给有些人相同的持存条件：人类动物与狗分享了他们的持存条件。这致使有些人提出了这样的异议：它根本不是一种关于人格同一性的观点（贝克，2000：124；也参见洛，1989：115）。这一抱怨有些道理。身体方法并不是打算给出所有人或某些人的持存条件。它甚至意味着我们只是暂时的和偶然的人（根据这个术语通常的定义）。但为什么这是一种反驳呢？如果有些人是动物，那么就没**有**任何持存条件会必然地适用于所有人，就像没有任何持存条件会必然地适用于所有学生或少年一样。这并不意味着成为一个人和成为一个学生都不是一种重要的属性。它仅仅意味着：一种东西成为一个人和它成为一个学生一样，与其在时间过程中的同一

性没有任何关系。在对什么使有些人（即我们自己）的持存作出了说明的意义上，以及在和其他观点——如心理方法，它们对所谓的严格的人格同一性作了解释——竞争的意义上，身体方法是对人格同一性的一种解释。

366

有些人反对这种观念：我们**仅仅**是动物。我们真的不仅仅是动物吗？但是，为什么我们是动物就该意味着我们“仅仅”是动物呢？笛卡尔是一位哲学家，但他不仅仅是一位哲学家，他也是一位数学家和一个法国人。为什么某种东西不能既是一个动物又是一个人、一位祖母、一位社会主义者以及其他很多东西呢？尽管“动物”可能是一个骂人的词语（把某个人称作动物是不礼貌的），但“我们是动物”在最平实的动物学意义上，并不必然意味着我们像畜生一样，或者我们和其他动物没有任何区别，或者我们只有“动物”属性。我们是非常特殊的动物，但我们仍然是动物。

看起来很明显，我们是动物，这与心理方法是不一致的：动物的持存并不是由于心理的连续性。我们还不太清楚是什么使一个动物持存。一个活机体就是某种有**生命**的东西：即这样一个复杂的生物学事件——尽管有大量物质在循环代谢，但它仍保持着一个有机体的结构。这就致使洛克等人相信，仅当其生命继续时，一个有机体才能持存（洛克，1975：330f.；范·英瓦根，1990：142-158；奥尔森，1997：131-140；威尔逊，1999：89-99）。这一点所具有的惊人推论是：一个有机体在死后就不再存在而且不能复活。严格地说，没有像死的动物这样的东西；无论如何，没有什么能既是一个有生命的动物，又是一个死的、腐烂的动物。有些人认为，有生命的动物在死后能作为尸体继续存在（费尔德曼［Feldman］，1992：89-105；卡特，1999；麦凯［Mackie］，1999）。

如我所理解的，活的有机体与尸体是根本不同的。活的东西就像一汪清泉，它的存在是通过不断吸收新的东西，将其特有的形式强加给新的东西，并将残余物排泄出去（米勒［Miller］，1978：140f.）。尸体就像一座大理石雕像，它保持其形式仅仅是由于其材料固有的稳定性。一个有机体死时所发生的变化，比随后其无生命的遗体所发生的变化要大得多。我从未见过一种关于什么使一个动物持存的可信解释允许一个有生命的动物作为一具腐烂的尸体继续存在。不过，这些都是很困难的问题。

15.8 结 论

我相信，心理方法的流行在很大程度上要归功于这一事实，即哲学家们对人格同一性的探究通常都是从这个问题开始的：是什么使得我们在时间过程中持存？（如我们在 15.2 节所看到的，另一个因素就是这个问题通常提出的方式。）但同样重要的问题是：我们是什么？我们是不是动物？如果我们不是动物，我们可能是什么？我们如何与这些动物——有些人把
367 它们称作我们的身体——相联系？这个问题经常被忽视，或者只是作为一个事后的想法（afterthought）来处理。这就是哲学家不理解关于能思动物的问题的原因。也许他们一开始就应该问我们是什么，只有这样，他们才能转向我们跨时间的同一性和别的问题。对此，许多人最终常有不同的看法。

参考文献

引导性的讨论标上了星号。

Ayer, A. J. (1936). *Language, Truth, and Logic*. London: Gollancz.

Ayers, M. (1990). *Locke*, vol. 2. London: Routledge.

Baker, L. R. (2000). *Persons and Bodies*. Cambridge: Cambridge University Press.

Campbell, J. (1994). *Past, Space, and Self*. Cambridge, MA: MIT Press.

Carter, W. R. (1989). "How to Change Your Mind." *Canadian Journal of Philosophy*, 19: 1–14.

—— (1999). "Will I Be a Dead Person?" *Philosophy and Phenomenological Research*, 59: 167–172.

Chisholm, R. (1976). *Person and Object*. La Salle, IL: Open Court.

Feldman, F. (1992). *Confrontations with the Reaper*. New York: Oxford University Press.

* Garrett, B. (1998). "Personal Identity." In E. Craig (ed.), *The Routledge Encyclopedia of Philosophy*. London: Routledge.

Hume, D. (1888). *Treatise on Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press. Original work 1739. Partly reprinted in Perry (1975).

Lewis, D. (1976). "Survival and Identity." In A. Rorty (ed.), *The Identities of Persons*. Berkeley: California. Reprinted in his *Philosophical Papers*, vol. I. New York: Oxford University Press (1983).

Locke, J. (1975). *An Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Clarendon Press. Original work, 2nd edn., first published 1694. Partly reprinted in Perry (1975).

Lowe, E. J. (1989). *Kinds of Being*. Oxford: Blackwell.

—— (1996). *Subjects of Experience*. Cambridge: Cambridge University Press.

Mackie, D. (1999). "Personal Identity and Dead People." *Philosophical Studies*, 95: 219–242.

Martin, R. (1998). *Self Concern*. Cambridge: Cambridge University Press.

McDowell, J. (1997). "Reductionism and the First Person." In J. Dancy (ed.), *Reading Parfit*. Oxford: Blackwell.

Merricks, T. (1998). "There Are No Criteria of Identity Over Time." *Noûs*, 32: 106–124.

Miller, J. (1978). *The Body in Question*. New York: Random House.

Nagel, T. (1971). "Brain Bisection and the Unity of Consciousness." *Synthèse*, 22: 396–413. Reprinted in Perry (1975) and in Nagel, *Mortal Questions*. Cambridge: Cambridge University Press (1979).

* Noonan, H. (1989). *Personal Identity*. London: Routledge.

—— (1998). "Animalism Versus Lockeanism: A Current Controversy." *Philosophical Quarterly*, 48: 302–318.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Olson, E. (1997). *The Human Animal*. New York: Oxford University Press.

—— (2002). "What Does Functionalism Tell Us About Personal Identity?" *Noûs*, 36: 682–698.

—— (2003). "Was Jekyll Hyde?" *Philosophy and Phenomenological Research*.

—— (forthcoming). "There Is No Bodily Criterion of Personal Identity." In F. MacBride and C. Wright (eds), *Identity and Modality*. Oxford: Oxford University Press.

Parfit, D. (1971). "Personal Identity." *Philosophical Review*, 80: 3–27. Reprinted in Perry (1975).

—— (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Penelhum, T. (1970). *Survival and Disembodied Existence*. London: Routledge.

Perry, J. (1972). "Can the Self Divide?" *Journal of Philosophy*, 69: 463-488.

—— (ed.) (1975). *Personal Identity*. Berkeley: University of California Press.

Puccetti, R. (1973). "Brain Bisection and Personal Identity." *British Journal for the Philosophy of Science*, 24: 339-355.

Quinton, A. (1962). "The Soul." *Journal of Philosophy*, 59: 393-403. Reprinted in Perry (1975).

Rigterink, R. (1980). "Puccetti and Brain Bisection: An Attempt at Mental Division." *Canadian Journal of Philosophy*, 10: 429-452.

Rovane, C. (1998). *The Bounds of Agency*. Princeton: Princeton University Press.

Russell, B. (1918). "The Philosophy of Logical Atomism." *Monist*, 28: 495-527; *Monist*, 29: 32-63, 190-222. Reprinted in R. Marsh (ed.), *Logic and Knowledge*. London: Allen and Unwin (1956), and in D. Pears (ed.), *The Philosophy of Logical Atomism*. La Salle, IL: Open Court (1985); page numbers from the latter.

Shoemaker, S. (1970). "Persons and Their Pasts." *American Philosophical Quarterly*, 7: 269-285.

* —— (1984). "Personal Identity: A Materialist's Account." In S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Blackwell.

—— (1997). "Self and Substance." *Philosophical Perspectives*, 11: 283-319.

—— (1999). "Self, Body, and Coincidence." *Proceedings of the Aristotelian Society*, Supplementary volume 73: 287-306.

Sider, T. (1996). "All the World's a Stage." *Australasian Journal of Philosophy*: 433-453.

Snowdon, P. (1990). "Persons, Animals, and Ourselves." In Christopher Gill (ed.), *The Person and the Human Mind*. Oxford: Clarendon Press.

* Swinburne, R. (1984). "Personal Identity: The Dualist Theory." In S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Blackwell.

Thomson, J. J. (1997). "People and Their Bodies." In J. Dancy (ed.), *Reading Parfit*. Oxford: Blackwell.

Unger, P. (1979). "I do not Exist." In G. F. MacDonald (ed.), *Perception and Identity*. London: Macmillan. (Reprinted in M. Rea (ed.), *Material Constitution*, Lanham, MD: Rowman and Littlefield.)

—— (1990). *Identity, Consciousness, and Value*. New York: Oxford University Press.

Van Inwagen, P. (1980). "Philosophers and the Words 'Human Body'." In van Inwagen (ed.), *Time and Cause*. Dordrecht: Reidel.

—— (1985). "Plantinga on Trans-World Identity." In J. Tomberlin and P. van Inwagen (eds.), *Alvin Plantinga*. Dordrecht: Reidel.

—— (1990). *Material Beings*. Ithaca: Cornell University Press.

Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.

—— (1980). *Sameness and Substance*. Oxford: Blackwell.

Wilkes, K. (1988). *Real People*. Oxford: Clarendon Press.

Wilson, J. (1999). *Biological Individuality*. Cambridge: Cambridge University Press.

第16章　自由意志

伦道夫·克拉克（Randolph Clarke）

369 我们通常认为，我们在作出决定和采取行动时是自由的[1]，并且由于若干原因，我们享有的这种自由对我们非常重要。自由地决定和行动，其实就是拥有对自己所做的事情的一种有价值的控制权。我们认为，拥有这一权利在一定程度上构成了人类的尊严。人们普遍认为，只有当自主体能够控制她①所做的事情时，她的决定以及其他的行动才能归属于她，进而她才能为其所做的事情承担责任，如基于她的决定以及其他的行动的道德性质，或受到表扬或受到批评，或受到褒奖或受到惩处。此外，我们希望出现这样的情况，即我们通过自由地行使控制权能够影响这个世界上发生的事情，包括我们想成为哪种人。当我们慎思时，我们一般认为，不止一种选择向我们开放，并且我们可以自由地追求我们正在思考的选择。如果这种想法整个都是错误的，那么我们常常会陷入一种意想不到的幻觉之中。

事实上我们是否享有在这些方面所看重的自由？一些人认为，我们没有这种自由，因为（他们认为）我们的世界是决定论的。[2]（如果自然规律使得这个世界在任何一个给定的时间点上的状态充分地决定了它在后来任何一个时间点上的状态，那么这个世界就是决定论的；下面我们将要更仔细地探讨决定论。）决定论的世界中不可能有自由意志的这种观点被称为不相容论。虽然一些不相容论者肯定了决定论，但是其他的人，即那些被称为自由主义者的人，否定了决定论而肯定了自由意志。值得注意的是，不相容论者可能既相信这个世界是非决定论的，同时又相信我们仍没有自由意志。

许多哲学家拒斥不相容论而支持相容论。相容论者认为，即使在决定

① 作者克拉克通篇都以“她”举例。——译者注

论的世界中，自由意志也能存在。尽管一些相容论者相信，我们的世界是决定论的，但是其他的人仍坚持认为，这个世界是非决定论的，或者没有必要对这个世界是否是决定论的问题作出承诺。

16.1 相容性问题

自由意志与决定论相容吗？在处理这个问题之前，我们需要理解什么是决定论。我们或者将其理解为这个世界可能有也可能缺乏的特征，或者将其理解为我们的世界实际上具有的那种特征。这两种方式都可以用来理解决定论，但是，这里所说的那种特征又是什么呢？[3]

16.1.1 决定论

决定论有时被简单地理解成每一个事件都有一个原因。然而这种说法是不正确的。正如下面我们将要看到的，可能存在非决定论的因果关系，有一些事件可能是被引起的却不是被决定的。如果是那样的话，也许每个事件都有一个原因，但这个世界却不是决定论的。[4]

决定论可以通过下述方式予以恰当地描述，即各个世界如果有相同的自然规律并且在某个时间点上是相似的，那么它们可能会是什么样的状态。以这种方式看，我们的世界仅在下述条件下才（在两种不同的时间方向上）是决定论的，即任何一个可能世界如果有与我们的世界完全相同的自然规律并且在任何一个时间点上与我们的世界极其相似，那么它在所有的时间点上都将与我们的世界极其相似。将范围稍微缩小一些，导向未来的（future-directed）决定论仅在下述条件下才存在于我们的世界之中，即任何一个可能的世界如果有与我们的世界完全相同的自然规律并且在任何一个给定的时间点上与我们的世界极为相似，那么它在**后来的**每一个时间点上将都与我们的世界极其相似。

如果这个世界在任何一个给定的时间点上的状态可以由一个命题完备地描述，并且如果所有的自然规律同样可以由一个命题完备地陈述，那么，我们可以通过一些命题和（广义的）逻辑必然性（即在每一个可能世界中都是真的）给出决定论的等价描述（equivalent characterizations）。导向未来形式的决定论仅在下述条件下才存在于我们的世界之中，即对于下述三个命题：命题 p：完备地描述了这个世界在任何一个给定的时间点上

的状态；真命题 q：（只是部分地）描述了这个世界在后来的某个时间点上的状态；命题 l：完备地陈述各种自然规律。如果（p 和 l）那么 q 在逻辑上是必然的。我们以模态逻辑符号将这个句子写作：□［（p&l）⊃q］。

16.1.2 后果论证

支持自由意志无法相容于决定论这一观点的论证，就是各种形式的所谓后果论证（Consequence Argument），近年来这些论证得到了最为广泛的讨论。在后果论证的文本中，我们将看到这样的观点，即自主体仅在下述条件下才是根据自由意志来行动的，即她对于是否采取那种行动作出了选择，或者说是她要做什么取决于她，抑或她同样有能力做她所做的那种行动之外的事情。这一论证可以作如下非正式的表述：

> 如果决定论是真的，那么我们的行动就是自然规律和很久以前发生的事件的结果。但是，我们出生之前发生了什么不取决于我们，自然规律是什么也不取决于我们。因此，这些事件（包括我们目前的行动）的结果也不取决于我们。[5]

使得这种非正式的论证变得更加精准的方法有很多。我们在这里将重点讨论其中一种备受人们关注的方法。[6]

有一种版本的后果论证运用了一个模态算子“N”，当将其附加到任何一个句子 p 上时，我们就会得到一个新句子，它表示的是：p 且人类自主体无论是现在还是过去对是不是 p 都不能作出任何选择。（有时也可以这样表述：“Np”表示的是，p 在任何时间对于所有的人类自主体而言都具有强必然性。）例如，“P”表达的命题是，地球围绕太阳转。而“NP”表示的就是：地球围绕太阳转且人类自主体无论是现在还是过去对地球是不是围绕太阳转都不能作出任何选择。

这一论证依赖于以下两个与强必然性有关的推导规则：

（α）□p ⊢Np

（β）N（p⊃q），Np ⊢Nq

规则（α）表示的是：p 在（广义的）逻辑上是必然的这一前提衍推了 p 具有强必然性；如果 p 在逻辑上是必然的，那么就可得出这样的结论：p 且人类自主体无论是现在还是过去对是不是 p 都不能作出任何选择。（β）表示的是：p⊃q 具有强必然性和 p 具有强必然性这两个前提衍推了 q 也具有

强必然性；如果 p⊃q 且人类自主体无论是现在还是过去对是不是 p 都不能作出任何选择，那么就可得出这样的结论：q 且人类自主体无论是现在还是过去对是不是 q 都不能作出任何选择。

现在假设“H”代表的句子所表达的命题完备地描述了这个世界在人类诞生之前的某个时间点上的状态。假设“L”代表的句子所表达的命题完备地陈述了所有的自然规律。假设“A”代表的句子所表达的真命题完备地陈述了这个世界在“H”所涉及的时间点之后的某个时间上的状态（例如，“A”代表的可能是克拉克同意撰写这篇论文）。现在假设（两种中的任意一种）决定论存在于我们的世界中。基于先前对决定论的描述，根据这个假设可以得出：

（1）□［（H&L）⊃A］。

步骤（1）在逻辑上等价于

（2）□［H⊃（L⊃A）］。

通过将规则（α）运用到步骤（2）上，我们可以得出

（3）N［H⊃（L⊃A）］。

现在这一论证将下述假定作为前提

（4）NH。

接着将（β）运用到步骤（3）和（4）上，我们得到

（5）N（L⊃A）。

现在这一论证再将以下假定作为第二个前提

（6）NL。

接着通过将（β）运用到步骤（5）和（6）上，我们得出结论：

（7）NA。

这一论证如果是正确的，那就表明：如果这个世界是决定论的，那么，基于克拉克实际上同意撰写这篇论文，克拉克同意撰写这篇论文且人类自主体无论是现在还是过去对克拉克是否同意撰写这篇论文都不能够作出任何选择。（这令我非常震惊，因为在这个问题上我认为**我**过去是能够作出选择的！）由于“A”可以被任何一个这样的句子取代，即这个句子正确地表

达了这个世界在“H”所涉及的时间点之后的某个时间点上的状态，因此，任何一个人类自主体所采取的任何行动也是如此。这个论证如果是正确的就表明了：如果这个世界是决定论的，那么，对于任何一个人类自主体是否能够采取任何这样的行动而言，所有自主体无论是现在还是过去都不能作出任何选择。这一论证主要是为了表明，决定论完全排除了自由意志，完全排除了我们对我们要做的事情所具有的选择权。

16.1.3 对后果论证的评价

这一论证依赖于两个前提。第一个前提是论证中的步骤（4），即 NH，大概说的是，我们对很久以前（在我们所有人存在之前）发生的事情不能作出选择。第二个前提是论证中的步骤（6），即 NL，大概说的是，我们对自然规律是什么也不能作出选择。在很多人看来，这两个前提显然是正确的。[7]但是，由于这两个前提正确与否依赖于如何解释某些事件能够“作出选择”，因此，否定其中的一个前提并不像初看上去的那样难以置信。

根据我们对决定论的描述，假设这个世界是决定论的，那么，如果任
373 何一个自主体本来可以做她实际上没有做的事情，那么，要么就是这个世界在每一个稍早的时间点上有所不同（因而“H”是错误的），要么就是自然规律有所不同（因而“L”是错误的）。“多重过去相容论者”（Multiple-Pasts Compatibilists）选取了第一个析取支（disjunct）。他们主张：如果这个世界是决定论的，那么，我们就有能力做一些事情；而如果我们真的做了，那么过去（在每一个时间点上）就有所不同。[8]在这种意义上，他们承认了，我们据说对久远的过去能够作出选择。但是，他们认为，这一主张不同于他们并**没有**承诺的那种更强的主张，即我们有能力做一些**在因果上**能够影响过去的事情。他们声称，一旦我们将这两种主张区别开来，并且看清楚他们所承诺的是哪种主张，那么，他们观点中的那种不可思议性的情况就会消失。

“局部奇迹相容论者”（Local-Miracle Compatibilists）选取了上面所说的第二个析取支。他们宣称，如果这个世界是决定论的，那么，我们就有能力做一些事情；而如果我们真的做了，自然规律就会有所不同。[9]（他们说，如果一个自主体过去做了一些她实际上没有做过的事情，那么，在这种可选择的行为产生之前可能存在使这个行为发生的某个违反规律的事件——某种奇迹。这个奇迹事件可能违反了某个现实的自然规律，但没有

违反它所在世界中的任何一个规律。那个世界包含了这种奇迹事件，但是，过去它在别的方面与我们的世界相似，因而它的规律不同于现实的规律。）这些相容论者在这一意义上承认了我们对自然规律是什么能够作出选择。但是，他们认为，这一主张不同于他们并**没有**承诺的那种更强的主张，即我们有能力采取这样的一些行动，即它们要么**是**违反规律的事件，要么**引起**了违反规律的事件。虽然这种更强的主张让人难以置信，但是，他们所承诺的这种稍弱的主张，据说也是有争议的。

很多人甚至发现，这两种相容论者所**承诺**的主张让人难以信服。[10]无论如何，这种后果论证的两个前提仍然是争论的焦点。这一论证所依赖的推导规则（β）也引起了广泛的争论。[11]

（β）是否是一条有效的推导规则依赖于如何解释算子“N”，这一问题反过来又依赖于如何解释“作出选择”。假设我们根据多重过去相容论者和局部奇迹相容论者所提出的思想来理解“作出选择”。那么，我们会说，一个自主体仅在下述条件下才能够对是否 p 作出选择，即她有能力采取某种行动，而如果她真的这样做了，p 就不是真的。那么“Np”表示的是：p 且所有人类自主体在任何时候都没有能力采取任何这样的行动，以至要是她采取了这种行动，p 就不是真的。碰巧的是，这些例子表明了，如果这样解释“N”，那么（β）就是无效的。

这儿有这样的一个例子。[12]假设只有一个人类自主体山姆存在。① 山姆有一小块儿镭，镭这种东西有时会放射出亚原子微粒；不过这个给定的一小块儿镭，在一个给定的时间里是否会放射出微粒是不确定的。山姆在时间 t 之前摧毁了这一小块儿镭能够确保其不会在 t 时放射出微粒，这也是山姆确保这一点的唯一方式。山姆对于他在 t 时是否摧毁这一小块儿镭 *374*
能够作出选择；他也能够禁止这样做。假设“R”表示的是，这一小块儿镭在 t 时没有放射出微粒；假设“S”表示的是，山姆在时间 t 之前摧毁了这一小块儿镭。于是我们有关于（β）的下述例证：

（1）N（R⊃S），
（2）NR，因此
（3）NS。

如果条件句“R⊃S”中的前件和后件都是真的，那么这个条件句也是真

① 假设只有一个自主体存在是为了排除其他自主体干涉的可能。——译者注

的。假设这个例子中的前件和后件都是真的，如果山姆能够做某件事情，以至要是他真的做了，R&¬S，那么，他就能够做某件事情，以至他要是真的做了，这个条件句就是假的。① 但是，山姆做任何事情都不能够确保R&¬S。（他能够禁止摧毁这一小块儿镭，但是要是他真的这样做了，这块镭就可能在t时放射出微粒。）因此，关于（β）的例证的第一步就是正确的。

第二步也是正确的。这一小块镭儿没有在t时放射微粒。由于它是否放射出微粒是不确定的，因此山姆不论做什么事情都不能确保这些镭在t时放射出微粒。（他能够禁止摧毁这一小块儿镭，但是要是他真的这样做了，这一小块儿镭也有可能在t时放射不出微粒。）

不过，正如我们假设的那样，山姆可以在时间t之前禁止摧毁这一小块儿镭。因此这一结论即步骤（3）就是错误的。因此，根据当前对“N”的解释，我们有一个关于规则（β）的反例，即在关于（β）的一个例证中，所有的前提都是正确的，但结论却是错误的。如果这样来解释“N”，那么规则（β）就是无效的。

有人提出一种取代（β）的不同推导规则[13]，或对（β）中的“N”作出一种不同的解释，来回应这种不相容论论证的辩护者。根据后一种方法，我们很容易理解：稍微修改一下我们对“N”的早期解释，就完全可以使（β）避免当前的这种反例。例如，自主体只有在下述条件下才能够对是不是p作出选择，即她有能力采取某种行动，以至要是她采取了那种行动，p就**可能**不是真的。“Np”现在表示的就是：p且任何一个人类自主体在任何时候都没有能力采取任何这样的行动，以至要是她真的做了，p就可能不是真的。[14] 如果这样解释“N”，那么，在关于镭的这个例子中，（β）的例证中的两个前提就是错误的。（山姆可以在时间t之前禁止摧毁这一小块儿镭；要是他真的这样做了，R&¬S就可能是真的；并且要是他真的这样做了，¬R就可能是真的。）因此，如果这样来解释“N”，那么这一小块儿镭的例子就不是（β）的反例，其他的反例似乎也就不存在了。如果这样来解释“N”，那么（β）似乎就是一个有效的推理规则。[15]

我们所考察的这个不相容论论证十分有力。基于对前一段提到的关于“N”的解释，这种论证所依赖的推理规则（β）似乎是有效的。这种论证

① 在条件句中，除了前真后假为假之外，其余皆真。因此，在条件句R⊃S中，除R&¬S为假之外，其余皆真。——译者注

的两个前提（正如我们看到的那样）似乎是十分合理的，但是，这两个前提中的任何一个是不是正确的，仍有怀疑的余地。

16.2 相容论者的说明

虽然一些相容论者坚持主张：即使这个世界是决定论的，在一般的情况下，当我们行动时，我们原本仍有能力采取别的行动，但是，其他的相容论者则认为，这种决定论或许排除了这种能力。回想一下，在评价自由意志时，我们对某种控制权感兴趣，因为我们相信它与下述这些东西有关，即人类的尊严、道德责任、影响的发挥以及选择的开放性。一些相容论者认为，我们或许享有各种各样的控制权，但它们只能满足这些东西中的某些部分，而不能满足全部。尤其是一些相容论者持如下观点，即如果这个世界是决定论的，我们或许常常缺乏采取别的行动的能力，但是一般而言，我们能根据（在一定程度上构成了人类尊严的）道德责任所需的那种控制权来行动。某些支持者将这种观点命名为“半相容论”（semicompatibilism）。[16]

16.2.1 法兰克福的例子

半相容论者拒斥了那个长期以来被人们广泛接受的与责任有关的观点，我们可以将其表述为如下的“可供取舍的可能性原则”（principle of alternate possibilities）：

> （PAP）自主体只有在下述条件下才对她所做的事情承担道德责任，即她本来能够做别的事情。

希拉里·G·法兰克福（1969）所列举的一些例子是导致许多人（如相容论者和不相容论者等）拒斥 PAP 的最主要原因。

法兰克福注意到，一个自主体可能在这样的一些情况下行动，即这些情况不仅是她采取某一行动的充分条件，还使得她不可能采取别的行动，但是实际上又没有产生她的行动。法兰克福认为，当自主体在这些情况下行动时，她不可能采取别的行动这一事实不能免除她的责任。在法兰克福提出的一些例子中，有一个例子可以阐明这些主张：

> 假设某个人，比如说布莱克，想要琼斯采取某个行动。布莱克打

> 算竭尽全力实现他的目的，但他更倾向于避免不必要地表现出他的意图，因此他会一直等待，直到琼斯即将决定自己将要做的事情。然而，除非布莱克知道琼斯即将决定要做的事情并不是他想要琼斯做的事情（他擅长判断这种事情），否则就不会采取任何措施。一旦布莱克知道琼斯即将决定做某件别的事情时，他就会采取有效措施以确保
> 376 琼斯决定做并且真的做了他想要琼斯做的事情。无论琼斯最开始的偏好和倾向是什么，布莱克都将实现他的目的……
>
> 现在假设布莱克之所以一直都没有表明他的目的，是因为琼斯是由于自身的缘故决定做并且真的做了布莱克想要他做的那件事情。(1969：835-836)

法兰克福在这里宣称，在这个例子中，所有的条件——布莱克的出现和他所作的干涉准备——都具备了，但是，琼斯除了做其实际上所做的事情之外，不可能做别的事情。不过，琼斯没有意识到这些条件，这些条件也未丝毫影响到他的决定和行动；要是布莱克没有出现，琼斯就会做他原本要做的那些决定和行动。在这个例子中，我们不能由于琼斯不能够做别的行动就该免除他对他做过的事情必须承担的责任。因此，PAP 是错误的。

对这个反对 PAP 的例子的讨论已相当广泛。[17]在这里，由于篇幅的限制，让我们简要地强调一下与法兰克福论证的重要性有关的两个要点。第一，仅根据 PAP 的错误性还不能得出结论说责任与决定论相容。因为决定论可能排除责任，即便它排除责任不是因为它排除了做别的事情的能力。[18]第二，如果 PAP 是错误的，那么在评价相容论者的说明时，我们需要留意这些说明意图阐明的东西是什么。事实上，关于行动自由的最先进的相容论说明近来已被提出，用以说明与控制权有关的什么因素才是道德责任所必需的。如果法兰克福是对的，那么，人们不能仅通过证明决定论排除了做别的事情的能力（如果这一点能够被证明的话）就来证明这些说明就是错误的。

16.2.2 分层说明

我们现在来研究一些最有名的相容论说明。法兰克福本人提出了一种观点，这种观点运用了态度分层的思想，这种思想也被其他几个相容论者所利用。[19]法兰克福指出，人们不仅能够具有采取（或不采取）某些行动的愿望，即他所谓的一阶愿望，还能够对他们自己所具有的这些一阶愿望

进行反思和作出批判性的评价。基于这种反思性的自我评价，我们就能够形成二阶或高阶愿望，诸如想要具有（或不具有）某些低阶愿望的愿望。在这些高阶愿望中具有特殊利益的一些愿望，就是法兰克福所谓的二阶意愿（volition），它们是想要使某些一阶愿望成为（或不成为）导致人们行动的愿望的愿望。当具有相互冲突的一阶愿望的自主体所形成的一个二阶意愿想要其中的某个愿望成为指引她行动的愿望时，她自己就因此“认同”了那个愿望，法兰克福称那个成功地导致自主体行动的一阶愿望为自主体的意志。

法兰克福在自由意志与自由行动之间作出了区分。在他看来，一个人的意志是自由的，仅当：

> 对于他的一阶愿望中的任何一个，他要么自由地使这个愿望成为他的意志，要么自由地使其他某个一阶愿望成为他的意志。因此，无论他的意志是什么，这一意志自由的人的这一意志都可能是别的意志；他除了能够形成他已形成的意志之外，还能够形成别的意志。（1971：18-19）

相反，法兰克福最初主张，自由行动的充分条件是，自主体“已做了他想做的事情；他做这件事情是因为他想做；当他这样做时，他据以行动的意志由于是他想要的从而成为他的意志”（1971：19）。根据这种观点，即使一个自主体没有能力做别的事情，即使她没有自由意志，她也可以自由地行动。根据法兰克福所说，道德责任所需的是自由行动，而不是自由意志。

法兰克福对自由行动所作的这种早期说明已经产生了若干问题。第一，正如法兰克福本人所注意到的（1971：21），就像自主体的一阶愿望之间可能存在冲突一样，在这个层次结构中，任何一个较高层次的内部也会存在冲突。如果发生这样的高阶冲突，那么关于行动的自由将从何谈起？第二，也是更根本的，高阶愿望毕竟也仅是愿望，并且就自主体的认同和自由而言，人们尚不清楚，与一阶愿望相比，二阶愿望如何可能有更高的权威性。[20]第三个问题是，法兰克福没有说明这些高阶愿望是如何形成的。[21]这些高阶愿望的形成在某些情况下似乎是源于破坏自由的强迫，或受到了破坏自主体自由的那种形式的外在控制；因而，法兰克福所说的对于自由行动而言是充分的各种条件，事实上可能是不充分的。[22]最后，自主体在某一时刻不想按照某个愿望来行动，但是，由于任性、意志软弱

或屈服，又按照那个愿望自由地行动。在自由行动的某个例示中，有些人可能没有运用他们的反思能力。因此，对于自由而言，法兰克福最初提出的各种条件似乎并不是必要的。

为了解决这些问题中的某些问题，法兰克福已经对他的早期说明作了几点修改。最初的修改（1976）是，自主体通过**决定**她想根据哪个一阶愿望来行动来使自己认同那个愿望。法兰克福表明，这种决定与高阶愿望不同，它本身不能成为自主体所认同的东西。这里他似乎想说明的是：在某人让自己认同某个一阶愿望方面，这样的一个决定似乎不可能没有权威性，并且通过作出这样的一个决定，她有可能解决高阶愿望中出现的任何冲突。[23]在后来的著作（1987）中，法兰克福提出，支持某个一阶愿望的决定仅在下述条件下才能有效地使这个自主体认同那个愿望，即这个决定使得这个自主体处于“全心全意的”（wholehearted）状态之中。在最近的

著作（1992）中，法兰克福一直强调并进一步阐发全心全意的观念。[24]全心全意的观念要求，如果自主体的诸高阶态度中存在着任何冲突的话，她也不会陷入矛盾的心境。对于这一冲突，她完全有办法解决她的立场问题。而且法兰克福主张，一个全心全意的自主体不想改变自己的承诺；她不想改变承诺的原因不是她没有思考，而恰恰是她对她的心理状态的理解和评价。

有关全心全意的规范，解决了上面注意到的那些问题中的第一个问题，因为它排除了某些类型的高阶冲突。不过，尽管这一要求或许适用于说明认同，但是，对于说明根据道德责任所需的那种控制权所采取的行动来说，又太严格了；矛盾心境通常不是免责条件。这一要求同样被认为解决了高阶愿望的权威性问题。但是，这种主张是值得怀疑的，尤其是根据第三个问题，即与自主体高阶态度的起源有关的问题。自主体全心全意地支持她据以行动的那个愿望，但是，如果她的这种支持与她的全心全意一样是被强迫或操控的结果，那么，这种支持似乎不会使得她的行动是自由的。法兰克福坚决地否定，关于高阶意愿的因果历史的任何事实与自由意志之间有任何相关性；他坚持认为，重要的是自主体态度的这种结构。[25]我们可以承认，在自主体全心全意地支持某个一阶愿望时，她对那个愿望以及据以采取的行动“承担责任”（1975：121）。但是，在这种意义上说，即使自主体实际上没有对她所做的事情**承担**责任，即使随之而来的行动没有真正地使她**受到**赞扬或责备，她也仍要承担责任。[26]

16.2.3 能力说明

法兰克福的说明中被确定为最后的一个难题是，即使我们没有根据一个高阶支持（higher-order endorsement）来运用我们的能力展开行动，我们似乎也可以自由地行动。这里的问题源于如下事实，即法兰克福要求，人们有效的一阶愿望要与某个高阶态度相一致。我们所谓的各种能力说明可以避免这一难题。根据这些观点，自由能动性（free agency）要求，人们要有一般的能力或才能来理解一些实际的理由，并根据实际的推理来支配他们的行为。（根据某些形式的能力说明，自由能动性同样要求他们有能力合理地反思他们的理由，有能力改变他们的状态，例如他们的愿望，因而有能力通过这种反思来改变他们的行为。）但是，有人认为，即使人们在某个时刻没有运用这种能力，人们也可以在那一时刻自由地行动。[27]

这样的观点有很多，我们将要在这里思考一个它们共同面对的问题。自由的行动就是受到某种控制的行动。它似乎不仅要求人们根据（反思性的）合理的自我支配能力来行动，还要求人们在一个给定的时刻控制是否运用以及如何运用那种能力。相容论版本的能力说明不得不解释：如果这个世界是决定论的，那么自主体如何可能控制是否运用以及如何运用她合理的自我支配能力。

16.2.4 反应观

379

由约翰·马丁·费希尔（John Martin Fischer）和马克·拉维扎（Mark Ravizza，1998）提出的理由反应（reasons-responsiveness）观与能力说明关系密切，并且对上面提到的问题作出了回应。根据费希尔和拉维扎的观点，责任所需的那种形式的控制权，就是他们所谓的导向控制权（guidance control）。[28]给定行动的导向控制权的特征不是根据**自主体**或她的能力来描述，而是根据行动据以进行的**机制**（或过程）来描述：这种机制既必须具有充分的理由反应性，又必须是自主体自己的机制。[29]让我们依次来讨论这两个要求。

费希尔和拉维扎（1998：69）指出了理由反应性的两个方面：接受性与反应性。第一个方面与理解理由或认知理由有关，第二个方面与基于对理由的认知作出的决定以及其他的行动有关。由于反应性是一种倾向或模式特征，因此可以通过下述方式进行描述，即在某一时刻，对一个自主体起作用的机制在各种假设的（或非现实的）情景中是如何导致行动的。

所需的接受性是一种在最低程度上根据实在来认知理由的可理解的模式。这就是说，这个自主体一定“在实在的本质上确实没有被欺骗”（费希尔和拉维扎，1998：73），并且一定存在这样的一种情境，在这种情境中，由于这里所说的机制正在运行，自主体可能会显示出一种理由认知模式以指明了她“认知了各种理由是如何相互适应的，明白了一种理由为什么比另一种理由更有说服力，同时理解了承认一种理由是充分的就意味着，更有说服力的理由也一定是充分的”（1998：71）。（例如，如果贝丝说了谎，要使导致其说谎行动的机制能够充分接受道德理由；而在必须存在的假设情境中，贝丝有许多不说谎的理由。相同类型的机制运行着，并且在这些情境的适当形式中，她认知了这些不说谎的理由。）此外，这里所说的机制一定是乐于接受**道德**理由等理由；一般而言，就接受理由来说，对道德理由的接受一定表现为一种可理解的模式。

反应性要求稍微弱一些，满足它的前提条件是，至少有一种这样的情境，即在这种情境中，这个自主体有充分的理由采取别的行动，这里所说的这种机制也运行了，并且自主体由于有采取别的行动的理由而这样做了。根据费希尔和拉维扎的观点，道德责任并不要求必须存在下述情境，即在这种情境中，由于这里所说的这种机制在运行，这个自主体就会按照道德理由来行动。一个认知了各种道德理由的自主体若是坚决地拒绝按照这些道德理由来行动，就会因为她的不道德行为而受到责备（1998：79-80）。[30]

380

现在开始讨论第二个主要要求，即根据费希尔和拉维扎的观点，一种机制仅在自主体要对由这种机制产生的行动“承担责任”时才是自己的机制。在他们看来，（对由某些类型的机制产生的行动）承担责任是这样的一个过程，在这个过程中，这个自主体开始意识到她自己是一个自主体——她的选择和行动在这个世界中是有效的；她承认了她是反应态度（如感激与愤慨）的合适目标，也是某些实践活动（如实施奖赏和惩处）的目标，因为她的行动是由这些类型的机制产生的；关于她自己的这些观点都恰当地建立在这种证据的基础上。[31]

如果一个自主体的行为是由她自己充分的理由反应机制产生，那么她就会根据导向控制来行动。因而有人可能认为，她可以根据合理的自我支配能力来行动，并且对在这个时刻是否运用以及如何运用那种能力进行控制。

然而，既然费希尔和拉维扎根据自主体对她自己的态度来描述所有

权，那么他们的说明就容易受到下述反对意见的抨击。[32]假设从自主体（艾伦）诞生之日起，他的行动据以产生的各种机制每时每刻都受到某个神经科学家（尼娜）的影响。尽管尼娜没有直接地改变艾伦想要的**东西**，也没有使他的理解力比我们中的任何一个普通人要差，但是，尼娜常常改变与艾伦愿望相关的动机力量，因而从因果上决定他选择和采取尼娜所选取的那些行动；并且要是艾伦的行动没有受到尼娜的干涉，他的决定以及其他的行动可能会完全不同。此外，尼娜喜欢理由反应性；在一些假设的情境中，她可能影响艾伦，从而使他可能显示出一种不可理解的理由认知（包括对道德理由的认知）模式，并且当他有某个采取别的行动的充分理由时，他至少在某些时候会由于那个理由而采取别的行动。再假设艾伦没有意识到尼娜的干涉，他开始持有费希尔和拉维扎所要求的那些关于他自己的观点，并且这些观点恰当地建立在他知道的这种证据的基础上。[33]他不知不觉地对由他人控制的某种机制产生的行动“承担责任”。艾伦似乎满足了如下要求，即根据他自己充分的理由反应机制来行动，但是他的行动似乎是不自由的，他似乎不应为他所做的事情承担道德责任。[34]

费希尔和拉维扎宣称，在这种情况中，尽管一个自主体受到一再的干涉，但是这个自主体“从来也没有形成一个融贯（coherent）的自我。这就是说，在这种假设的情境中，根本就不存在自我或真正的个体”（1998：234-235，注释 28）。他们的意思是，由于不存在真正的自我或个体，所以责任主体不存在。这种回答似乎又向自由行动提出了一种新的（即第三个）要求，即对于真正自我的要求；同时我们需要知道这一要求意味着什么。为了强调这种需要，我们可能要注意到：在目前的例子中，我们根本就不清楚为什么不能说艾伦已形成了一个融贯的或真正的自我。毕竟，尼娜就像其喜欢理由反应性一样喜欢融贯性。[35]

16.2.5 遗留的争论

381

我们已完整分析了几个例子，在相容论者和不相容论者等人看来，这些例子表明，某个相容论说明在破坏责任的影响和不破坏责任的影响之间没有作出正确的区分。对于一些特定的相容论说明而言，这种分析可能得出了消极的结论。但是，人们似乎不能从根本上解决这个一般性问题，即责任所需的控制权是否与决定论相容。无论我们的分析取得了多么大的成功，相容论者按道理可能会继续寻找一种不会碰到反例的相容论说明。如果他们能够提出这种理论，那么不相容论者会再一次切合实际地

认为，所有被认为是有效的决定论的例子，都是这里所说的这种观点的反例。

解决这个一般性问题的更有效方式可能有两种，一是研究与责任有关的某些基本原则[36]，二是研究关于责任是什么的理论。[37]朝着这两个方向努力，即使没有解决问题，至少也会澄清争论的焦点。但是，既然以上所提出的与责任有关的原则或理论本身也是有争议的，那么我们对这个问题的根本解决方案就似乎是遥不可及的。

16.3 自由主义者的说明

如果自由决定和自由行动与决定论不相容，那么，原因要么是这种自由是不可能的，要么就是非决定论以某种方式使这种自由成为可能。非决定论是如何使自由成为可能的呢？近来不相容论者（或自由主义者）关于自由行动和自由意志的说明对这个问题提出了三种不同的回答。一些人认为，自由决定以及其他的自由行动从根本上讲一定（或至少可能）是没有原因的；另外一些人认为，它们一定是由某些先前的事件不确定地产生的；第三种人认为，自由决定或其他的自由行动一定是由自主体这个实体所产生的。[38]

在考察每种观点最具有代表性的说法之前，让我们简单地思考一下自由主义者的自由和心性的本质之间的关系。从历史的角度看，很多自由主义者一直都是心身二元论者，他们要么认为心灵是非物质的实体，要么认为心理属性和心理事件都是非物理的属性和事件。但是，如果自由意志和自由行动与决定论不相容，二元论似乎也不能帮助那些要在这个世界中为自由找到位置的人。原因有三点，第一，回想一下，上面提到的决定论的
382 那些特征并不局限于物理事件；如果这个世界是决定论的，那么（假设二元论是正确的）非物质的心理事件像任何其他事件一样，也完全是由先前的事件所决定的。第二，即使只有物理事件是完全被决定的，并且我们身体的所有运动也是完全被决定的，但是，如果自由与决定论不相容，那么我们充其量只能作出这样的一些自由决定，即它们丝毫不能影响我们正在进行的身体行为。第三，也是最重要的，如果自由与决定论不相容，一个二元论者仍然会遇到如下的问题，即解释非决定论如何可能在非物质领域使自由成为可能。如果我们能解释一个不确定的非物质的决定如何是自由

的，那么，似乎我们也只能解释一个碰巧是物理事件的不确定（undetermined）的决定如何是自由的。因此，二元论对自由主义者是否有用尚不清楚；相反，如果唯物论是一种关于心灵的更好观点，那么这个问题似乎就不会成为自由主义者的问题。[39]

16.3.1 非因果的观点

一些自由主义者的说明要求，自由行动根本就没有原因；另一些自由主义者的说明要求，自由行动，要么是没有原因的，要么是被不确定地产生的。由于这两种观点都相信，要使行动是自由的，没有任何正面的原因条件需要满足，所以我们称这两种观点为“非因果的观点”（non-causal views）。

卡尔·吉内特（Carl Ginet，1990）提出过一种最为精致的非因果的自由主义说明。[40]在他看来，每一个行动在因果上只不过是（或始于）一个简单的心理事件，即一个内部没有因果结构的心理事件。决定和意愿可以说是这些基本行动的例示；意愿常被认为是自主体运用其自身的愿望和想法。使某个心理事件成为基本的行动而没有成为自主体消极经历的变化的东西，不是心理事件是如何被引起的，而是心理事件所具有的某种内在的特征，即“行动性的现象性质”（actish phenomenal quality）（1990：13）。吉内特建议，最好将这种性质描述为，在自主体看来，似乎是由她直接地产生或决定这里所说的这种心理事件。（于是非基本的行动就被认为是一个行动产生或引起某个别的事件，或被认为是一些行动的集合。[41]）

如果某个事件是一个行动，那么使这个事件成为一个自由的行动还需要什么东西呢？根据吉内特的观点，没有其他的正面条件需要满足，附加条件全部是负面的。行动一定不是被因果地决定，并且在自主体采取行动的过程中一定没有受到不可抵制的强迫（毒瘾所引起的就是一种不可抵制的渴望）。

这种观点产生了两个问题，所有非因果的说明都会碰到这两个问题。第一，根据自由意志所采取的行动就是某人对其行为行使了某种控制权，但是一些非因果的说明似乎没有充分地说明这种控制权是什么。一种明显 *383*
可能的答案是：控制权就是如下这种行动，即自主体或与她有关的某些事件（如她具有某些理由和某个意图）以一种适当的方式引起。[42]尽管吉内特认为，每一个基本的行动在自主体看来似乎都是由她直接产生的，但是他又坚持主张，严格来讲，自主体引起了她们的行为这种说法是错误

的。[43]至于行动性的现象性质本身，似乎令人怀疑的是，就经历了一个心理事件的个体而言，心理事件如何可能**成为**这个个体对那一事件所施加的控制，而没有成为这种控制（多少值得信赖）的标志。这种怀疑被如下事实强化了，即在吉内特（1990：9）看来，就这个“自主体”而言，有行动感（actish feel）的心理事件在没有任何相关的愿望或意图的情况下，可能是由外部的大脑刺激所引起的。这样产生的事件似乎没有施加任何积极的控制，虽然在那个个体看来，这个事件似乎施加了积极的控制。

第二，自由行动就是根据合理的自我支配能力来行动，并且决定自己在一个给定的时刻是否运用以及如何运用这种能力。因此，一个自由的行动很可能是出于某种理由而行使的，因而可以获得合理的解释。能够说明这些现象的候选理论要求，理由状态（例如愿望）与行动之间必须有因果关联：自主体只有在下述条件下才会出于某个理由而展开行动，即这个相应的理由状态（或这个自主体具有这个状态）以一种适当的方式引起了她的行为；并且只有在下述条件下列举理由状态才有助于对其行动进行合理的解释，即这个理由状态（或自主体具有这个状态）以一种合适的方式导致了那一行动的产生。[44]虽然各种非因果的观点拒绝了这些建议，但是，它们提供的这些选择是否充分也是值得怀疑的。

吉内特（1990：143）对于列举一个愿望进行合理性解释给出了如下说明。如果凯特想要戴维高兴起来，并且相信讲个笑话就能逗乐戴维，那么，她就会讲个笑话给他听。在吉内特看来，列举那个想要戴维高兴起来的愿望只有在下述条件下才能解释她为什么讲那个笑话，即（a）凯特在讲那个笑话之前有一个想要戴维高兴起来的愿望；（b）凯特在讲那个笑话的同时，始终记得她的这个愿望，并且意图讲这个笑话能够实现（或有助于实现）她的这个愿望。[45]首先要注意，这里所需的同时发生的意图是一种二阶态度：对她自己的各种态度中的另一个（或别的）态度（或某个愿望）的态度。但显而易见的是，即使自主体行动时没有任何一种这样的二阶态度，她也有可能由于某个别的理由而行动，并且列举愿望可以合理地解释她的行动。例如，凯特可能按照她的愿望来行动，从而使戴维高兴起来（并且列举那个愿望可以合理地解释她的行动），即便她唯一的意图就是通过讲笑话想要他高兴起来。此外，吉内特的说明是否为合理地解释提供了充分条件也是值得怀疑的。假设凯特讲这个笑话时同样有别的一些理由，这些理由在因果上导致了她这样做，并且当她讲这个笑话时，她清楚地知道那些理由，那么，她想使戴维高兴起来的愿望在产生（引起）她的

行为的过程中就根本没有起任何作用。可见，她是否真的按照那个愿望来行动，以及列举那个愿望是否真的可以解释她所做的事情，（在最好的情况下）也是值得商榷的。[46]

16.3.2 非决定论的事件因果关系的观点

针对非因果的说明所提出的这两种反对意见表明：基于一种适当的自由主义观点，自由行动被认为是有原因的。一些自由主义者坚持主张，必须诉诸非决定论的事件因果关系。当一个事件引起了另一个事件时，因果关系的这种例示（根据某些关于因果关系的观点一定）受到了一条因果律的支配。但是因果律可能是决定论的，也可能是非决定论的。对前一种观点的陈述意味着第一类事件常常引起第二类事件；对各种非决定论规律的陈述则意味着第一类事件可能引起第二类事件。这些规律可能是或然性的，对它们的陈述意味着第一类事件（在某种程度上）概率化第二类事件，或者说，第一类事件发生时，它引起或不引起第二类事件存在一定的概率。当一个事件非决定论地引起了另一个事件时，我们说第一个事件产生了第二个事件，不过仍然有这样的机遇（chance），即第一个事件可能没有产生第二个事件。[47]

最简单的事件因果关系的自由主义观点不仅要满足一个好的相容论说明应达到的各种要求，还要承认某些事件（例如自主体有某些理由）如果引起了决定或其他的行动，那么就一定是非决定论地引起它们的。例如，如果自主体支持第一种行动是由第一组理由非决定论地引起的，那么她可以自由地决定与支持第一种行动。但是仍存在这样的机遇，即如果她支持第二种选择的决定是由第二组理由引起的，那么她就可以决定支持第二种选择。如果这些条件被满足，那么这一行动就可能得到某些理由的支持而做出，并且（支持者会说）这一行动就是基于某种形式的控制权而做出的，不过，自主体仍然有可能做出别的行动。[48]

反对这种观点的一种普遍意见是，这种观点所需的非决定论具有毁灭性，它会削弱自主体据以行动的控制。这种反对意见经常通过所谓的运气问题提出。例如，假设某个自主体如伊莎贝尔一直在思考要不要遵守一个承诺。她认为在道德上应该遵守它，尽管她已认识到有一些不遵守这一承诺的自利理由（并被诱使按照这些理由来行动）。她决定遵守这个承诺，那么这个决定是由先前的一些慎思（包括道德判断）非决定论地产生的。但是，在她作出决定之前，她慎思过程的结果仍有机会决定不遵守这个承

诺，这个决定是由伊莎贝尔自利的理由非决定论地产生的；先于这个决定
385 的所有事情（包括关于伊莎贝尔的所有事情）可能完全相同，但她还是作出了不同的选择。因此，根据这种反对意见，如果伊莎贝尔做了她认为在道德上是正确的事情，那么这只不过是运气问题。（据说，伊莎贝尔在其他的可能世界有这样的一些复制品，在作出这个决定之前，它们都与她极其相似，然而它们决定不遵守那个承诺；伊莎贝尔遵守那个承诺完全是由于好的运气。）换言之，某个事件的发生是运气问题，反对意见会说，这个事件的发生不受任何人的控制。因而所需的非决定论削弱了伊莎贝尔据以作出那个决定的控制权。[49]

由于要回应这种反对意见，一些事件因果关系的自由主义说明的支持者已经在一定程度上对我们在前面两段所思考的这种观点的简单版本作出了修改。[50]修改最详细的观点是由罗伯特·凯恩（Robert Kane，1996）提出的，它与这种简单版本有两点不同。[51]第一，如果诸如伊莎贝尔的决定是自由的，那么，在凯恩看来，在作出这个决定之前存在着意志的努力，这种努力就这个自主体来说是为了实现她的目的或澄清她的目的。[52]在道德冲突的情况下，这个自主体努力地抵制诱惑，并决定做她认为在道德上应该做的事情。凯恩要求，这种努力是“不确定的”，就像一个亚原子粒子的位置和动量根据量子力学规律是不确定的一样。的确，正是由于这种努力的不确定性，凯恩认为，这个自主体所做的决定也将是不确定的。

第二，在凯恩看来，如果这个决定是自由的，那么，通过作出这种决定，这个自主体使她据以作出决定的那些理由成为她最想据以行动的理由。在伊莎贝尔的例子中，根据一些道德理由作出决定，使伊莎贝尔的这些道德理由成为她最想据以行动的理由。

这些修改中的第一处修改，即要求有意志的努力，被认为从两个方面解决了运气问题。这个问题之所以在前面被提出，是因为人们注意到伊莎贝尔有这样的一些复制品，在作出这个决定之前，它们都与她极其相似，但它们决定不遵守那个承诺。凯恩宣称，如果存在不确定性，正如他认为意志有不确定的努力一样，那么，就不可能存在两个完全相同的世界。因此，在他看来，伊莎贝尔不可能有这样的复制品，即它的意志做出了完全相同的努力，在作出那个决定之前，它与她都极其相似，但它决定不遵守那个承诺。因此，凯恩证明运气论证是无力的。

但是这个问题并不是那么容易消解。尚不清楚的是，如果存在着不确

定性，那么两个世界实际上为什么不可能相同。在物理学中，粒子位置的不确定性可以用波函数来描述（这种函数能够说明某个粒子在各个确定的位置上被观察到的概率），一个粒子和它的复制品也许都可以用实际上完全相同的波函数进行正确的描述。此外，即使这种实际上的相同性不存在，但是这个问题依然存在。因为伊莎贝尔的决定事实上仍然是机遇过程出现的结果，因而这个机遇过程也可能产生一个不遵守那个诺言的决定。下述这种反对意见仍然会出现，即相信她的决定是运气问题，与这一决定在因果上由她的慎思所引起相比，运气不那么受控制。

凯恩认为，所需的这些意志的努力之所以有助于解决运气问题的第二个方面，是因为它与下述事实有关，即这些努力是自主体做某种特殊的事情所付出的积极努力。在他看来，如果一个自主体，例如伊莎贝尔，决定做她认为在道德上应该做的事情，那么她努力作出那个决定的一种结果是她成功地作出了那个决定，她成功地做了她正在努力做的事情，虽然是非决定论的。凯恩指出，如果真的是这样，那么该案例所涉及的这种非决定论通常没有破坏责任（因而它并没有减少控制从而使得满足责任的控制不够多）。他举了一个例子（1999b：227）进行类比，在这个例子中，一个男人敲击玻璃桌面，并试图将其打破。即使这个男人的努力是否成功是不确定的，但凯恩指出，如果他成功了，那么他同样要对打破桌面承担责任。

凯恩近来（1999a、1999b、2000）将这种策略推广至两类决定上，一是自主体决定做其被诱使做的事情，二是自主体决定做其相信应该做的事情。[53]他提出，在诸如伊莎贝尔的例子中，在她的决定之前有两种意志的努力同时发生，这两种意志的努力都是不确定的。这个自主体试图作出符合道德的决定，与此同时，她也试图作出符合自己利益的决定。[54]不论她作出哪种决定，她成功地做了她努力做的事情不过是非决定论的。因此，凯恩认为，无论她作出哪种决定，她都应像打破桌面的那个男人一样，为她所做的事情承担责任。

不过，也应该注意到，在凯恩的例子中，那个男人据以行为的控制权只有在下述条件下才能满足责任的需要，即他打破桌面的努力本身是自由的。因此，一个努力产生的决定只有在这个努力本身是自由的情况下才能以相同的方式使得这个决定也是自由的。因此，需要做的就是说明自主体在做出这些努力时是自由的。[55]凯恩在提出这种说明时碰到了一个两难困境。如果意志的努力先于诸如伊莎贝尔的决定，那么对意志努力的自由的

说明就要求这种努力本身源于先前的自由努力，因此，恶性倒退就出现了。另一方面，如果对意志努力的这种自由的说明不用诉诸任何先前的意志的自由努力，那么，对自由决定本身的说明，似乎同样不需要诉诸先前的意志的自由努力。总之，自由的决定，例如伊莎贝尔的决定，应该在意志的不确定的努力之后。但是这一要求似乎不能解决任何问题。

凯恩所赞成的第二处修改似乎也无助于解决运气问题。这一问题与自主体对其所做的事情进行控制有关，而控制似乎是一种因果现象，是关于什么引起决定以及其他行动的问题。但是，在凯恩看来，自主体更想根据某些理由来行动，它是作出决定的结果，而不是作出决定的原因。因此，自主体更想根据哪些理由来行动似乎完全无助于自主体对她作出那个决定进行控制。

对这种简单的事件因果关系的修改似乎无助于解决运气问题。但是，对于比其更简单的那种观点，这一问题又有多严重呢？

第一，显而易见的是，伊莎贝尔的决定不**完全**是运气问题。因为她的
387 决定（我们假设以某种适当的方式）产生于她对她的那些理由（包括对她应该遵守那个承诺的判断）的理解。相容论者应该会同意，如果决定以这种方式产生，其自身就会受到某种程度的控制。[56]

第二，有人质疑伊莎贝尔的决定究竟是不是一般意义上的运气问题。“运气”这个词在日常用法中，带有某些事情不受自主体控制的含义，但是，事件因果关系的自由主义观点所需的非决定论导致了削弱控制的运气的出现。为了理解这一点，我们可以区分两种完全不同的情况：一种情况是，非决定论存在于基本行动和本身不是行动的意向结果之间；另一种情况是，非决定论存在于基本行动本身的因果关系中，如伊莎贝尔的情况。至于第一种情况，假设你掷球以试图击中目标，并且你成功地这样做了。那么球击中目标本身不是一种行动，你只是通过掷球这一先前的行动来对这个事件进行控制。现在假设，由于球和风的某些属性，你掷球和球击中目标之间的过程是不确定的。（我们可以假设）你正在（自由地）努力产生一种稳定的结果，而这里的非决定论却抑制你成功地产生这种结果，并且由于这一原因，非决定论的确明显地削弱了你对这一结果的控制。[57]但是，伊莎贝尔例子中的非决定论与简单的事件因果关系中的自由主义观点所需的非决定论相比，所处的位置是不同的。它不是处于行动和本身不是行动的意向结果之间，而是处于本身**就是**行动的那个决定的直接因果关系之中。伊莎贝尔对那种决定行使控制权不只是（或根本就不需要）通过她

采取某种先前的行动。因此，这里的非决定论与它所处的第一种情况不同，不是一种抑制因素。如果伊莎贝尔例子中的非决定论还是会削弱控制，那么，对非决定论为什么会削弱控制的解释将不同于对它为什么适用于第一种情况所作的解释。但是，由于我们尚不清楚这种可供选择的解释是什么，因而伊莎贝尔例子中的非决定论事实上是否真的产生了削弱控制的“运气”，也是难下定论的。

这种关于事件因果关系的自由主义说明的运气反对意见，似乎不能让人信服。不过，第二种反对意见还需要考虑。即使所需的非决定论没有削弱控制权，也可能有人反对其增加了某些有价值的东西。

为了评析这一主张，让我们回到自由之所以对于我们来说非常重要的那些理由上面。我们重视作为尊严和责任的基础的自由，在我们行使自由的过程中，我们改变了世界运行的方式，并且自由与我们慎思时所发现的开放性这一现象一致。我们可以区分这种自由的两个方面：第一个方面是关于选择的余地或开放性；第二个方面是关于施加到行动上的控制。正如我们注意的那样，当我们思考（16.2.1 中）法兰克福对可供选择的可能性原则的抨击时，由于前面提到的某些理由，我们感兴趣的自由可能只涉及这两个方面中的一个，而与另外一个无关。与此相同，事件因果关系的自由主义观点所需要的非决定论也可能只与其中的一个方面有关，而与另外一个无关。

自主体在行动过程中行使控制权，就是对她决定做什么事情施加积极 *388*
的力量。我们有理由认为，行使控制权是这样的一种问题：这个行动是由自主体或某些与她有关的事件（如她的某些理由和意图）以一种适当的方式引起的。由于事件因果关系的自由主义观点没有给相容论说明所需的内容增加一些新的原因，因而前者似乎没有在自主体的积极力量之上增加任何东西以决定她所做的事情。就自由的这一方面而言，关于非决定论的要求（在很大程度上）似乎是多余的。

但是，至于另一方面，即开放的行动不止一种，却不是这样。如果（在 16.1.2 和 16.1.3 中思考的）后果论证是正确的，那么，在任何决定论的世界中，从来都不可能存在这种开放性。事件因果关系的自由主义说明所需的非决定论足以保证这种余地或开放性，并且这一点由于某些理由对于我们来说非常重要。有些个体至少会发现，当他们慎思时，他们不得不假定不止一种行动向他们开放。如果这个世界事实上是决定论的，那么这些个体就会陷入一种不可避免的幻觉之中（因为我们不可避免地要慎思）。

他们有理由断定，基于这一原因，如果事情就像事件因果关系的自由主义观点所说的那样，情况或许更好一些。与此相同，有些个体也有理由断定，如果事情就像这种观点所说的那样，那么，在我们采取行动的过程中，这一点在我们影响世界运行的方式方面将会更好一些。即使这个世界是决定论的，也会存在一种我们在行动的过程中据以发挥影响的方式：如果我们过去没有做我们已做的事情，那么事情将会非常不同。如果事情就像事件因果关系的自由主义说明所说的那样，那么我们一般仍可以以这种方式发挥影响。同时，我们也能以第二种方式来发挥影响，即在行动的过程中，我们可以通过对行动施加控制的方式在正在展开的或然的历史中开创其分支。还有可能存在这样真实的机遇，即事情没有按照某一种方式发展，我们的行动可能是使事情按照第二种方式发展的事件。有人可能有理由断定，就改变某人的行动来说，第二种方式要好于第一种。如果世界是决定论的，我们就不能以第二种方式发挥影响，因此一些个体可能会发现，事件因果关系的自由主义观点所需的非决定论，并非毫无益处，而是增加了某种有价值的东西。

我们是否可以得到与责任有关的东西？这是不清楚的。如果责任与决定论不相容，那么，一种好的相容论说明是否没有满足责任所需要的东西？事件因果关系的自由主义观点所保留的这种余地，似乎不是其所需要的附加物；如果法兰克福是对的，那么这种余地根本就不需要。产生决定或其他的行动的实际因果过程是不确定的，这种因果过程是否能发挥关键性的影响也尚不清楚。事实上，根据相容论说明，它仍然是这样的一种过程，即在这个过程中，产生决定或其他的行动的所有原因都是事件，这些事件由别的事件产生，因而可以一直追溯到宇宙大爆炸。正如上文所表明
389 的那样，我们尚不清楚，根据这种观点，自主体能否施加任何积极的控制力量。如果与责任有关的结论是不同的，那么这似乎就是我们需要的东西。如果责任与决定论不相容，那么它同样也不能被事件的因果关系的自由主义观点所保证。

16.3.3 自主体因果关系的说明

如果根据事件因果关系的自由主义说明自主体施加的控制不比他们根据相容说明施加的控制更加积极，那么哪种自由主义的观点能够保证自主体施加的控制更为有力呢？一些自由主义者一直主张：自由决定或其他的自由行动，即使不由事件因果地决定，也是由自主体所引起的。[58] 自主体

因果关系不同于也不等于事件因果关系（例如，自主体有某些理由）。[59]自主体是一种持存物或实体，因而不是某种本身就是结果的事物（尽管她生命中的各种事件本身可能是结果）。那么，根据这些自主体因果关系的说明，从严格的字面意义上讲，自主体是她自由行动的起因（originator），是她行为的无前因的原因（uncaused cause）。人们认为，非决定论与起因二者的结合最佳地把握了我们想要的这种与尊严、责任、影响的发挥以及开放性的现象关联在一起的自由。

自主体因果关系的辩护者遇到了两个主要问题，其中一个问题与自主体因果关系的概念有关，另一个问题与自由决定或其他的自由行动根据这种观点所作的解释的合理性有关。

理论家们如果接受了对施加到自主体行动上的控制所进行的因果分析——包括大多数相容论者和许多自由主义者——就会认为自主体在某种意义上引起了他们的自由行动。不过，大多数人认为，自主体因果关系只不过是由某些与这个自主体有关的事件因果关系（例如，自主体有某些理由和某个意向）组合而成的。但是，正如我们所看到的那样，自主体因果关系的说明对自己的论证与事实很不相符。大多数自主体因果关系理论家说，它从根本上讲不同于事件因果关系。这就导致了一个问题，即我们是否可以得到关于自主体因果关系的任何可理解的说明。甚至一些自主体因果关系的支持者似乎都怀疑这一点，他们宣称自主体因果关系很怪异甚至很神秘。[60]

不但如此，即使自主体因果关系这一观念能够被理解，我们仍存在下述的问题，即实体或持存物的因果关系这种事情本身是不是可能的。重复的论证表明，它是不可能的。每一个事件，包括每一个行动，可在任意时间中发生。进一步论证这个论证，如果一个行动产生了，那么，它的总原因中的某个部分一定是一个事件，这个事件本身是在某个时间中发生的事情。否则，如果这个行动发生了，它就绝对不能对它的发生作出说明。因此，如果自主体引起了某个行动，那么她也一定做了某件事情，或经历了某种变化，从而引起了那个行动。既然要么自主体所做的事情可能是一个事件，要么她经历的变化可能是一个事件，那么，事实就不会是大多数自主体因果关系所坚持的那样，自由的行动是由自主体而不是由某种事件引起的。[61] 390

第二个自主体因果关系的主要问题是，自由的行动是由于某些理由而采取的，因而可以得到合理的解释，但是，如果如大多数自主体因果关系

论者所认为的那样，自由的行动没有事件原因，那么这个合理的自由行动似乎就是不可能的。我们先前已看到，基于某些理由的行动和合理解释并需要诉诸行动的说明是由于自主体有某些理由而产生的，并且他们对这些现象所作的非因果的说明似乎是不充分的。像大多数自主体因果论者那样，这些理论家在否认自由行动有任何事件原因时，似乎也排除了任何合理的自由行动。

为了回应第二个问题，我已经提出了（克拉克，1993、1996）一种自主体因果关系的说明，根据这种说明，一个自由的行动由自主体产生，**并且**由某些与自主体有关的事件不确定地产生，例如自主体有某些理由。基于这种观点诉诸了理由因果关系，因此它可以提出一个与事件因果关系相同的关于合理行动和合理解释的说明。由于所提出的事件因果关系被要求是非决定论的，因此，即使这种观点与决定论不相容，它也保证了选择的开放性。最后，自主体因果关系本身还被认为是不同于也不等于任何事件因果关系，因此这种观点确保了自由行动的起因，这种起因似乎是更传统的自主体因果关系说明的一个具有吸引力的特征。[62]

对传统自主体因果关系所作的这种修改也解决了先前所描述的关于自主体因果关系可能性的反对意见。那种反对意见作出的下述推断就无法成立，即自由行动是自主体引起的而不是事件引起的；据说如果一个自主体引起了一个行动，那么，与那个自主体有关的某个事件一定引起了那个行动，并且当行动发生时，也能说明了那个行动的发生。根据这种观点，与这个自主体有关的某些事件的确引起了所有的自由行动，并且当一个行动发生时，也说明了那个行动的发生。

但是，与自主体因果关系的可理解性和可能性有关的问题依然存在。我（1993、1996）和蒂莫西·奥康纳（Timothy O'Connor，1995a、1996、2000）虽然在细节上不同但都提出了如下观点，即如果事件因果关系有一个非还原的说明，我们就可以用描述事件因果关系的方式来描述自主体因果关系。人们熟知的还原说明是通过恒常结合、反事实依赖和概率增加来描述事件因果关系的；如果通过这种方式来描述事件因果关系，那么自主体因果关系当然与之有着根本的不同。但是，如果因果关系是这个世界不可还原的基本特征，那么，由于具有相同的可理解性，我们也可以将这些实体和事件当作原因。

即使我们可以理解自主体因果关系这一观念，即使先前关于它的不可能性论证思考是无效的，我们仍有理由怀疑，一个实体是否可能引起某些

事物。举个例子来说明，从表面上看，虽然因果关系不能还原为概率的增加，但是，如果先于结果发生的原因直接引起了某个结果，那么它一定是影响这个结果出现概率的事物。事件就是一种能够这样影响概率的东西，之所以如此，是因为事件是在时间中发生的。所有的实体都不会发生（与它们有关的事件才会发生），这些实体不是那种以已指出的方式来影响概率的事物。以上考虑虽然不是关键性的，但似乎不利于论证实体因果关系的可能性。

16.3.4 存在的问题

即使这些自由主义观点中的某一个能够很好地描述我们所看重的自由，即使那种说明所描述的东西是可能的，但是，如下问题依然存在，即是否有充分的证据证明那种说明所假定的东西真的存在。答案似乎是否定的。

自由主义的各种说明首先要求决定论必须是错误的。除此之外，它们不仅要求必须存在某种类型的非决定论（例如某些事件要么根本就没有原因，要么由其原因非决定论地引起，要么由自主体引起，要么由事件非决定论地引起），还要求这种非决定论必须处于特定的位置（例如，处于决定的作出与行动的发生之间）。我们有什么证据证明这些要求已被满足了呢？

有人主张，当我们做决定和行动时，我们的经验构成了这种证据，它能证明所要求的这种非决定论就处在必要的地方。[63]我们可以区分这一主张的两个构成部分：一是在我们做决定或行动时，事物以某种自由主义者所说的方式向我们显现；二是这种显现证明了事物事实上就是以那种方式存在。一些相容论者否定了第一部分。[64]不过，即便第一部分是对的，第二部分似乎也是令人怀疑的。如果事物是某种自由主义说明所说的那个样子，那么自然规律，如物理学规律、化学规律和生物学规律，也一定是那个样子。[65]在我们看来，让人难以相信的是，在我们做决定或行动时，事物在我们看来是什么样子的，能够让我们深入地了解自然规律。因此，我们不得不研究自然和自然规律，以找到所需的非决定论的证据。

据说量子力学的科学证据有时表明了决定论是错误的。量子理论的确已被很好地确证了。不过，人们在以下两个问题上还没有达成共识：一是如何解释量子理论；二是在这个世界中的事物是什么样的状态这一问题上，它向我们表明了什么。非决定论解释与决定论解释都被提出来，但是我们还不清楚正确的解释是否就在它们之中。[66]在这里充其量只能说，如

果经典力学和电磁理论是错误的，那么就没有很好的证据来证明决定论是正确的。

这种论据对于如下问题也不是关键性的，即非决定论是否处于某种自
392 由主义说明所要求的精确位置上。就自由决定而言，除非心理事件完全独立于物理事件，否则某种特定类型的非决定论必定处于大脑过程的某个交汇处。在一些自由主义论著中，对于为什么情况如此有一些非常有趣的猜测[67]；但是，我们现在对大脑的理解无法证明情况究竟是不是如此。对于这一问题，我们似乎暂且最好持不可知论的态度。

即使自由主义的自由需要自主体因果关系，并且即便自主体因果关系是可能的，我们也没有证据证实非决定论必定存在于大脑中。的确，我们尚不清楚，是否可能有任何经验证据赞成或反对自主体因果观的这一要求。[68]

16.4 结 论

自由意志是否与决定论相容？我们是否有自由意志？这两个问题经常被看作是两个非有即无的（all-or-nothing）问题：对于每一个问题而言，答案不是肯定的就是否定的。但是，我们对自由的兴趣，源于我们对各种各样的东西的关注。这些东西中的某些成分或它们的变体所需的控制权与决定论相容（且与事件因果关系非决定论相容），但其中的其余成分则与它们不相容；我们或许只具有这些东西中的某些成分，而不具有其他成分。我们没有受到像尼娜这样的神经科学家的控制，我们中的大部分人也没有受到强迫症和毒瘾的影响。我们对理由的认知与完全可理解的模式相一致，而且我们从根本上没有脱离与实在的联系。谁也不能否认我们因此具有能给予我们某种程度的尊严且有价值的控制权。

即使我们有做别的事情的能力这一点与决定论不相容，但我们也有理由认为，这种能力不是责任所必需的。即使责任的某些方面被决定论破坏了（或被事件因果关系的非决定论破坏了），其他方面可能并没有被破坏。即使决定论是真的，行动也可归之于自主体；即使从来没有人由于做错了事而受到严厉的惩罚，但我们对冒犯者表现出某种反应态度（例如愤恨）
393 从而保护自己，这也是适当的。另外，即使决定论是正确的，但一般而言，我们在行动时也会以某种方式对世界的状态发挥影响，虽然不能以别

的方式发挥影响；同样，我们在慎思和决定时至少能以某种方式发挥影响，虽然每当我们慎思时，都不幸地受制于幻觉。

如果我们事实上只拥有我们由以看重自由的这些东西中的一部分而非全部，那么明智的做法就是认清这一事实并且接受它。要是这样做了，我们就能避免一种过度的悲观。这种做法既拒斥了如下观点，即我们一直以来想要的自由原本是一种让人沮丧不已的自由，又拒绝了下述这种毫无根据的顽固信念，即我们具有我们能想象得到的那种最强势的自由。[69]

【注释】

我要感谢查尔斯·克罗斯（Charles Cross），约翰·马丁·费希尔，罗伯特·凯恩，艾尔弗雷德·米尔和布鲁斯·沃勒，他们对我这一章的初稿提出的意见很有帮助。

[1] 作决定就**是**行动，就是采取了一个心理行动。在这里，我为了强调作决定，从而将区别对待决定和行动。在我们所有的行动中，决定似乎尤为重要，因为它审慎地行使了我们的积极控制权。

[2] 据称，神圣的预知可能会排除我们享有自由意志，一些赞成这种观点的论证在结构上类似于证明决定论与自由意志不相容的论证。参见费希尔（1994：1-6）关于强调它们之间相似性的讨论。由于篇幅的限制，这里我们将着重讨论所谓的决定论威胁。

[3] 对决定论的充分讨论，可以在伊尔曼（Earman，1986）的著作中找到。虽然这本书中的部分章节多少有些技术指向，但是其中的第 2 章对这一问题的说明提供了出色且易懂的说明。

[4] 对决定论与普遍的因果关系的进一步讨论，可参见范·英瓦根（1983：2-5）和伊尔曼（1986：5-6）。

[5] 范·英瓦根（1983：16）。

[6] 我这里采用范·英瓦根（1983：93-105）提出的论证。关于其他的不相容论论证，都完全可以看作是不同形式的后果论证，它们在威金斯（1973）、兰姆（Lamb，1977）、范·英瓦根（1983：68-93）和吉内特（1990：第 5 章）中提出。关于这些论证的一般讨论，可参见费希尔（1983；1988；1994：第 1～5 章）、费林特（Flint，1987）、维维因（Vihvelin，1988）、甲必丹（Kapitan，1991）、希尔（Hill，1992）和奥康纳（O'Connor，2000：第 1 章）。对这些论证的一些特定方面的讨论，可参见下面的注释。

[7] 考虑一下范·英瓦根的设想。以 P_0 代替我所说的“H”，他写道：

> 命题 P_0 是一个关于久远过去的命题。如果我们愿意，我们可以规定，它是一个关于初期的总统星云（presiderial nebulae）中的原子和其他粒子的分布和动量的命题。因此，肯定没有任何人对是否 P_0 能够作出选择。命题 L 是一个“记录”

了所有自然规律的命题。如果角动量守恒是一个自然规律，那么任何人都不可能对角动量是否守恒作出选择，并且一般而言，既然L是一条自然规律，那么没有任何人能够对是否是L作出任何选择。(1983：96)

[8] 加洛伊斯（Gallois，1977）和纳维森（Narveson，1977）是这一观点的代表；在同一卷中，附在他们论文后面的就是范·英瓦根的回应。那里的讨论关注的是一种在形式上稍有不同的后果论证，我已采纳了某些主张，以至这些主张适用于这儿正在考虑的这种形式。我借用了费希尔（1994：第4章）提出的“多重过去相容论者”这个名称（和“局部奇迹相容论者”——见下文）。

394 [9] 刘易斯（1981）是这种观点的倡导者。此外，他的讨论针对的是一种不同形式的后果论证，我对他的某些论断已作出必要的改动。霍根（1985）、费希尔（1988；1994：第4章）和吉内特（1990：111-117）讨论了刘易斯的观点。

[10] 是否可能否定（6），即NL，似乎部分地依赖于自然规律是什么，尤其依赖于这些自然规律是否包含了任何不可还原的必然性。关于通过诉诸规律是非必然性的观点来辩护相容论，参见斯沃茨（Swartz，1985：第10章）和贝罗夫斯基（Berofsky，1987：尤其是第8章和第9章）。

[11] 对于（β）的有效性，已有广泛的讨论。参见，如斯洛特（Slote，1982）、费希尔（1983；1986b；1994：第2章）、威德克尔（Widerker，1987）、维维因（1988）、奥康纳（1993）、甲必丹（1996）、麦凯和约翰逊（McKay and Johnson，1996）、卡尔森（Carlson，2000）、克里斯普和沃菲尔德（Crisp and Warfield，2000）。

[12] 这个例子改编于威德克尔（1987：38-39）。

[13] 关于这两条建议，参见威德克尔（1987）和奥康纳（1993）。

[14] 麦凯和约翰逊（1996）建议以这种方式来理解有算子的推导规则。

[15] 关于这种规则有效性的论证，参见卡尔森（2000：286-287）。

[16] 参见费希尔（1994）、费希尔和拉维扎（1998）。

[17] 参见，如布卢门菲尔德（Blumenfeld，1971）、内勒（Naylor，1984）、斯顿普（Stump，1990、1996、1999a、1999b）、洛（Rowe，1991：82－86）、威德克尔（1991、1995a、1995b、2000）、哈吉（Haji，1993；1998：第2章）、兰姆（Lamb，1993）、齐默尔曼（Zimmerman，1993）、费希尔（1994：第7章；1995；1999：109-125）、费希尔和霍夫曼（Fischer and Hoffman，1994）、吉内特（1996）、亨特（Hunt，1996、2000）、凯恩（1996：40-43，142-143）、威德克尔和卡佐夫（Katzoff，1996）、科普（Copp，1997）、麦肯纳（McKenna，1997）、怀曼（Wyma，1997）、德拉·罗卡（Della Rocca，1998）、米尔和罗布（Mele and Robb，1998）、大冢（Otsuka，1998）、戈茨（Goetz，1999）、奥康纳（2000：18－22，81－84）、维维因（2000）、佩里布默（Pereboom，2001：第1章）。

[18] 关于决定论排斥责任并不依赖于PAP的各种论证，参见范·英瓦根（1983：161-188；1999）。关于对它们的讨论，参见费希尔（1982）、海涅曼（Heinaman，

1986)、沃菲尔德（1996）、费希尔和拉维扎（1998：第 6 章）、斯顿普（2000）、斯顿普和费希尔（2000）。

[19] 莱勒（Lehrer，1997：第 4 章）提出了另外一种新的分层说明，关于这些观点的详尽讨论，参见沙茨（Shatz，1986）。

[20] 这种观点最早由沃森（Watson，1975：218）提出，法兰克福承认了它，因为他写道：

> 在分层结构中，一个愿望所处的层次比另一个愿望所处的层次更高，这一事实一般而言似乎还不足以将更大的权威性或任何基本的合法性赋予这个愿望。换言之，将一些愿望分配到分层结构的不同层次之中并不能单独地解释：在某人自己的愿望中，他何以认同其中的这一愿望而不是另一愿望。（1987：166）

[21] 他写道：

> 一个人在他形成二阶意愿时可能是任意的或不负责任的，并且没有认真思考什么才是至关重要的。二阶意愿只有在它们都是偏好这一意义上说才表达了评价。即使二阶意愿的形成有任何基础，也没有必要对这种基础作出限制。（1971：注释 6）

此外，“[一个自主体] 的各种行动以及他对这些行动的动机的认同是如何产生的，这一问题与以下这两个问题没有关系：一是他采取的行动是否是自由的；二是他是否对他采取的这些行动承担道德责任”（1975：122）。 *395*

法兰克福的确主张，“正是由于个人的理性能力，他才能逐渐明确地知道自己的意志和正在形成的二阶意愿”（1971：17）。我们将要在下面考虑一些相容论说明，它们强调要有实际推理的能力和理性行动的能力。

[22] 关于对法兰克福的说明（以及任何其他类似的观点）所碰到的该类问题的讨论，参见费希尔和拉维扎（1998：第 7 章）。米尔（1995：第 9 章）认为，一种适当的相容论说明要对自主体态度的这种历史提出某些要求。

[23] 注意，采取某些行动时的自由，现在可以根据作出某些决定来说明。尽管法兰克福提出，没有任何决定可以“外在于”这个自主体，但是，一般说来，决定可能是不自由的。因此，需要对现在诉诸的这些决定的自由进行说明。不过，既然法兰克福后来不再诉诸这些决定，那我们也不必再深究这一点。

[24] 他注意到，与全心全意相对的矛盾心境“不能通过意愿（voluntaristically）来克服。一个人不能仅凭一个‘意志行动’从意愿上就确定了因而创造了从未存在过的真理。换言之，他不可能仅通过完全受其意愿直接控制的物理运动来使自己处于全心全意的状态”（1992：10）。决定在自主体形成的认同中几乎是无足轻重的。

布拉特曼（Bratman，1996）批评了法兰克福，因为他否定了决定对认同有重要作用，布拉特曼由此提出了一种将决定和全心全意（他称其为满足）合二为一的观点。他所提出的观点虽然不是用来说明自由行动，但是，如果这种观点有可能适用于那个

目的，那么，正如上面注意到的那样，我们不得不对所需的这些决定的自由作出说明。

[25] 参见，例如，法兰克福（1987：注释 13）。

[26] 沃勒（Waller，1993）提出了这种反对意见。

[27] 华莱士（Wallace，1994）和沃尔夫（Wolf，1990）提出了能力说明。米尔（1995）提出一种诉诸自主体当下的理性能力和她心理态度的历史的相容论观点。

[28] 导向控制被认为满足了责任中“与自由相关”的部分。费希尔和拉维扎（1998：26）承认，可能有其他的要求（例如一种认识要求或知识要求）与责任有关。

[29] 费希尔和拉维扎称所需的这种反应性类型为“适度反应性”，并将它与他们描述的弱反应性与强反应性区别开来。

[30] 这种反应性要求似乎太弱了。思考一个自主体卡拉（Karla），她经常有做某种事情的强迫性愿望（例如，去盗窃的强迫性愿望）。卡拉适当地接受了一些理由；她倾向认知一种关于不去盗窃的各种理由（包括抵制盗窃的道德理由）的可理解的模式。例如，卡拉似乎像许多有盗窃癖的人一样，由于某些正面理由而不去盗窃，例如有一
396 个警察一直在盯着她；因而她满足了这种反应性要求，虽然不存在由于一些理由而不去盗窃的其他情境。但是，由于附近没有警察，当她盗窃时，她正在进行的这种被迫行为是由她的强迫愿望所导致的，并且她没有按照道德责任要求的那种方式来控制她所做的事情。对责任所需的这些态度的反应，与根据这种说明所需的这种反应相比，似乎要强烈一些。（这种反对意见是由米尔在 2000 年提出的。）

费希尔和拉维扎都可能反对，在卡拉对警官的出现作出反应的情境中，这时运行的机制不同于她的强迫性盗窃愿望导致其盗窃时所运行的机制，因此，根据他们的观点，导致她盗窃行为的这种机制不能算作对理由的充分反应。（参见费希尔和拉维扎，1998：74；在这个例子中，某种理由给予自主体更多的“能量或关注”。）但是，如果他们这样回应，那么我们就没有必要更多地知道如何区别这些机制。否则，这一举措看起来非常奇怪。

有人也可能反对，在卡拉的例子中，导向控制的第二个要求——这个机制必须是自主体自己的——不能被满足。（这个要求将在下文中讨论。）作为回应，简单地讲，这种所有权据说是这个自主体有某些与她自己有关的态度的问题，并且似乎没有理由可以说明，卡拉为什么不能具有所需的这些态度。

[31] 在费希尔和拉维扎看来，如果一个自主体根据道德责任所需的自由来行动，那么她一定有关于自己的某种信念。盖伦·斯特劳森（Galen Strawson，1986）也同意这一点，他认为，相信某人自己是一个自由的自主体，是她成为一个自由的自主体的必要条件。

[32] 文中所举的例子表明，一个自主体即使满足费希尔和拉维扎观点所提出的全部要求，也不用承担道德责任。（艾尔弗雷德·米尔在交流中所描述的）另外一个不同的例子表明，一个自主体即使没有满足这种观点所提出的要求，也要承担道德责任。假设山姆偶尔行为乖张（akratically）：他判断某种行动是最好的，但是由于做别的事

情的愿望更强烈一些（有最强的动力），他做了别的事情。鉴于山姆存在这样的问题，一群善意的神经科学家偷偷在他的大脑中植入一块电脑芯片，这块电脑芯片可以发挥以下作用：每当山姆判断某种行动是最好的时候，这块芯片确保他采取这种行动的愿望是最强的。这块芯片所做的唯一事情就是，帮助山姆克服意志的软弱，帮助他采取他判断是最好的那些行动。在这里，即使这些帮助没有被山姆意识到，我们也无需取消山姆对他行为所承担的责任。但是，在费希尔和拉维扎看来，这种帮助似乎会取消了他应当承担的责任。至少，在植入芯片初期，这块芯片帮助山姆产生其行为时所运行的机制，在类型上不同于任何这样的机制，即它产生了那些需要山姆承担责任的行为，因此那种机制似乎不是他自己的机制。

[33] 有人可能认为，艾伦没有意识到尼娜对他的影响，导致他要对他那些显然不是基于这种证据的行为承担责任。不过，正如费希尔和拉维扎认识到的那样，要完全认识我们的行为据以产生的机制似乎是不太可能的，因为有很多我们没有意识到的因果影响作用于我们的行为。因此，这种证据要求可以被这样的一个自主体满足，即这个自主体（例如艾伦）没有意识到他的行为据以产生的机制的某些特征。正如费希尔和拉维扎所述："一个人要对产生于这种机制的行动承担责任，好比他因为看到了冰山的顶部就要对整个冰山承担责任一样。"（1998：216-217）

[34] 注意，艾伦没有意识到的那些影响来自于另外一个智能主体的蓄意干涉，而我们经常没有意识到的那些影响则是由一些没有考虑到的原因产生的。但是让人怀疑的是，这种区别能否说明艾伦的不自由。的确，我们假设对他的例子进行改动，在这个改动过的例子中，某个无生命的对象在相关方面发挥了与尼娜相同的作用。例如，假设在艾伦的一生中，每当他行动时，蒙古境内的一块陨石就放射出 M 射线，对他和他的行为产生了与前面假设尼娜所作的干涉相同的作用。此外，即使不清楚费希尔和拉维扎的观点所提出的任何要求是否被违反了，仍让人怀疑的是，在这里我们是否有一个能自由行动的自主体，他是否需要对他所做的事情承担责任。

[35] 正如费希尔和拉维扎（为回应法兰克福对类似的反对意见提出的辩护时）所说的那样："连续的操控与连贯性和可理解性相容。一个自主体的历史在其内容上是连续的还是间断的，与这种历史是在内部产生的还是在外部产生的，是两个十分不同的问题。"（1998：198-199）

[36] 上面注释 18 所列举的讨论就利用了这个策略。

[37] 采用这种方法的两个相容论者分别是华莱士（1994）和斯坎伦（Scanlon，1998：第 6 章）。

[38] 正如下面将要解释的那样，第三种类型的观点认为，自主体因果关系不同于事件因果关系。

[39] 关于一种不同的观点，参见科弗（Cover）和奥利里-霍索恩（O'Leary-Hawthorne，1996）。他们认为，某种类型的自由主义观点，如自主体因果观，与关于人和心性的二元论观点十分一致。

[40] 非因果关系的说明同样由麦考尔（McCall，1994：第 9 章）、戈茨（1997）和麦卡恩（McCann，1998）提出。

[41] 吉内特对非基本行动的说明（尤其是对产生过程的说明）非常复杂。有兴趣的读者应该仔细阅读他 1990 年著作之第 1 章。

[42] 这里包含的表达式“以一种适当的方式”是用来排除所谓的“异常的”或“不规则的”因果关系。行动的因果理论的支持者认为，行动的特征表现为下述事实，即它们是由与自主体有关的某些类型的事件所产生的。但是应当承认，身体的运动可能正是由这种类型的事件所产生，如果因果路径是异常的或不规则的，那么它将不能成为一个行动。关于对这个问题和对提出的解决方案的讨论，参见戴维森（1973：153-154）。

[43] 因此韦尔曼（Velleman，1992：466，注释 14）反对说，根据吉内特的观点，每个基本行动据说具有行动性的现象性质都是引入歧途的幻觉。不过，吉内特将他关于人们行动时具有的经验的描述误认为是隐喻；他认为，这种经验并没有准确地向这个自主体指出，她正在使这一事件发生。

[44] 关于对由于某个理由而采取的行动所作的细致的因果说明，参见奥迪（Audi，1986）。

[45] 吉内特（1990：143）宣称，对于列举一个愿望所作的说明的真理性而言，这样一些条件是充分的。但他似乎将这些条件（或至少有一个相关的同时发生的意图）
398 同样看作是必要的。因为他坚持认为（1990：45），这个自主体所具有的愿望是实施某个行动的理由，并且在这个自主体行动时，如果她没有同时发生的相关意图，那么，她就不会意识到她所具有的这个愿望不是她行动的理由。

[46] 这一反对意见由米尔（1992：250-255）提出。

[47] 关于对非决定论的因果关系的说明，参见，如刘易斯（1973［1986］：附录 B）、图利（Tooley，1987：289-296）和伊尔斯（Eells，1991）。

[48] 近来描述这种简单的事件因果关系观点的人有威金斯（1973）、索拉布吉（Sorabji，1980：第 2 章）和范·英瓦根（1983：137-150）。诺齐克（1981：294-316）提出了类似的观点，不过，附带了一个条件，即至少某些自由决定是“自我涵摄的”（自我解释的）。

[49] 运气论证由哈吉（1999）和米尔（1999a、1999b）提出。

[50] 丹尼特（Dennett，1978）、米尔（1995：第 12 章）和埃克斯特龙（Ekstrom，2000：第 4 章）提出了事件因果关系的自由主义观点，根据这种观点，只有在慎思过程的早期阶段才需要非决定论。根据他们的观点，在慎思过程中，某些不确定的事件在因果上决定了自由决定。关于这种观点的批判性讨论，参见克拉克（2000）。

[51] 关于凯恩观点的讨论，我是从我 1999 年的论文中抽取出来的。

[52] 我在这里假设，伊莎贝尔的决定就是凯恩所谓的“形成自我的行动”，这种行动在因果上不是由先前任何事件因果地决定，因而这种行动的自由不是由在因果上

决定它的那些先前自由行动的自由派生出来的。在凯恩看来，事实或许是，形成自我的行动的自由是由关于意志的努力的自由所派生的，这种行动由意志的努力非决定论地产生。这种观点随后会在本节中讨论。

[53] 这种新的建议是用来回应米尔（1999a：98-99；1999b：279）所提出的反对意见。

[54] 在说明自由决定时，加倍的意志的努力引入了令人困扰的非理性。在道德冲突中，自主体的各种动机互不融贯；但是这种冲突为自由所熟悉，对自由也没有明显的威胁。不过，让这个自主体同时积极尝试做两种明显不相容的事情——她同时做这两件事情显然是不可能的——提出了与自主体的合理性有关的严重问题。这种附加的不融贯性与其说对自由有帮助，不如说对自由有威胁。

[55] 提出这种说明的任务之所以被耽搁了，是因为有人认为，这些努力的自由是间接的，是从先前那些自由行动的自由中派生出来的。但是，这种策略没有避开以上所提出的问题——对这些先前行动的自由的说明为什么不能直接运用到说明由意志的努力所产生的决定上？

[56] 事实上，许多当代相容论者（如费希尔，1999：129-130）认为，满足责任的这种控制权，不仅与关于决定以及其他的行动的非决定论因果关系相容，还与关于它们的决定论因果关系相容。这些相容论者承认，如果事件因果关系的自由主义说明所需的这种非决定论削弱了控制权，那么，若是非决定论没有削弱控制权，就相当于说它破坏了责任。

同样值得注意的是，非决定论的因果关系不能构成所谓的异常或不规则的因果关系。因为后者关注的是因果过程的路径，且非决定论因果关系可能遵循与决定论因果关系相同的路径。

[57] 然而正如凯恩（以打破玻璃桌面的那个人为例）所指出的那样，这里的非决定论在下述意义上甚至不一定会削弱控制权，即这个自主体对其产生的这种结果不用承担责任。 *399*

[58] 一些自主体因果关系理论家坚持的不是自由行动是由自主体产生的，而是自主体产生某个行动是一个自由行动。这种区别与我们这里的思考没有什么直接关系。

[59] 近几年来，对自主体因果关系作出说明的人有，奇泽姆（1966、1971、1976a、1976b、1978）、泰勒（Taylor，1966、1992）、索普（Thorp，1980）、齐默尔曼（1984）、多纳根（Donagan，1987）、洛（1991）、克拉克（1993、1996）和奥康纳（1995a、1996、2000）。

[60] 参见索普（1980：106）和泰勒（1992：53）。

[61] 这种反对意见源于布罗德（Broad，1952：215），吉内特（1990：13-14）也提出了这种观点。

[62] 根据这种自主体因果关系观点，自由行动除了受到与自主体有关的一些事件的因果作用之外，还受到这个自主体所发挥的因果作用，自由行动由这些事件非决定

论地引起。但是，她施加的控制力比所有原因都是事件所施加的控制力要更加积极（参见克拉克，1996：27-30）。因此，与非决定论的事件因果关系说明相比，这种观点能更好地解决运气问题。不过，还需要说明的是，这种解决方式是否绝对充分还需要论证。

[63] 坎贝尔（1957：168-170）和奥康纳（1995a：196-197）诉诸这种经验以证明自由主义的自由意志。

[64] 例如，参见米尔（1995：135-137）。

[65] 之所以如此，不仅是因为不论心与身之间的关系如何，身体行动都是外显的，还因为决定和其他的心理行动阻碍了心理事件完全独立于物理、化学和生物的事件。

[66] 关于对这些与自由意志的各种理论有关问题的简明讨论，参见洛伊尔（1996）。除了考虑关于量子力学一些比较有名的解释，洛伊尔认为，自由主义要求某些事件缺乏客观的概率，而许多自由主义者可能拒绝这种论断。

[67] 参见，如凯恩（1996：128-130，137-142）列举的其他来源。

[68] 关于持反对意见的观点，参见佩里布默（2001：第3章），他认为，我们现在有证据反对自主体因果关系的存在。

[69] 关于对我们缺乏自由意志（或自由意志的某些有价值的形式）是什么的细致讨论，参见洪都里奇（Honderich，1988：第3部分）、斯密兰斯基（Smilansky，2000）和佩里布默（2001）。

参考文献

Audi，Robert（1986）. “Acting for Reasons.” *The Philosophical Review*，95：511-546.

Berofsky，Bernard（1987）. *Freedom from Necessity：The Metaphysical Basis of Responsibility*. New York：Routledge and Kegan Paul.

Bishop，John（1989）. *Natural Agency：An Essay on the Causal Theory of Action*. Cambridge：Cambridge University Press.

Blumenfeld，David（1971）. “The Principle of Alternate Possibilities.” *Journal of Philosophy*，67：339-344.

Brand，Myles（1984）. *Intending and Acting：Toward a Naturalized Action Theory*. Cambridge，MA：Bradford Books.

Bratman，Michael（1996）. “Identification，Decision，and Treating as a Reason.” *Philosophical Topics*，24（2）：1-18.

Broad，C. D.（1952）. *Ethics and the History of Philosophy*. London：Routledge and Kegan Paul. Campbell，C. A.（1957）. *On Selfhood and Godhood*. London：

George Allen and Unwin.

Carlson, Erik (2000). "Incompatibilism and the Transfer of Power Necessity." *Noûs*, 34: 277–290.

Chisholm, Roderick M. (1966). "Freedom and Action." In Keith Lehrer (ed.), *Freedom and Determinism*. New York: Random House: 11–44.

—— (1971). "Reflections on Human Agency." *Idealistic Studies*, 1: 33–46.

—— (1976a). "The Agent as Cause." In Myles Brand and Douglas Walton (eds.), *Action Theory*. Dordrecht: D. Reidel: 199–211.

—— (1976b). *Person and Object: A Metaphysical Study*. La Salle, IL: Open Court.

—— (1978). "Comments and Replies." *Philosophia*, 7: 597–636.

Clarke, Randolph (1993). "Toward a Credible Agent-Causal Account of Free Will." *Noûs*, 27: 191–203. Reprinted in O'Connor (ed.) (1995b): 201–215.

—— (1996). "Agent Causation and Event Causation in the Production of Free Action." *Philosophical Topics*, 24 (2): 19–48. Reprinted in abbreviated form in Pereboom (ed.) (1997): 273–300.

—— (1999). "Free Choice, Effort, and Wanting More." *Philosophical Explorations*, 2: 20–41.

—— (2000). "Modest Libertarianism." *Philosophical Perspectives*, 14: 21–46.

Copp, David (1997). "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility." *Noûs*, 31: 441–456.

Cover, J. A. and O'Leary-Hawthorne, John (1996). "Free Agency and Materialism." In Jordan and Howard-Snyder (eds.) (1996): 47–71.

Crisp, Thomas M. and Warfield, Ted A. (2000). "The Irrelevance of Indeterministic Counterexamples to Principle Beta." *Philosophy and Phenomenological Research*, 61: 173–184.

Davidson, Donald (1973). "Freedom to Act." In Honderich (ed.) (1973): 139–156. Reprinted in Davidson, *Essays on Actions and Events*. Oxford: Clarendon Press (1980): 63–81.

Della Rocca, Michael (1998). "Frankfurt, Fischer and Flickers." *Noûs*, 32: 99–105.

Dennett, Daniel C. (1978). "On Giving Libertarians What They Say They Want." In Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, VT: Bradford Books: 286–299.

Donagan, Alan (1987). *Choice: The Essential Element in Human Action*. London: Routledge and Kegan Paul.

Earman, John (1986). *A Primer on Determinism*. Dordrecht: D. Reidel.

Eells, Ellery (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.

Ekstrom, Laura Waddell (2000). *Free Will: A Philosophical Study*. Boulder, CO: Westview Press.

Fischer, John Martin (1982). "Responsibility and Control." *Journal of Philosophy*, 79: 24-40. Reprinted in Fischer (ed.) (1986a): 174-190.

—— (1983). "Incompatibilism." *Philosophical Studies*, 43: 127-137.

—— (ed.) (1986a). *Moral Responsibility*. Ithaca: Cornell University Press.

—— (1986b). "Power Necessity." *Philosophical Topics*, 14 (2): 77-91.

—— (1988). "Freedom and Miracles." *Noûs*, 22: 235-252.

—— (1994). *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell.

—— (1995). "Libertarianism and Avoidability: A Reply to Widerker." *Faith and Philosophy*, 12: 119-125.

—— (1999). "Recent Work on Moral Responsibility." *Ethics*, 110: 93-139.

Fischer, John Martin and Hoffman, Paul (1994). "Alternative Possibilities: A Reply to Lamb." *Journal of Philosophy*, 91: 321-326.

Fischer, John Martin and Ravizza, Mark (eds.) (1993). *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press.

—— (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Flint, Thomas P. (1987). "Compatibilism and the Argument from Unavoidability." *Journal of Philosophy*, 84: 423-440.

Frankfurt, Harry G. (1969). "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy*, 66: 828-139. Reprinted in Fischer (ed.) (1986a): 143-152; in Frankfurt (1988): 1-10; and in Pereboom (ed.) (1997): 156-166.

—— (1971). "Freedom of the Will and the Concept of a Person." *Journal of Philosophy*, 68: 5-20. Reprinted in Fischer (ed.) (1986a): 65-80; in Frankfurt (1988): 11-25; in Pereboom (ed.) (1997): 167-183; and in Watson (ed.) (1982): 81-95.

—— (1975). "Three Concepts of Free Action." *Proceedings of the Aristotelian Society*, Supplementary vol. 49: 113-125. Reprinted in Fischer (ed.) (1986a): 113-23; and in Frankfurt (1988): 47-57.

—— (1976). "Identification and Externality." In Amélie O. Rorty (ed.), *The Identities of Persons*. Berkeley: University of California Press: 239-251. Reprinted in Frankfurt (1988): 58-68.

—— (1987). "Identification and Wholeheartedness." In Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press: 27 – 45. Reprinted in Fischer and Ravizza (eds.) (1993): 170–187; and in Frankfurt (1988): 159–176.

—— (1988). *The Importance of What We Care About*. Cambridge: Cambridge University Press.

—— (1992). "The Faintest Passion." *Proceedings and Addresses of the American Philosophical Association*, 66: 5–16.

Gallois, André (1977). "Van Inwagen on Free Will and Determinism." *Philosophical Studies*, 32: 99–105.

Ginet, Carl (1990). *On Action*. Cambridge: Cambridge University Press.

—— (1996). "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing." *Philosophical Perspectives*, 10: 403–417.

Goetz, Stewart (1997). "Libertarian Choice." *Faith and Philosophy*, 14: 195–211.

—— (1999). "Stumping for Widerker." *Faith and Philosophy*, 16: 83–89.

Haji, Ishtiyaque (1993). "Alternative Possibilities, Moral Obligation, and Moral Responsibility." *Philosophical Papers*, 22: 41–50.

—— (1998). *Moral Appraisability: Puzzles, Proposals, and Perplexities*. New York: Oxford University Press.

—— (1999). "Indeterminism and Frankfurt-type Examples." *Philosophical Explorations*, 2: 42–58.

Heinaman, Robert (1986). "Incompatibilism without the Principle of Alternative Possibilities." *Australasian Journal of Philosophy*, 64: 266–276. Reprinted in Fischer and Ravizza (eds.) (1993): 296–309.

Hill, Christopher S. (1992). "Van Inwagen on the Consequence Argument." *Analysis*, 52: 49–55.

Honderich, Ted (ed.) (1973). *Essays on Freedom of Action*. London: Routledge and Kegan Paul.

—— (1988). *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes*. Oxford: Clarendon Press.

Horgan, Terence (1985). "Compatibilism and the Consequence Argument." *Philosophical Studies*, 47: 339–356.

Hunt, David (1996). "Frankfurt Counterexamples: Some Comments on the Widerker–Fischer Debate." *Faith and Philosophy*, 13: 395–401.

—— (2000). "Moral Responsibility and Unavoidable Action." *Philosophical Stud-*

ies, 97: 195-227.

Jordan, Jeff and Howard-Snyder, Daniel (eds.) (1996). *Faith, Freedom, and Rationality: Philosophy of Religion Today*. Lanham: Rowman and Littlefield.

Kane, Robert (1996). *The Significance of Free Will*. New York: Oxford University Press.

—— (1999a). "On Free Will, Responsibility and Indeterminism." *Philosophical Explorations*, 2: 105-121.

—— (1999b). "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *Journal of Philosophy*, 96: 217-240.

—— (2000). "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson." *Philosophy and Phenomenological Research*, 60: 157-167.

Kapitan, Tomis (1991). "How Powerful Are We?" *American Philosophical Quarterly*, 28: 331-338.

—— (1996). "Incompatibilism and Ambiguity in the Practical Modalities." *Analysis*, 56: 102-110.

Lamb, James W. (1977). "On a Proof of Incompatibilism." *The Philosophical Review*, 86: 20-35.

—— (1993). "Evaluative Compatibilism and the Principle of Alternate Possibilities." *Journal of Philosophy*, 90: 517-527.

Lehrer, Keith (1997). *Self Trust: A Study of Reason, Knowledge and Autonomy*. Oxford: Clarendon Press.

Lewis, David (1973). "Causation." *Journal of Philosophy*, 70: 556-567. Reprinted with postscripts in Lewis (1986): 159-213.

—— (1981). "Are We Free to Break the Laws?" *Theoria*, 47: 113-121. Reprinted in Lewis (1986): 291-298.

—— (1986). *Philosophical Papers*, vol. II. New York: Oxford University Press.

Loewer, Barry (1996). "Freedom from Physics: Quantum Mechanics and Free Will." *Philosophical Topics*, 24 (2): 91-112.

McCall, Storrs (1994). *A Model of the Universe*. Oxford: Clarendon Press.

McCann, Hugh J. (1998). *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca: Cornell University Press.

McKay, Thomas J. and Johnson, David (1996). "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics*, 24 (2): 113-122.

McKenna, Michael (1997). "Alternative Possibilities and the Failure of the Counterexample Strategy." *Journal of Social Philosophy*, 28: 71-85.

Mele, Alfred R. (1992). *Springs of Action: Understanding Intentional Behav-*

ior. New York: Oxford University Press.

—— (1995). *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.

—— (1999a). "Kane, Luck, and the Significance of Free Will." *Philosophical Explorations*, 2: 96-104.

—— (1999b). "Ultimate Responsibility and Dumb Luck." *Social Philosophy and Policy*, 16: 274-293.

—— (2000). "Reactive Attitudes, Reactivity, and Omissions." *Philosophy and Phenomenological Research*, 61: 447-452.

Mele, Alfred R. and Moser, Paul (1994). "Intentional Action." *Noûs*, 28: 39-68.

Mele, Alfred R. and Robb, David (1998). "Rescuing Frankfurt-Style Cases." *The Philosophical Review*, 107: 97-112.

Narveson, Jan (1977). "Compatibilism Defended." *Philosophical Studies*, 32: 83-87.

Naylor, Margery Bedford (1984). "Frankfurt on the Principle of Alternate Possibilities." *Philosophical Studies*, 46: 249-258.

Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Belknap Press.

O'Connor, Timothy (1993). "On the Transfer of Necessity." *Noûs*, 27: 204-218.

—— (1995a). "Agent Causation." In O'Connor (ed.) (1995b): 173-200.

—— (ed.) (1995b). *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press.

—— (1996). "Why Agent Causation?" *Philosophical Topics*, 24 (2): 143-158.

—— (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

Otsuka, Michael (1998). "Incompatibilism and the Avoidability of Blame." *Ethics*, 108: 685-701.

Pereboom, Derk (ed.) (1997). *Free Will*. Indianapolis: Hackett.

—— (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.

Rowe, William L. (1991). *Thomas Reid on Freedom and Morality*. Ithaca: Cornell University Press.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Belknap Press.

Shatz, David (1986). "Free Will and the Structure of Motivation." *Midwest Studies in Philosophy*, 10: 451-482.

Slote, Michael (1982). "Selective Necessity and the Free-Will Problem." *Journal of Philosophy*, 79: 5-24.

Smilansky, Saul (2000). *Free Will and Illusion*. Oxford: Clarendon Press.

Sorabji, Richard (1980). *Necessity, Cause, and Blame: Perspectives on Aristotle's Theory*. Ithaca: Cornell University Press.

Strawson, Galen (1986). *Freedom and Belief*. Oxford: Clarendon Press.

Stump, Eleonore (1990). "Intellect, Will, and the Principle of Alternate Possibilities." In Michael D. Beaty (ed.), *Christian Theism and the Problems of Philosophy*. Notre Dame: University of Notre Dame Press: 254 - 285. Reprinted in Fischer and Ravizza (eds.) (1993): 237-262.

—— (1996). "Libertarian Freedom and the Principle of Alternative Possibilities." In Jordan and Howard-Snyder (eds.) (1996): 73-88.

—— (1999a). "Alternative Possibilities and Moral Responsibility: The Flicker of Freedom." *The Journal of Ethics*, 3: 299-324.

—— (1999b). "Dust, Determinism, and Frankfurt: A Reply to Goetz." *Faith and Philosophy*, 16: 413-422.

—— (2000). "The Direct Argument for Incompatibilism." *Philosophy and Phenomenological Research*, 61: 459-466.

Stump, Eleonore and Fischer, John Martin (2000). "Transfer Principles and Moral Responsibility." *Philosophical Perspectives*, 14: 47-56.

Swartz, Norman (1985). *The Concept of Physical Law*. Cambridge: Cambridge University Press.

Taylor, Richard (1966). *Action and Purpose*. Englewood Cliffs: Prentice-Hall.

—— (1992). *Metaphysics*, 4th edn. Englewood Cliffs: Prentice-Hall.

Thorp, John (1980). *Free Will: A Defence Against Neurophysiological Determinism*. London: Routledge and Kegan Paul.

Tooley, Michael (1987). *Causation: A Realist Approach*. Oxford: Clarendon Press.

Van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Clarendon Press.

—— (1999). "Moral Responsibility, Determinism, and the Ability to Do Otherwise." *The Journal of Ethics*, 3: 341-350.

Velleman, J. David (1992). "What Happens When Someone Acts?" *Mind*, 101: 461-481.

Vihvelin, Kadri (1988). "The Modal Argument for Incompatibilism." *Philosophi-*

cal Studies, 53: 227-244.

—— (2000). "Freedom, Foreknowledge, and the Principle of Alternate Possibilities." *Canadian Journal of Philosophy*, 30: 1-23.

Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Waller, Bruce N. (1993). "Responsibility and the Self-made Self." *Analysis*, 53: 45-51.

Warfield, Ted A. (1996). "Determinism and Moral Responsibility are Incompatible." *Philosophical Topics*, 24 (2): 215-226.

Watson, Gary (1975). "Free Agency." *Journal of Philosophy*, 72: 205-220. Reprinted in Fischer (ed.) (1986a): 81-96; and in Watson (ed.) (1982): 96-110.

—— ed. (1982). *Free Will*. Oxford: Oxford University Press.

Widerker, David (1987). "On an Argument for Incompatibilism." *Analysis*, 47: 37-41.

—— (1991). "Frankfurt on 'Ought Implies Can' and Alternative Possibilities." *Analysis*, 51: 222-224.

—— (1995a). "Libertarian Freedom and the Avoidability of Decisions." *Faith and Philosophy*, 12: 113-118.

—— (1995b). "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities." *The Philosophical Review*, 104: 247-261.

—— (2000). "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look." *Philosophical Perspectives*, 14: 181-202.

Widerker, David and Katzoff, Charlotte (1996). "Avoidability and Libertarianism: A Response to Fischer." *Faith and Philosophy*, 13: 415-421.

Wiggins, David (1973). "Towards a Reasonable Libertarianism." In Honderich (ed.) (1973): 31-61.

Wolf, Susan (1990). *Freedom within Reason*. New York: Oxford University Press.

Wyma, Keith (1997). "Moral Responsibility and Leeway for Action." *American Philosophical Quarterly*, 34: 57-70.

Zimmerman, Michael J. (1984). *An Essay on Human Action*. New York: Peter Lang.

—— (1993). "Obligation, Responsibility and Alternate Possibilities." *Analysis*, 53: 51-53.

索 引*

acacia tree/antelope example，金合欢树/羚羊例子，161-162
Adams，Fred，弗雷德·亚当斯，159，161，162
Adler，D.，阿德勒，247
affect programs，情感程序，294-295，397n42
agency：causality，能动性：因果性，389-392，397n42，399n62；Consequence Argument，后果论证，372-374；environment，环境，256-257；freedom of will，自由意志，378，399n62；interventions，干涉，397n34；moral decision-making，道德决策，299-300，386；moral responsibility，道德责任，379-380，396n31，396n32；non-causal views，非因果的观点，382-383；wholeheartedness，全心全意的，377-378，395n24
aggression，攻击，297
Aizawa，K.，艾扎瓦，159，161，162
Alexander，Samuel，萨缪尔·亚历山大，18
Allen，Colin，科林·艾伦，161
alternate possibilities principle（PAP），可供取舍的可能性原则，375-376，394n18
AMPA receptors，AMPA受体，327
analyticity，分析性，207-208，209n9
angels，天使，365
anger，愤怒，297
animals：behavior，动物：行为，293，297；conditioning，条件作用，151-152；human animals，人类动物，354，362-364；motivation，动机，297；thought，思想，363
Annas，J.，安娜斯，35n32
antelope/acacia tree example，羚羊/金合欢树例子，161-162
anti-individualism，反个体主义，256-257，260
anti-materialism，反唯物主义，56
anti-physicalists，反物理主义者，79-81
anti-reductionism：conceptual，反还原论：概念的，13，18-19；ontological，本体论的，13-17；teleofunctions，目的论功能，332
arachidonic acid，花生四烯酸，327

* 索引中页码系原书页码，即本书边码。——译者注

Aristotle，亚里士多德，14，231n4，257-258
Armstrong，David，大卫·阿姆斯特朗，16，25，51，58，90
artificial intelligence，人工智能，309，317，346
associative learning，联想学习，324-325，328
assumptions，假设，274，275，277-278
astronomers，天文学家，182-183
atomic representationalism：cognitive representations，原子表征主义：认知表征，172；compositionality，构成性，181；Gödel numerals，哥德尔数，183-187；systematicity of thought，思维系统性，176-178，179-180
Atomism：Conceptual，原子论：概念的，202-205，206，207；content，内容，201-202；eliminativism，取消主义，40-41n83；Fodor，福多，203；individuation，个体化，203-204
attention，注意，340-341，343-344
attitudes，态度，4-5，289，290；propositional，命题的，4-5，55，57
attribute theory，属性理论，85，98n3
automatic appraisal mechanism，自动评价机制，303
autonomy，自主性，10，75-76，215-216，261

bachelor concept，单身汉概念，191，192
Baker，L.，贝克，160
batrachians，两栖类脊椎动物，175-176
behavior：animals，行为：动物的，293，297；appetitive，食欲的，293；consciousness，意识，138n26；folk psychology，民间心理学，29；linguistic，语言的，145-146；mental states，心理状态，11；rationality，理性，合理性，314-315
behaviorism：analytical，行为主义：分析的，49-50；criterial，标准的，21；feelings，感受，288；innateness，天赋性，293；instinct，本能，291-292，293；logical，逻辑的，21，37n58；Pavlovian conditioning，巴甫洛夫条件作用，151-152；philosophy，哲学，49-50，236；psychology，心理学，48-50；ritual，礼仪的，292；semantics，语义学，309-310；translational，翻译的，21
belief：ascriptions，信念：归属，58-59；desire，愿望，4-6，55，58-59，242，289，290；discrepant，矛盾的，249-251；false，错误的，假的，4，32n5；Fodor，福多，317；functional，功能的，111-112；Hume，休谟，128；intentionality，意向性，3-4；knowledge，知识，191，192；perseverance，坚定，249-250；post-event interventions，事后干预，250-251；public policy implications，国家政策意蕴，250-251
Berkeley，George，乔治·贝克莱，16
Bickle，John，约翰·比克尔，348n7，348n8
Bierbrauer，G.，比尔布劳尔，246
Biogeography，生物地理学，175-176
biological systems，生物系统，54
Black，Max，马克斯·布莱克，137n19

Blackburn, Simon, 西蒙·布莱克本, 99n13
Block, Ned, 内德·布洛克, 22, 23, 38n66, 56, 137n17
Body, 身体, 1, 216, 365; 也参见 dualism, 二元论; mind-body problem, 心身问题
Brain, 大脑, 295, 329
brain damage, 脑损伤, 315-316
brain-imaging techniques, 脑成像技术, 333-334
Bratman, Michael, 迈克尔·布拉特曼, 395n24
Brentano, Franz, 弗兰兹·布伦塔诺, 3, 55, 57
bridge laws, 桥梁法则, 19, 37-38n62, 75, 87
Broad, C. D.: emergentism, 布罗德: 突现论, 18, 19, 125; against mechanism, 反机械论, 136n9; mental states, 心理状态, 102; The Mind and its Place in Nature, 《心灵及其在自然中的位置》, 102-103; vitalism, 活力论, 110
Brooks, Rodney, 罗德尼·布鲁克斯, 313
Burge, Tyler, 泰勒·伯奇, 259, 260, 265-266, 267
Buss, David, 大卫·巴斯, 296

Campbell, F. W., 坎贝尔, 277
Caramazzo, Alfonso, 阿方索·卡拉马佐, 347
carburetor example, 汽化器例子 203
Carnap, R., 卡尔纳普, 48 - 49, 136n16, 258-259
categorization, 分类, 范畴化, 191, 198-199
causality: agency, 因果性: 能动性, 389 - 392, 397n42, 399n62; consciousness, 意识, 126; contingency, 偶然性, 24; events, 事件, 24, 384-389, 390 - 391; explanation, 解释, 225 - 226; functionalism, 功能主义, 22; interactionism, 交互作用论, 15; mental events, 心理事件, 217 - 218, 226 - 228, 230 - 231; nexus, 关系, 125; physical state, 物理状态, 77 - 78, 82n14; physicalism, 物理主义, 81; reference, 指称, 157
central nervous system, 中枢神经系统, 325
cerebral hemispheres, 大脑半球, 353; transplanted, 移植的, 361-362
Chalmers, D. J., 查默斯, 57, 111, 135n1, 137 - 138n21, 138n26, 229 - 230
chaos theory, 混沌理论, 35-36n46
Chomsky, Noam, 诺姆·乔姆斯基, 260-261, 262
Churchland, P. M.: connectionism, P. M. 丘奇兰德: 联结主义, 317, 318; content similarity, 内容类似性, 210 - 211n17; eliminativism, 取消主义, 28 - 29, 60, 61; folk Psychology, 民间心理学, 238; luminescence, 荧光, 136n13; neuroscience, 神经科学, 346; Quinean, 蒯因主义的, 123; statespace semantics, 状态空间语义学, 200
Churchland, P. S., P. S. 丘奇兰德, 346; connectionism, 联结主义, 201;

consciousness，意识，119；emotion，情绪，316；neuroscience，神经科学，60-61，322，337，346
Clark，A.，克拉克，318，319
Clarke，Randolph，鲁道夫·克拉克，390
Classical Theory：analyticity，经典理论：分析性，209n9；concepts，概念，191，193，209n5，209n7；primitive concepts，原始概念，209n5；psychology，心理学，195，205；typicality effects，典型性效应，194
classicism，古典主义，144，172-174，182，313
Clifford，W. K.，克利福德，16
co-consciousness，共意识，97
coevolutionary research ideology，协同进化研究意识形态，329-330
cognition：computational approach，认知：计算方法，272-273；decision-making，决策，242，243；Fodor，福多，173-174；Pylyshyn，派利申，173-174；systematicity，系统性，172，178-180；unconscious inference，无意识推理，32n11
cognitive psychology，认知心理学，288-289，347
cognitive science，认知科学，261，346；emotions，情绪，302-303，305；individualism，个体主义，262；social psychology，社会心理学，246-247
collapse principles，坍缩原则，125-126，138n25；quantum mechanics，量子力学，127，138n27
combinatorialism，functional，功能组合主义，172，183-187
commissurotomy，连合部切开术，353
common-sense psychology，常识心理学；*see* folk psychology，参见民间心理学
compatibilism：determinism，相容论：决定论，375，381；Frankfurt cases，法兰克福的例子，375-376；hierarchical account，分层说明，376-378；local-miracle，局部奇迹，373；multiple-pasts，多重过去，373；responsibility，责任，398n56
Compatibility Question，相容性问题，370-374
complex systems，复杂系统，12
compliance，服从，246
compositionality，构成性，175，180-181，182
computational approach，计算方法，275-276；cognition，认知，272-273；individualism，个体主义，276；Marr's theory of vision，马尔的视觉理论，273-274；mental states，心理状态，263；psychology，心理学，261
computer technology，计算机技术，22
conceivability arguments，可想象性论证，105-106，119
concepts，概念，4；acquisition，获得，192，204；atomism，原子论，202-205；autonomy，自主性，10；Classical Theory，古典理论，191，193，209n5，209n7；definitions，定义，192-193；Dual Theory，双重理论，197；explanatory role，解释作用，203；mental，心理的，8-9，28-29；nativism，天赋论，204；phenomenal，现象的，118；physical，物理的，118；primitive，原始的，209n5；prop-

erties，属性，7；Prototype Theory，原型理论，195－198；psychology，心理学，208；structure，结构，206－207，210n13；theory of，……的理论，190－191
Connectionism：activation values，联结主义：激活值，38n63；Churchland，丘奇兰德，201；classicism，古典主义，313；inferences，推理，175，317；rationality，合理性，312；robotics，机器人学，315，319
consciousness：attention，意识：注意，340－341，343－344；attribute theory，属性理论，98n3；behavior，行为，138n26；causal role，因果作用，126；conceivability argument，可想象性论证，105－106；Descartes，笛卡尔，2；epistemic gap，认识的间隙，认识的鸿沟，107－108，118；experience，经验，103－104；explanatory argument，解释论证，104－105，115；functionalism，功能主义，109，136n12，229；Huxley，赫胥黎，19；identity，同一性，113，114；intentionality，意向性，5，34n30；knowledge argument，知识论证，106－107；materialist solution，唯物主义的解答，104；mental states，心理状态，3，5；metaphysics，形而上学，102－103，135n1；natural world，自然界，102；neurophysiology，神经生理学，323；phenomena，现象，102－103，122；physicalism，物理主义，65－66，119，333－334；privacy，私人性，85；problems with，……的问题，103－104；qualia，感受性质，55－57；reductionism，还原论，104，135n4；token reductions，个例还原，89；truth，事实，123
Consequence Argument：agency，后果论证：能动性，372－374；determinism，决定论，370－372；event-causal libertarian view，事件因果关系的自由主义观点，388；inference rule，推理规则，373
content：atomism，内容：原子论，201－202；disjunctive，析取的，278－279；externalism，外在主义，334；identity，同一性，200；individualism，个体主义，265－266；intentional，意向的，279－280；meaning，意义，153；names，专名，156－157；narrow，窄的，207，210n12，211n20，279；similarity，相似性，200，201；thought，思维，143，146－147，165－166，219；wide，宽的，210n12
contingency，偶然性，24，73，74
continuity：identity，连续性：同一性，353，354－355；mental，心理的，361，362，363，364；psychological，心理学的，360
Cosmides，Leda，勒达·考斯麦茨，262，296
Counterfactuals，反事实条件句，97－98，100n15
“crackdow” example，“裂影”例子，278
creationism，特创论，176
culture，文化，294
Cummins，R.，卡明斯，54，159－160

Damasio，Antonio，安东尼奥·达马西奥，293，302－303，304，315
Dancy，J.，丹西，192

Darwin, Charles, 查尔斯·达尔文, 175, 288, 290-291, 302
Davidson, Donald: anomalous monism, 唐纳德·戴维森: 异常一元论, 18-19, 59-60; intentional states, 意向状态, 217; skepticism, 怀疑论, 282; supervenience, 随附性, 33n19; Swampman, 沼泽人, 152
DeAngelis, G. C., 迪安杰利斯, 339
deception cues, 欺骗暗示, 240
decision-making, 决策, 242, 243, 319, 369, 386, 393n1
deducibility, 可演绎性, 137n17
Democritus, 德谟克利特, 40-41n83
Dennett, D. C.: artificial intelligence, 丹尼特: 人工智能, 346; beliefs/desires, 信念/愿望, 58-59; heterophenomenology 异己现象学, 111; psychological explanation, 心理学解释, 54; Quinean, 蝴因主义的, 123; real patterns of reality, 实在中的真实模式, 89; vitalism, 生机论, 110
Descartes, René: causal interactionism, 勒内·笛卡尔: 因果交感论, 15; cogito, 我思故我在, 92; consciousness, 意识, 2; dualism, 二元论, 13-15, 47-48, 85, 98n2, 124-125, 231, 235-236; epiphenomenalism, 副现象论, 15, 86; God, 上帝, 14; mental causation, 心理因果关系, 230-231; mind, 心灵, 28, 216; motion, 运动, 231n1; phantom pain, 幻肢痛感, 214-215; sensation, 感觉, 143; substance, 实体, 物质, 14, 86, 214; thought, 思想, 14, 31n3
description, 描述, 91, 122, 271-272
descriptivism, 描述主义, 257, 258
desires: ascriptions, 愿望: 归属, 58-59; attribution, 归属, 246-249; belief, 信念, 4-6, 55, 58-59, 242, 289, 290; first-order, 一阶的, 377; prediction, 预测, 预言, 247-248; simulation theory, 模仿论, 248-249
determinism, 决定论, 381-382; compatibilism, 相容论, 375, 381; Consequence Argument, 后果论证, 370-372; Earman, 埃尔曼, 393n3; freedom of will, 自由意志, 96-97, 370, 392-393; libertarian accounts, 自由主义的解释, 391; responsibility, 责任, 388-389, 394n18
disjunction problem, 析取问题, 147-148, 153-154, 158-161
displacement activity, 代替活动, 292
disposition, 倾向, 32n10
divine foreknowledge, 神圣的预知, 393n2
doppelgangers, 孪生人, 257, 269, 278-279
double aspect theory, 两方面论, 16-17
double-occupancy view, 双重占有观, 361-362
Dretske, Fred: externalism, 弗雷德·德雷斯基, 334; historical instantiation condition, 历史的例示条件, 164-165; information theory, 信息理论, 166n10; knowledge, 知识, 147; learning period, 学习时期, 149, 151, 158-159; meaning, 意义, 153-156; misrepresentation, 错误表征, 150; names, 专名, 157; Swampman, 沼泽人, 162-163; teleology, 目的论,

54；thought/symbols，思维/符号，145，146，162；uninstantiated properties，无例示的属性，157
Dual Theory，双重论，197，199，206－207，210n12
dualism，二元论，35n38；bundle，捆绑，85－86，92－95；Descartes，笛卡尔，13－15，47－48，85，98n2，124－125，231，235－236；modal，模态的，118；predicate，谓词，86，87－89；property，属性，36－37n54，127；substance，实体，物质，85－86，98n2，124－125，127；type-D，D类，124－127，134；type-E，E类，124，127－129，134；type-O，O类，133；也参见 mind-body problem，心身关系

Earman，John，约翰·厄尔曼，393n3
Egan，F.，伊根，268，271－274，275，278－279
Ekman，Paul：automatic appraisal mechanism，保罗·埃克曼：自动评价机制，303；deception cues，欺骗暗示，240；display rules，展示规则，299；emotions，情绪，294－295，298；evolutionary explanation，进化的解释，296；neurocultural theory，神经文化理论，300；universality of emotions，情绪的普遍性，301－302
eliminativism：atomism，取消主义：原子论，40－41n83；entities，实体，实在，222－223；folk psychology，民间心理学，91，238，241；intentionality，意向性，225；mental states，心理状态，25，28－29，252n6；mind-body problem，心身问题，13；rejection of，对……的否定，61；representationalism，表征主义，187－188n1；type-A materialism，A类唯物主义，109
Elizabeth of Bohemia，Princess，波希米亚的伊丽莎白公主，215，216
emergentism：bridge laws，突现论：桥梁法则，37－38n62；Broad，布罗德，18，19，125，127；emergent materialism，突现的唯物主义，18，36n51；epiphenomenalism，副现象论，19；neutral，中立的，18
emotions，情绪，58；brain damage，脑损伤，315－316；cognitive science，认知科学，302－303，305；culture，文化，294；Darwin，达尔文，288；Ekman，埃克曼，294－295，298；ethology，动物行为学，291－293；evaluative judgments，评价性判断，289；evolutionary theories，进化论，290－291，295，296，297，302；experience，经验，288；facial expressions，面部表情，290－291，294，298；feeling，感觉，1，215，288，302－303；folk psychology，民间心理学，297，298，304；game theory，博弈论，296；mood，情绪，304；moral agency，道德动因，299－300；neurocultural theory，神经文化理论，300；philosophy，哲学，288－289；rationality，理性，合理性，315－316；social constructions，社会的建构，299，300－301；sociobiology，社会生物学，295－296；transactional theory，交往理论，297－299；twin－pathway models，双路径模型，303；universality，普遍性，301－302

Endogenesis，内源，337，338
endowment effect，馈赠效应，247
entities，实体，实在，66-67，90，216，222-223
environment，环境，256-257
epiphenomenalism：Descartes，副现象论：笛卡尔，15，86；dualism type-E，E类二元论，127-129；emergentism，突现论，19；intuition，直觉，139n29；non-reductivism，非还原论，134；physicalism，物理主义，19
epistemic gap：consciousness，认识性鸿沟：意识，107-108，118；materialism，唯物主义，112-113，114，119-121
epistemology，认识论，74-75
error，错误，193-194，318
essentialism，本质主义，198-199
ethology，动物行为学，291-293
evaluative judgments，评价性判断，289
event-causal libertarian view，事件因果关系的自由主义观点，384-389，390
events，事件，31n4；anomalous monism，异常一元论，36n49；causal relations，因果关系，24，384-389，390-391；mental，心理的，2；nonphysical，非物理的，2，78，80-81
evolution，进化，176
evolutionary biology，进化生物学，264
evolutionary psychology，进化心理学，296
evolutionary theories of emotions，进化的情绪理论，290-291，295，296，297，302
existence，存在，354，391-392
experience，经验，20；consciousness，意识，103-104；decision-making，作出决定，319；emotion，情绪，288；human animals，人类动物，363；mental episodes，心理情景，50；pain，疼痛，6，217；qualities，性质，227-228，231n2；representation，表征，231n2
explanation，解释，54，154，163，225-226，296
explanatory argument，解释论证，104-105，115，217
explanatory role，concepts，概念的解释作用，203
extension，广延，14
externalism：assumptions，外在主义：假设，274；content，内容，334；Dretske，德雷斯基，334；functionalism，功能主义，27-28；individualism，个体主义，257；metaphysics，形而上学，264-265；Newsome，纽瑟姆，338；phenomenal，现象的，334，338，340；reductionism，还原论，28；self-knowledge，自我知识，281-282，285；social，社会的，260
extras，problem of，额外之物问题 70-71
eyewitness testimony，目击者的证词，250-251

facial expressions，面部表情，290-291，294，298
Farkas，Katalin，卡塔琳·法卡斯，99n13，99n14
feelings，感受，1，215，288，302-303；hybrid，混合的，289，290
Feigl，Herbert，赫伯特·费格尔，19，25

fetuses, early-term, 早期的胎儿, 356, 363
Feyerabend, Paul, 保罗·费耶阿本德, 40-41n83, 60
finite state machines, 有限状态机, 22
Fischer, John Martin, 约翰·马丁·费希尔, 379 - 380, 393n2, 395n29, 396n30, 396n32
fission, Persistence Question, 持存的分体问题, 361-362
Fodor, Jerry: asymmetrical dependency, 杰里·福多: 非对称的依赖性, 160-161, 165; atomism, 原子论, 203; beliefs, 信念, 317; classicism, 古典主义, 172 - 174; cognition, 认知, 173 - 174; cognitive representations, 认知表征, 178 - 180; concept nativism, 概念天赋论, 204; functional combinatorialism, 功能组合论, 183; functionalism, 功能主义, 346; historical instantiation condition, 历史的例示条件, 157 - 158, 159, 164 - 165, 183; identity theory, 同一论, 52; language, 语言, 145 - 146, 147; learning period, 学习时期, 149; meaning, 意义, 150 - 153; methodological solipsism, 方法论的唯我论, 260 - 261; names, 专名, 156 - 157; prototypes, 原型, 197; psychosemantics, 心理语义学, 57; rationality, 理性, 合理性, 310, 311, 319; "Special Sciences", "专门科学", 87 - 88; Swampman, 沼泽人, 163; symbols, 符号, 162; systematicity/compositionality, 系统性/构成性, 175, 180 - 181; teleology, 目的论, 54; Twin Earth, 孪生地球, 159; unicorns, 独角兽, 158; uninstantiated properties, 未被例示的属性, 157; vacuity of theory, 理论的空洞无物, 164-165
folk psychology: behavior, 民间心理学: 行为, 29; Churchland, 丘奇兰德, 60-61; description, 描述, 91; eliminativism, 取消主义, 91, 238, 241; emotion, 情绪, 297, 298, 304; functionalism, 功能主义, 237; individualism, 个体主义, 273; Lewis, 刘易斯, 51, 237; mental states, 心理状态, 24, 60, 241; mindreading, 心灵阅读, 239, 251; philosophy, 哲学, 237-238; platitudes, 老生常谈, 239, 240; simulation theory, 模仿论, 241; thought, 思维, 145
form/meaning, 形式/意义, 309-311
four-dimensionalism, 四维主义, 362
Frank, Robert A., 罗伯特·弗兰克, 295-296
Frankfurt, Harry G., 哈里·法兰克福, 375-378, 387, 394-395n21, 394n20
freedom of will: agency, 自由意志: 能动性, 378, 399n62; decision-making, 作出决定, 369; determinism, 决定论, 96 - 97, 370, 392 - 393; divine foreknowledge, 神圣的预知, 393n2; Frankfurt, 法兰克福, 376 - 377; libertarian accounts, 自由主义者的说明, 381-382; mind-body problem, 心身问题, 10; self-governance, 自我支配, 383
Frege, Gottlob, 戈特罗布·弗雷格, 257
Frege cases, 弗雷格案例, 146

Fridlund, Alan, 阿兰·弗里德伦德, 298-299, 304
functional combinatorialism, 功能组合主义, 172, 183-187
functional magnetic resonance imaging, 功能性核磁共振成像, 346
functionalspecifier, 功能指示, 231-232n5
functionalism, 功能主义, 251n2; analytic, 分析的, 21-22, 23, 24-25, 109; Aristotle, 亚里士多德, 231n4; causal role, 因果作用, 22; consciousness, 意识, 109, 136n12, 229; counterexamples, 反例, 56; externalism, 外在主义, 27-28; Fodor, 福多, 346; folk psychology, 民间心理学, 237; homuncular, 小人的, 53-54; identity theory, 同一论, 231-232n5; individualism, 个体主义, 273; machine, 机器, 52-53, 58; meaning, 意义, 237, 240-241; mental states, 心理状态, 23-24, 238-239; physicalism, 物理主义, 262-263; propositional attitudes, 命题态度, 57; Putnam, 普特南, 22, 220, 346; Pythagoreans, 毕达哥拉斯学派, 37n61; realizability, 可实现性, 232n5; supervenience, 随附性, 9; teleological, 目的论的, 58; zombies, 怪人, 56

Gage, Phineas, 菲尼亚斯·盖奇, 315-316
Gallistel, Randy, 兰迪·加利斯特尔, 347
Gallois, André, 安德烈·加洛伊斯, 393n8
game concept, 游戏概念, 192
game theory, 博弈论, 296
"gant" example, "巨大无比的蚂蚁"例子, 158
Gazzaniga, M., 加扎尼加, 334, 345, 346, 347
generalizations, psychological, 心理学归纳, 200
Gettier, Edmund, 埃德蒙·格蒂尔, 192
ghost in the machine, 机器中的幽灵, 99n8, 236
Gibson, J. J., 吉布森, 261
Ginet, Carl, 卡尔·吉内特, 382-383, 397-398n45, 397n41, 397n43
glutamate, 谷氨酸盐, 327
God/gods, 上帝, 14, 123, 365
Gödel numerals, 哥德尔数, 183-187, 188n17
Godfrey-Smith, P., 戈弗雷-史密斯, 164
gold, 金子, 194
Goldman, A., 戈德曼, 248
Gordon, R., 戈登, 248-249, 252n8
grandmother concept, 祖母概念, 196-197
Greenspan, Patricia, 帕特丽夏·格林斯潘, 289
Grice, H. P., 格莱斯, 146, 147, 148
guidance control, 导向控制权, 379, 395n28

haecceitas, 个性, 93-94, 96
Hare, R. M., 黑尔, 33n19
Harris, P., 哈里斯, 244

Hart, W. D., 哈特, 57
Haugeland, J., 豪格兰, 310
Hauser, Mark, 马克·豪泽, 299
Hawkins, R. D., 霍金斯, 327, 328
Hebb, D. O., 赫布, 325
Heil, John, 约翰·海尔, 228, 282
Heinroth, Oskar, 奥斯卡·海因罗斯, 292
hemispherectomy, 脑半球切除手术, 361
heterophenomenology, 异己现象学, 111
Hildreth, Ellen, 埃伦·希尔德雷思, 266
Hill, Christopher, 克里斯托弗·希尔, 118, 137n21
Hinde, Robert, 罗伯特·海因德, 293, 297, 298
historical instantiation condition, 历史的例示条件, 152, 153, 157-158, 159, 164-165, 183
homology, 同源性, 302
homuncular functionalism, 小人功能主义, 53-54
hope, 希望, 289
Horgan, T., 霍根, 61
human animals, 人类动物, 354, 362-364
Hume, David, 大卫·休谟, 35n45, 85-86, 125, 128, 354
Huxley, T. H., 赫胥黎, 18, 19

idealism, 唯心主义, 15-16, 133
identity: bodily criterion, 同一性: 身体的标准, 365; consciousness, 意识, 113, 114; content, 内容, 200; continuity, 连续性, 354-355; counterfactual, 反事实的, 反事实条件句, 94, 96, 100n15; deduced, 演绎的, 114; diachronic, 历时性的, 353; empathy failure, 移情的缺失, 95-96; memory, 记忆, 353; non-physical, 非物理的, 75; numerical, 量上的, 93, 355-356, 361; persistence, 持存, 355-358; personal, 个人的, 86, 352-355; personhood, 人格, 353-354; philosophy, 哲学, 352; physical state, 物理状态, 113, 353; psychology, 心理学, 352; synchronic, 共时性的, 353; time, 时代, 92, 353; token identity, 个例同一性, 36n48, 52-53, 56, 72, 81n11, 231n3; twins example, 孪生人例子, 94-95; type identity, 类型同一性, 56, 72, 232n3; vague/partial, 大致的/部分的, 93-94
identity theory: Fodor, 同一论: 福多, 52; functionalism, 功能主义, 231-232n5; instrumentalism, 工具主义, 58-60; Kripke, 克里普克, 137n19; Lewis, C. I., 刘易斯, 40n74; Lewis, D., 刘易斯, 24, 27, 51, 76; mind, 心灵, 50-52; pain, 疼痛, 52; properties, 属性, 230; psychophysical, 心理物理的, 21, 24-26, 76-77; Putnam, 普特南, 52
imprinting, 刻, 149-150
incompatibilism, 不相容论, 369, 374, 393n6
inconsistency, 不一致性, 282-283
indeterminism, 非决定论, 385, 387, 388
indication, 指示, 147, 148, 155
indiscernability claims, 不可辨别性主张,

33-34n20
individualism: anti-individualism, 个体主义：反个体主义, 256-257; Burge, 伯奇, 259; cognitive science, 认知科学, 262; computational approach, 计算的方法, 276; externalism, 外在主义, 257; folk psychology, 民间心理学, 273; functionalism, 功能主义, 273; locational, 定位的, 276; mental content, 心理内容, 265-266; physicalism, 物理主义, 261, 262-263; Putnam, 普特南, 260; rejection, 反对, 265; representational primitives, 表征的原语, 270-271; self-knowledge, 自我知识, 280; vision, 视觉, 267-268
individuation, 个体化, 203-204, 263-264, 334
inference: abductive, 推理：诱导的, 317, 319; conjunction fallacy, 合取谬误, 252-253n12; connectionism, 联结主义, 175, 317; Consequence Argument, 结果论证, 373; logical, 逻辑的, 174; prediction, 预测, 预言, 244-245; rational, 合理的, 311, 312, 317; systematicity, 系统性, 173-178
information, 信息, 147, 148, 166n4, 166n10
information-processing, 信息加工, 266-267
innate behavior, 天赋的行为, 293
inner-cause thesis, 内因论题, 59
instinct, 本能, 291-292, 293
instrumentalism, 工具主义, 58-60
intentionality: belief, 意向性：信念, 3-4; consciousness, 意识, 5, 34n30; content, 内容, 279-280; eliminativism, 取消主义, 225; mental states, 心理状态, 5, 32n8, 32n9, 38-39n71, 85, 217-219, 232n9; physicalism, 物理主义, 65; problems with, ……的问题, 57-58; representation, 表征, 270-271; teleology, 目的论, 54-55
interactionism: causal, 交感论：因果的, 15; Chalmers, 查默斯, 138n26; dualism type-D, D类二元论, 134; microphysics, 微观物理学, 124; physics, 物理学, 125, 138n26; rejected, 被否定的, 126
interdisciplinarity, 跨学科性, 323, 345, 347
internalism, 内在主义, 257
interpretations, 解释, 138n25
interventions, 干涉, 397n34
introspection, 内省, 283, 285
intuition, 直觉, 92, 112, 137n21, 139n29
inverted spectrum, 颠倒光谱, 23
irrealism, 非实在论, 13, 28-29

Jackendoff, R., 杰肯道夫, 262
Jackson, Frank, 弗兰克·杰克逊, 56, 106-107
jade example, 翡翠的例子, 148, 149-150, 168n43
James, William, 威廉·詹姆斯, 16, 20, 132, 261, 288
jealousy, 忌妒, 296
de Jong, H. Looren, 卢伦·德扬, 329, 330, 331, 332

Kahneman, D., 卡尼曼, 252-253n12
Kandel, E. R., 坎德尔, 324, 327, 328, 349n12
Kane, Robert, 罗伯特·凯恩, 385-386, 398n52
Kant, Immanuel, 伊曼努尔·康德, 16
Katz, J., 卡茨, 209n9
Kim, Jaegwon, 金在权, 70
Kitcher, P., 基切尔, 61
kitten example, 小猫例子, 210-211n17
knowledge: a posteriori, 知识: 后验的, 8; a priori, 先验的, 8; belief, 信念, 191, 192; consciousness, 意识, 106-107; disjunction problem, 析取问题, 147-148; Dretske, 德雷斯基, 147; experiential, 经验的, 9-10
Kosslyn, Stephen, 斯蒂芬·科斯林, 346
Kripke, Saul: God, 索尔·克里普克: 上帝, 123; identity theory, 同一论, 137n19; materialism type-B, B类唯物主义, 114, 115-118; names, 专名, 40n78, 193, 258; natural kind terms, 自然种类术语, 40n78, 193, 202; table example, 桌子例子, 92-93; type/token identity, 类型/个例同一性, 56

language: behavior, 语言: 行为, 145-146; community of, ……的共同体, 259-260; evolutionary perspective, 进化的视角, 347; Fodor, 福多, 145-146, 147; naturalization, 自然化, 146-149; 也参见 natural language, 自然语言
language of thought, 思想语言, 144-146, 161-162
learning: associative, 学习: 联想的, 324-325, 328; central nervous system, 中枢神经系统, 325; neuron differentiation, 神经元的差别, 349n12
learning period, 学习时期, 149, 151, 158-159
LeDoux, Joseph, 约瑟夫·勒杜, 303
Leibniz, G. W., 莱布尼兹, 16, 39-40n73
Leibniz's Law, 莱布尼兹定律, 56
Leucippus, 留基伯, 40-41n83
Levine, Joseph, 约瑟夫·莱文, 118, 217
Lewes, G. H., 刘易斯, 17
Lewis, C. I., 刘易斯, 37n58, 40n74
Lewis, David, 大卫·刘易斯, 97; folk psychology, 民间心理学, 51, 237; identity theory, 同一论, 24, 27, 51, 76; pain, 疼痛, 39n72, 51; theory of mind, 心灵理论, 58
libertarian accounts: agent-causality, 自由主义者的解释: 自主体因果性, 389-392; determinism, 决定论, 391; event-causality, 事件因果性, 384-389; existence question, 存在问题, 391-392; freedom of will, 自由意志, 381-382; noncausal views, 非因果的观点, 382-383
linguistics, 语言学, 261
Llinás, Rodolfo, 鲁道夫·利纳斯, 337
Loar, Brian, 布赖恩·洛尔, 118, 137n21
Locke, John, 约翰·洛克, 50, 354, 356, 357, 359-360, 366
Loewenstein, G., 洛温斯坦, 247
Loftus, E., 洛夫特斯, 250-251

logical behaviorism，逻辑行为主义，21，37n58
logical positivism，逻辑实证主义，34n23，48，87
Lorenz，K.，洛伦兹，292-293，297
Lormand，Eric，埃里克·劳曼德，200
LTP，长期计划；*see* potentiation，long-term，参见长期势差现象
luck，幸运，385-387
luminescence，荧光，136n13
Lycan，William G.，威廉·莱肯，57，60

McAdams，C. J.，麦克亚当斯，340，341-344
macaques，猴子，340-344
McClintock，Barbara，芭芭拉·麦克林托克，333
McGinn，C.，麦金，32n7，119，139n33
Mach，Ernest，恩斯特·马赫，20
machine functionalism，机器功能主义，52-53，54，58
McKinsey，M.，麦金西，283-285
McLaughlin，B.，麦克劳林，36n51，175，183
MacLean，Paul D.，保罗·麦克莱恩，295
Maddell，Geoffrey，杰弗里·麦德尔，94
Manfredi，P.，曼弗雷迪，161
manifestation condition，表现条件，32n10
Marr，David，大卫·马尔，265，266，277-278，313
Marr's theory of vision，马尔的视觉理论，265，266，267-268；computational approach，计算的方法，273-274；Egan，伊根，271-274；perception，知觉，276-277；Segal，西格尔，268-271，278
Martin，C. B.，马丁，228，231
match-to-sample task，匹配样本任务，341-344
materialism：central state，唯物主义：中枢状态，25-26；consciousness，意识，104；emergent，突现的，18，36n51；epistemic arguments，认识论证，108；epistemic gap，认识性鸿沟，112-113，114，119-121；explanatory argument，解释论证，104-105；false，错误，107，108；reductive，还原的，35n41；type-A，A类，108-112，123，136n16；type-B，B类，112-119，123，136n15；type-C，C类，119-122
Matthews，R.，马修斯，268
Maunsell，J. H. R.，曼塞尔，340，341-344
meaning，意义，149-150；asymmetrical dependency，非对称的依赖性，165；content，内容，153；Dretske，德雷斯基，153-156；Fodor，福多，150-153；form，形式，309-311；functionalism，功能主义，237，240-241；information，信息，166n4；natural，自然的，147，148，151，153-154；natural language terms，自然语言术语，257；semantic promiscuity，语义的混杂，161-162；verification，证实，48，236-237
mechanical systems，机械系统，13-14，54，310

Medin, D., 梅丁, 198
memory, 记忆, 325, 328－329, 353, 359－360
mental episodes/experience, 心理情景/经验, 50
mental particle theory, 心理粒子理论, 13
mental states: attitudes, 心理状态：态度, 289; attribute theory, 属性理论, 85; behavior, 行为, 11; Broad, 布罗德, 102; causality, 因果性, 217－218, 226－228, 230－231; computational processes, 计算过程, 263; consciousness, 意识, 3, 5; continuity, 连续性, 361, 362, 363, 364; eliminativism, 取消主义, 25, 28－29, 252n6; folk psychology, 民间心理学, 24, 60, 241; functionalism, 功能主义, 23－24, 238－239; intentionality, 意向性, 5, 32n8, 32n9, 38－29n71, 85, 217－219, 232n9; pain, 疼痛, 221; physical state, 物理状态, 125; Putnam, 普特南, 53; second-order, 二阶的, 280; self-knowledge, 自我知识, 281
Mervis, Carolyn, 卡罗林·莫维斯, 194, 196
metaphysics, 形而上学, 16, 102－103, 135n1, 264－265
methodological solipsism, 方法论的唯我论, 260－261
methodology in neuroscience, 神经科学方法论, 329－331
mice/shrews example, 老鼠/鼩鼱例子, 159－160
microphysics, 微观物理学, 124, 133, 137n17
microstimulation studies, 微刺激研究, 337, 338－339
middle temporal cortex area, 中颞皮层区域, 334－337, 338－339
Milgram, S., 米尔格拉姆, 246
Millikan, R. G., 米利肯, 54
mind: dependence, 心灵：依赖性, 163; Descartes, 笛卡尔, 28, 216; entities, 实体, 实在, 216; identity theory, 同一论, 50－52; theory of, ……的理论, 58; 也参见 dualism, 二元论
mind－body problem, 心身问题, 10－12, 31n1, 216－217; continental divide, 大陆分水岭, 29, 31; Descartes, 笛卡尔, 235－236; eliminativism, 取消主义, 13; logical space of solutions, 解答的逻辑空间, 30; properties, 属性, 7－8; reduction, 还原, 8－9; supervenience, 随附性, 9－10; 也参见 dualism, 二元论
mindreading: deception cues, 心灵阅读：欺骗暗示, 240; desire attribution, 愿望归属, 246－249; folk psychology, 民间心理学, 239, 251; information-rich accounts, 信息丰富解释, 243－244; simulation theory, 模仿论, 241－242, 243－244, 245－246, 252n8
misrepresentation, 错误表征, 150
modularity theory, 模块性理论, 239
molecular genetics, 分子遗传学, 332－333
monism: anomalous, 一元论：异常的, 18－19, 36n49, 59－60; neutral, 中立的, 13, 20－21, 35n45, 131; physicalism, 物理主义, 32, 139; type-F,

F 类，117 - 118，124，129 - 133，134，139n32
monkeys：macaques，猴子，340 - 344；match-to-sample task，匹配样本任务，341-344；middle temporal cortex area，中颞皮层区域，334-337，338-339
Moore，G. E.，摩尔，2
moral agency，道德动因，299-300
moral responsibility，道德责任，379 - 380，396n31，396n32
Morgan，C. Lloyd，劳埃德·摩根，18，19
motion，运动，231n1，334-335，337
motion direction task，运动指导任务，345
multiple spatial channels theory，多空间通道理论，277

Nagel，Ernest，恩斯特·内格尔，75
Nagel，Thomas，托马斯·内格尔，3，16，32n7，56，119，136n7
names：causal theory of reference，专名：因果的指称理论，157；contents，内容，156 - 157；descriptivism，描述主义，258；Dretske，德雷斯基，157；Fodor，福多，156 - 157；Kripke，克里普克，40n78，193，258；Putnam，普特南，40n78，193
Narveson，Jan，简·纳维森，393n8
natural kind terms，自然种类术语，40n78，88，114，193，202，258
natural language，自然语言，144 - 145，168n35，257，258-259
natural selection，自然选择，337
natural world，自然界，1，102
necessitation，必然蕴含关系，108
necessity：a posteriori，必然性：后验的，115；conceptual，概念的，8 - 9，70；metaphysical，形而上学的，8，9-10，26，27，70；nomological，法则学的，8，9，70
neo-dualism，新二元论，57
neural networks，神经网络，200 - 201，311，312，322，333-334
neurobiological mechanisms，神经生物机制，60-61，324，344
neurocultural theory，神经文化理论，300
neurons，神经元，340-344，349n12
neurophysiology，神经生理学，61，323，334，340
neuroscience，神经科学，322 - 323，329-331，333
neurotransmitters，神经递质，327-328
Newell，A.，纽厄尔，311
Newsome，William，威廉·纽瑟姆，334-337，338，345
Nisbett，R.，尼斯比特，250
NMDA receptors，NMDA 受体，327，330-331
non-causal views，agency，非因果的观点，能动性，382-383
non-mentalism，非心理主义，11
non-reductionism，非还原论，75，134
Noonan，H.，努南，364
Nordby，Knut，克努特·诺比，136n8

O'Connor，Timothy，蒂莫西·奥康纳，390
odometer example，里程表的例子，274-275
Olson，Eric T.，埃里克·T·奥尔森，

363
Ortony, A., 奥托尼, 198

pain: Armstrong, 疼痛：阿姆斯特朗, 51; experience, 经验, 6, 217; identity theory, 同一论, 52; introspection, 内省, 283; Lewis, D., 刘易斯, 39n72, 51; mental state, 心理状态, 221; phantom, 幻肢痛感, 214-215; realizability, 可实现性, 221-222
pain predicate, 疼痛谓词, 224-225
panprotopsychism, 泛原心论, 117-118, 131-132, 134
panpsychism, 泛心论, 16, 35n41
Pappas, G., 帕帕斯, 60
parallelism, 平行论, 15, 16-17, 86
Parfit, Derek, 德里克·帕菲特, 99n13, 362
Pavlovian conditioning, 巴甫洛夫条件作用, 151-152
Peacocke, Christopher, 克里斯多佛·皮科克, 209n9
Peirce, C. S., 皮尔士, 16
people, inorganic, 无机的人, 365
perception, 知觉, 5, 276-277
Persistence Question, 持存问题, 353; double-occupancy view, 双重占有观, 361-362; fission, 分裂, 361-362; identity, 同一性, 355-358; Lockean View, 洛克主义观点, 357, 358; Psychological Approach, 心理的方法, 358, 359-361, 362-364, 366-367; Simple View, 简单观点, 358-359; Somatic Approach, 身体的方法, 358, 364-366
persistent vegetative state, 持存的植物状态, 356-357, 363
personhood, 人格, 353-354, 356-357, 364, 366
phenomena: consciousness, 现象的：意识, 102-103, 122; mental, 心理的, 2, 65-66, 235-236; physics, 物理学, 12; reductionism, 还原论, 135n4, 136n13; states, 状态, 133-134; straightforward/Gestalt, 直接的/格式塔, 89; teleology, 目的论, 88
phenomenal individuals, 现象的个体, 56-57
phenomenalism, 现象主义, 16
phenomenology, 现象学, 132
philosophy of mind: artificial intelligence, 心灵哲学：人工智能, 309; behaviorism, 行为主义, 49-50, 236; emotions, 情绪, 288-289; folk psychology, 民间心理学, 237-238; identity, 同一性, 352; neuroscience, 神经科学, 322-323; physics, 物理学, 66; realism, 实在论, 223-224
physical state, 物理状态, 69; causal closure, 因果的封闭性, 77-78, 82n14; identity, 同一性, 113, 353; mental state, 心理状态, 125; predicates, 谓词, 8-9; qualities, 性质, 6-7, 216-217
physicalism, 物理主义, 6-7; abstract/concrete entities, 抽象/具体实体, 66-67; causality, 因果性, 81; consciousness, 意识, 65-66, 119, 333-334; contingency, 偶然性, 73, 74; counterexamples, 反例, 78-79; counterintuitive, 违背直觉的, 67-68;

epiphenomenalism，副现象论，19；epistemology，认识论，74－75；false，错误，71；functionalism，功能主义，262－263；individualism，个体主义，261，262－263；intentionality，意向性，65；justifying，为……辩护，76－78；logical behaviorism，逻辑行为主义，21；mental phenomena，心理现象，65－66；mental properties，心理属性，9；monism，一元论，139n32；neuroscience，神经科学，333；non-reductionism，非还原论，75；objections，异议，反驳，78－81；ontology，本体论，262－263；particulars，殊相，特殊事物，71；predicate dualism，谓词二元论，86；properties，属性，71；rationality，理性，合理性，65；realizability，可实现性，68，72；reductionism，还原论，75；supervenience，随附性，69－71，73－74，261
physics：closed systems，物理学：封闭系统，215；commonsense physics，常识物理学，198；entities，实体，90；indeterminacy，不确定性，385；interactionism，交互作用论，125，138n26；phenomena，现象，12；philosophy，哲学，66；properties，属性，130
Pinker，Steven，史蒂文·平克，347
Place，Ullin，阿林·普赖斯，25，50
plague concept，黑死病概念，199－200
plate tectonics example，板块构造例子，88
platitudes，老生常谈，238－239，240
Plato，柏拉图，35n33
Poggio，Tomas，托马斯·波吉奥，266
Popper，Karl，卡尔·波普尔，29
positron emission tomography，正电子发射成像术，346
potentiation，long-term，长期势差现象，325－327，330－331，348n8
predicates，谓词，8－9，232n6
prediction，预言，244－245，247－248
principle of alternate possibilities，可供取舍的可能性原则，375－376，394n18
properties，属性，2，14；concepts，概念，7；dispositional，倾向的，228；dualism，二元论，36－37n54，127；identity theory，同一论，230；mental，心理的，7－8，9，10，13，16；narrow/wide，窄的/宽的，57；non-mental，非心理的，34n24；non-relational，非关系的，34n25；physical/phenomenal，物理的/现象的，133；physicalism，物理主义，71；physics，物理学，130；protophenomenal，原现象的，123－124，129－131；qualitative，质的，228；uninstantiated，无例示的，157－158
propositional attitudes，命题态度，4－5，55，57
Prototype Theory，原型理论，195－198，199，203，209n11
proximal projections，近端投射，162
psychological explanation，心理学解释，54
psychological laws，心理学规律，34n21
Psychological Approach，Persistence Question，持存问题的心理方法，358，359－361，362－364，366－367
psychology：autonomy，心理学：自主性，75－76；behaviorism，行为主义，

48－50；Classical Theory，经典理论，195，205；cognitive revolution，认知革命，288－289；computational revolution，计算革命，261；concepts，概念，208；continuity，连续性，360；evolutionary，进化的，296；generalizations，概括，200；identity，同一性，352；reductionism，还原论，99n8；unavoidability，不可避免性，90－92；也参见 cognitive psychology；folk psychology；social psychology，认知心理学；民间心理学；社会心理学
psychoneural inter-theoretic relations，心理神经的理论内关系，329
psychophysical identity theory，心理物理同一论，21，24－26，76－77
psychosemantics，心理语义学，57
public policy implications，国家政策意蕴，250－251
Putnam，Hilary：functionalism，希拉里·普特南：功能主义，22，220，346；identity theory，同一论，52；individualism，个体主义，260；jade example，翡翠的例子，148；“The Meaning of ‘Meaning’”，《“意义”的意义》，257；mental states，心理状态，53；methodological solipsism，方法论的唯我论，260；names，专名，40n78，193；natural kind terms，自然种类术语，40n78，202，258－259；natural language，自然语言，258－259；person/robot，人/机器人，38n66；“Psychological Predicates”，《心理学谓词》，220；Twin Earth example，孪生地球例子，57
Pylyshyn，Zenon：classicism，齐诺·派利申：经典主义，172－174；cognition，认知，173－174；cognitive representations，认知表征，178－180；functional combinatorialism，功能组合论，183；sensory neurons，感觉神经元，262；systematicity/compositionality，系统性/构成性，180－181
Pythagoreans，毕达哥拉斯，37n61

qualia：consciousness，感受性质：意识，55－57；individuation，个体化，334；materialism type-A，A 类唯物主义，111，135n3；mental causation，心理因果关系，226－228；motion，运动，334－335，337；reification，具体化，3
qualities，性质，227－228，231n2；mental，心理的，215，216－217，220；physical，物理的，6－7，216－217
quantum mechanics，量子力学，34n31，124，126，391－392；collapse theories，坍缩理论，127，138n27
quark example，夸克例子，130，203
quasi-memory，准记忆，360
Quine，W. V.，蒯因，123，193，195，207－208，345

radium example，镭的例子，373－374
Raichle，Marcus，马库斯·雷切尔，346
rationality：behavior，合理性：行为，314－315；connectionism，联结主义，312；ecological，生态学的，318－319；emotions，情绪，315－316；errors，错误，318；Fodor，福多，310，311，319；mechanical，机械的，310，317，320；physicalism，物理主义，65；representation，表征，313－314

Ravizza, Mark, 马克·拉维扎, 379, 380, 395n29, 396n30, 396n32
reactivity, 反应性, 379, 395-396n30
realism, 实在论, 10, 97-98, 166n2, 223-224
reality, 实在, 89, 262
realizability: functionalism, 可实现性: 功能主义, 232n5; multiple, 多样的, 22, 68-69, 88, 217, 221-223, 232n5, 329; pain, 疼痛, 221-222; physicalism, 物理主义, 68, 72
reasons-responsiveness, 理由反应性, 379-380
recognition heuristic, 识别启发法, 318-319
reduction, 还原, 10, 87, 89, 323-324, 327
reductionism, 还原论, 88, 89; conceptual, 概念的, 13, 20-28; consciousness, 意识, 104, 135n4; event causality, 事件因果性, 390; externalist, 外在主义的, 28; natural kind terms, 自然种类术语, 88; neuroscience, 神经科学, 322-323, 329-330; phenomenal, 现象的, 135n4, 136n13; physicalism, 物理主义, 75; psychology, 心理学, 99n8
reference, direct theory, 直接指称理论, 157
Reichenbach, Hans, 汉斯·赖辛巴赫, 345-346
relativity laws, 相对律, 124
Renaissance philosophy, 文艺复兴时期哲学, 16
representation: abstraction, 表征: 抽象, 190; algorithmic level, 算法层次, 266; cognitive, 认知的, 172, 178-180; combinatorial mental, 组合的心理的, 172; compositionality, 组合性, 180-181; experience, 经验, 231n2; functional, 功能的, 111, 121; implementational level, 执行层次, 266; intentionality, 意向性, 270-271; internal, 内部的, 314; mental, 心理的, 172; molecular, 分子的, 172; phenomenal, 现象的, 111; rationality, 合理性, 313-314; reality, 实在, 262; Segal, 西格尔, 269-271
representational primitives, 表征基元, 270-271, 280
representationalism, 表征主义, 138n23, 187-188n1; *see also* atomic representationalism, 也参见原子表征主义
Rescorla, R. A., 雷斯科拉, 325
responsibility, 责任, 388-389, 394n18, 398n; *see also* moral responsibility, 也参见道德责任
retrocognition, 倒摄认知, 360
Rey, Georges, 乔治斯·雷伊, 41n83
rhesus monkeys, 猕猴; *see* monkeys, 参见猴子
ritual, 礼仪的, 292
Robinson, W. S., 鲁宾逊, 57
robot/cat example, 机器人/猫的例子, 155-156, 160-161
robots, 机器人, 38n66, 310, 313, 315, 319, 365
Robson, J. G., 罗布森, 277
Rorty, Richard, 理查德·罗蒂, 40-41n83, 60
Rosch, Eleanor, 埃莉诺·罗施, 194, 196

Ross, L., 罗斯, 250
Russell, Bertrand, 伯特兰·罗素, 4-5, 20, 34n29, 130, 257
Ryle, Gilbert, 吉尔伯特·赖尔, 21-22, 32n10, 48-49, 136n16, 236, 288

Salzman, C. D., 萨尔兹曼, 337, 339
Schank, R., 香克, 311
Schopenhauer, Arthur, 亚瑟·叔本华, 16
Schouten, Maurice, 莫里斯·斯考滕, 329, 330, 331, 332
Schrödinger evolution, 薛定谔进化, 125, 126
Searle, John, 约翰·塞尔, 4, 38n65
Segal, G., 西格尔, 268-271, 278
Sejnowski, T., 塞耶诺斯基, 60-61
self, 自我, 35n45
self-forming action, 自我形成的行动, 398n52
self-governance, 自我支配, 380, 383
self-knowledge: externalism, 自我知识: 外在主义, 281-282, 285; inconsistency, 不一致性, 282-283; individualism, 个体主义, 280; introspection, 内省, 283, 285; McKinsey, 麦金西, 283-285; mental states, 心理状态, 281
Sellars, Wilfrid, 威尔弗里德·塞拉斯, 60, 237, 273, 345
semantics, 语义学, 200-201, 309-310
semicompatibilism, 半相容论, 375-376
sensation, 感觉, 6, 143
sensory neurons, 感觉神经元, 262
sensory performance, 感觉性能, 340-341
serotonin, 血清素, 327-328
sexual jealousy, 性忌妒, 296
Shapiro, James, 詹姆士·夏皮罗, 332-333
Shapiro, L., 夏皮罗, 274
Shepherd, Gordon, 戈登·谢泼德, 323-324, 344, 347
Shoemaker, S., 休梅克, 363
Shope, R., 肖普, 163, 164
similarity, 相似性, 200, 201
Simon, H., 西蒙, 311
Simple View, Persistence Question, 持存问题的简单观点, 358-359
simulation theory: desire detection, 模仿论: 愿望探测, 248-249; folk psychology, 民间心理学, 241; mindreading, 心理阅读, 241-242, 243-244, 245-246, 252n8
single unit approach, neurophysiology, 神经生理学的单一单元方法, 334
skepticism, 怀疑论, 10, 280-281, 282
Skinner, B., 斯金纳, 99n8
Smart, J. J. C., 斯马特, 25, 50
Smolensky, Paul, 保罗·斯莫伦斯基, 183
social constructions, 社会的建构, 299, 300-301
social interaction, 社会相互作用, 295
social psychology, 社会心理学, 249-250
sociobiology, 社会生物学, 295-296
solipsism, 唯我论, 260-261
Somatic Approach, Persistence Question, 持存问题的身体方法, 358, 364-366

somatic markers，躯体标记，315－316，319
soul，灵魂，1，354
space，sentential，句子空间，311
sphericity example 球形的例子，228
Spinoza，B.，斯宾诺莎，15，16－17
split personality，人格分裂，353
Stalnaker，R.，斯托尔内克，137n17
Stampe，D.，斯丹普，146
state，状态，2，5；*see also* mental states；physical states，也参见心理状态；物理状态
stereopsis，立体影像，277
stereoscopic depth，立体视觉的深度，339
Stich，Stephen，斯蒂芬·斯蒂克，28－29，59，261
stimulus：conditioned/unconditioned，刺激：条件作用的/非条件作用的，327
Stoljar，D.，斯托尔加，132
story-understanding，故事理解，311
Strawson，Galen，盖伦·斯特劳森，57，396n31
Strawson，P. F.，斯特劳森，15
substance，物质，实体，14，86，214
substance dualism，实体二元论，85－86，98n2，124－125，127
Summerfield，D.，萨默菲尔德，161
supervenience，随附性，33n19；Davidson，戴维森，33n19；functionalism，功能主义，9；global，整体的，70，71，73；metaphysical，形而上学的，10；mind-body problem，心身问题，9－10；multiple，多样的，329；nomological，法则学的，10，19；physicalism，物理主义，69－71，73－74，261；strong，强的，33－34n20，70
sustaining mechanism，syndrome-based，基于特征群持续机制，204－205
Swampman，沼泽人，152，162－163
symbols，符号，145，146，160，161，162，168n35，311
synaptic plasticity，突触可塑性，324，325－327，328－329，330－331，348n2
synaptic transmission，突触传递，348n2
systematicity：artificial intelligence，系统性：人工智能，317；classicism，经典主义，182；cognitive representations，认知表征，178－180；compositionality，组合性，175，180－181；functional combinatorialism，功能组合论，183－187；inference，推理，173－178；thought，思维，176－178，179－180

taxonomy，分类法，263－264
teleofunctions，目的论功能，332，333
teleology，目的论，54－55，58，88，91
Tensor Product Theory，张量积理论，183
territorial displays，领土宣示行为，297
Theory Theory，理论论，198－200，202，210n13，239
thought：abstraction，思想：抽象，190；animals，动物，363；content，内容，143，146－147，165－166，219；Descartes，笛卡尔，14，31n3；folk psychology，民间心理学，145；human animals，人类动物，363－364；identical，同一的，153；language of，……的语言，144－146，161－162；natural world，自然界，1；private，私人的，

215；sensation，感觉，143；symbols，符号，145，146，160，161，162；systematicity，系统性，176－178，179－180；transitions，转换，311
thought experiments，思想实验，11，23，38n65，136n7
Tinbergen，Niko，尼科·廷伯根，292，297
tokens，个例，36n48；identity，同一性，36n48，52－53，56，72，81n11，231n3；reductionism，还原论，88，89；robust，强健的，151
Tomkins，Silvan S.，西尔万·S·汤姆金斯，295
Tooby，John，约翰·图比，262，296
transactional theory，交往理论，297－299
"triune brain" theory，"三位一体脑"理论，295
truth，事实，真理，108－109，121，123
Tulving，Endel，恩德尔·塔尔文，347
Turing，Alan，阿兰·图灵，49
Turing Machine，图灵机，309－311，317
Turing Test，图灵测试，49，50
Tversky，A.，特韦尔斯基，252－253n12
Twin Earth example，孪生地球的例子，334，338；"crackdow"，"裂影"，278；doppelgangers，孪生人，257，278－279；Fodor，福多，159；meaning，意义，144；Putnam，普特南，57；water，水，144，152－153，219，258，278
twins example，孪生人例子，94－95，278－279
Two Factor Theory，两因素理论，210n12，211n20
type identity，类型同一性，56，72，231n3
typicality effects，典型性效应，194，196，206

Ullman，S.，厄尔曼，274
understanding，理解，259
unicorn example，独角兽的例子，157－158
universals，共相，90，99n10，301－302

vacuity of theory，理论的空洞无物，164－165
Van Gulick，R.，范久利克，54，119
van Inwagen，Peter，彼得·范·英瓦根，393n6，393n7
verification，证实，48，49，118，236－237
vision：individualism，视觉：个体主义，267－268；information processing，信息加工，266－267；Marr，马尔，265，266，267－274，278
vitalism，活力论，110

Wagner，A.R.，瓦格纳，325
water：microphysics，水：微观物理学，137n17；natural kind terms，自然种类术语，87－88，114，258；Twin Earth，孪生地球，144，152－153，219，258，278
Watson，Gary，加里·沃森，394n20
Watson，John B.，约翰·华生，288
Whitman，Walt，沃尔特·惠特曼，1

wholeheartedness，全心全意，377－378，395n24
Wilson，E. O.，威尔逊，12
Wilson，Robert A.，罗伯特·威尔逊，265
Witmer，D. Gene，吉恩·威特默，71
Wittgenstein，L.，维特根斯坦，40－41n83，143，192，289
Woodward，J.，伍德沃德，61

Zajonc，Robert，罗伯特·扎伊翁茨，303
zero-crossings，零相交，270，278
zombies，怪人，56，105－106，119，228－230

译后记

本书系根据斯蒂芬·斯蒂克和特德·沃菲尔德主编的 *The Blackwell Guide to Philosophy of Mind* 第一版（即 2003 年版）译出。本书译者分工如下：

高新民：第 4、6 章；

刘占峰：导论，第 1、2、3、5、10、11 章，索引；

陈　丽：第 12、13、14、15 章；

王世鹏：第 7、8 章；

张卫国：第 9、16 章。

全书由高新民、刘占峰负责统稿，陈丽参与了索引的翻译和其他的技术性工作。

由于译者水平有限，错误之处难免，恳请专家和同道不吝批评指正。

译者

2013 年 11 月 12 日

The Blackwell Guide to Philosophy of Mind edited by Stephen P. Stich and Ted A. Warfield Copyright © 2003 by Blackwell Publishing Ltd
Simplified Chinese version © 2012 by China Renmin University Press.

All Rights Reserved.
Authorised translation from the English language edition published by Blackwell Publishing Limited. Responsibility for the accuracy of the translation rests solely with China Renmin University Press and is not the responsibility of Blackwell Publishing Limited. No Part of this book may be reproduced in any form without the written permission of the original copyright holder, Blackwell Publishing Limited.

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.