

THE
PHILOSOPHY
OF
MIND

心灵哲学



高新民 储昭华 / 主编



商务印书馆



THE
PHILOSOPHY
OF
MIND

ISBN 7-100-03323-3



9 787100 033237 >

ISBN 7-100-03323-3 / B·506

定价：56.00 元

心 灵 哲 学

高新民 储昭华 主编

商 务 印 书 馆

2002年·北京

图书在版编目(CIP)数据

心灵哲学/高新民,储昭华主编. —北京:商务印书馆,2002

ISBN 7-100-03323-3

I. 心… II. ①高… ②储… III. 心灵学-研究
IV. B84

中国版本图书馆 CIP 数据核字(2001)第 030780 号

所有权利保留。

未经许可,不得以任何方式使用。

XIN LING ZHE XUE

心 灵 哲 学

高新民 储昭华 主编

商 务 印 书 馆 出 版

(北京王府井大街36号 邮政编码 100710)

商 务 印 书 馆 发 行

北 京 冠 中 印 刷 厂 印 刷

ISBN 7-100-03323-3/B·506

2002 年 7 月第 1 版

开本 850 × 1168 1/32

2002 年 7 月北京第 1 次印刷

印张 36 1/4

定价: 56.00 元

前 言

心灵哲学是一门古老而年轻的哲学分支学科。说它古老,是因为其渊源甚至可以追溯至原始思维,在那里已有“丰富的”“灵魂”观念,并形成了关于灵魂与身体的关系的“前哲学”。说它年轻,是因为作为一门独立的、相对成熟和完善的哲学分支学科,严格说来,是现代哲学向纵深发展以及与有关科学互动的产物,而步入健康、迅猛发展的“快车道”则只有约 50 年的历史。尽管年轻,尽管只是与语言哲学、道德哲学等处在同样层次的分支,但它“后来居上”,一跃而成为哲学领域的具有基础意义和举足轻重地位的学科。如果说近代哲学向现代哲学的运动经过了所谓的“哥白尼式革命”,语言哲学由此取代认识论在近代哲学中的“重心”“基础”地位而成为现代哲学的重心和标志,那么在当今哲学的百花园中,心灵哲学则成了重心之重心、基础之基础。对此,美国著名哲学家 J. 塞尔作了精辟的概括:心灵哲学是语言哲学的基础,“语言哲学是心灵哲学的分支之一。”(《意向性》,剑桥大学出版社,1983,Ⅶ.)如果说本体论、认识论问题的解决离不开语言哲学的话,那么同样可以说,语言哲学问题的解决则有赖于有关心灵哲学的探讨和发展,例如,要说明语言的意义、指称、性质、特征等,当然要诉诸于心理状态的意向性,尽管不是唯一地由之所决定的。J. 塞尔的断言是否言过其实,只要随便翻翻西方尤其是英美最近一、二十年出版

2 心灵哲学

的哲学书籍和杂志就清楚了。心灵哲学方面的论著数量之多、问题之深广、见解之新颖别致、争论之激烈、进展之快确实令人震惊。至少可以肯定,心灵哲学是当代西方哲学家关注的重心和热点之一。

心灵哲学研究的主要是心理现象的形式、范围、本质、特征、心与身的关系、心理内容及其根源等,并对常识心理解释模式和心理哲学进行哲学反思。随着认识的深入,心灵哲学已改变了或正在改变传统的心理观。由于心理现象是宇宙结构图景中的重要组成部分,因此心灵哲学的最新探索也触及到了一些根本的宇宙观问题,如随附性、依赖性、决定、协变、还原、规律等。有迹象表明,心灵哲学向纵深的发展将是引起未来哲学变革的重要根源和动力之一。由此看来,心灵哲学不是一门狭隘的心灵学问,而是一个有着相对稳定的硬核、模糊的边界、开放的性格和远大的前途的广博的研究领域。

为了了解西方心灵哲学的最新发展,推进我国心灵哲学的研究,我们选编了这本文集。本选辑由于篇幅所限仅从该学科的大量研究领域挑选了最近研究较多、属前沿和焦点问题的九大问题。它们分别是:一、心身问题,即心理现象的本质、心与身的关系问题,当前的争论主要集中在还原主义、功能主义、“实现”和物理主义的二难困境等问题上。二、感受性质,即人在经历感觉等心理状态时所体验到的主观的特征或现象学性质。非物理主义者认为:物理主义已能同化各种“反例”,但是不能同化感受性质。感受性质是心理世界的“新大陆”,是物理主义无法说明的非物理的东西。而物理主义者针锋相对,在反击中将物理主义艰难地向前推

进。三、随附性。这是心灵哲学中的一个新问题、新范式,指的是心理现象由物理现象所决定、依存的伴随或随带发生的特征。当然也可泛化到其他关系、属性之上,从而成为一个具有普遍意义的哲学范畴。这一研究是人们试图进一步把握心理现象的本质、特征以及在宇宙结构图景中的地位、在还原论和二元论之间寻找中间出路的产物,涉及到决定、还原、必然性、心物规律等心灵哲学和宇宙观的重大问题。四、意识。这里所说的意识是狭义意义上的意识,即贯穿于人的各种有意识心理现象中的共同的“觉知”“知晓”特征。围绕意识问题已形成了一个专门的、独立的研究领域,有的还提出了“意识学”的概念。意识问题林林总总,现在有人把它区分为两大类问题,即“困难问题”和“容易问题”,前者指“经验”问题,是倍受关注的问题。五、思想的语言和心理表征。前者指的是不同于自然语言的“机器语言”,是人脑思维的真正媒介,因为在许多人看来,自然语言有形、声等特点,根本不可能为人脑所加工。后者指的是人们认识世界和自身所得的信息的存储、加工的表达形式。这些研究真正深入到了心灵世界的内部结构、运作及机制。六、意向性和内容理论。这既是一个古老的问题,又是一个崭新的问题,语言哲学和心灵哲学正是在此相交汇的。许多人认为:自然语言的意义与心理状态的意向性有关,而后者又根源于思想的语言、心理表征的语义性。但思想的语言的语义性是什么,与心理状态的功能作用是什么关系,与心理语句的句法是什么关系,其语义性又由什么决定?这些是当前争论的焦点。七、他心知问题,即我心以外有无他心?如果有,如何认识、如何证明?认识他心的基础、根据、过程是什么?判断对象如机器人有无心灵、智能的根据

4 心灵哲学

是什么? 争论主要集中在他心知怀疑论和类比论证两大问题上。八、行动哲学。这是在传统的自由意志问题的基础上派生出来的一个专门研究行动的领域。当前的研究主要集中在行动的特征、标志、结构、解释和评价等问题上, 触及到理由与原因的关系、有无涵盖心物关系的规律、心理有无原因作用等根本哲学问题。九、心理解释与民众心理学的地位。这也是一个新的研究领域。民众心理学是潜藏在我们每个人心理生活和日常解释活动中的、关于行动之原因的常识心理学理论。一般人包括很多哲学家在内所持的心理观、心身观以及由此所构成的宇宙观实际上是常识心理学的观点。这一研究不仅涉及到如何解释心理、心理学的问题, 而且涉及到心理现象的地位和命运、心理学有无自主性、是否会被还原、取消等根本问题。

关于本书的选编、译注, 有几点需要予以说明: (一) 编者在选编过程中查阅过国外较多版本的“心灵哲学原著选辑”, 尽力取各家之所长, 但又努力追求自己的风格和特点, 因此编者依据多年的积累及研究, 结合我国哲学发展的需要及现状, 对体例和内容选裁作出了新的尝试; (二) 书中涉及到大量国内少见或未见的人名, 已有通行译法的一律照旧, 凡新译人名一般据《英语姓名译名手册》(商务印书馆 1997) 和《简名不列颠百科全书》(中国大百科全书出版社 1986-8) 译出; (三) 大多数文章或著作后均附有较详细的“注释”和“参考文献”, 酌情译出, 有些作了较大删节, 参考文献的顺序一般根据出现于文中的先后顺序排列; 如作者较多, 一般只译第一人, 然后加“等”字样; 文章篇名用引号、著作作用书名号分别予以标示; (四) 为了帮助读者理解原文中的有关概念, 译者酌情加了

前 言 5

一些注释,并用*予以标示。

本选辑由我们两人共同选编,并组织翻译、校改及统稿。具体分工是这样的:储昭华译第八部分、编人名对照表,熊哲宏译“心理语义学或者真值条件从哪里来”,杜保友译第七部分前三篇,殷筱译“随附性的种种概念”,周克武译“心理表征”,徐弢译“人”和“上帝心灵中的还原”,其余内容均由高新民所译。

本书从构思到具体翻译自始至终得到陈小文先生的指导及帮助;牟博先生对本书的选编和翻译提出了很好的建议,并为联系外国版权付出了辛劳;张大松教授审读了与现代逻辑联系较为密切的译文,刘红平博士审读了第四部分涉及到现代物理学的译文,均提出了有价值的修改意见及建议;刘雅琴及我们的研究生舒年春、董强、季岸先、戴振宇等作了大量校对工作,有的还结合学习我们所开的“专业英文翻译”课程、对照原文指出了一些明显的错误,在此一并谨致谢意。

心灵哲学的译介在我国刚刚起步,许多术语、内容第一次表现为中文,加上编译者学识所限,错漏一定不少,敬请读者批评指正。

高新民 储昭华

2000年12月

目 录

第一部分 心身问题

心身问题、描述心理学与心理物理学问题

..... 罗德里克·M. 齐硕姆(3)

心理与物理的偶然等同论 D.M. 阿姆斯特朗(11)

心理事件与大脑 保罗·K. 费耶阿本德(27)

自然层次的连续性 威廉·G. 利康(30)

把功能还给功能主义..... 埃里奥特·索伯(61)

第二部分 感受性质

副现象的感受性质..... 弗兰克·杰克逊(79)

玛丽不知道什么..... 弗兰克·杰克逊(97)

成为一只蝙蝠可能是什么样子 托马斯·内格尔(105)

感受性质与物理主义、取消主义 金在权(123)

功能主义与感受性质 悉尼·休梅克(137)

感受性质、表征论与自然主义..... 弗雷德·德雷特斯克(165)

第三部分 随附性

随附性的种种概念..... 金在权(203)

本体论的随附性 约翰·豪格兰德(236)

2 心灵哲学

上帝心灵中的还原 丹尼尔·博恩瓦克(256)

物理主义、随附性与依赖性

..... 保罗·K. 莫泽 J.D. 特劳特(279)

第四部分 意识

“意识”一词的两种用法 诺曼·马尔科姆(323)

勇敢地面对意识难题 大卫·J. 查默(360)

意识、信息与泛心论 威廉·西格尔(396)

意识、民众心理学与认知科学 阿尔文·I. 戈德曼(427)

关于心灵起源的四个假说 朱利安·杰恩斯(457)

第五部分 思想的语言与心理表征

为什么仍必须有思想的语言,其意何在?

..... 特伦斯·霍根 约翰·廷森(479)

心理语义学或者:真值条件来自哪里? J.A. 福多(515)

表征与心灵理论 莫顿·韦格曼(556)

心理表征 哈特里·H. 菲尔德(603)

第六部分 意向性与心理内容理论

心灵能做什么:非意向世界中的意向性

..... 普莱尔·雅各布(665)

计算心理学与解释理论 H. 普特南(694)

什么是心理表征理论 斯蒂芬·斯蒂克(716)

内容理论 J. 福多(742)

关于心理内容的种种理论

..... 桑福德·戈德堡 安德鲁·佩辛(796)

第七部分 他心问题

类推 伯特兰·罗素(873)

情感的类比 斯图尔特·汉普希尔(879)

关于他心的知识 诺尔曼·马尔科姆(893)

关于他心的类比推理 亚历克·希斯洛普(906)

人 P.F. 斯特劳森(925)

第八部分 行动哲学

行动、理由与原因..... 唐纳德·戴维森(959)

行动的心理原因与解释..... 辛西娅 麦克唐纳(983)

第九部分 心理解释与民众心理学的地位

民众心理学及其职责..... 威廉·利康(1007)

心灵哲学中的工具主义..... 加里·富勒(1027)

联结主义、取消主义与民众心理学的未来

..... 斯蒂芬·P. 斯蒂克等(1035)

民众心理学不会消失..... 特伦斯·霍根等(1068)

联结主义与心灵哲学概论..... 威廉·贝希特尔(1101)

解决心灵哲学问题的控制论方案..... 肯尼思·M. 塞耶(1134)

人名译名对照 (1139)

第一部分

心 身 问 题

心身问题、描述心理学 与心理物理学问题*

罗德里克·M. 齐硕姆

第9届维特根斯坦国际专题讨论会的主题是心灵哲学和心理学哲学。详细地讨论过的问题包括：心身问题、心灵哲学的传统问题和心理学哲学。

1. 心身问题

什么是心身问题？“心灵”一词有多种非常不同的用法，这一事实导致了近来讨论中的某种混乱。对这个词，我们至少可以区分出五种不同的用法，因此心身问题至少有五种意义。

(1)我们可以像笛卡尔使用“mens”一词那样使用“心灵”一词，让它指称具有心理属性的东西，亦即指称那种进行思想、感知、相信和欲求的东西。如果是这样，那么我们通常是用“心灵”这个词意指“人”这个词所意指的东西，因而该词表示的就是你和我这

* 本文是 R. 齐硕姆为《心灵哲学和心理学哲学》(第9届维特根斯坦国际专题讨论会记录汇编, 1984, 奥地利)一书所写的“序言”, 标题为译者所加。

4 第一部分

样一些实体。如果我们用“心灵”这个词意指“人”这个词所意指的东西,那么心身问题的一种表现形式就是这样一个问题:人与他的身体之间有何关系?例如在我和我的身体之间存在着一种什么样的关系?这里主要存在着两种可能的情况:或者我与我的身体是同一的,或者两者是不同一的。如果我与我的身体没有同一性,那么又会有两种可能:要么我与包括我的身体之一部分的某物是同一的,要么我与包括我的身体之一部分的任何东西没有同一性。在后一种情形之下,我又是一种什么样子的东西呢?

近年来有一种意见认为:“心灵”与身体以计算机的抽象图表(the abstract diagram)与计算机相联系的那种方式密切相关。但是如果“心灵”一词在上述所区分的那种意义上被当作指称具有心理属性的东西来使用,那么上述主张就是不能成立的。因为计算机的抽象图表本身就是一种属性,就是一种抽象物,而那具有心理属性的东西,即进行感知、思维、感觉和欲望的东西,则是一个独立的个体事物。

(2)我们也可以像现今许多人那样用“心灵”一词表示人的智能。如当人们说“那个人具有优良的心智”,意即说这个人很聪明时,“心灵”一词就是以这种方式而加以使用的。在“心灵”一词的第二种意义上,“对心灵的考察”就是指对人的智能的心理学考察。在这种情况下,作名词用的“心灵”就可能引起人的误解,因为它可能被人们用来指示一种与人及其身体一同存在的独立的个体事物。正是在这种情况下,人们就会错误地认为,存在着三种相互关联的独立的个体事物,即人、人的心灵和人的身体。

在第二种意义上使用“心灵”一词,我们还能说存在着心身问

心身问题 5

题吗？也许有这样的问题：“身体怎能具有智能”？如果我们假定存在着具有智能的某物，那么这种形式的心身问题就可归结为第一种形式：一个人与他的身体之间有何关系？因为大概正是人才具有智能。

(3)我们也可以将一个人的心灵说成是这样一种存在，正是借助于它，他才能进行思维。倘若如此，“心灵”一词指示的确实是一个个体的事物。一个人赖以进行思维的那种东西显然是大脑，或者至少是包括大脑一部分的某种东西。如果我们在这个意义上使用“心灵”（意即人们靠它才能思维），而不是在第一种意义上（意即人）或第二种意义上（指智能）使用该词，那么我们所说的“对心灵的研究”就会是对大脑的神经生理学研究。而“心身问题”这一表述所涉及的就是神经生理学的某些问题。对这些神经病学问题的回答本身并不能为我们提供解决“心身问题”的方案，因为在这里，“心身问题”这一表达式是根据上面所区分的第一种意义上的“心灵”而加以解释的。

(4)“心灵”一词有时还用来指示一种精神实体，即一种具有非物质本质的独立的个体事物（也许在这里使用“灵魂”一词更恰当一些）。在讨论人—体（person-body）问题时，如果我们断定人与他的身体或人与包括他的身体一部分的任何事物是不能等同的，那么我们可能要考虑这样的可能性，即按“心灵”一词的第四种意义，人是一个灵魂或心灵。（如果我们断言人是一种灵魂或心灵，那么我们就应小心谨慎，千万不要用“人有灵魂或心灵”这一说法去表述这一论断。因为那样的话，我们可能再次陷入这样的错误泥潭，即认为我们涉及的是三类独立的个体事物：人、心灵和身体。）

6 第一部分

(5)最后“心身问题”有时被认为是这样一个问题,即探究心理的与物理的属性之间的关系,并由此去获得一种不仅是对心理世界的本质的,而且也是对物理世界的本质的更好的理解。这里涉及的是这样一些问题:每一种心理属性是与某种物理属性相同一的吗?每一种心理属性是在某种特定意义上依赖于物理属性,或某组物理属性的吗?在这里,“心身问题”这一表达式可能再一次被人误解;我觉得,使用“心理物理学问题”这一表述似乎更好一些。

当代涉及到心身问题的大多数讨论关心的主要就是心理物理学问题。

2. 描述心理学

通常被称为“心灵哲学”的这门学问关心的主要是对人的心理状态与属性的描述,也就是说像处在这些状态之中和具有这些属性的人体验到它们时那样描述它们。这样做的一个目的就是解决上述某些心理状态和属性所引起的哲学上的困惑,以便由此更好地理解作为这些属性和状态的主体的人。这种类型的心灵哲学不仅包括布伦塔诺的描述性心理学,而且也包括胡塞尔以及后来的存在主义者的“现象学”,同样,这类研究也存在于维特根斯坦的《哲学研究》中。

某些自称为“现象学者”的哲学家曾经说过,他们的描述的目的就是要把握这些现象的本质——比如说像悲伤这种现象的本质。一个独立的个体事物的本质就是一组由该事物而且仅仅是由

该事物所必然地具有的属性,其他任何事物都不可能具有。既然悲伤是一种属性而不是一个独立的个体事物,那么我们可以说,悲伤这种属性的本质就是悲伤这种属性的属性吗?——是一种必然属于而且仅仅属于悲伤而与其他任何事物都不相容的属性吗?于是可以毫无困难地说,悲伤的本质就是这样一种属性,它必然如此,以致可以表现于一切具有悲伤性的那些事物之中,并且只表现于它们之中。

当哲学家们说他们描述了悲伤的本质时,他们的意思可能是说,他们提供了对悲伤的一种分析。但那分析又是什么呢?

大概是这样的:那些哲学家们关心的是一种熟悉的现象,他们揭示了这一现象的一些显露出来但常被人忽视的本质属性。如果你说“悲伤应当被分析为如此这般和如此那般”,同时如果我理解了你说所说的话,那么有两种情况:要么我已经认识了悲伤;要么我认为你是说,你所认识的悲伤是如此这般和如此那般的。不管是哪一种情况,我都能够独立地分辨出被分析的对象(analysandum)。更一般地说,如果某人声称对一种状态或属性的本质提供了有信息内容的陈述,那么他便预先假定:那一属性是能够独立地予以分辨的。但是某些声称向我们表明了心理的“真正本质”的哲学家并未预先假定这一点。

这样的哲学家会说:描述性心理学家所采用的那些材料本身是值得怀疑的,完全可以与“民众心理学”的无知老妇式的无稽之谈相媲美。那些哲学家关注的分明是:(1)如果心理学在物理学和化学被说成是一门科学的那种意义上也是一门科学,不过,(2)如果心理学是要研究心理状态和心理属性,那么心理状态和属性究

8 第一部分

竟会是什么。如此一来,他们关注的便是这样的问题,即如果心理学是一门自然科学,那么什么样的事情可以恰如其分地称之为“相信”、“情感”、“感觉”等?有时还有人主张,那可能被称之为“心理的”或“精神的”东西好像与我们事实上很熟悉的那些心理状态与属性非常不同。假如是这样的话,对于我们亲知的那些心理属性的本质,那些哲学家们就不可能有什么理论。因为他们声称,他们根本就不相信存在着这样一些属性。因此既然我们不清楚他们采用的材料是什么,或者也不清楚他们对自己想说的话究竟肯定或否定什么,那么就很难对他们的陈述做出评价。

3. 心理物理学问题

《科学世界观》*的作者们和那些追随他们的物理主义传统的人,希望证明心理状态和属性与物理状态和属性的同一。但是与什么样的物理状态和属性相同一呢?这个纲领相对于感觉来说似乎是有前途的,因为在感觉中,据说存在着每一类感觉所独有的物理关联物。这样一来,有人就认为可能提出一个一般的假说:每一感觉都同一于其物理关联物。但事实上很难找到与各种有意向性的现象(如相信、希望、疑问、愿望等)相对应的关联物。例如想要占有一个 23 英尺的小帆船,我们能找到与此愿望相同一的什么物理状态呢?似乎没有独一无二的、与这种愿望相关联的、可辨认的

* 即 维也纳学派出版的一本宣言式的小册子:《科学世界观——维也纳学派》。——译者

物理状态。

所以,人们终于认识到:这样一些有意向性的现象可能同一于行为的倾向,但不能同一于物理状态或属性。这种观点的一种粗糙的表现形式就是:你想要拥有一个 23 英尺的小帆船,当且仅当问你是否想要拥有一个 23 英寸的小帆船而你做出肯定的回答。但此双重条件句需要加以限定。例如,如果你并不想要这样一个帆船,而是企图欺骗某人,让他认为你想要一个帆船,那么这一条件句就是假的。而如果你的确想要这样一个帆船却又不明白那位向你提问的人的语言,那么这一条件句也是假的。我们能为这个双重条件句找到恰如其分的限定条件吗?这里的困难在于,我们为了处理这样的实例而必须施加的限定条件又将涉及其他的有意向性的现象,例如涉及你想望或不想望的其他事情,涉及你相信或不相信的什么东西,涉及你感知到的什么东西,和涉及你记忆的什么东西。既然有关的行为倾向只能参照其他心理现象才能予以描述,那么就不能说这种理论将心理现象还原为行为的倾向。

能否不诉诸于心理物理的相关论(correlation thesis)而为心理物理的同一论作一点辩护呢?诚然,人们不能说,一件事在一定时刻可以与某件事相同一,而在另一时刻又与另一件事相同一。但一直有这样的主张:任何特定的心理属性的独特的显现(标记)可以是这样的,其中一些显现与一种物理属性的独特显现相同一,而其他的则与某种其他物理属性的独特显现相同一。但有什么根据来支持这种形而上学的论点呢,照此论点看来,存在着诸属性的这样一些独特的显现(标记)?就我看来,上述论点毫无根据,除了能为我们提供心理物理学问题这一可能性以外,别无其他。为了维

10 第一部分

护心理物理同一论而接受这种形而上学学说的形而上学代价似乎是太高了,尤其是当我们认为这种理论的“标记-标记”形式是最卓越的时候,这就有了一种既不能证实又不能证伪的理论。

事实上,某些“功能主义者”曾说过,即使心理物理的东西不可能零星地还原为物理的东西,但整个地还原还是可能的。这个意思主要是说,不是以某些确定的方式将特定的心理状态(如想要一个 23 英尺的小帆船)与神经病学的状态或行为的倾向相同一,而是将一个人任何时候的全部心理状态与这个人在那个时候的全部物理状态相同一。这种理论不再带有《科学世界观》一书的精神实质,因为它像已考察过的“标记-标记”理论一样,似乎是那既不能证实又不能证伪的理论的典型例证。

心理与物理的偶然等同论*

D.M. 阿姆斯特朗

导 言

人是什么？可以明确地说出的一点是：他是某种类型的物质对象。人的躯体可以以远较其他已知的任何物体——自然的和人工的——更复杂、更奇特的方式发生作用。但它又是一种物质对象。这样一来，问题便出现了：“人只不过是物质躯体？我们能用纯粹物理学的术语给予人以完全的解释吗？”

过去，对于给予人以纯物理学的解释似乎有两种重要的责难。其一，人具有一种只为他与动植物所共有的、而一般的物质对象则没有的性质，那就是：他是有生命的。但生命仅仅只是一种纯物理的性质吗？其二，人具有一种只为他与许多动物所共有的、而物理世界的其他东西所没有的性质，即他有心灵。他能知觉，有情感，能思维，有目的。心性仅仅只能是一种纯物理的性质吗？

日益增长着的科学知识已给予第一个责难以部分的答复。现在即使不能肯定生命是一种纯物理—化学现象，但它非常有可能

* 译自作者的专著：《唯物主义的心灵理论》（1968，伦敦，劳特利奇与基根·保罗出版公司）中的“导言”与第17章。标题为译者所加。

12 第一部分

是这种现象。为了解释生命,我们已没有必要假定“植物灵魂”或“活生生的隐德莱希”了。第二个责难又怎样呢?越来越多的心理学家和神经生理学家已明确地或含蓄地接受了这样的观点,即就心理过程而言,在人的中枢神经系统中除了纯物理过程以外,完全没有必要再假设存在其他任何东西。如果我们让“心灵”一词意指“心理过程由以发生的東西”或“有心理状态的東西”,那么我们就能够这样简明地表述这个观点而又不致引起误解:心灵不过是大脑而已。如果科学的进一步发展支持这一观点,那么人似乎只是一仅具有物理性质的物质对象。

然而大多数哲学家则坚信:拒绝物理主义者关于心灵的上述理论是绝对有理由的。本书的写作正是由此所促发的。它由一位哲学家所写,并主要是为哲学家们而写的。它的目的是想表明:没有任何真正的哲学理由去否定心理过程就是中枢神经系统中的纯物理过程,这意思也就是说:没有真正的哲学理由去否定人仅仅只是一物质对象。

本书并不企图证明这一关于心灵的物理主义命题的真理性。如果证据确有其来源的话,那么它必定是来自于科学,特别是神经生理学。本书力图说明的就是:没有什么哲学的和逻辑的理由能否定心与脑的等同论。像洛克一样,我将从否定的方面设想我的任务。我是一名清除道路上的垃圾的清洁工,我设想或猜想,科学就是沿着这条道路大步向前的。

基于这一理由,此项工作对于心理学的重要性(如果有的话)主要是间接的。对洛克的隐喻稍作修改,我所做的只是设法保护那些公开地或隐蔽地把心与脑等同起来的心理学家,以使他們免

遭骚扰性的袭击,也许他们中的一些人还未严肃地对待这种骚扰。如果我所说的有助于心理学家们处理那些在思考我们的日常心理概念时所碰到的逻辑难题,那么我将非常荣幸。我所设想的,除了对心理学的这项工作有特别的重要性之外,这还将是一种理智的赠品。

本书前五章考察和批评了我视之为关于心物等同论的一些重要的、可供选择的方案。我把环绕这些方案的困难看作是迫使我相信我自己的理论没错的最重要的理由。当然,某些读者在一本已够长的书籍中略去这些章节不看也是可以的。即使后面行文中所涉及到的一些参考书目附在前面的章节中,但如果读者从第六章开始阅读,也应当不会有什么太大的妨碍。

我也应提请读者注意:我在前五章所论述的某些内容是一些铺垫性的东西。这样做是因为:如果第六章略述的观点是正确的,那么关于心与身的某些古老的观点和论证便将以新的面目表现出来。

心灵的本质:心理的东西 同一于物理的东西

这最后一章将是很简要的。理由有多方面。首先,在第一章中,对心理的东西与物理的东西的等同已给予了大量的论证。第二,许多作者已为我们论证中的经验步骤作了充分的辩护,而且比我所能作的要精彩得多。(特别值得一提的是J.J.C. 斯马特的《哲学和科学实在论》,劳特利奇,1963。)对于他们所述的,我几乎

14 第一部分

没有什么可补充的。因此我觉得更适合我做的是:竭尽全力去完成他们似乎不那么注意的任务,即阐释心灵概念以及种种特殊的心理概念。第三,如果第二部分中的论证是成功的,对心灵的解释是正确的,那么主张心理的东西与物理的东西的偶然(contingent)等同似乎既不特别地荒谬,又不特别地冒失。

不过,我又让本章享有本书最后一部分的显要地位。其理由是:尽最大可能地重申这最后一步在论证上对于前面两部分的逻辑上的独立性。在第二部分中,除了偶尔的提及,我所论述的并不能推出唯物主义心灵理论的真理。对于第二部分,充其量只能说:它为等同论铺平了道路。

现在我可以提出这样的问题:对于把心理状态与中枢神经系统的物理-化学状态相等同的观点,人们会提出什么责难呢?如前所述,心理状态就是人的状态,就是一种多少有点复杂的物种的状态,而这里的人只能根据他与产生心理状态的对象或条件以及构成它们的“表现形式”的身体行为的因果关系而界定。同样,基因也只能从它们与遗传特性的因果关系而予以定义。(遗传特性可以说成是基因的“表现形式”。)有可靠的理论上的(与观察上的相对)科学根据把基因与位于活细胞中央的DNA分子等同起来。假如我们对表示心理现象的概念的解释是正确的,难道就没有同样真实的根据支持我们把心理事件与中枢神经系统的物理-化学状态等同起来吗?

对等同论的反对意见可以区分为两种情况。第一,等同论之所以遭到抵制,是因为人们认为:中枢神经系统中发生的物理-化学过程事实上不足以解释人类的所有一切行为。第二,假如

物理-化学过程足以解释人类的行为,但把心灵与大脑等同起来仍可以基于其他理由而加以抵制。第一种责难似乎是需要理智地、认真地予以对待的,第二类反对意见理智地说则没有什么意义。作出上述区分的理由在下面再予陈述。

在考察第一类反对意见时,又宜于作出进一步的区分,即把它们分为两类,第一类是从每个人都承认存在的心灵表现形式中推论出来的,另一类是从那些存在尚有争议的表现形式中推论出来的。第一类表现形式包括这样一些东西,如“知觉到在我们面前有一棵树”、“把5与7相加”以及“写作莎士比亚的剧本”、“作出牛顿的发现”等。第二类是那类异乎寻常的表现形式。它包括心灵感应和超常视觉这类所谓的事实。我们先来考察前一类或关于心灵的无可争辩的表现形式。

我们正在科学和经验事实的王国里而不是在那适宜于概念分析的逻辑可能性王国里驰骋。因而必须没有疑义地承认:中枢神经系统中发生的物理-化学过程是那些决定人和更高级的动物的行为的因素中的一种。现在值得探讨的是:这些过程作为原因是否完全决定了行为。

如果我们注意一下人的已知的活动,那么十分明显,像理智的发明和艺术的创造这类事情特别是牛顿的科学发现或莎士比亚的剧本这类超验的事实就是一些给当今唯物主义者以最大困难的事实。因为理智的发现和艺术的创造在一种完全客观的意义上可以说是人的“更高级”的活动,较之于人类的其他活动,它们具有更大的复杂性和奥秘。因而特别难以弄清楚的是:它们怎么可能是纯物理结构的产物。

16 第一部分

对此,反唯物主义者可能采取较强烈或较温和的论战的架势。首先他可能会坚持说:任何物理结构从经验上来看都不可能产生这样的表现形式。其次,他可能退一步承认物理结构有产生这类现象的可能性,但又认为:人的身体事实上并不含有完成这一任务的结构。

第一个论点由于能重复越来越多的人类行为的机器的设计和制造而逐渐丧失其基础。毫无疑问,迄今为止还没有制造出能显示独造性或创造性的、并以之发挥其作用的机器。例如,现在的机器可以解决数学问题,不过只有当这些问题成为那种能被一组循序渐进的程序回答的问题时才能如此。但是说这体现了机器行为的最大作用,说设计出一种可以表现出独创性的机器的设想已经存在,则又是非常不可信的。(试比较:斯马特:《哲学和科学实在论》,第六章,第三节,“解决问题的独创性”。)如果“独创性屏障”被突破,那么主张心灵具有机器所不能产生的其他表现形式则仍然是非常武断的。

然而,一个反唯物主义者一方面承认模拟人的全部心理活动的表现形式的机器在物理上是可能的,而另一方面又仍可能坚持认为:心理状态不是中枢神经系统的物理状态,因为在人类躯体中事实上不可能找到有这种能力的结构。不过这种辩护方式在实践中是很少见的。一旦反唯物主义者设想存在着能产生某种类似心灵的表现形式的结构的物理可能性,一般来说他便承认这种物理结构或某种相差不太远的东西在中枢神经系统中起着作用。(其实,产生类似心灵的表现形式的结构的构想或设计已被证明是关于中枢神经系统的工作的富有成效的假说的最好源泉之一。)

不过,有的人也许会论证说:反唯物主义者之所以让自己如此轻易地得到满足,仅仅是因为我们关于中枢神经系统的知识仍是微不足道的。一种关于大脑作用的真正周详的知识可能会表明:它尚不足以说明引力定律之发现或《李尔王》之创作。

人们可以随心所欲地承认可能性,也可以无拘无束地承认可能性不过是一种赤裸裸的逻辑可能性。但有什么理由认为它就是一种有很高概率的可能性呢?对此,我看不出有什么理由。

在关于心灵的纯物理-化学的唯物主义和关于心灵的属性理论之间出现了一种有趣的调和的可能性。可能有这样的论证:人的全部行为从因果关系上说根源于他的中枢神经系统中的物理过程,但也可以说:那些物理过程中至少有一些是那种不能根据物理学和化学规律予以解释的东西。这也就是说,人们可以认为:中枢神经系统中的某些过程是按突现的规律而起作用的。这种规律不可能是,甚至根本就不是从物理学和化学的规律中推导出来的。由此规律所决定,行为发生了,它不可能是某种按纯物理化学原则而起作用的东西产生的。这种观点仍然是一种唯物主义,因为它不需要任何突现的性质,更不需要突现的实体,但它不是一种物理-化学的唯物主义。

此处再自然不过的假说就是:那些突现的规律是在已具有一定复杂程度的相互关联的全部物理系统中显现出来的某种东西。以这种方式,人们可以避免把突现规律与只有生物系统中才有的更高程度的复杂性联系起来的任意性。人们可能会寄希望于具有必需复杂程序的人工机器,它完全像中枢神经系统那样按突现的规律发挥作用。

18 第一部分

再则,如果人们真地接受了这类假说,那么就更有必要支持关于一般的突现规律的观念。人们也许有望看到,包含有比一般物理系统的组成部分更复杂的相互关系的一切生物系统按照超越了物理学和化学原则的规律运行。如果物理-化学规律适用于物质的一切普通集合,并正好存在着这样一点,向突现规律的跳跃恰恰就发生在此处,心灵的突现也在这点一同出现,那么上述假说就将是科学规律体系的一种令人惊异的任意的特征。

如前所述,我深知没有什么使人非相信不可的科学理由促使我们认为:哪怕是牛顿的发现或莎士比亚的作品也要求我们假定突现规律在中枢神经系统中起着作用。不过假如这类理由存在,或后来会被发现,那么求救于这种“突现规律”的唯物主义便将是自然的事情,从本体论上来说也是相当经济的步骤。

(这里可以附带地说一下,随便哪种类型的平行论者都一定同意说:大脑的活动本身足以解释人类的全部行为,虽然他对大脑活动所遵循的规律坚持的也许是一种突现论者的观点。因为按照平行论者的观点,心理状态在行为的因果关系中没有任何作用。如果人们想说:大脑不足以从因果关系上解释人类的全部行为,那么他就不得不做一位关于心灵的二元论者和交感论者。)

对于“突现规律”的唯物主义,可能有这样的责难:它与交感论者的二元论可能没有什么区别。假如中枢神经系统中的一个物理事件在短暂的间隙之后为另一物理事件所伴随,但第二个物理事件的发生无法用物理-化学术语予以解释。为了解释这种现象,“突现规律”的唯物主义者便假定了物质的突现规律。但一位二元论者则可能这样解释所发生的事情,即说第一个物理事件在精神

实体中引起了变化,而精神实体反过来又引起了第二个物理事件。在这两个假说之间我们应怎样作出抉择呢?

在答复上述责难时,必须指出的是:二元论的观点可能以两种不同的方式予以阐发。如果按第一种方式,那么在二元论和“突现规律”的唯物主义之间就不可能有任何实验的判决。经济的考虑无论如何会赞成后一观点。如果按第二种方式亦即似乎更有理的方式加以阐发,那么便有可能从经验上对问题作出裁决。

假设中枢神经系统处于一确定的物理状态,此系统中的一物理事件引起了一精神事件,此事件又引起了一物理事件。在这种情况下,精神实体的因果有效性是唯一由第一个物理事件决定的吗?首先假定是这样,那么就存在可被发现的规律,假如中枢神经系统处于通常的物理状态,那些规律原则上允许我们作出这样的预言:第二个物理事件来自于第一个物理事件。假如是这样,我们怎么可能确定一精神事件在因果链中所起的作用呢?我们怎么可能排除这样的观点,它认为第一个物理事件即使遵循了突现规律,但却直接地引起了第二个物理事件呢?绝对不能。在理论之间没有可观察到的差异,正像它们所表现的那样。然而每一本体论上的经济考虑都不利于二元论者的看法。

这样设想似乎更合理一些,即假设在精神实体中所发生的事情和第二个物理事件的本质不仅是由第一个物理事件所决定的,而且也是由过去发生在精神实体中的东西所决定的。因此根据第一个物理事件不可能作出关于第二个物理事件的预言。这种预言的不可能性就会成为交感论者的二元论证明其真理的证据。自然,假如精神实体的发生和随后的历史完全是由中枢神经系统中

20 第一部分

的物理因素决定的,那么仍有可能作出这样的论证:第二个物理事件遵循“突现”规律直接从过去与其他物理事件连结在一起的第一个物理事件中派生出来,而不需要一精神实体作为其中介。但是这一假说变得如此复杂难解,以致假定与大脑相互作用的精神实体显然就成为一更好的假说。因此突现规律的唯物主义与交感论者的二元论便泾渭分明了。

面对那似乎与唯物主义者关于心灵的观点相矛盾而其实更有利于他的证据,他也有另一条有效的退却路线。可以这样设想:人类的全部行为可以解释为大脑作用的一种结果,不过那种作用本身还不能用我们现有的物理原则去解释。当然也许能用重新改铸过的物理学解释。即使不承认突现规律的存在,我们仍有可能为物理学找到一新的基础,据此,大脑运作的明显的特殊方式就变成了在原则上可从物理学规律中引申出来的东西,它们适用于大脑的特定物理结构。

如果为物理学所提供的新的基本原则在最低限度上能够解释和预言一般的(即是说非心理的)现象,至少是那些通常被承认的现象,此外,如果它们能够预言中枢神经系统的异常行为,那么为了一种统一的解释图式起见,我们就可能转向新物理学。当一个人没有碰到那个实际上正在产生困难工作的困难工作时,自然容易畅快地谈论新物理学。但这正好是唯物主义者可能以此为他的二元论观点辩护的方式。

不过,可以设想:不管是根据中枢神经系统运作的突现规律还是通过改造我们的物理学去解释人类的某些活动——大概是“更高级的”活动——这都是没有可能性的。假如是那样的话,中枢状

态的唯物主义就将是虚妄的,即使本书第二部分所提供的对心理概念的分析——因果关系分析——依然是真实的。

关于人们承认已得到证明的心灵的那些表现形式就说这么多。现在我们转入对“特异”现象的考察。它们主要包括:心灵感应、超常视觉、预知、心灵致动(心灵直接作用于物质对象)和离体的存在。

我们这里所碰到的第一个大问题就是:这些现象是否是真实存在的。如果不存在,它们便不妨碍中枢状态唯物主义,因为它们纯逻辑的可能性不是反对心与脑的偶然等同的理由。现在,所有严肃而周密地研究过这一课题的人都不相信发生过这类现象。(M. 斯克里芬的“大脑的新领域”一文,载《超心理学杂志》,第 25 卷,1961,pp.305-318,给予物理学家以极大的安慰。)

在考察这一问题时,似乎有理由集中在下述问题上:是否能找到在实验上重新产生的这类现象。人们公认这是一条严格的标准,不过既然它几乎是对那类现象的科学探究的一种逻辑上的前提条件,那么它也是方法论上似乎可以正确地确立的前提条件。

有启发性的实验材料的数量目前还不太多,但有少数的结果无论如何是难以解释的,除非承认某种特异功能正在发生作用。特别是存在着某种关于心灵感应和(或)超常视觉以及(或)先知先觉的证据。(我之所以以这种模棱两可的方式阐述我的结论,是因为那类证据一般都容许作出不同的阐释。这里不妨作一非常简要的说明。假如一被试有这样的预知的知识,即一副纸牌中什么牌会两次向上翻。如果我们涉及到的是一副在两个轮次之间没有洗过的牌,那么只要假定有对于那张上面为 2 的纸牌的超常视觉知

22 第一部分

识,这个事实便能得到同样圆满的解释。倘若有这种超常视觉知识,那么一种完全普通的推理就足以产生那显然是预知性的知识。)一当承认实验材料也提供了肯定的结果,那么许多没有在实验条件下发生的所谓现象就会变成特异现象的似乎合理的“候选人”。

于是便可假定:事实上存在着特异现象。由之而起的问题便是:它们与中枢状态的唯物主义是否处于必然的矛盾对立状态。就拿心灵感应来说,我们可以把心灵感应定义为获得关于另一心灵中所发生的事情的非推论的知识。要不就采用这样的、与实验事实非常一致的定义,即心灵感应就是在没有任何根据的情况下对另一心灵中当下正发生的事情作出的猜测,这些猜测必然比那些可归之于运气的结果更成功。如果此类现象存在,那么它就是“特异的”,因为发现他人心中正发生什么的正常方式是根据他的身体的所作所为去作出推论。但是这类现象为什么与中枢状态的唯物主义相矛盾呢?难道就不可能有这样的现象,它们是尚未发现的物理过程(在别的情况下,也许是一类非常熟悉的过程),正是这些过程将一个人与另一个人的中枢神经系统联系起来以致让信息相互传递?对于其他的超常现象也可作出同样的猜想。

这类猜想最终可能会驱散由那些现象所引起的问题,而无论如何又不至于妨碍中枢状态的唯物主义。不过如果我们注意到那些所谓的现象的特殊本质,如果它们实际上发生了,那么很难在我们所知的物理学框架内为其找到解释。例如有这样一个原本就不是没有前途的设想,即心灵感应的信息交流是以某种为一个中枢神经系统所发射的、而又为另一中枢神经系统所接受的物理辐射

为中介的。这一假说可以通过做心灵感应实验而予以验证,在实验中可用防辐射外壳将被试相互隔离开。这种实验在苏联实际上已经做出来了。据称,在这种条件下所作出的关于另一人心中发生的事件的猜测仍比凭运气所作的预期要好得多。(试比较:L.L. 瓦西利夫:《心理暗示实验》,英译本:心像研究协会,1963。)如果实验能够重复,那么用辐射对心灵感应所作的解释似乎是非常令人难以置信的。但是基于小心谨慎地设计的现代实验,能设想在被试之间存在着其他什么样的物理联系的通道呢?我们能说的是;有关的辐射极不像是为“防辐射”罩所阻挡的那种物理辐射。但这一来,我们便开始背离已知的物理学体系。

如果预知是一事实,那么更难予以解释。诚然,在理解对于未来的非推论知识的抽象可能性时,的确没有什么特殊的困难。假设一系统S正向某一方向运动。让我们假定它可能不会碰到什么障碍,如果是这样,它最终会进到状态 S_1 。不难想象,在系统进入 S_1 状态之前,它的小部分能量分离出去了,作用于心灵,在那个心灵中产生了这样的非推理知识,即S在 S_1 状态中将逗留很短的时间。任何警告系统诸如卡车上的油灯,都是根据同样的原则而起作用的,即使它产生的是不太复杂的结果:灯的打开,表示没有非推理的知识。肯定存在这样的可能性,即我们有时不基于任何根据而只是由于那种给予我们关于我们未来躯体状况的非推理知识的机制的作用,便得到了像“我会生病”这类信息。

但是如果我们考虑到实际的事例,据说预知能力就存在于其中,那么就很难运用上述解释。例如假定一被试以比运气更好的结果猜测到什么纸牌会翻开,而在每次猜测之间,那副纸牌均被随

24 第一部分

机的程序彻底地洗过了。使一特定纸牌翻开的物理原因多不胜数,而且它们具有非常复杂的相互关系。于是弄清“信息”怎么可能从相互关联的原因中传送出来,几乎变成了不可能的事情。因为那些原因在任何程度上与原因将要使其变成现实的结果是相互关联着的。因此这类预知如果发生了,那么它似乎向各种公认的解释模式提出了挑战。

不错,这还不是争论的终结。在讨论正常的现象与作为其对立面的异常现象的联系时,唯物主义者有可能仍采取两种回避方式中的一种。他会设法杜撰出一种突现的唯物主义,它包含有适合于中枢神经系统的、甚至原则上不是从普通物理学中推演出来的特殊规律。或者他可能寻求一种新物理学,它内在地包含有一套统一的原则,不仅可以解释、预言物质和人的通常能力,而且也能解释、预言人的超常技能。但是如果这些逃避方式证明是不尽人意的,那么中枢状态的唯物主义就不可能是关于心灵的全部真理。

我认为:心灵研究*的种种论断是笼罩在唯物主义心灵理论地平线上的一团小小的乌云。如果对特异现象的考察没有什么疑义,那么对于心灵状态与中枢神经系统的物理-化学状态的完全等同论来说也几乎不存在什么了不起的障碍。(这自然承认了遭到逻辑上的反驳的可能性。)我们所说的等同大概不像基因与DNA分子的等同那样确定。明显地存在的特异现象必定会为人们留下一点存疑的空间。拥护任何科学原理的人都有一种理智的

* 超心理学(parapsychology)或特异现象学(psiology)的别名。

责任,那就是应一丝不苟地考察那些似乎最有可能瓦解他的观点的基础的根据。因而唯物主义者也有—种理智的责任,即应小心谨慎地考察心灵研究所提供的所谓结果。

最后,我要简要地考察—下这样—些人的观点,他们—方面承认:中枢神经系统中的物理—化学作用是人类—切行为的充分原因,但另—方面又抵制心灵状态与大脑状态的等同论。这就是我在本章开头视之为理智上无意义的观点。

这里有必要再次强调—下:在本章中,我假定第二章对心灵概念的说明是具有普遍的真理性的。因之我可以否定那些人的观点,例如他们论证说:可以有根有据地反省到心理状态是某种不同于人的状态的东西,或者说是某种多于人的状态的东西,因为人的状态适合于产生某些类型的行为。如果人们接受根据心灵状态与行为的因果关系对心灵状态所作的说明,同时人们又赞同说大脑的物理作用足以引起人类的一切行为,可是又想极力诋毁心理状态与大脑状态的等同论,那么人们能采取的似乎只有—个观点。人们必定会说,大脑中的物理过程引起非物质类型的心理过程,它反过来又引起行为。必须在某—点上把心理过程镶入因果链之中,即使该链实际上是伸展的,就像没有这种镶入,它也会伸展—样。(试比较“突现规律”的唯物主义和相互作用论者的二元论之间的争论,在那里,相互作用都以这样的方式发生,以致两种理论实际上作出了相同的预言。)

这—种逻辑的可能性。它在逻辑上可与已观察到的事实和—谐共存。但是科学上的每—具有简单性的原则都反对采取这—立场。为了说明这—点,有必要考察另—种逻辑上的可能性。也许

26 第一部分

事实是这样的:在它发生作用过程中的某一点上,DNA 分子使一非物质的元素得以形成。此非物质元素反过来又有进一步的作用:传递遗传特性。整个因果链实实在在地起作用,好像就是一物理链一样,而事实上它包含有一非物质的环节。人们能清楚地看到:这一假说即使是一逻辑上可能的假说,并在逻辑上与已观察的事实相容,但它并没有什么可取之处。对于关于心灵的平行假说也可以作出同样的说明。

至此,我们的冗长的论述已趋近尾声。那些必须予以解决以期形成科学世界观的重大问题中有这样一个问题,即怎么可能让人在 worldview 内提出世界观的问题。把人、包括他的心理过程作为一纯物理的对象,以为他像所有其他自然事物一样遵循着实际上相同的规律,这个目标因有最大可能的、理智上的经济性,因而实现了。认知者仅仅是由于他的物理器官具有更高的复杂性才有别于所认知的世界。人是拥有自然的事物。

然而我们必须意识到:即使本书所提出的心灵学说是正确的,但物理主义者的哲学并不是它的问题的终结,而毋宁说只是它的问题的开始。排除心灵问题只会使我们碰到更深层的、与物质相关联的问题。在我们仅用实体、原因、规律、时间、空间之类的概念去说明心灵的空间和时间上的现象时,这些概念依然像过去一样含糊不清。物理主义的心灵理论不过是物理主义形而上学的导言而已。这种形而上学正如同心灵理论一样无疑是科学探讨和哲学反思相结合的产物。

心理事件与大脑*

保罗·K. 费耶阿本德

谢弗的评述^①及其所提及的先前的讨论都再清楚不过地表明:任何关于心理事件与大脑过程的同一假说都包含有二难推理。这些假说通常是由那些还想成为经验主义者的、具有生理学倾向的思想家所阐发的。由于有生理学倾向,他们试图断言:心理过程具有物质特征。而作为经验主义者,他们又想使他们的断言成为一种关于心理过程的可检验的陈述。他们试图在下述经验陈述中把两种倾向结合起来:

x 是类别 A 的一种心理过程 $\equiv x$ 是类别 α 的一种中枢过程
(H)

但是这种假说产生了适得其反的结果。它不仅如其所愿地意味着:心理事件具有物理特征;而且它似乎还暗含有这样的意思(如果从右看到左):某些物理事件例如中枢过程具有非物理的特征。因而它以关于特征的二元论取代了关于事件的二元论。而且这个结论似乎是根源于生理学家用以系统阐述其论点的方法。即使他是一个公认的一元论者,他也会因其一元论论点的那个内容

* 该文最初发表于《哲学杂志》,60(1963):295-296,后收入 W. 利康编:《心灵与认知文选》,1990,巴兹尔·布莱克韦尔。中文依据后者译出。

28 第一部分

而被迫承认一种二元论观点的正确性。

对于一个二元论者来说,这种尴尬的困境足以证明一元论是站不住脚的。但是他在得出这个结论时的确太草率了! H包含着二元论。因此假如 H 是真的,那么二元论也是真的。然而如果一元论是正确的,那么 H 就是虚妄的:这样一来,在通常的(非唯物主义的)意义上,根本就不存在心理过程。这说明,对被当作一种经验假说的 H 的内容的讨论丝毫无助于对二元论与一元论之间的争端的解决。它还说明:一元论者在为 H 辩护时错误地陈述了自己的理由。

他应采取的正确的程序是:不求助于任何现存的术语去阐发他的理论。如果他真的想利用 H(即是说,如果他想继续使用古老的术语),那么他就应当用它去重新定义心理过程。他的理论的经验特征并不会因此而陷入危机。关于癫痫病的生理学理论毕竟不会成为一种空洞的同义反复,因为事实是,它并没有使用“魔鬼附体”这类短语或概念,在这里“魔鬼”是神学意义上的。存在着独立的、有效的预言,事实上,除了心灵主义者所能提供的、甚或打算提供的预言之外,还存在着许许多多的预言(只要想一想知觉生理学的广博的领域就清楚了)。

然而它常常遭到驳难,除非与先前的语言联系起来,否则我们就不知道我们正在谈论什么,进而我们也不能清楚表述我们的观察结果。这种驳难假定的是,一般观点的语词和对应语言的语词只有通过相关于某种别的众所周知的、习以为常的观点的语词才能获得其意义。如果确实如此,那么后一种观点和后一种语言怎样变成众所周知的呢? 如果它像它过去所是的那样,不借助于“外

在的力量”，而成为人们所熟知的，那么就没有理由假定：一种不同的观点不能同样为人所熟知。（此外，在我们很小的时候，我们就学会了日常的惯用语；难道能断言：成人生理学家不能完成小孩能很好完成的事情吗？）而且观察结果永远必须相对于某种理论背景（用时髦的术语说是相对于某种语言游戏）而予以阐述。没有任何理由说，生理学本身为什么就不能形成这样的背景。因而，我们不得不得出结论说：对人类的纯生理学探讨的合理性以及成功完全不依赖于对 H 分析的结果。

像 H 这类“桥梁规律”在流行的解释和还原理论中起着最重要的作用。如果我们的上述评论是正确的，那么就可得出结论说：这些理论是不够格成为理论建构之成功的标准的。

注 释

- ① 谢弗：“心理事件与大脑”，《哲学杂志》，60（1963），160—166。

自然层次的连续性*

威廉·G. 利康

心灵哲学中的当代功能主义发端于对作用和占用者的区分。就我们所知,对人(或他们的大脑)与计算机的饶有兴味的比较把我们的注意力引导到下述对比之上,即机器程序(抽象地看待的)与那机器碰巧由以从物理上构成的、实现那程序的特殊材料之间的对比。正是前者而不是后者,才使我们有可能面对面地去说明、解释、预言和利用机器的“行为”;人们制造计算机以执行程序,利用一切物理材料以最大限度地服务于那个目的。

当普特南和福多揭露同一理论的沙文主义要害时,“程序”与“实现材料”或更熟悉的“软件与硬件”之间的区分便幸运地帮了心灵哲学的大忙。“C-纤维”等之类所做的事情可能由某种物理学上不同的结构来完成——这种作用可能由那种结构来完成。可以肯定的是,如果同一种作用为有神经化学基础的硅片而非有神经化学基础的碳所完成,同一种功能如此被实现,或者说,如果我们的个别神经元逐渐地为完成了相同工作的电子假体所替代,那么从直观上说,我们的智力仍不会受什么影响。至关重要的是

* 译自 W. 利康编:《心灵与认知文选》(巴兹尔·布莱克韦尔,1990)。原为作者的《意识》(麻省理工学院出版社,1987)一书的第4、5章的部分内容。

功能,而非工作人员;是程序而非实现材料;是软件而非硬件;是作用而非占用者。因此功能主义,以及有机体的“功能的”与“结构的”状态或属性之间的区分便应运而生了。

在我准备(即使未获特许)诛戮的全部哲学中,功能主义是唯一富有建设性的学说。^①我知道,“作用”与“占用者”的区分(某些人津津乐道地说)对形而上学来说是必不可少的。但我又坚持认为:近来心灵哲学对这一区分的补充既是错误的又是有害的。在本章中,我的目的就是批评“软件”与“硬件”、“功能与结构”之间的以通常的哲学形式表现出来的二分法,进而指出某些根本性的混乱,纠正从它们中派生出的某些错误。

等级系统

总的来说,我的反对意见是:“软件”与“硬件”的说法有利于大自然由两部分构成的观念,该观念把它分成两个层次,大致是物理化学的和随附性的“功能的”或高级—组织的层次——这是违反事实的,因为大自然是一个由自然层次构成的多样性的等级系统,在那连续体上,每一层次具有一系列像自然法则一样的普遍原则,并随附于它之下的所有的层级。^②以这种方式把大自然看作是从等级上组织起来的東西,“功能”与“结构”的区分就变成了相对的:某物只有对模被指定的自然层次来说才是一种对立于占用者的作用,是一种对立于实现者的功能状态,反之亦然。下面我将予以阐述。

生理学和微观生理学中的例子俯拾即是:细胞——一般认为

32 第一部分

是一个非常明显的功能术语(!)——由包含着膜、核、线粒体等更小项目的合成成员所构成:这些项目本身又是由更小的、合成的要素所构成的系统。就此而言,自然的更低级的层次也是数不胜数的、有显著差别的:化学的、分子的、原子的、(传统的)亚原子的、微观物理的。层次是由有趣的、类规律的一般规则所连成的系列,是按照有关类型的一般规则而个体化的。细胞沿着该等级系统往上升便形成组织,而组织相结合又构成器官,器官本身又结合为器官系统,这些系统结合在一起——奇迹般地——组成人这样的完整的有机体。甚至有机体本身又结合成为有组织的(组织化的)群体。在我们一般所认识的单个有机体与以显而易见的、真实的方式共同起作用的有机体群体——我们大概可以说“群体有机体”^③——之间,不存在泾渭分明的类别差异。

与这种关于自然等级组织的自下而上的整体图景相应的是众所周知的自上而下的解释战略。^④如果我们要知道废物和有害物质怎样从人体中排出,那么我们就得寻找并发现与消化和循环系统有联系的排泄系统。如果我们直接注意那个系统,那么我们便会发现(用不着大惊小怪):它区别对待水溶性的和不可溶的废物。尤其是我们会发现肾,它特别能对水溶性废物发挥作用。如果我们进一步沿着层次的等级向下深入探讨其细节,那么我们会看到肾分为肾脏皮质(过滤器)和髓质(合成器)。皮质主要由肾单位构成。每个肾单位都有一个小球,它与传入小动脉和控制压力的可收缩的肌肉腕是相通的(这种压力让水和溶解质沿着毛细膜进入鲍曼囊,而让血液细胞和更大的血液蛋白质存留下来)。再吸附之类可以用细胞术语如根据沿着肾单位的长小管排列的上皮细胞的

特殊属性予以解释;这些特殊属性进而又可根据细胞膜的物理化学予以解释。

大脑并不是有机体及其器官的这一等级图景的例外的东西。神经元就是细胞,由包含着核和细胞质的细胞体加上附着于其上的纤维所构成,而那些纤维很显然有独立的功能;甚至我们还知道,像离子泵这样的更小的功能项目,它们在内部维持着钾的高浓度。神经元本身结合为神经网络和其他结构,如柱状构形,这些结构进而又结合在一起形成大脑的更大的、更清晰的功能(即使不是如此明显的模件)部分。听觉系统是很好的例子。有证据表明:听觉皮质表现为双维度柱状组织;^⑤当沿着一条轴线排列的各种特化细胞的柱大致垂直于那些从一个耳朵进入的与另一耳朵的输入并列的输入时,这些柱便有选择地对由听觉神经向内的刺激指示的频率作出反应。特化细胞的这种特殊的敏感性应进一步根据通过了细胞膜及其后面的要素的离子传输来加以解释。因为就其本身而言,听觉皮质与其他高层次构造*——丘脑、上丘和别的皮质区域——是相互作用的,这些相互作用是极有条理性的。

因此一种关于自然的作为整体的本体论和自上而下的认识论便结合在一起了。阿特尼夫(1960)、福多(1968)和丹尼特(见本书)有说服力地论证说,这种结合有利于心理学科学。我将以一定

* agency 以及同源词 agent,在中文中无贴切的对应词。在英文中,相当于 intelligent agent architecture,即智能动力构造,指的是一种信息加工系统模型,其特点是有重要的子系统、功能作用、信息加工以及对它们的控制。现根据它们的意义及上下文,姑且把 agency 译为“构造”,把作名词用的 agent 译为“行动者”(在有的地方,酌情译为“动原”)。

34 第一部分

的篇幅,根据我的 1981a,对这种观点作出阐发。

小人*功能主义

丹尼特(参阅《心灵与认知文选》,pp. 69 – 70)效法的是某些 AI 研究方案的方法论。^⑥

AI 研究者从有意向地被描述的问题(如我怎样让一计算机理解关于英语的问题?)出发,把它分解为同样是有意向地被描述的子问题(如我怎样让那计算机识别问题,把主词与谓词区分开来,搁置无关的语法分析?),然后再对这些子问题进一步分解,直至最后去研究显然是机械论的问题或目标描述。

丹尼特把这种方法论步骤推广到了人类心理学的实例上,而我认为,它主张的不过是:我们应把人看作是协同一致地完成许许多多复杂功能——其中有的种类常被称之为精神的或心理的功能——的合成的实在。赞成福多和丹尼特受 AI 启发的方法论的心理学家会根据下述流程图描绘人,这种流程图描绘了那个人的直接的亚人(sub-personal)构造以及它们的五花八门的、能相互通达的路径,这些路径又使它们协同作用,以完成那个包容结构或有机体的目标,而此有机体正是那个人。原始流程图上由“黑箱”所表征的每个直接的亚人构造进而又可为它自己的流程图所描述,后者把

* homuncular,原指炼金术士在蒸馏瓶中人工制成的小人,转义有“矮人”、“小人”、供解剖用的人工模型。在这里,指人内部的、实现整体功能不可缺少的部分。参照上下文,故译。

它进一步分解为协力完成它的目标的亚-亚人构造,如此等等。根据这种观点,人的心理能力与协作组织的各种执行单元具有相同类型的功能等级结构,并正好具有同样的意义。

以我的这种方式描述心理学家们的探索其实是认为,在材料或现象的层面上,他们优先注意人类主体的从意向方面或别的方面描述的能力,进而假设——作为理论实在的——小人或亚人构造,这些构造是解释主体拥有那些能力所必不可少的。然后,那些心理学家进一步假设更小的小人,以便解释先前假设的原初小人的克分子行为,如此等等。正是阿特尼夫/福多/丹尼特模型的这一特征巧妙地回击了对心理学中的小人理论的无穷倒退的标准的赖尔式责难。^⑦我们对某个小人的成功的活动的解释,不是借助于空洞地假设它之内的、成功地完成那个活动的第二个小人,而是通过假设由几个更小的、分别有点才能的、更专门化的小人所构成的小组——进而详述那个小组的成员由以协作以形成它们的结合或联合输出的方式。

认知和知觉心理学家对各种亚人构造有合理的、美妙的想法,要说明人能完成它常常完成的动作和别的功能,就必须假定:这些构造在人类身上起着功能作用。丹尼特(1978,第9章)曾说过,直接的亚人层面有:“印出”部分或言语中心,^⑧“高级的执行或控制部分”,“短期记忆储存或缓冲记忆”,“知觉分析部分”和“问题解决部分。”丹尼特(1978,第11章)还从某种临床细节上考察了模拟人类表现自己疼痛行为的多级亚人结构。“行为”在这里的含义一定极为丰富,因为他从细节上作了说明,不仅有哲学行为主义者和常识心理学的热心倡导者们津津乐道的常见行为类型,还有许多微

36 第一部分

妙的现象:在我们关于疼痛的现象学描述中的非常微细的差异;很罕见的现象,诸如在我们感觉到我们被烧伤了与我们感觉到烧伤的剧痛之间的被感觉到的时间间隙;(从小人功能主义的观点看,最有趣的是)大量各种不同种类的麻醉药和别的药物对病人生活和关于疼痛的内省报告的各种影响。那些心理学家(以及丹尼特)正好把这些五花八门的考察当作关于中枢神经系统的有关功能组织中的复杂性的真实的线索,认为它们显示的是处在我们必须以我们的有等级性的流动曲线图表征的机构组织的各种层面——如我们必须假设的各种接受器、抑制器、过滤器、减幅机、触发器等——上的不同的黑箱成分,以及那些把这些成分相互关联起来、把它们与它们的所有者如知觉分析器、信息储存和言语中心的更粗糙的功能成分关联起来的、可比的各种各样的路径。

从目的论上来看,小人探讨方案有许多优点。当我对目的论稍作说明之后,我会列举这些优点的。同时,我还会对心理的东西与明显非心理的东西的类型同一的一般形式公开我的想法:我打算让心理状态从类型上同一于这样的属性,即在某人的一种(或更多的)相应的小人功能部分或子构造中得到的、具有如此那般的、从结构上描述的事态的属性。(子构造指这样的东西,它们是用关联于处在各种结构抽象层面上的、它们的所有者的流程图予以详述的。)这种观点也适用于心理事件、过程和属性。我们可以说,处在 T 类疼痛中就是某人的亚……亚-人的 ϕ -者处在一种特定的状态 $S_T(\phi)$ 中,或者说一种特定的活动 $A_T(\phi)$ 在某人的 ϕ -者中发生了。

小人与目的论

可以断言,“ ϕ -者”和“ $S_T(\phi)$ ”这样的描述本身只能用那个有机体的目的论图式含蓄地加以定义,进而对它们的阐述从根本上包含着对该有机体的其他从目的论上被描述的构造和状态的不可排除的参照。这是合理的,但只是相对无害。作为心灵哲学家,我们要做的就是用还原的(且非循环的)方式阐释心理的东西,这就是我正在从事的工作,亦即把心理描述还原为小人结构描述,这种描述是各种功能抽象层面上的目的论描述。我没有必要附带地把结构上的描述还原为“更精密”、更富有结构性的描述;如果说存在着从结构类型比如说向生理学类型的还原,那么根据小人功能主义,同一论就是真实的。结构类型(在任何特定的抽象等级层次上)是不可还原的,即使我自始至终假定结构标记在严格同一的意义上是可还原的,即完全可还原到亚原子水准。

事实上,结构类型的不可还原性对把小人功能主义当作关于心理的东西的哲学理论起了推波助澜的作用。正如 D. 戴维森和 W. 塞拉斯两人所述的那样,关于心灵的充分的理论其首要的任务就是解释心身问题的存在本身;它涉及到这样一些解释:心理的东西为什么似乎如此不同于物理的东西,以至于使笛卡尔主义以天真的形式出现?就是对行家里手来说,为什么系统阐述心理的东西对物理的东西的合理还原已历史地被证明是困难的?我们的心理概念作为一个家族为什么看起来像构成了一个“天衣无缝的整体”,在概念上与生理学的或物理学的家族没有太大的关系?⑨

38 第一部分

小人功能主义提供了这些解释的雏形。心理的东西的显而易见的不可还原性就是结构类型向较少目的论类型的地地道道的不可还原性。^⑩之所以难以勾画心理的东西哪怕是向结构的东西的有根据还原,主要是因为我们不知道抽象的、足够低的层次上的结构本身的组织作用。结构类型向生理学类型的不可还原性并不是屏障,只要我们关于结构范畴的系统、关于生理学范畴的系统、关于物理学范畴的系统正好是相同标记的可供选择的群集。

某些哲学家可能会发现,小人功能主义的还原是非常令人心寒的安慰剂。它肯定会使前此已根据愿望或意图之类的心理项目理解关于事物的目的论描述的人厌烦不已。正如前述讨论所暗示的,我自然不是这样理解目的论的言论的;毋宁说,我认为心理类型形成了目的论类型的一个小小的子类,它们大部分发生于功能抽象的高级层面上。但是如果是这样,那么我是怎样理解目的论的东西的呢?

对于这个一般性的问题,我自己几乎没有什么建树。我希望,同时也倾向于相信:小人功能主义所需要的目的论描述能够用进化论的术语独立地加以解释。这种可能性主要得益于 K. 波普尔、W. 温萨特、L. 赖特、K. 林德尔和其他生物学哲学家的工作。^⑪我不可能改进他们的技术讨论。不过,我想提出一种理论上的观点,然后用一个例子予以证实。

这个理论观点就是:描述的目的性有程度的差别;对一事物的某些描述比其他的描述具有更多的目的性。同一时空份额可以由下述东西所占据,如一组分子集合体,一块坚硬的材料,一具有铰接凸缘的金属条带,转向轮的转动器,一把钥匙,门上的一把锁,进

入旅馆的通行证,私通关系的支持者,灵魂的毁灭者。因此我们不能整齐划一地把我们关于自然的理论区分为各机通用的、纯机械论部分和暧昧的、凌乱的、最好是予以忽略或抛弃的生机论部分。基于这个理由,我们就不能认为:心理的东西向目的论的东西的还原在本体论的探索中不可能有什么进展;高级的目的论描述,不同于原始的、已得到阐释的心理描述,具有逐渐地、平稳地演变成(更)纯粹关于肉体的物理描述的优点。^⑫

我愿对此作出与心理学有关的说明。试考察一种能识别面容的有机体(就拿丹尼特关于可编程的心理能力的一个俏皮的例子来说)。对于有机体怎样发挥其作用这一问题,已有许多观点;那种生物可能是根据许多完全不相似的功能程序表进行建构从而实现它的面部识别的。假设它使用的一个特定程序表是这个样子:只有当它在正前方、右侧面或左侧面得到输入时,它才会接到分辨的命令。执行路径会让观察点探测器去观察知觉显现,进而观察点探测器又会把输入归入三种可能的方位范畴中的一种。接着显现物就会提供给相应的分析器,该分析器可形成作为那显现物的内容编码的输出。管理器则会对照已储存在有机体记忆中的、相似的被编码视觉报告检查这种被编码的公式;如果它发现了匹配,它就会注视加之于匹配编码公式之上的分辨附加语,并把该附加语显示给有机体的公共关系主管,后者又向运动子路径发出音位指令,最后这些子路径使有机体公开而有声地说出一个名字。

如果明白了这就是我们的特定的面部识别器由以履行其职责的方式,那么我们就可以进一步去探寻其细节。我们也许想知道:观察点探测器是怎样工作的(它是一种简单的模板吗?),或者说公

40 第一部分

共关系主管机构是怎样被组织的,或者说分析器用的是什么样的子-组件。假设分析器被发现是由把方格加于视觉显现物之上的投影器和每次通过那方格浏览一正方形、并产生二进制编码数据的扫描器所构成的。我们还可以进一步追问:扫描器怎样工作,然后我们可能被告知:它主要由在方格中记录黑暗程度的曝光表和关于“0”、“1”的报告所构成;我们还可以问:那个曝光表是怎样工作的,然后我们又被告知关于感光化学等方面的某些信息,如此等等。一言以蔽之,在这种沿着结构等级的下降(从认知器到扫描器,再到曝光表,最后到感光实体,只要愿意,人们还可以进一步后溯)中,我们的描述到达哪一点才不再是目的论性质的描述呢?在哪一点上才开始作为纯机械的描述呢?我认为:显然不存在这样的点,只存在着更精致的连续体,是它把抽象的、高级的目的论的东西与砂砾般具体的、仅有勉强的目的论性质东西联系起来。心理的东西之所以看起来是完全不同的东西,之所以与物理化学的东西判然有别,以致在本体论上根本就没有这种东西,这就是其理由。^⑬

我相信赤裸裸地阐述的目的论,对此我最后要说的是,我并不是要主张:赤裸裸地阐述的目的论是好的或有希望的。我本人一点也不喜欢它。我的观点不过是:心理的东西的秘密并不比心脏、肾脏、汽化器或袖珍计算机的秘密大到哪里。作为一个本体论的观点,它是一个非常令人心慰的观点。^⑭

目的论方法的优点

读者不会不注意到;我像地地道道的自然目的论那样,非常严肃地、按字面意义去理解功能。^⑤从目的论上理解“功能”的政策有一些至关重要的好处:(1)正如我们已看到的那样,对“功能”的目的论理解有助于说明心理的东西的被知觉到的无缝隙性,即在与化学和物理学概念没有明显联系这种意义上的。心理概念的互锁。^⑥(2)通过把目的论要求加于功能实现的概念之上,我们完全可以避免机器功能主义所面临的标准反例,而且我断言,还可避免其他任何形式的功能主义所面临的标准反例;参见后面的详细论述。(3)目的论功能主义还有助于我们理解生物学、心理学规律的本质,尤其是面对戴维森关于后者的怀疑主义时更是如此(利康 1981b;卡明斯 1983)。(4)如果目的论描述本身可用进化论术语予以阐释,那么我们的心理状态本身的性能就更有把握借助最终原因予以阐释,我们为什么有疼痛、信念、愿望等也就更好理解了。^⑦(5)目的论观点提供了说明意向性的出发点,这种说明能避免其他自然主义说明的标准难题,尤其是容许大脑状态和事件拥有虚假的意向内容。而关于意向性的因果的、法则学的理论在面对后一个任务时常常举棋不定(参阅利康 1989)。

在上面,我已论证了:我们需要一种目的论概念,它以不同的程度出现,或至少允许描述具有不同程度的目的论性质,而且我们已有这样的概念,尽管很难加以阐明——试回忆面部识别器和钥匙的例子。对于这种程度的目的论概念的正确分析,哲学家们当

42 第一部分

然可以见仁见智——就我自己而言,我倾向于认为,这里的程度是由服从“诉诸于终极原因的解釋”所决定的,在这里,“诉诸于终极原因”的解釋又可重构为一种进化論解釋(尽管其中的一些细节仍有待澄清)。但是两个主要观点已经清楚了:(1)至少就单个的有机体而言,描述的目的論性质的程度与自然的层次是相当吻合的。^⑧(2)无论是在具有目的論性质的连续体上,还是在各种自然层次中(在这里作出明确区分是理所当然的),都不存在这样的、单一点,在该点上,关于自然的描述可以匀称地区分为各机通用的、纯“结构性的”、纯机械論的样式与更抽象、更模糊、有意图或许还有点活力論色彩的样式——肯定没有还能与下述区分即心理的东西和纯化学的东西之间的任何直观区分相符合的点,因为在它们之间有太多、太复杂的生物学。

我本人的泛心論或至少是泛目的論傾向至此已昭然若揭了。许多固执的哲学家会觉得它们充其量是幻想,而我(在我神志清醒时)自然乐意承认:要弄清把原子論层次的描述当作达到任何程度的目的論描述有什么用途,这是很困难的;^⑨求助于终极原因的解釋肯定是不长久的。但毫无疑义的是,目的論描述(即显然是达到了某种无论多么低的程度的目的論描述)是会如此长久地持续下去的,以至于可能适用于心理学(例如对后面的神经解剖学有用)。作用/占用者的划分仍在进一步蔓延。因此哲学家们通常所设想的、被吹得天花乱坠的“功能”/“结构”划分并未把握它所依存的人类心理学。……

我迄今所述的每一件事情似乎既隐晦又明了。但愿如此。我试图促使人们关注:什么是我所考察的关于物理世界的结构的大

实话,因为我认为:忽视这个真理,无视自然的等级本性已在意识和感受性质问题上导致了重大失误。在本文剩余部分,我将对此作一简要的讨论。

布洛克(1981,以及本《文选》中所选的第16章)、利康(1987)和其他人都曾提出了各种各样的反例,其意图是要表明:具有功能组织,不管多么复杂,对于拥有质的、现象上可感的状态都是远远不够的;这些事例中最著名的、经常被讨论的大概是布洛克的“小人—头脑”和“中国人口”(参看下面第7部分)。如果这些反例应予反驳,那么功能主义者就必须表明某种合理的、而这些反例又未满足的要求,尽管它们以这样那样的方式细致地模拟了一个真正有感觉的生物的功能组织。

从目的论上理解的小人功能主义不费吹灰之力地实现了预期的目的。因为在反例中设想的那些系统并没有以适当的方式从目的论上去组织;大多数甚至连有机体都算不上(参阅利康1987,第3章和第5章)。

即使谜一般的事例并没有驳倒小人功能主义,但沙文主义和自由主义的某些问题仍有待解决。福多和布洛克认为:普特南在从同一论倒退而转向行为主义时,走得太远了,不管这个看法是不是对的,但功能主义者肯定有这样的责任,即找到心理状态描述的基准线,它既不是如此抽象的或行为主义的,以致排除了颠倒光谱之类的东西的可能性,又不是那样具体或结构上的,以致陷入了沙文主义。布洛克本人进一步论证说,这个问题是不可解决的。

尤其是,他指出了输入与输出描述的二难困境。直截了当地说,输入与输出不可能用人类中枢术语予以描述;这一来就沙文主

44 第一部分

义式地阻止我们把心理描述用之于机器、火星人和其他在生物学上有别于我们的造物,不管它们在为它们的感知能力辩护时可能提供什么令人信服的凭证。另一方面,输入与输出也不能用纯抽象的术语(即仅仅作为“输入”和“输出”)予以描述,因为这样会导致那种极端的自由主义。对此,布洛克一直是嗤之以鼻的,他所根据的既有较早的例证,又有新的事例,如有非常复杂的输入输出和内在状态但肯定没有心理特性的经济系统。我们也不能诉诸于有感知能力的物种通过输入输出所完成的、与其环境的任何特殊类型的相互作用,因为在一些事例(如麻痹病人、玻璃试验管中的大脑之类)中,我们试图把心理描述归之于不能用任何方式成功地与它们的环境相互联系的对象上。布洛克断言:

有这样的输入与输出描述吗,特殊地说,它足以避免自由主义,而一般地说,它足以避免沙文主义?对此我表示怀疑。

我所看到或所想到的、关于输入与输出的描述的每一个建议,不是犯了自由主义的错误就是犯了沙文主义的错误。即使这篇论文的焦点是自由主义,但沙文主义更成问题。

……绝不会有这样的物理描述,它们适用于所有心理系统的输入与输出。因此想用关于输入与输出的物理描述系统表述功能描述的任何企图,要么不可避免地会排斥某些具有智能的(可能)系统,要么会包括某些没有智能的系统。

……另一方面,正如你会记起的那样,把输入与输出简单地描述为输入与输出,这未免太自由了。举例来说,我不明白:功能主义怎样描述输入与输出才不会与自由主义或沙文主义发生冲突,才不会放弃原先用非心理术语描述智能的计

划。我无意于主张,这是反功能主义的决定性论证。更确切地说,像功能主义者反物理主义的论证一样,也许最好把它当作是一种为论证举证的责任。(本文选,pp. 464-465)

即使我一直赞成说,根据对“举证责任”的温和理解,布洛克挑战的东西是,功能主义背负的正好是这个责任,但我仍不敢肯定布洛克要求功能主义者的是多么详细的计划。问题在于:这个负担是否像布洛克所假定的那样艰难沉重。我认为,至少有三个因素可使之大大地减轻,并为我们赞成乐观主义提供某种根据:

第一,有这样一条论证路线,即它至少为主张沙文主义和自由主义的二难困境没有得到解决提供了某种有点确实性的理由或自然的动机(这里的二难困境是:功能主义者要让其同一于我们的心理状态的东西要么是输入和输出,要么是内在状态)。它一开始就表现为可能导致可怕后果的论证。布洛克以没有半点妥协的方式陈述那个二难论证,指出:人们的选择只能是(a)从生理学上描述输入和输出,从而成为一个沙文主义者,或者(b)“纯抽象地”描述输入与输出,从而成为一个伪善的同情者。但是这种关于选择项目的赤裸裸的陈述忽视了这样的事实……即功能抽象是一个程度问题。纯生理学描述是一种极端,位于那个系列的较低的或更富结构性的末端;而“纯抽象”的描述是相反的极端,位于较高的或更富功能性的末端。须注意的是……存在着这样的描述,它们甚至比生理学描述更富“结构性”,如微观物理学的描述,相对于它们而言,生理学描述是“功能性的”;同样,事实上还有比输入、输出更抽象的描述,诸如“传输”、“运动”甚或“偶然出现”。如果“纯抽象的”描述和生理学描述差不多只是处于功能抽象连续体的两端,如果

46 第一部分

是这样,像看起来所是的那样,那么就有理由预期:存在着居中的抽象层次,它会产生这样的描述,它们排除了波利维亚经济制度、阿布尼戈银河系、关于大沼泽地*的微生物学之类的东西,而为人类、软体动物、火星人的玻璃试验管内的大脑腾出了地盘。真理存在于(像通常所是的那样)两者之间的某一处,而且取决于哪一种心理状态的哪一方面对其中之一的利益,并不总是在两者之中的同一个点上。等一下就会看到,在各种中间层次上将会得到什么样的资源。^⑩……

此外(这里是我回答布洛克挑战的第二个观点),还得记住:没有什么强迫我们假定:各种不同种类的心理状态发生在同一功能抽象的层次上。直观上“更富有行为色彩的”各种心理状态,如信念、愿望和意图,大概出现在相对高的抽象层次上,这便促使我们轻而易举地把信念、愿望和意图归之于火星人,因为它们的外显的行为和表面的心理功能与我们的是相匹配的;这也适用于高级的“提供信息的”心理活动,如记忆和(字符)计算。从直观上看,“较少行为性的”、更富有质的特征的心理状态大概出现在更低的抽象层次;具有某些特殊类型的质的特征的信号感觉可能是物种独有的(当发现真的如此时我们至少不会大惊小怪),而且相当有可能的是,我们的火星人的、具有人的特性的行为是为他所拥有的、有点不同于我们的感觉(或可能“似人的感觉”)的感觉所推动的,尽管他的表面的行为相似于我们。

我不知道,有谁曾对两层次论本身作过明确的辩护。^⑪但两层

* Everglades,美国佛罗里达州南端的广大沼泽地。

次论似乎就是直接存在于像布洛克的“输入输出难题”那样明显的二难困境后面的东西。

类似的考察适用于意向性问题。我们认为:有机体的状态要么是有意向的状态,要么不是,就是这么回事,进而我们想知道的是:意向性的功能或结构定位可能是什么。我认为,意向性不可能是纯功能的属性,其理由在现在是众所周知的,^②但是就其本身而言,我认为,我们最好是承认意向性本身有程度的差别。^③意向性或关于性的“标志”还不太清楚,但基于反思,现在看起来清楚的是:存在着中间层次的功能描述,它提供了一种对可能的、非存在的对象或类型的指向性,然而这种指向性尚不及人类心灵所表现出的那种丰富的、纯正的意向性。在这个中间层次,我们常常从系统理论的角度讲到“分辨器”、“扫描器”、“过滤器”、“抑制器”等,其意思是说,这些术语完全是字面上的,而没有实际地输进思想或输进可被称之为“偶然发生的”关于性。但是我不能在这里阐述这些观察资料,而得另找机会。^④

第三,我们直接支持关于输入和输出的“纯抽象”描述,使沙文主义与自由主义的整个问题重新回到我们对内在状态和事件的描述之中,这可能是有益无害的。在这种描述所适用的大脑的功能等级系统中,存在着如此多的可能性,如此多的、相互有别的抽象层次(其中许多相互交错、重迭),以至于可以相当合理地预期:对于每一心理状态-类型来说,在沙文主义和自由主义之间应该存在着某种中间道路——对于每一状态-类型来说,并非必然是相同的中间道路。以为所有心理现象一定在功能上定位在相同的层次,或以为任何单个的心理状态一定完全局限在一个层次,这完全

48 第一部分

是错误的。至于“更具功能性的”、近乎于行为主义式的心理状态，如果有必要的话，我们甚至不反对说：一种经济系统或中国的人口能有这种状态（比如说倾向性的信念）。至少在那连续体的功能一端，甚至可能存在着符合于同一论的心理状态—类型，即使很难设想象“质的”状态那样的任何心理状态。

前述的评论使人想到了对布洛克的“缺席感受性质”论证的一个最终的附加回答，我认为它实际上是结论性的。我原先以为，布洛克对功能主义的直觉上的忧虑近似于这样的感觉，即觉得功能主义的诸阐释的亲缘关系和它们的有待阐释的话语的同质的、原始单子论的质的特征之间有不协调性；我推测，这种不协调性对他来说是绝对的。须知，他显然没有对同一论提出同样的反对意见；像其他唯物主义者一样，他轻描淡写地指责同一论者的沙文主义错误，而又没有提出更进一步的控诉根据。毕竟，该理论的主要优点之一是，它能说明颠倒光谱或别的没有相应外在表现的内在变故的可能性。但是如果我们赞成我的下述主张：对人的状态的小人功能描述和生理学描述反映的只是邻近的功能等级系统或连续体之内的不同的抽象层次，那么我们就不能再以任何绝对的方式把功能主义者与同一论者区别开来。例如，“神经元”要么是被理解为一个生理学术语（指的是人的一种细胞），要么是被当作一个（目的性的）功能术语（指的是电极的分段传递器）；无论是根据哪一种阐释，它表示的都是一种可例示的东西——即由一组更基本的对象所发挥的一种作用，如你愿意这样想的话。因此，同一论者甚至就是功能主义者——即把心理实在定位在非常低的抽象层次

的人。这里的启示在于:如果布洛克想一意坚持:功能主义心理学因我所述的那种根本不协调性而陷入了困境,根据关系作用或可例示物说明心理项目的心灵哲学在原则上不可能涵盖感受性质的棘手的单子性,那么人们就不得不对同一论提出同样的责难,而且我敢肯定,他没有感觉到这样做在直观上的压力。^⑤事实上,布洛克是要让该理论对物种沙文主义避重就轻地认错,甚至承认这大概也适用于某些心理属性。

有一种观点是由难以理喻的两层次论所引起的,它认为:功能主义在概念上或结构上有点不同于同一论,正因为如此,便遭到了种种抨击。我已说过,同一论正好是功能主义的经验上的一个特例,它(不合理地)把所有心理状态定位在同一个非常低级的结构抽象层次之上——即神经解剖的层次。因此就不可能出现这样一些纯概念或哲学的非难,它们适用于功能主义、而不适用于同一论,反之亦然,即使一个在经验上没有另一个那样合理。然而布洛克这样一些哲学家一直声称已看到了这样的责难。如果我关于自然连续性的理论是正确的,那么这里就一定有某种错误;因为神经解剖术语即使处在较低的抽象层次,但仍是功能的,因此也是关系性的,正如高级组织术语那样。如果在关系描述和现象感受性质的内在性之间存在着根本的不一致性,如果那种不一致性使功能主义如履薄冰,那么它一样会妨碍同一论。^⑥

试考察这种非难的第二个例子。布洛克进一步争论说,功能主义不可能容许颠倒光谱或其他类型的、内在地转换的、而没有反事实地反映在行为中的感受性质的可能性——而在这方面同一论则不是不可能的,因为同一论就是为了陈述颠倒感受性质的事例

50 第一部分

而定制的。但是如果我对自然层次的连续性的反思是正确的,那么这里一定有某种错误。而且事实就是如此。正如设想作为颠倒光谱基础的、不可捉摸地转换的神经生理结构是易如反掌的一样(参阅利康,1973),设想更抽象地被描述的功能要素的转换也是如此。(即使这无疑是有限制的,而且相当有可能的是,人们不会上升到更高层次的抽象,不让那种颠倒在行为中觉察不到。)

问题的真相为“颠倒感受性质”概念中的实用主义的模棱两可弄得模糊不清了,我认为,这种模棱两可即使一点也不精湛,但给予布洛克以修辞上的助益。也就是说,有一种隐蔽的参项:对于什么来说是“颠倒的”? (试比较随附性的相关关系:随附于什么?)就传统而言,“颠倒光谱”意味着这样的(颜色)感受性质,它只是相对于实际的、反事实的输入-输出关系才是颠倒的。要么是出于责任,要么是由于倾向,分析行为主义者和维特根斯坦主义者都否认那种颠倒的可设想性,但大多数人的日常的模式直觉则予以肯定,而且同一论和功能主义的理论也轻易地承认了它;它对功能主义绝对没有提出任何威胁。对功能主义有害的东西是相对于输入-输出关系加上内在的功能组织而言的颠倒的感受性质的可设想性。这种颠倒假说更强、更大胆。说得婉转一点,其可能性是有争议的。其实,肯定它无异于否定功能主义的真理——不用证明就可以说:两个有机体可能在质的状态方面相互有别,即使它们在整体的功能组织方面、在这里所说的一切结构抽象的层次上事实上一模一样。当然,一直有这样一些哲学家,他们武断地坚持这样的形而上学的可能性,即即使就那点来说诸有机体是克分子复制品,但是它们在质的状态方面仍有可能彼此不同。即使关于心灵的有

关理论最终被证明是错误的,但这种断言并没有内在的可信性。仅相对于输入-输出关系而颠倒的光谱,其可能性是有可靠的根据的,也是应予重视的,即使我假设模态直觉是可以废除的;还有这样一种光谱,它相对于输入-输出关系以及无论多么低的、其倡导者觉得可合理地予以陈说的抽象层次上的内在功能组织而言,其颠倒的可能性一点也不明朗,而且与某些直观上合理的随附性命题是格格不入的。

(某些注重关系的理论家也许会发现:假定开端上有某种特许的抽象层次,这是很自然的。例如“分析功能主义者”,或如我喜欢称呼的:常识关系论者,就认为:心理术语的意义决定于由常识或民众心理学而与那些术语相关联的因果作用,他们还由此拒绝求助于任何比常识所能理解的层次更低的功能组织层次。^②撇开民众心理学不说,“高教会派”(high church)计算主义者^③甚至在对认知和行为的纯科学的说明之内也蔑视求助于人类生物学,即使他们自己所选择的自然层次没有得到太清楚的说明。^④开辟这样一种特许的组织层次的理论家自然可以承认相对于那种被选择的层次的“颠倒光谱”,只要他或她愿意在类型上把感受性质与仍属低层次的术语等同起来。^⑤)

两种可供选择的战略

我一直倡导一种方法,它主张在关于心理的东西的功能主义本体论的范围之内解决关于沙文主义和自由主义的问题。存在着多种可供选择的、可能的战略。一种备选方案是把我们关于心理

52 第一部分

的东西的观点一分为二,如直接承认心理状态与其质的特征的区别,用功能术语阐述那些状态,用相当广泛的物理术语阐明那些特征,并默认下述结论,即颠倒光谱或者说感受性质中的较少人际间的差异性可能比我们所想的更普遍(即是说,实际上像人间生理结构的、具有可比量值的差异一样普遍)。^④

疼痛为这第二种承认感受性质的、被举荐的方式提供了一种有用的检验实例。关于疼痛的一个饶有兴味的、独特的事情是:(不像其他大多数心理状态那样)它既有与行为紧密相联的模式,又有可为内省逼真地感觉到的属性。根据现在的建议,这意味着疼痛状态可接受多层次分析。例如(稍作推测)我们最终可能愿意把有机体的起着疼痛的常见而“明显”行为作用(由伤害引起并导致移开)的任何内在状态归结为疼痛,但又愿意根据那些状态的生理基础对疼痛的感觉作出区分。^⑤由此可以说,即使软体动物和火星人有疼痛,但它们感觉到它们的疼痛可能在方式上不同于我们感觉到我们的疼痛。进而还可以说,一种感觉起来像我的疼痛状态的状态,在有不同组织的造物中可能是某种有别于疼痛的心理状态;某些哲学家也许会觉得,这完全是反直觉的。

附带地说,二分的观点在过去的岁月里已相当流行,^⑥并常得到这样的表述:(如)“当疼痛本身是功能性的时候,疼痛的特殊感觉就是神经生理学的感觉。”但是后一表述再一次预先假定了两层次论。把“功能的/结构的”的区分当作层次上的相关物,进而被二分的理论就会萎缩成下述论点的毫无意义的、特殊的翻版,这一论点(我希望它是自明之理)就是:心理状态和它们的质的特征最好不要根据同一自然层次加以阐述(尤其是,质的特征的位置在等级

系统中可能比一般被考察的心理状态的位置要低)。正如我已指出过的,我特别赞成后一点,但它不是与功能主义相对抗的选择。

第三种选择方案是针对躯体感觉的事例而提出来的(即使我怀疑,能不费吹灰之力地把它用之于知觉感受性质)。可以设想,在现象上似乎是简单的感觉实际上是复杂的,与某种类型的感觉有联系的特定感受性质其实是许多不同的、可个别地加以处理的小人功能特征的巧合或附加物。我认为,这条路线,而不是前一段所述的路线,是说明那个疼痛实例的最合理的路线,因为它已得到了丹尼特(1978,第11章)所搜集和简述的麻醉学材料的有力支持。这些材料所表明的似乎是:化学上不同的麻醉剂和止痛剂在不同功能的交界处上使被试的正常的“疼痛”子程序遭到破坏,引起被试对它们的效果作出相当不同的语言报告。在碰到类型与强度大致相同的疼痛的一群被试中,有一被给予药物A的子群体可能会报告说:疼痛减轻或完全消失了,而用了药物B的子群体则会报告说:即使他们知道疼痛仍然在那里,但他们不能感觉到;用了药物C的子群体则说:即使他们能感觉到疼痛像先前一样强烈,但他们并不在意;如此等等。这些报告中的一些在我们看来是滑稽可笑的(维特根斯坦主义者把它们作为“不合理的”而加以嘲笑),这一点自然反映了这样的事实,即被试的正常的内在运作由于药物而受到了破坏,他们关于疼痛的正常的内在经验也因此而发生了改变。药物所起的作用就是通过把被试的经验的相当复杂的功能基础分割成元件子程序,而把被试关于疼痛的现象经验分割成元件。如果是这样,那么就可以说,我们关于疼痛的现象经验有构成元件——它是复合的,由出现在相当不同的结构抽象层面

54 第一部分

上的欲望、愿望、冲动和信念所构成。如果这些元件能够由于药物而单个地相互分离开,那么我们就可以完成下述理想实验,在其中,我们假设接受了一个有麻烦的被试,通过给予药物 A,从他的疼痛中分离出了一个元件,进而通过给予药物 B,又分离出了另一元件,然后重复这一过程,同时让其在我们的所作所为时作出报告。我认为可以合理地设想:如果我们继续这样做下去,消解一个又一个的访问路径,逐一排除欲望、愿望、信念之类的要素,那么我们或迟或早会成功地消除疼痛本身;再者,如果颠倒那个过程——通过逐渐停止各种药物而使那些路径得到恢复——那么被试就必然重新感觉到完全一样的疼痛(假如他的被伤害的组织并未同时得到修复)。我相信这便使下述假定成为合理的,这假定是:与布洛克的反自由主义精神恰恰相反,有关复合功能行为的某些多层次的严格后果,对于疼痛之发生既是必要的又是充分的。

在我已述的三个备选方案中,我不知道怎样作出最终的抉择,或者说我们可能找到哪些类型的进一步证据。我概述了某些观点,目的只是想表明小人功能主义有着相当丰富的资源,这些资源对于沙文主义和自由主义的二难问题以及说明感受性质的实证任务都可能产生影响。我相信,以这些资源为基础,我们有权利断言:布洛克关于感受性质的悲观主义是不能得到证实的。……

如果我的连续性学说像已述的那样是清楚明白的,那么它对某些主要的心灵哲学家来说就是不够清楚的。我希望,前述的证明还有助于赋予小人功能主义作为关于心理的东西的理论以更强的魅力。

注 释

① 在伦理学中,我正好坚定地相信某种形式的行动-功利主义,但大多数人的最大幸福这一神圣原则本身阻止我对你这样说,更不用说在名义上承认(公开的)谋杀。

② H.A. 西蒙(1969)曾强调并表述过这种多层次等级结构;我不知道该观点在他之前是否已出现了。W.C. 温萨特也曾给予了明确的论述(1976)。

我注意到,这是由福多(1968)和丹尼特(本文选)最先应用于心理学中的,参阅后面的参考文献。

③ 我记得,刘易斯·托马斯(1974)曾讨论过昆虫社会以及人与其自己的线粒体的关系。关于“有机体”的部分学与利益休戚相关。应注意,我们必须承认自然层次之间的不同还原关系的多元主义。

④ 关于该战略详细的阐释与辩护,请看卡明斯(1983)。不过,里查森(1983)则泼冷水。

⑤ 哲学中的相关讨论及文献,请看 P.M. 丘奇兰德 1986, P.S. 丘奇兰德 1986。

⑥ 在包含着下面一段的著作中,丹尼特关心的主要是对意向性的阐述。那种关心与我这里的关心不可同日而语;我感兴趣的只是小人的进一步细分。

⑦ 事实上,正如 D. 阿姆斯特朗向我指出的那样,现在的策略是在心灵哲学中阻止大量典型的无穷后退论证,包括赖尔对决策的意志理论的抱怨。丹尼特本人则用它来反对关于自我理解表征的“休漠问题”(1978, pp. 122 ff)。

⑧ 对言语中心的实际的、递进的小人分解,可参阅利康 1984, p. 262, 图 1。

⑨ 对这一点的坚定的支持,可参阅戴维森 1970。

⑩ 因此,斯马特的下述例子比他想象的也许更加合理,这例子说明的是“民族”陈述的逻辑有别于“公民”陈述的逻辑。

⑪ 波普 1972;温萨特 1972;赖特 1973;R. 米利肯 1984;林德尔 1981, 1983。后者的进化论说明是我所知的说明中最好的。它也受到了别人的批

56 第一部分

评。在我看来,真理似乎存在于两者之间的什么地方。

J. 贝内特(1976)对源于 A. W. 麦肯齐(1972)的目的论作出了一种不同的自然主义探讨。

⑫ 对我们的时-空部分的内容描述因而可整理成一种连续体,从最少目的论的部分到最多(最高)目的论的部分。这个连续体正好与功能例示或实现的等级系统密切相应。分子结合在一起实现或起着金属组成部分的作用;金属组成部分又起着钥匙的作用;钥匙可用作我们开门的工具;等等。我相信,这类普遍的功能等级系统就是推动本体论还原的东西,也是“一切从根本上可归结为物理学问题”这一观点的动力。关于从进化观、功能等级系统、本体论角度所看的目的论与科学还原的方法论之间的关系,可再看温萨特 1976。我从阅读梅利克 1973 中也受益匪浅。

⑬ 正如 J. 福多在讨论中向我指出的那样,存在着一种尚较清楚、双校准器可能考虑到的、确凿无疑的区分。它是两类对象之间的区分,一类对象的部分对它们是必不可少的,而另一类对象的部分则相反。

⑭ A. 罗蒂向我推荐了亚里士多德的观点,它根据有机体的那些构件功能对于那个有机体的种的物质条件的适用性来解释有机体的构件功能。这个观点与我想赞同的关于功能的原因论说明完全一致。

⑮ E. 索伯赞赏这种态度,称之为“把功能还给功能主义”;试比较利康 1981a, p. 27 的说明,它涉及到普特南和福多对“功能”的一语双关的用法。

⑯ 详情请看利康 1981b。

⑰ 为什么疼痛使人感到痛苦?我们为什么不可能有一种这样的系统,它发出伤害的信号、促成修复、而又不会使人不舒服?答案很简单。假设我正好有这样一个系统,像我的自动机上的红警示灯。我好像习惯性地(即使是非理性地)忽视了那警示灯,模模糊糊地希望它离去,如果人的警示灯不能内在地为我提供做出关于它的某事的紧迫的动机,那么我就会忽视人的警示灯。

⑱ R. 范·久里克(在通信中)为我提供了一些气象学和地质学方面的实例,在那些实例中,目的性的(明显的)程度并不伴随自然的层次。

⑲ 在这个问题上,R. 布洛克与我的看法大相径庭,(在谈话中)他曾经说过,“我可以让你的神经元、细胞等成为功能性的,但当你得到氢气和氧气

时,当你下降到化学层面时,那里则完全没有功能的或目的论的东西!”噢,不是吗?“氢”有什么目的?“氧”有什么目的?(这种推测很廉价,但颇令人满意。)

② “等到明年吧!”J. 塞尔在不同的但又颇相似的上下文(1980)中嘲笑说。当然得等到明年!

③ 也许拯救了“分析功能主义者”,但我拒绝其观点(参看注⑥)。

④ 普特南 1975;福多 1980;斯蒂克,见本文选;伯奇 1979;利康 1981c;
.....

⑤ 德雷特斯克(1981)前此已部分提出了这个观点。还可参阅范·久里克(本文选,1982)。

⑥ 我也注意到,在认知科学界,某些常见的争论是以两层次主义者的方式错误地设想的。

⑦ W. 塞拉斯感觉到了。但那又是另一码事……;参阅利康 1987,第 8 章。

⑧ 布洛克本人没有强调关系/单子比较,但却让人觉得他的特殊的直觉不够成熟,因此他对我前述的对人不对事的直截了当的观点——具有像我们一样的神经化学结构是体验到感受性质的(不论是不是关系的)的必要条件——仍然无动于衷。不过,我不解的是,哲学家怎么可能知道这一点?它焕发出了自然之光吗?

⑨ 我与 S. 休梅克就此的有价值的通信使我受益匪浅。就我自己而言,我不赞成分析功能主义,理由有两点:(1)我拒绝这一理论所坚持的关于意义的非此即彼的概念分析或隐含定义理论。(2)对下述观点我表示怀疑,这观点认为:常识或“民众心理学”包含有关于心理实在的足够信息,以致能淋漓尽致地描述它们的本质,并避免反例。民众心理学的衣夹模型唾手可得,且用不着为真实的智能归因作担保的极度的复杂性和目的论组织。

⑩ 这一术语源自丹尼特(1986)。(译者补注——原意是指特别强调以教士权力、礼仪、传统、天主教基本原则为生活准则的教派,在这里是指正统的计算主义者。)

⑪ 在这里,我依照某些作者新近的做法,假定:真的存在着高教会派计算主义者;我不能肯定的是:真正的功能主义者在自我意识中想要的是不是

58 第一部分

这种观点。根据 Z. 皮利辛和 J. 福多关于多样实现的说明,通常把这种观点归之于他们。也许布洛克真的坚持该观点,否则他就不会一如既往地抵制我在 1981a 中就开始了的对两层次主义的批评。

⑩ 尤其可参阅贝希特尔 1985。

⑪ 布洛克在第 460 页暗示:他并不觉得这个建议是完全不令人惬意的。参阅注⑩。

⑫ 这种改进使我所理解的、为 D. 刘易斯在 1980 年所作的反功能主义论证易于为人所接受。

⑬ 正如我已述及的,布洛克暗示了这种观点。我在 1981a(pp. 47-48)中阐发了这个建议。H. 普特南(1981)、S. 休梅克(1981)、P. 基切尔(1982)、T. 霍根(1984)和 G. 谢里登(1986)等也接受了这一观点(休梅克称之为“选择狭隘主义”)。

参考文献

D.M. 阿姆斯特朗(1968):《唯物主义的心灵理论》,劳特利奇和基根·保罗。

F. 阿特尼夫(1960):“对小人的辩护”,载 W. 罗森塔尔编:《感性沟通》,麻省理工学院出版社。

P.W. 贝希特尔(1985):“当代联结主义:新的平行分布加工模型是认知性的还是联想主义的?”《行为主义》,13,53-61。

J. 贝内特(1976):《语言行为》,剑桥大学出版社。

N.J. 布洛克(1981):“心理主义与行为主义”,《哲学评论》,90,5-43。

T. 伯奇(1979):“个体主义与心理的东西”,载 P. 弗伦奇等编:《中西部哲学研究》,第 4 卷,《形而上学研究》,明尼苏达大学出版社。

P.M. 丘奇兰德(1986):“认知神经生物学中的某些还原战略”,《心灵》,95,223-38。

R. 卡明斯(1983):《心理学解释的本质》,麻省理工学院出版社。

D.C. 丹尼特(1986):“计算探讨的逻辑地理学:从东极的观点看”,载 M. 布兰德等编:《知识和信念表征》,亚里桑那大学出版社。

F. 德雷特斯克(1981):《知识与信息流》,麻省理工学院出版社。

J. 福多(1968):“在心理学解释中诉诸于不言而喻的知识”,《哲学杂志》,65,627-40。

J. 福多(1980):“认知心理学中作为研究战略的方法论的唯我论”,《行为与脑科学》,3,63-73。

T. 霍根(1984):“功能主义、感受性质与颠倒光谱”,《哲学与现象学研究》,44,453-70。

P. 基切尔(1982):“同一论的两种观点”,《认识》,17,213-28。

W. 利康(1973):“颠倒光谱”,《理智》,15,315-19。

W. 利康(1981a):“形式、功能与感觉”,《哲学杂志》,78,24-50。

W. 利康(1981b):“心理学规律”,《哲学论坛》,12,9-38。

W. 利康(1981c):“关于信念的小人理论”,《认知与大脑理论》,4,139-59。

W. 利康(1984):《自然语言的逻辑形式》,麻省理工学院出版社。

W. 利康(1987):《意识》,麻省理工学院出版社。

A.W. 麦肯齐(1972):“目的性行为之分析”,康奈尔大学博士论文。

D. 梅利克(1973):“行为阶层”,俄亥俄州立大学博士论文。

R.G. 米利肯(1984):《语言、思想与其他生物学范畴》,麻省理工学院出版社。

K. 林德尔(1983):“异态心理生物学”,拉特罗布大学博士论文。

K. 波普(1972):“云与钟:关于合理性和人的自由问题的探讨”,载《客观知识》,牛津大学出版社。

H. 普特南(1975):“‘意义’的意义”,载 K. 冈德森编:《明尼苏达科学哲学研究》,第7卷,《语言、心灵与知识》,明尼苏达大学出版社。

H. 普特南(1981):《理性、真理与历史》,剑桥大学出版社。

R. 里查森(1983):“心灵的计算模型”,未发表的专题论文。

G. 谢里登(1983):“在物理主义的宇宙中有道德主体吗?”《哲学与现象学研究》,43,425-480。

G. 谢里登(1986):“选择狭隘主义与休梅克对功能主义的论证”,打印稿,西密执安大学出版社。

S. 休梅克(1981):“功能主义的某些变种”,《哲学论坛》,12,93-119。

60 第一部分

H. 西蒙(1969):“复杂性的结构”,载《人工科学》,麻省理工学院出版社。

L. 托马斯(1974):《细胞的生活》,矮脚鸡丛书。

范·久里克(1982):“心理表征——一种功能主义观点”,《太平洋哲学季刊》,63,3-20。

W.C. 温萨特(1972):“目的论与功能陈述的逻辑结构”,《科学史与科学哲学研究》,3,1-80。

W.C. 温萨特(1976):“还原论、组织层次与心身问题”,载 G. 格洛布等编:《意识与大脑》,会议论文集。

L. 赖特(1973):“功能”,《哲学评论》,82,139-68。

把功能还给功能主义*

埃里奥特·索伯

功能主义一开始就给人以不好的印象。问题在于：功能是模棱两可的，该学说是根据心中的错误意义而予以阐发的。数学上的功能指的是从某些对象向另一些对象的映射；每一输入都有一种唯一的输出。例如我们可以谈论加法的功能，在那里，两个数映射到它们的总数上。另一方面，还有关于功能的目的论概念——就像我们说心脏的功能是泵血时所说的那样。与第一个概念相应，在心灵哲学中就出现了一个我称之为图林机功能主义的观点。与第二个概念相应的是我将称之为目的论功能主义的观点。功能主义作为对心身问题的一种解决方案，理应阐发为目的论功能主义的形式。然而相反，功能主义通常被理解为意指图林机功能主义。

由于这一前期的失误，功能主义被认为面临着许多致命的责难。我要论证的是：一当功能主义从目的论上加以理解，这些责难便会烟消云散。这不是说，这样做就能解决心身问题，因为目的论观念有其自身的难题。但是如果我们能从理论的失误中吸取教

* 节译自“潘格罗式的功能主义与心灵哲学”，载《综合》第64卷，第2期，1985，pp. 165-193。

62 第一部分

训,那么我们一定能明白,是什么使一种理论的最好阐释陷入夭折。

功能主义是作为同一论的反动而登上历史舞台的。其否定的见解是:心理属性并不在类型上同一于物理属性。心理状态的类型是可多样实现的(普特南 1967;布洛克和福多 1972)。但是这种关于心理状态不是什么的主张一定得辅之以关于心理状态的本质是什么的某种正面的说明。

功能主义者在他们的启阀器和捕鼠器中找到了正面主张的出发点(福多 1968)。粗略地说,使捕鼠器之成为捕鼠器的东西是:它起着某种因果作用。假如一逃跑的老鼠作为输入,那么它就把逮住老鼠作为输出产生出来。一特定的功能状态对其他状态将具有某种原因与结果的关系。这样说自然也适用于特定的物理状态;我们可以根据酸所起的因果作用描述酸之为酸究竟是什么(利康 1981)。功能主义者希望根据心理状态与行为、与刺激的因果联系描述心理状态。只有当这些描述为心理状态怎样个体化提供了基本的限制条件,功能主义才会成为一种有价值的学说。

与这种目的论观点——心灵是一种功能装置,正如捕鼠器是一种功能装置一样——大相径庭,另一建设性的观点悄然生起。它是这样的思想,即心灵是计算机。其核心是主张:说个体有不同的认知能力就等于说该个体能够使用某些计算机程序。因此知觉到某物或得到了一特定信念就被理解为执行一系列的计算。

这依然不能算作是对心身问题的一个取而代之的解决方案。如果心理状态应当从计算上加以个体化,那么人们应怎样理解计算概念呢?正是在这里,隐藏在图林机功能主义后面的观点便粉

墨登场了。计算机借助例示机器模拟运算表进行计算,这正好是从输入到输出的一种功能。该机所处的状态决定了它应当输出什么,它下一步将进入什么状态。

将机器模拟运算表个体化的原则就是把任何数学功能个体化的原则。大致地说,它纯粹就是抽象同型性的要求。当我 14 次成功地把 1 加到 1 上时,我与大海在 14 天中每天加一层沙到海滩执行的是“同一”的程序。

这种对功能主义观点的图林机阐发带来了种种问题。假设我们具有某一心理属性 P。图林机功能主义就会这样理解,即当任何个体抽象地同型于我们时,它显然也具有 P。然而再明显不过的是:这一主张既非必要,又非充分。

这种不充分性首先得到了戴维斯(1974)、丹尼特(1978a, 1978b)和布洛克(1978)所讨论的一种观点的说明。假设有一大群人,他们被安排用下述方式行动,它在结构上同型于我想到“我要一只冰淇淋”时我的大脑所采取的那种行动方式。当一神经元给另一神经元发送了一信号时,一个佣工就用锤子击打另一个的头。那群人在结构上同型于我的大脑,然而我有信念,而那群人则(可能)没有。

结构同型对于具有相同的心理状态并不是必要的。这一事实是由布洛克和福多(1972)所提出的一个要点。给同一论以致命打击的东西是:两个个体即使处在不同的物理状态,但仍有可能处在相同的心理状态。同样,两个个体即使处在不同的计算状态,但似乎仍有可能处在相同的心理状态。两台计算机能以非常不同的方式让 35 乘以 44。如此类推,两个有知觉力的存在即使其心理功

64 第一部分

能与非常不同的程序有关,但仍有可能具有某些相同的心理属性。令人震惊的是,关于心灵的计算机模型由此倡议的是,心理状态不应同一于计算状态。

即使这些责难针对图林机功能主义而发时似乎极富判决性,但转向目的论功能主义时情况则判然有别。为了弄清为什么,试把它类推到心身问题上:消化过程与身体的物理过程之间是什么关系?消化的状态、属性与物理的状态、属性之间是什么关系?

一种可能的主张就是同一论:对食物的加工并不在于某种非物质实体对食物项目的操作。毋宁说,完全现实的物理-化学过程就是所涉及到的全部。这一观点比下述观点有所前进,这观点是:食物包含着神秘的精神实体,而精神实体又是由有机体用深奥难解的精神方法选出来备用的。另一方面,同一论有某些明显的限制。多样实现性正好是消化的一种属性,就像它是知觉与认知的属性一样。其他物种通过显然有别于在我们自己物种身上所用的那些物理-化学过程而消化食物;此外,还可以设想:未来某一天的先进的计算机自动机将会搜寻食物,并予以消化,它们这样做的方式将会更明显地有别于我们自己的方式。

因此关于消化的同一论就让位于功能主义了。后一学说应怎样加以阐发呢?图林机变种受着前面在心身问题中述及的同样问题的困扰。正像可以让一群人抽象地模仿我的思维过程一样,也可以让一群人抽象地模仿我的消化过程。但是那群如此去模仿的人既不是在消化,也不是在思维。一种更好的方案就是关于功能主义的目的论变种。(目的论的)功能同一性与抽象程序的同一性不可同日而语:两个有机体可以用截然不同的方式消化食物。在

它们的消化过程之间可能存在着物理上的差异,而且在它们之间还有抽象的差异。一系统完成消化过程就是它完成具有某种(目的论的)功能的过程。

功能主义的另一问题,与已述的群体模仿中的问题有关联,可借助目的论探讨加以澄清。不再让一群人在一小时或一天抽象地模仿我的大脑(或肠子)中的微观过程,而这样来设想:在这群实在与我之间发现了一种相当不同的关系。这里我想到了布洛克(1978)关于小人头脑的观点。假设我的大脑拿走了,代之以无线电发射器和接收器。进入大脑的各种不同的输入通道,现在可以把它们的信号送至一群人。每一个体模仿我的一个神经元的行为。其中的一些个体直接与我头脑中的无线电发射器接通;而另一些则只与该群体中的其他个体接通;还有一些个体把信号送回到我身体中的接收器,它接着再把信号分程传递给我的身体的其他部分,直至引起行为。

如果是这样,那群实在就不仅仅是进行抽象的、非推论性的模仿,而毋宁说是作为我的补充性大脑而发挥功能作用(还可参阅利康 1979)。它们之对于我的正常的心理功能作用就像肾机器之对于我的正常的肾脏功能一样。当我启用一种肾机器时,我的血液就会变清洁。在小人情形之下,我与那一群体接通后,知觉和认知就可进行。两种情形是相似的。

这里有一道难解之谜,它源于严肃地接受缸中之脑有心理状态这一观点。即使我的大脑允许我以各种方式与环境相互作用,但大脑与身体的这种联系对我的大脑拥有心理状态似乎是可有可无的。即使我承认信念是一种广泛的心理状态[正如普特南

66 第一部分

(1975)极力主张的],以至于信念的个体化依赖于关于环境的事实,而不取决于心灵的内部状态,但是观点依然如故:坚持缸中之脑没有任何心理状态这一点似乎是不合理的。

如果这适用于正常的大脑,而此时它们与包容它们的身体分离开来了,那么这也一定适用于人工大脑。在前面的例子中,一群人人为我提供了一种人工大脑。现在让我们割断这群人与我的身体的联系,而让他们与传输、接收信号的计算机联系起来,就像计算机与缸中之脑相联系时很可能发送和接收信号一样。倘若如此,那群人就不会作为心理器官对任何有机体发挥功能作用;在模仿实验中,它正好是不重要但不可缺少的东西。我们还要说那群人有心理状态吗?

直觉强有力地告诉我们:缸中之脑有心理状态。也许,直觉对人们关于假体大脑会说什么是不那么可靠的,这种假体大脑与大脑正常地对之发挥功能作用的那些包容它们的系统并没有联系(或没有以适当方式相联系)。我认为,这并不是目的论功能主义的难题。其他人工器官——尤其是我们称之为“控制装置”的器官——也会产生同样的问题。

健康的心脏控制着它自身的收缩。起搏器就是能完成同样功能的装置。我们不妨假设,起搏器能够通过外科手术置入病人之中,或安放在他身外,而且能通过无线电路耦合使它们发挥作用。此外,一群人还能抽象地模仿起搏器的运作;如果恰到好处地联结于某人的心脏,那么它们本身也能成为起搏器。不过,即使在下述情况下,即那群人割断了与他们为其提供假体大脑的任何有机体的联系,也很难判断:人们对于能够起功能作用而事实上没有功能

作用的装置,会根据它缺乏与“正常”类型的包含系统的联系而说些什么。一个在物理上与连接于计算机的起搏器没有任何区别的装置能完成冠状动脉的调节机制吗?对了,显然没有什么心脏是它能调控的。另一方面,似乎有这样一层意思,即那装置在发挥调节作用。

我不奢望为回答这一类难解之谜指明根本的出路。我的想法是:那些异乎寻常的例子应受到怀疑的对待,而不应提高到检验理论的实例的层次。然而我想强调的是:功能主义在心身问题中面临的困难,从目的论上来理解,其实是关于功能归因的更一般的问题。这种困难出现了,并有其常有的逻辑特征,这样的事实与其说诋毁目的论功能主义,倒不如说是对它的支持。

当我们用目的论功能主义替代图林机功能主义时,另一指向功能主义的批评便一目了然了。这就是布洛克(1978)的“输入与输出难题”。当功能主义者描述下述信息加工过程时,例如当你形成知觉判断时你所利用的那些过程,你就会把这组过程当作有输入和输出的程序。在视知觉的情况下,我们可能要根据光的模式从物理上描述输入;而输出则可能被描述为表征的建构,进而这建构又会影响其他的表征和行为。正如布洛克所述,功能主义的否定的洞见就是要避免沙文主义。我们不应认为:具有心灵的有机体一定在物理上同我们一样。当然这个教训与功能主义怎样描述输入、输出有关联。没有什么理由能说明知觉或认知为什么必须通过我们在视觉中所用的那种光性信号加工而起作用,知觉表征为什么不一定非要在我们碰巧调用的神经元结构中实现。功能主义者应该抛弃这些物理的细微末节。

68 第一部分

这是怎么可能的呢？如果我们剔除的物理细节太多，那么我们最终得到的是关于我们自己心理机制的纯抽象的说明，我们就会回到雇来从事抽象模仿的那群人所碰到的难题。布洛克推测：这个问题是不可解决的——当功能主义者着手决定输入与输出的正确描述时，不是倒向沙文主义一边，就是落入自由主义一边。

目的论功能主义使这个问题具有看起来易于处理的外观。让我们还是来看消化。我们通过某些物理过程消化某些食物。但是消化能够且的确是通过其他物理过程而得以进行的，而且其他有机体在得到食物时是明显有别于我们自己的。正如我们与物理世界的知觉联系是以某些物理信号而不是别的东西为中介一样，我们与世界的能量联系也是以食品而非别的东西为中介。知觉与进食在其机制上可能是多种多样的，就能作为输入的东西而言，也是五花八门的。不过我不明白的是：这为什么意味着功能主义的消化理论的破灭。尽管有多样实现这样的事实，但似乎也存在着某种像共同的硬核一样的东西。粗略地讲，所有消化似乎都有可能涉及到从环境中吸收能量。对于什么是心灵的功能，即使我们没有什么根本的东西要说，但是对布洛克关于知觉中的输出与输入问题的（有点模糊的）回答还是值得考虑的。知觉涉及到从环境中吸取信息。输入应当根据它们携带信息的属性而加以具体说明。光—信息包是多余的，视网膜和神经元也是如此。这些只是构成了我们称之为知觉加工的许多物理实现中的一种实现。对信息流过程作出详细说明，这是科学的任务。我的观点是：在这个抽象的层次，理论化为什么是不可能的，对此我看不到任何先天的理由。

与此同时，我们必须注意到这样的可能性，即科学不可能对常

识心理学中所利用的某些非常抽象和普遍的属性提供理论描述。“知觉”也许会成为太宽泛的范畴,以致不值得加以理论化。如果是这样,批评功能主义心理学没有为作为一般范畴的知觉提供一种功能主义理论就是错误的。确切地说,更为切实可行的观点是:关于不同种类的知觉,可能存在着五花八门的功能主义理论。不管怎么说,事情最终会是怎样,这不是崇尚空谈、不切实际的哲学能予回答的。我的结论是:一当从目的论上解释功能主义,那么布洛克的输入和输出问题就迎刃而解了。

功能主义的另一问题根源于把普特南(1975)关于意义的观点应用于像信念这样的心理状态。孪生地球思想实验告诉我们:两个个体在物理上可能是同一的,但又有不同的信念。你的术语“水”指的是 H_2O , 而你的孪生地球人的术语“水”指称的则是 XYZ。如果功能主义被理解为图林机功能主义,那么看起来就不可能有关于信念的功能主义理论。你的术语“水”和你的孪生地球人的术语“水”有相同的因果敏感性和因果力。 H_2O 或 XYZ 在“正常”条件下的出现会使你们两人形成“存在着水”这样的句子。当那个句子出现在“信念盒”中时,它就会导致饮水,如果你和你的孪生地球人很口渴的话。

在讨论像信念这样的非狭义的心理状态时,功能主义的这种明显的无能并不会使布洛克(1978)和福多(1980)等人感到烦恼,因为他们认为:心理学关心的只是狭义的状态。但是对于像我本人(参阅索伯 1982)这样一些主张有机体的狭义和广义属性在心理学中均有作用的“多元主义者”来说,功能主义对非狭义的状态所表现出的无能为力则是值得注意的事情。

70 第一部分

然而,当再一次想到了目的论功能主义时,问题就大不一样了。关于(目的论)功能的一个简单的事实是:某物的功能是什么取决于它内在于其中的环境的类型(温萨特 1980)。如果把两片物理上同一的金属放在不同的机器中,那么它们就可能有不同的功能。你有你的关于水的概念;你的孪生地球人有关于孪生地球水的概念。如果你和你的孪生地球人在物理上是同一的,那么你们两个怎么可能有不同的概念?我们可以把这种差异看作是功能上的差异。再者,我得强调的是:我不敢妄称知道任何心理项目的功能是什么。但是为了说明目的论功能主义的来源,我得作这样的推测:你的术语“水”有表征水($=\text{H}_2\text{O}$)的功能,而你的孪生地球人的术语则有表征孪生地球水($=\text{XYZ}$)的功能。关于内容的自然主义理论面临着详尽阐述这一简单明了的建议的重任;这种理论怎样发展,仍值得拭目以待。但就现在的情况而言,我认为功能主义根本就不用担心非狭义的心理状态。

我要考察的对功能主义的最后一个责难与感受性质有关。布拉德利、坎贝尔和布洛克首先以不同的方式对之作了阐发。问题可以由颠倒光谱问题来说明。假设你和我在心理特性的其他方面是同一的,所不同的是,红的东西对你来说看起来像绿的,而绿的东西看起来又像红的。另一方面,从心理学上来说,我是“正常的”。对我来说,绿的东西看起来是绿的,红的东西看起来是红的。当我看到绿色的东西(在正常的照明条件下等)时,我就进到了 G 状态,而你进入的是 R 状态。当我看到红色的东西时,我进入的是 R 状态,而你进入的则是 G 状态。在这里,功能主义面对的难题是:即使你的 G(R)状态和我的 R(G)状态在质上是不同的,但

它们从因果关系上说则是一样的。使你进入 $R(G)$ 状态和使我进入 $G(R)$ 状态的东西实际上是同一个东西。我处在 $R(G)$ 状态的结果与你处在 $G(R)$ 状态的结果显然是相同的。例如看到了救火车,在我身上就产生 R 状态,而在你身上则产生 G 状态;我处在 R 状态使我对问题“那是红的吗?”说“是的”(如果其他的心理倾向是正常的),这实际上就是你处在 G 状态所引起的说法。所以那些状态在心理上是彼此有别的,即使它们在因果(功能)作用方面没有不同。因而功能主义无法说明感受性质。

和布洛克(1978)一样,我将假定“感受性质实在论”为真。在你的质的状态与我的质的状态之间的确可能存在着区别,即使这不能在我们的任何外在行为中觉察到。我与布洛克的分歧在于:对感受性质实在论者是否能够成为功能主义者这一问题有不同意见。

为了说明感受性质实在论为什么不一定成为功能主义的难题,试看下面对前述论证的拙劣可笑的模仿:

假设有两种极不相同的计算机程序,它们都把双数作为输入,把那些数字的乘积当作输出。这两种程序在无论多大的兼容计算机中都是可以互换的。因此功能主义不能描述两个程序之间的差异,因为根据假说,两程序起着同一的因果作用。

这个论证的缺点在于:它未能注意到在程序之内可能会发生什么。即使两个程序可能是微弱地等值的,但它们的内在过程在功能上则可以有所不同。

你的感受性质与我的相互有别,即使它们似乎起着相同的因

72 第一部分

果作用。由此是否可以得出结论说：功能主义没有抓住我的绿的感觉性质与你的红的感觉性质之间的差异呢？不，不能这样推论，因为可能有这样的情况，即在内部存在着把两者区别开来的大量隐藏的结构。我们也许把感觉性质体验为单一的东西，但那并不意味着它们就是单一的。就我们所知，可能存在着相当的复杂性，对此功能主义可以加以识别，并在此基础上澄明你的感觉性质与我的感觉性质之间的差异。

基于这个理由，我认为我们不能断定功能主义不能说明感觉性质。症结就在感觉性质是单一的东西这个假定上。尽管如此，我认为从感觉性质问题中我们仍可以吸取一个教训，即功能主义的确有局限性，即使是当那种理论得到了最强的（目的论）表述时也是如此。

在任何功能组织系统中，不管它是消化系统还是知觉系统，我们都可以预期：存在着功能上等值的状态或装置。正如功能主义者所意识到的，物理上不同的状态不仅确实能够在功能上等值，而且事实上心理上不同的状态也可能在它们的心理功能上同一。

因此我赞同的是布洛克的一般的观点，而不是他关于感觉性质的具体结论。我认为，没有任何令人信服的理由断言：功能主义不能说明感觉性质。不过，如果每一心理差异都相当于一种功能差异，那么我是会感到惊讶不已的。这常常成为心理的东西的一个标志，因为它把心灵放在一个完全不同于我们的其他功能系统如消化和呼吸的范畴之中了。

生物学家由于发现两种特性有同一的生物学功能，因此便不会说，在它们之间不存在生物学差异。爪哇岛的犀牛有两只角，而

非洲的犀牛只有一个。没有绝对理由认为它们在功能上存在着任何差异。难道能由此说:在有两只角和只有一只角之间没有生物学的差异吗?同样,在发现某种特性完全没有生物学功能时,我们为什么还要断定它根本就不是一种生物学特性呢?古尔德和莱旺丁(1979)说:人类的下颏没有任何功能;难道能由此说它不是一种生物学项目吗?

生物学提供了大量的事实。关于生物学功能的一个显著的事实就是:在有功能组织的地方,也会有功能组织的人工制品——完全没有功能的项目。当对象整合进一系统时,每一个由于有其独立的功能,因而都会有附带的结果;都会有系统的特性以及有相关的对象的特性,这些对象本身并没有功能,而只有结果——人工制品——事实上存在着的功能的结果。既然生物学不仅研究关于功能组织的事实,而且还研究功能组织的结果,因此生物学也会研究人工制品。

心理学家研究的是知觉与认知系统的功能组织。但是由于与对象的任何系统联在一起,因此功能之存在差不多证实了人工制品的存在。如果心理学家研究功能组织的结果,那么他们最终也会研究没有心理功能的特性。他们也希望对功能上等价的特性作出区分。

因此我的看法是:如果把功能主义当作关于伴随着每一心理状态的同一条件的一种主张,那么它就是一种稀奇古怪的学说。功能主义似乎设想:存在着完善的组织——每种心理状态都有其独一无二的、个体化的心理功能。步古尔德和莱旺丁(1979)的后

74 第一部分

尘,我把这种假定称之为潘格罗式的* 假定;试想伏尔泰的《老实人》一书中的潘格罗,他主张:一切都有作用,这种作用是它以最佳方式实现的。潘格罗式的功能主义具有简单性这一优点。但功能组织比这杂乱得多,也许必然如此。

既然我在本文中大言不惭地盗用了功能概念,那么我就应该对我怎样理解那个隐晦的观念作几句交待。有一种关于功能的最低限度的解释,它曾得到了卡明斯的说明(1975),根据这一解释,每一事物都有一种功能。一种包容系统中的某一部分的功能就是那个部分所完成的、有助于该系统具有它所具有的属性的一切事情。这是一个用装配线所作的隐喻:组成部分的功能就是它们所做的、使装配线履行它应履行的工作的一切事情。根据这种观点,说心脏的功能就是产生某种力量是没有什么错误的。心脏有一种特殊的力量,由此使身体能够以某种力量、有时是更大的力量让天秤倾斜。在卡明斯看来,这种对功能的归因没有什么错误。但是它在我们看来并不十分有趣。很明显,根据这种分析,没有事物不具有功能。而且任何两个具有不同类型的结果的装置都会有不同的功能。功能无处不在,在功能中,差异俯拾即是。

我无意于证明:这种分析是错误的。不过,我们但愿能够说,在心理学中如同在生物学中一样,某些项目没有功能,而其他的则有同一的功能,即使就它们的结果而言它们之间存在着明显的不同。这意思是说,我们必须对我们把功能归因于包容系统的某些

* 潘格罗系伏尔泰的《老实人》一书中的哲学家,他认为世界上的一切将臻于至善。意即过分乐观的。

占优势的状态这一点用相对论的术语加以描述。例如,如果我们想说阑尾没有功能,我们大概可以这样加以表述:就消化而言,阑尾的确没有功能。如果我们想说,我的红的感受性质和你的绿的感受性质有相同的功能,那么这一定意味着,就某些心理过程而言,它们有同一的功能。

没有这种相对论的描述和附加的细节,主张心理状态是功能状态就难以令人置信,就没有什么意义。每一心理状态都有其原因和结果,而且如果我们观察得足够仔细,那么我们就总是可以发现任何两个不同的心理状态之间的因果差异。事实上,如果没有相对论描述,功能主义可能与同一论完全同流合污了,因为物理状态也能根据它们的因果联系而加以个体化(利康 1981)。另一方面,一当功能主义完全用相对论术语加以描述,那么它就成了一种可疑的、潘格罗式的理论。任何生物学状态都不是依照它们的生物学功能而加以个体化的。我想心理学也不例外。

潘格罗式的假定对于一门学科来说可能是有用的,尤其是当该学科相对年轻、且理论相对不成熟时更是如此。每一特性都有一种功能、且最大限度地行使其功能,这一观点在进化论中是有用的,如果只是作为提出问题的向导的话(这并不否认该观点常被滥用)。但是功能理论成熟的一个标志是:对于一种特性是否有功能、两个特性是否真的在功能上等值这样的问题,它能够提供某种独立的原始证据。

在心理学及心灵哲学中,潘格罗式的观点常出现于许多情况之下。我这里考察的正好是其中的一种。值得提及的另一种是:关于理性地位的假定;许多哲学家认为:要把信念归因于动原就必

76 第一部分

须把它们当作是有理性的。据说,这不仅是一个实际的权宜之计,而且是一个先天的要求。我的看法是:它是一个潘格罗式的简单化,对于还不够成熟的理论化而言是有用的,但不是不可缺少的(索伯 1978,1985)。

由于潘格罗式的假定在生物学中已无可挽回地失去了其地位,因此生物学哲学中的类似问题就不那么令人感兴趣了。物活论以及关于生物状态、属性与物理状态、属性之间的关系的同类的理论过去的命运如何呢?以为唯物主义(即类型/类型同一论)在那些时候已取得了胜利,就大错特错了。功能主义的否定观点适用于生物学,就像适用于心理学一样。适者生存之类的概念不可能在类型上还原于物理概念(罗森堡 1978;索伯 1984b)。我曾说过:我们现在对作为范畴的生物状态及属性的本质尚没有适当的哲学理论;我还要补充的是:这种过于一般性的问题一开始就是索然无味的问题。生物学诸领域中的进展已消除了使物活论所提出的哲学问题如此令人钟情的神秘性。各别的生物学理论限定了自己的范围,并以相当异质的方式描述了它们自己的属性。这种生物学的类比表明:认知科学的进步除了放宽了对潘格罗式的假定的理解以外,还将改变我们关于功能主义观念原本要解决的某些问题的观点。

参考文献

N. 布洛克(1978):“功能主义的麻烦”,载 W. 萨维奇编:《知觉与认知:关于心理学基础的争论》,《明尼苏达科学哲学研究》,第 9 卷,明尼苏达大学出版社。

N. 布洛克和 J. 福多(1972):“心理状态不是什么”,《哲学评论》,81,159-82。

M. 布拉德利(1964):“斯马特的‘哲学和科学实在论’之批判性评论”,《澳大利亚哲学杂志》,71,262-83。

K. 坎贝尔(1970):《身与心》,纽约:道布尔戴。

R. 卡明斯(1975):“功能分析”,《哲学杂志》,72,741-64。

R. 戴维斯(1974):“功能定义以及怎样感觉处在疼痛中”,未发表手稿。

D. 丹尼特(1978a):“论意识的认知理论”,载《脑猝死》,蒙哥马利,佛蒙特:布拉福德丛书。

D. 丹尼特(1978b):“你为什么不能制造出可感觉到疼痛的计算机”,载《脑猝死》。

J. 福多(1968):《心理学解释》,纽约:蓝登书屋。

J. 福多(1980):“作为认知科学研究战略的方法论唯我论”,《行为与脑科学》,3,63-100。

S. 古尔德和 R. 莱旺丁(1979):“圣马科和潘格罗式图式的拱上空间:对适应主义纲领的批判”,《伦敦皇家协会论文汇编》,B205,581-598。

W. 利康(1979):“反机器功能主义的新小人国论证”,《哲学研究》,35,279-87。

W. 利康(1981):“形式、功能和感觉”,《哲学杂志》,78,24-50。

H. 普特南(1967):“心理谓词”,载 W. 卡皮坦等编:《艺术、心灵和宗教》,匹兹堡大学出版社。

H. 普特南(1975):“‘意义’的意义”,载 K. 冈德森编:《明尼苏达科学哲学研究》,第7卷,明尼苏达大学出版社。

A. 罗森堡(1978):“心理主义”,《社会行为理论杂志》,8,165-191。

E. 索伯(1982):“理性生物学和自然主义生物学”,《行为与脑科学》,5,300-2。

E. 索伯(1984a):《进化生物学中的概念争论选集》,麻省理工学院出版社。

E. 索伯(1984b):《选择的本质:哲学聚焦中的进化理论》,麻省理工学院出版社。

E. 索伯(1985):“潘格罗式的功能主义与心灵哲学”,《综合》,64,165-193。

W. 温萨特(1980):“目的论与功能陈述的逻辑结构”,《科学历史与哲学研究》,3,1-80。

第二部分

感 受 性 质

副现象的感受性质*

弗兰克·杰克逊

不可否认,关于我们生活于其中的世界和关于我们自身,物理的、化学的和生物的科学已提供了大量的信息。我将把这类信息以及习惯上伴随它的信息称为“物理信息”。例如,如果医学家详细地告诉我,关于我的神经系统中发生的过程,关于这些过程怎样相关于我周围的世界中的现象,怎样相关于过去发生了的东西和将来还可能发生的东西,怎样相关于在其他相同或不同的有机体身上可能发生的東西等,那么他或她就是告诉我——如果我很聪明,足以对之作出适当的整合——关于通常被称之为这些状态在我身上(以及在一般的有机体的相似的情况下)的功能作用的事情。这种信息及其家族,我也称之为“物理的”。

我的意思不是说,这些简略的说明构成了关于“物理信息”、关于物理属性、过程等有关概念的定义,而只是说,这些说明表达了我在这里所想到的东西。众所周知,对这些概念以及物理主义的下述论点,即所有(正确的)信息都是物理信息^①给出明确的定义,是有很多困难的。但是——不像某些人——我把定义的问题当作是解决我本文想讨论的中心问题的一种捷径。

* 译自《哲学季刊》,32(1982),pp. 127 - 136。

我有时就是以“有感受性质的畸形人”著称的。我认为,躯体感觉尤其有许多特征,当然某些知觉经验也是如此,它们在任何纯物理信息都不包含的。告诉我物理的东西,就等于告诉我:在活的大脑中发生了什么,有什么类型的状态,它们的功能作用,它们与别的时候、在别的大脑中发生的東西的关系,如此等等,不一而足,如果我如此聪明,能够把它们整合起来,那么你并没有告诉我:疼痛的伤害性、头痛的疼痛性、妒嫉的感觉,或者说尝柠檬汁、闻玫瑰花、听噪音或看天空时的有独特特征的经验。

存在着许多有感受性质的畸形人,其中一些说:他们对物理主义的拒斥是一种未经论证的直觉。^②我认为,他们对自己是不公正的。他们有下列的论证。例如关于物理的东西,你所能说的并不能把握玫瑰花的味道。因此物理主义是错误的。根据我们的标准,这是一个相当漂亮的论证。怀疑它的有效性显然是不妥当的,其前提从直观上说对他们和我显然都是真的。

不过,我必须承认,从辩论的观点看,它是脆弱的。对我们来说,不幸的是,有这样一些人,他们并没有看到,那个前提从直觉上说是明白无误的。这样一来,接下的任务就是提出一个论证,其前提对所有的人都是明白无误的,至少对多数人是这样。这就是我在第1节中想做的工作,我将把它称作“知识论证”。在第2节中,我将对知识论证与模态论证以及第3节的“可能是什么样子”的论证作出比较。在第4节中,我将集中讨论感受性质的因果作用问题。妨碍人们承认感受性质的主要因素就是这样的信念,即相信:就物理世界、尤其是大脑来说,它们一定被赋予了因果作用;^③如果不像信仰仙界的某人那样大声呼唤,是很难完成这个任务的。

在第4节中,我将转向这个反对意见,方法就是去证明感受性质是副现象这一观点完全是可能的。

I. 关于感受性质的知识论证

人们在分辨颜色的能力方面千差万别。假设在一项旨在对这种差异进行分类编目的实验研究中,找到了弗雷德。他有比别的、记录在案的任何人都强的颜色视觉;他作出了其他任何人曾作出的每一种分辨,而且他还作了一个我们不曾作出的分辨。给他一堆成熟的西红柿,他能把它们分成大致相等的两组,而且他能完全一模一样地重复这一工作。即是说,如果你蒙住他的眼睛,把西红柿弄乱,再松开他的眼睛,要求他再对其分类,结果他把它们分成了与以前一样的两组。

我们问弗雷德:你是怎样做到这一点的。他解释说:所有成熟的西红柿对他来说,其颜色看上去是不一样的,事实上这对我们归之为红的大量对象都是适用的。在我们看到只有一种颜色的地方,他看到了两种颜色,结果,他创造了自己的用法,让单词“红₁”和“红₂”标示其差异。他也许会告诉我们说,他常想把“红₁”与“红₂”之间的差异教给他的朋友,但行不通,因此他得出结论说:世界上其他的人都是红₁-红₂的色盲——或许他在他的小孩身上已取得了一部分成功;但那没有关系。无论如何,他对我们解释说:以为“红”表现为红₁和红₂,以致两种颜色是同一种颜色的样式,这可能是极其错误的。他只用了日常用语“红”,以便更方便地适合于我们有限制的使用。在他看来,红₁与红₂是彼此不同的,也不同

84 第二部分

于其他一切颜色,就像黄色不同于蓝色一样。他的分辨行为证明了这一点:在广泛的观察环境中,他易如反掌地把西红柿分成了红₁与红₂。而且对弗雷德的独特能力的生理学基础的研究表明:他的视觉系统能在红色光谱中明快地分出两组波长,就像我们能 把黄色与蓝色分开一样。^④

我认为,我们应当承认:弗雷德能看到、事实上已看到的颜色比我们看到的至少要多一种;红₁是一种不同于红₂的颜色。我们对于弗雷德就像一个完全的红-绿色盲对于我们一样。H. 威尔士的故事“盲人之国”讲的就是生活在十足的盲人共同体中的一个有视力的人。^⑤这个人从来不打算让他们相信:他能看物,即他有一种例外的感觉。他们把这种感觉当作完全不可想象的东西加以嘲笑,把他的能不掉进沟渠的能力、能取得战斗胜利的能力等恰好当作是那种不可想象的能力,不过如此而已。如果我们拒绝承认弗雷德能比我们多看到一种颜色,那么我们就犯了那些盲人的错误。

当弗雷德看到红₁和红₂时,他有什么样的经验呢?这新的颜色或类似的颜色是什么?我们很想知道,但做不到;关于弗雷德的大脑和视觉系统的任何物理信息似乎也不能告诉我们什么。我们也许会发现:弗雷德的视锥能对红的光谱中某些光波分别作出不同的反应,而这些光波对我们的视锥却是没有区别的(或许他有例外的视锥),我们也许还会发现:这在弗雷德身上引起了许许多多的大脑状态,这些状态使他能完成视觉分辨行为。但这并没有告诉我们:关于他的颜色经验我们真的想知道的东西。关于它,存在着我们不知道的东西。不过我们可以假设,我们知道他的身体以

及他的行为、他行为的倾向、他的内在生理学的一切,知道他的历史、他与他人的关系的一切,而这些又能在关于人的物理说明中被提供出来。我们拥有一切物理信息。因此知道了这一切并不等于就知道了关于弗雷德的一切。由此可以得出结论说:物理主义遗漏了某种东西。

为了加强这一结论,试想:由于我们研究弗雷德的内在活动,因而我们认识到了:怎样才能使每个人的生理机能在有关方面像弗雷德的那样;或许弗雷德愿把他的躯体捐献给科学,因此在他死的时候,我们就能把他的视觉系统移植到另一个人身上——再者,精密的处理都不成问题。关键在于:这样的事情会带来极大的好处。人们也许会说:“我们最终将知道:看到例外的颜色可能是什么样子,我们最终将知道:弗雷德怎样不同于我们,而在这点上,弗雷德对把它们告诉我们长期存有顾虑。”因此我们不可能始终如一地知道关于弗雷德的一切。但是关于弗雷德,假如我们始终知道那在物理主义图式中起着重要作用的一切;因此物理主义图式遗漏了某种东西。

可以这样加以表述:在手术之后,我们将知道更多关于弗雷德的东西,尤其是关于他的颜色经验。但是前此,对于他的身体、大脑以及那些曾在物理主义对心灵和意识的说明中起着重要作用的一切,我们已经得到了我们想得到的一切物理信息。可见应知道的比那一切还要多。因此物理主义是不完善的。

弗雷德和新的颜色自然是必不可少的修辞手段。对于正常的人和常见的颜色也可以如是说。玛丽是一位杰出的科学家,不知何故,她不得不在黑白的房间借助黑白电视监视器去研究世界。

86 第二部分

她精通关于视觉的神经生理学,而且我们不妨假定,当我们看成熟的西红柿或天空,使用“红”“蓝”等术语时,她得到了关于所发生的、应知的一切的物理信息。例如她发现,哪一些来自天空的波长组合刺激了视网膜,这实际上又是怎样通过中枢神经系统产生声带的挛缩、肺中气息的膨胀的,正是后者导致了“天空是蓝色的”这样的句子的说出。(几乎不能否认,从黑白电视中获取所有这些物理信息在原则上是可能的,否则电视大学就一定得用彩色电视机。)

当玛丽从黑白房间中被放出来了,或者被给予了一台彩色电视监视器,那么会发生什么呢?她学到了什么东西没有?似乎很明显,关于世界,关于我们对它的视觉经验,她将学到某种东西。但这样一来,她先前的知识不可避免地具有不完善性,当然她有一切物理信息。因此还有比那多得多的东西,这一来,物理主义就是错误的。

很显然,知识论证的这种方式可以一无例外地推广到味觉、听觉、躯体感觉以及一般所说的各种心理状态上,它们据说有(以不同方式表述的)原(raw)感觉、现象特征或感受性质。在每种情况下,结论都一样:感受性质是物理主义描述所遗漏的东西。知识论证的富于论战性的力量在于:要否认下述核心主张是很困难的,这主张是:人们可能有一切物理信息,而并没有一切应有的信息。

II. 模态论证

我所说的模态论证指的是具有下述风格的论证。^⑥关于他心

的怀疑论者在演绎逻辑中并未犯错误,不管他们的观点有其他什么错误。关于他人的任何物理信息在逻辑上并不能衍推出:她或他意识到或感觉到什么东西。因此有这样一个可能世界,它有这样的有机体,就它们没有有意识的心理生活而言,它们极不同于我们,而在其他一切物理方面(记住包括功能状态、物理历史等)实际上与我们毫无两样。但是这一来,我们有的、而它们没有的东西是什么呢?据假设,不是任何物理的东西。在所有物理方面,我们与它们实际上一模一样。因此就我们来说,除了纯物理的东西之外,我们还有更多的东西。因此物理主义是错误的。^⑦

有时有这样的反对意见,即认为模态论证以那学说被作为偶然的真理加以阐发为理由,错误地设想了物理主义。^⑧但是这样说只不过是说:物理主义者把他们的主张限制在某些可能世界之中,尤其包括我们的世界;模态论证只是指向了这种次要的主张。除了别的可能世界中的、我们的物理复制品的那些特征之外,如果我们在我们的世界,且不说在其他的世界的生物,还有一些特征,那么我们就有非物理的特征或感受性质。

模态论证的困难在于;它依赖的是有争议的模态直觉。因为它被争论,因而是有争议的。一些人诚实地否认:在没有意识的另外的可能世界中可能存在着我们的物理复制品。而且曾有过那种直觉的人,至少现在会表示怀疑。^⑨

民意调查对于模态论证的讨论可能是一种蹩脚的方案。但是当模态直觉是有争议的时候,我们常常没有更好的办法,而且我们可能还记得,我们最初的目标就是寻找那个论证的最大的辩论价值。

我们作为知识论证的倡导者很可能接受有争议的模式直觉；但是这是我们对感受性质是物理主义理论所遗漏的东西这一结论的已作出的论证的一个结果，而不是那个结论的根据。而且，问题由于下述可能性而变得复杂了，这个可能性是：物理的东西与感受性质的联系可能类似于这样的联系，即有时在美学性质与自然性质之间可发现的那种联系。在所有“自然方面”（包括有知觉力的生物的经验）一致的两个可能世界，在美学性质上也一定一致，但是可以合理地说：美学性质不能还原为自然性质。

Ⅲ.“成为……可能是什么样子”论证

在“成为一只蝙蝠可能是什么样子？”一文中，T. 内格尔论证说：物理信息无论如何不能告诉我们：成为一只蝙蝠可能是什么样子，因而我们人类的确不能想象成为一只蝙蝠会是什么样子。^⑩他的理由是：这会是什么样子，只能从蝙蝠的主观观点出发才能得到理解，而它的观点不是我们的观点，不是用物理的术语能够把握的东西，因为物理的术语必然是用许多观点都同样能够得到理解的术语。

把这个论证与知识论证区别开来极为重要。当我抱怨说，关于弗雷德的所有物理知识不足以告诉我们：他的特定的颜色经验可能是什么样子时，我并不是在抱怨：我们没有认识到成为弗雷德可能是什么样子。我抱怨的是：关于他的经验，存在着某种东西，即他的经验的一种属性，对此我们一无所知。等到我们最终知道这种属性是什么时，我们仍不知道成为弗雷德可能是什么样子，当

然我们将知道关于他的更多的东西。关于弗雷德的任何知识,不管是不是物理的,都不能算是“从内在的方面”考察弗雷德而得到的知识。我们不是弗雷德。因而存在着这样一整套知识单元,它们由下述语词形式表达出来:“那就是我自己,我自己是……”,弗雷德有而我们完全没有这种知识,因为我们不是他。^①

当弗雷德观看只能由他看到的颜色时,他知道的一件事就是:他对那颜色的经验区别于他看红色等的经验的方式;另一件事是,他自己看到了那颜色。物理主义者与有感受性质的畸形人一样会承认:别的人关于弗雷德的任何信息都不能认为是关于第二件事的知识。不过,我的抱怨是与第一个东西有关的,那就是他的经验的特殊的质肯定是一种关于他的经验的事实,这是物理主义所遗漏了的,因为任何物理信息都不能告诉我们:它是什么。

内格尔好像是说,他所提出的问题是这样的推断问题,即根据关于一种经验的知识推断另一种经验,它也是一个设想问题,即根据熟悉的经验设想不熟悉的经验可能是什么。拿休谟的例子来说,根据关于蓝颜色的某些样式的知识,我们能得知:观看蓝色的其他样式会是什么情形。内格尔论证说:蝙蝠等所面临的麻烦在于:它们也不同于我们。在这里很难看到对物理主义的驳难。物理主义对人类的想象力和推断力并没有作出特殊的断言,因此很难弄明白它为什么必须这样做。^②

无论如何,我们的知识论证对这一点并没有作出任何假定。如果物理主义是对的,那么关于弗雷德的足够的物理信息就会排除这样的必要性,即为了弄清关于他的特殊的颜色经验的一切,必须进行推断或必须运用特殊的想象或理解技能。这种信息已为我

们所掌握。但显然不是这样。那就是该论证的要害。

IV. 副现象论的幽灵

真的有充分的理由拒绝赞成这样的观点,即感受性质在因果上对物理世界是苍白无力的吗?我将为否定的回答作出论证,但是在论证时,对于与古典的副现象论有联系的两个观点,我不打算说什么。第一是认为,心理状态对物理世界是无效力的。我想作出辩护的不过是:坚持下述观点是可能的,即某些心理状态的某些属性即我称之为感受性质的属性,是这样的属性,有它们或没有它们对物理世界没有什么影响。第二个观点是:心理的东西完全没有因果效力。虽然如此,我要说的是:你不得不坚持认为:感受性质的例示即使对任何物理的东西没有用,但对其他心理状态会造成影响。而且只要一般地想一想你最终怎么可能知道感受性质的例示,就能使人想到这样的结论。^⑬

疼痛之伤害性这样的感受性质在物理世界必然有因果效力,因此例如它的例示对大脑中将发生什么有时一定会有影响,对这一观点,通常可能提供三个标准的理由。我们要论证的是:没有一个是真有力量的。(我非常感谢 A. 希思洛普和 J. 卢卡斯,是他们使我信服这一点的。)

(1)据说,疼痛之伤害性显然对主体试图避免疼痛负有部分的责任,比如说“很痛”等等。但是如果把休谟的话倒转过来,那就是,任何东西都不能引起任何东西。不管 B 怎样经常跟着 A 出现,不管联系的因果性在开始似乎多么明显,A 引起 B 这一假说可

能为下述总合(over-arching)理论所推翻,这理论说明,两者可表现为—共同的、根本性的因果过程的不同结果。

对于未受正规训练的人来说,屏幕上 L. 马文的拳头从左到右运动的图像之后紧接着的是 J. 韦恩的头沿着相同的方向运动的图像,这看起来像是一种因果关系。^⑭相似于第一类图像的无数西方人的图像之后伴随的是相似于第二类图像的图像。当我们懂得了总合理论所说的有关的图像怎样成为一个与幻灯、电影有关的根本的因果过程的两个结果时,所有这一切都一钱不值。副现象论者对比比如说伤害与行为之间的关系事实上也可以说出同样的话来。事实的结论显然是,两者均由大脑中的某些事件所引起。

(2)第二个反对意见与达尔文的进化论有关。按照自然选择,在时间中进化的特质是那些对肉体的生存有用的特质。我们可以假定:感受性质是在时间中进化形成的——我们有它们,最早的生命形式却没有——因此我们应当把感受性质当作对生存有用的东西。反对的意见是:如果它们对物理世界没有什么作用,那么它们对于我们生存几乎不能有什么助益。

求助于这个论证是无可厚非的,但对它也有漂亮的答复。北极熊有特殊的、厚而暖和的外套。对此,进化论是这样解释的(我们假定):有厚而暖和的外套是有益于在北极的生存的。但有厚外套与有重的外套是结伴而来的,而有重外套则不利于生存。它使动物的速度慢下来。

因为我们发现了一个进化的特质——有重外套——它不利于生存,就意味着我们驳倒了达尔文吗?显然不是。有重外套是有温暖的外套的一个不可避免的伴随物(在那个环境下,现代的绝热

92 第二部分

物是无效的),有温暖外套对生存的益处价值上大大超过了有重外套对生存的害处。总之,我们从达尔文的理论中吸取的东西不过是:我们希望任何进化的特征要么对生存有益,要么是对生存有益的东西的一个副产品。副现象论者坚持认为:感受性质属于后一范畴。它们是那些对生存极有助益的大脑过程的副产物。

(3)第三个反对意见是以关于我们怎样知道他心的观点为根据的。我们通过知道他人的行为而知道他心,至少部分如此。这个推理的本质是有某种争议的,但它从行为出发,这是没有争议的。这就是我们为什么认为石头没有感觉而狗有感觉的理由。但是根据那种反对意见,一个人的行为怎么可能为相信他有同我一样的感受性质甚或任何感受性质提供理由,除非这行为能被当作是那感受性质的结果。M. 弗莱德(鲁滨逊在孤岛上的唯一的人类朋友。——译注)的脚印是 M. 弗莱德的证据,因为脚印是从属于人的脚的、因果上的结果。副现象论者不能把行为甚或任何物理的东西当作是感受性质的结果。

但是试考察:我在《时代》杂志上获悉斯帕斯获胜。这为《电讯报》上报道的斯帕斯取胜提供了极好的证据,尽管有这样的事实,即(我相信)《电讯报》不是从《时代》上得到那个结果的。它们各自让自己的记者到比赛现场。《电讯报》的报道无论如何不是《时代》的结果,但是后者为前者提供了极好的证据。

因此有关的推理可以予以重构。我在《时代》上读到:斯帕斯取胜了。这使我有理由认为:斯帕斯取胜了,因为我知道斯帕斯的获胜是引起《时代》予以报道的最有可能的候选者。但我也知道:斯帕斯的获胜会有许多结果,包括在《电讯报》上的属实的报道。

我能根据一个结果,溯推它的原因,然后再从原因推论另一个结果。结果不再引起另外的东西这一事实是无关紧要的。副现象论者承认:感受性质是大脑中发生的東西的结果。感受性质不引起物理的东西,但是是由某种物理的东西所引起的。因此副现象论者能从别人的行为推论出别人的感受性质,方法是根据别人的行为溯推它在别人大脑中的原因,进而再由此推出他们的感受性质。

这个推理链比在新闻报道中它的模型更可疑,对此你很可能已感觉到了这样或那样的理由。你是对的。他心问题是重大的哲学问题,而其他新闻报道中的问题则不是。但在这里,对作为相互作用论的对立面的副现象论来说并没有什么特殊的难题。

对我所作的三个答复,有一很好理解的反应。“不错,对副现象的感受性质之存在的确不存在有分量的反驳。但事实依然是:它们是多余无用的。它们什么也不能做,什么也不能解释,它们只能用来抚慰二元论者的直觉,而它们怎样与科学的世界观相一致,这是留下来的一个绝对的秘密。总之,对于它们怎么样、为什么这样的问题,我们没法理解,也不可能理解。”

这一点也不假;但它不是对感受性质的驳难,因为它依赖于一种关于人类、动物及其力量的过分的乐观主义观点。我们是进化的产物。为了生存,我们理解和感觉到什么是我们必须理解和感觉的。副现象的感受性质与生存完全无关。在我们进化的任何阶段,自然选择不会偏袒这样的人,他们懂得他们是怎样被产生的,知道主宰着他们的规律,或者说知道他们事实上为什么存在。这就是我们为什么不能理解的理由。

94 第二部分

尚未足够估计到的是:物理主义是关于我们的力量的一种极端的乐观主义观点。如果它是真的,那么在公认的、非常广泛的纲要中,我们对我们在事物图景中的地位就有一种把握。而某些极其复杂的问题使我们未能如愿——存在着数不胜数的神经元——但在原则上我们还是把握了。试看前件的或然性:宇宙中的每一事物都是在某方面或别的方面与人类的生存密切相关的东西。其或然性肯定是很低的。但是这一来人们一定会承认:存在着这样的可能性,即事物的整个图式有这样一部分,也许是很大的一部分,任何进化都不会让我们接近关于它的知识或对它的理解。因为理由很简单,那就是这种知识及理解与生存没有关系。

物理主义者常常强调:根据他们的观点,我们是自然的一部分,这是足够公正的。但是如果我们是自然的一部分,那么我们就是这样的一部分,即自然把我们甩在后面了,不管自然的进化花了多少年。在那个进化阶梯上,每一步正好是由保护或提高生存价值的需要所约束的几率问题。奇怪的是,我们理解的实际上就是我们所理解的,毫不奇怪的是,存在着大大超出我们理解力的问题。副现象的感受性质怎样与事物的这种图式一致也许正好就是这样的问题。

这似乎是关于我们系统阐述我们的世界以及我们在其中的地位的真正广泛的图景的能力的一种极度的悲观主义观点。但是假设我们发现:在最深海底生活着一种具有智力的海参;也许在那种条件下生存需要智力。尽管它们有智力,但这些海参与我们的智力相比却只有关于世界的非常有限的概念,而我们能为这种存在提供关于其周围环境的本质的解释。即便如此,它们发明了科

学,这些科学以那些有限的术语令人惊讶地发挥着作用。它们也有哲学家,即所谓的海参家。一些称自己是意志坚强的海参家,另一些则自认为是意志薄弱的海参家。

意志坚强的海参家认为:有限的术语(或极像这样的术语,即随着它们的科学的进步而被引进的术语)在原则上足以毫无遗漏地描述一切。这些意志坚定的海参家在懦弱的时刻也承认这样的感觉,即它们的理论遗漏了某种东西。它们抵制这种感觉,抵制它们的敌手即意志薄弱的海参家,指出——绝对正确地——任何海参家在详细阐明这种神秘的残渣怎样与它们的科学所拥有的、它们的科学所阐发的、关于它们的世界怎样运作的极其成功的观点相一致时,都不曾成功过。

我们的海参是不存在的,但它们可能如此。可能还有这样超级的存在,它们对我们是存在的,就像我们对那些海参是存在的一样。我们不能采取这些超级存在的观点,因为我们不是它们,但我认为,这样一种透视法可能是极端乐观主义的解毒剂。

注 释

① 例如可看 D.H. 梅勒(1973):“唯物主义与现象性质”,载《亚里士多德学会增刊》,47, pp. 107 - 119。以及 J.W. 柯曼(1971),《唯物主义与感觉》,纽黑文与伦敦。

② 正在讨论中,但可参阅,如 K. 坎贝尔(1976):《形而上学》,贝尔蒙特, p. 67。

③ 例如可看 D.C. 丹尼特(1978):“心灵哲学中的流行争论”,载《美国哲学季刊》,15, pp. 249 - 261。

④ 参阅, E.H. 兰德(1959):“对颜色视觉的实验”,载《科学美国人》,200 (1959, 5, 5), pp. 84 - 99。

96 第二部分

⑤ H.G. 威尔斯:《盲人王国与其他故事》,伦敦,无日期。

⑥ 参阅 K. 坎贝尔(1970):《身与心》,纽约;R. 柯克(1974):“知觉能力与行为”,载《心灵》,83,pp. 43-60。

⑦ 我是用世界间而不是通常所说的世界内的这种形式表达这一论证的,旨在避免与随附性、异态性等有关的不必要的复杂性。

⑧ 例如可参阅, W.G. 利康(1979):“反机器功能主义的新小人国论证”,载《哲学研究》,35,pp. 279-287, p. 280;唐·洛克(1976):“僵尸、患精神分裂症者和纯物理对象”,载《心灵》,85,pp. 97-99。

⑨ 参阅 R. 柯克(1979):“从物理的明晰性到血肉丰满的唯物主义”,载《美国哲学季刊》,29,pp. 229-239;还可参阅反模态直觉的论证,如 S. 休梅克(1975):“功能主义与感受性质”,载《哲学研究》,27,pp. 291-315。

⑩ 《哲学评论》,83(1974),pp. 435-450。关于这篇文章,有两点需说一说。其一,尽管我与此相左,但我还是感激不尽的。其二是要强调文章观点的变化,到最后,与其说内格尔反对物理主义,不如说他反对所有不承认他所说的“观点”的心灵理论,包括承认感受性质(不可还原)的那些理论。

⑪ 用 D. 刘易斯的术语,是关于实在的知识,见“关于名称与关于实在的态度”,载《哲学评论》,88(1979),pp. 513-543。

⑫ 参阅 L. 尼米罗对“……是什么”的评论,见他对 T. 内格尔发表在《哲学评论》(89(1980),pp. 473-477。)上的“可朽问题”的评论。这里我尤其感谢与 D. 刘易斯的讨论。

⑬ 参阅我对 K. 坎贝尔的《身与心》的评论,载《澳大利亚哲学杂志》,50(1972),pp. 77-80。

⑭ 试比较 J. 皮亚杰(1977):“儿童关于物理因果性的概念”,重印于《皮亚杰基本读本》,伦敦。

玛丽不知道什么*

弗兰克·杰克逊

玛丽被关闭在一个只有黑白两色的房间里,对她的教育是通过只有黑白两色的书和黑白电视机转播的讲座来进行的。这样她学会了关于世界的物理本质应知道的一切。在“物理的”一词的宽泛意义上,她知道了关于我们自己及其环境的一切。这里所谓的“物理的”包括的是全部的物理学、化学和神经生理学中的一切,以及所有应该知道的关于原因的、由之而产生的关系的事实,当然也包括功能作用。如果物理主义是真的,那么她知道了所有应该知道的一切,因为假设别的东西就无异于假设应知道的还有每一物理事实之外的东西,而这正是物理主义所否认的。

物理主义不是这样的没有争议的命题,即现实世界主要是物理的,而是下述富有挑战性的命题:它完全是物理的。这就是物理主义者为什么必然坚持全部物理知识绝对就是全部知识的原因。因为假若它不是全部的,那么我们的世界必定不同于这样一个世界,即 $W(P)$,对它来说,它是全部的,而且这个不同一定存在于非物理的事实中;因为我们的世界和 $W(P)$ 在所有的物理的质料方

* 我极大地受惠于与 D. 刘易斯和 R. 帕吉特的讨论,谨致谢忱。(译者按:本文译自《哲学杂志》,LXXXIII,1986 年,第 5 期。)

面是一致的。因此,对我们的世界而言,物理主义是虚妄的(虽然偶然如此,因为对于 $W(P)$ 来说,它会是真的。)^①

但是,玛丽似乎并不知道所有应知的东西。因为当她从黑白房子里被放出来时,或者在被给予了彩色电视机时,她将会得知,比如说看见某物之红色可能是什么样子。这可以恰如其分地描述为学习——她不会说“嗨,哼。”因此,物理主义是虚假的。这是以其诸种形式之一表现出来的反物理主义的知识论证。^②这篇随笔是对 P.M. 丘奇兰德* 所提出的关于它的三个责难的答复。

I. 三点澄清

知识论证并不依赖于下述含糊不清的主张:从逻辑上讲,你不能想象感知红色可能是什么,除非你曾感知过红色。想象的力量无济于事。关于玛丽的争论不在于:虽然她对于神经生理学和其他一切物理的东西极为精通,但她不能想象感知红色可能是什么;争论恰恰在于,实际上她不知道那一点。但如果物理主义是真的,那么她就会知道;而且不用诉诸于想象的巨大力量。想象是缺乏知识的人必须求助的一种能力。

其次,知识的内涵性无济于事。这个论证并不依赖于下述错误的假定:如果 S 知道 a 是 F,且 $a=b$,那么 S 知道 b 是 F。它涉及到玛丽被放出来之前她的整体知识的本质:它是全部的呢,还是有

* “还原、感受性质和大脑状态的直接的内省”,《哲学杂志》,LXXXII,1(1985,1): 8-28,除非另有说明,后面的引文页码均出自于此文。

些事实被它漏掉了？中肯的观点是：S 可能知道 a 是 F ，而且知道 $a=b$ ，不过大概不知道 b 是 F ，因为在逻辑上没有足够的理由完成这些推论。如果玛丽缺乏的知识的的确似这个样子，那么其中就没有对物理主义的威胁。但是，令人难以置信的是，她缺乏知识只有借助对她的广泛的物理知识的清晰的、合乎逻辑的推论才能加以弥补。赋予她以巨大的逻辑睿智和恒心，这本身不足以填补她知识上的空白。在她被放出来时，她不会说：“我此前就能通过一些更纯的逻辑推理，搞清楚所有这一切的本性。”

第三，玛丽所缺乏的、对于反对物理主义的知识论证具有特别意义的知识，是关于别人经验的知识，而不是关于她自己的知识。在她被放出来时，她有了新的经验，即她从前从没有过的颜色经验。因此，在她被放出来时，她学到了某种东西，这不是对物理主义的驳难。在她被放出来之前，她可能不知道关于红色经验的事实，因为不存在这样的、要知道的事实。物理主义者和非物理主义者对此都会举双手赞成。在她被放出来之后，情况就变了；物理主义可能高兴地承认：她学会了这一点；毕竟，某些物理的事物将要发生变化，比如她的大脑状态及其功能作用。物理主义的麻烦在于：在她看见第一个成熟的西红柿后，她会意识到，她关于别人心理生活的概念一直是何等的贫乏。她会意识到，尽管她一直就别人的神经生理学、他们的内部状态的功能作用进行着艰辛的研究，但那些人身上还是存在着她完全没有认识到的某种东西。他们的经验（或者其中很多，如从西红柿、天空、……中得到的那些经验）一直有一个对于他们来说是显著的特征，但是在此之前它对她一直藏而不露（实际上，而不是逻辑上）。当然她始终知道关于它们

100 第二部分

的所有物理事实;因此,直到她被放出来时才知道的东西不是关于他们经验的一个物理的事实。但它却是关于它们的一个事实。这就是物理主义的难题。

II. 丘特兰德的三个责难

(一)丘特兰德的第一个责难是,知识论证有一个“本身很明显”的错误(23)。这个论证在“知道关于”的意义问题上含糊其辞。何以如此呢?丘奇兰德认为,下面是关于知识论证的一个方便而浓缩的表述:

(1)关于大脑状态及其属性,玛丽知道一切应知道的东西。

(2)事实是,玛丽并不知道关于感觉及其属性的一切应知道的东西。因此,根据莱布尼兹定律:

(3)感觉及其属性 \neq 大脑状态及其属性(23)。

丘奇兰德非常合理地注意到,前提(1)中所涉及到的知识类型或种类与前提(2)中所涉及到的知识种类是截然不同的。我们可以按他的方法把第一种知识称作“描述的知识”,把第二种知识称作“亲知的知识”;但是无论用什么标签,他都是正确的,因为,那样展开的论证包含着对于莱布尼兹定律的十分含糊的运用。

我的回答是,那样展开的论证可能是便利的,但却是不准确的。它不是知识论证。以前提(1)为例。知识论证的全部要旨在于,玛丽(在被放出来之前)并不知道关于头脑状态及其属性的应知道的一切东西;因为她不知道与它们有联系的某些感受性质。根据这个论证,所有的东西就是她关于物理事物的知识。该论证

的便利而准确的展开是：

(1)'玛丽(在她被放出来之前)知道关于他人的、应知道的一切物理的东西。

(2)'玛丽(在她被放出来之前)并不知道关于他人的一切应知的东西(因为在她被放出来时,她学到了关于他们的某种东西)。

因此,

(3)'存在着关于他人(关于她自己)的、为物理主义描述所遗漏的真理。

关键不在于玛丽所具有的知识种类、方式或类型,而在于她知道什么。据假定,以前她所知道的是一切应知的物理的东西,但是它是应知的一切东西吗?这是问题的要害。

可是,存在着一个涉及到知识种类问题的相关的质疑。它与前提(2)'的证据有关。前提(2)'的理由是,玛丽在她被放出来时,学到了某种东西,获得了知识,这可衍推出:她从前的知识(她所知道的东西,不管是通过描述、亲知还是别的方式)是不完全的。D. 刘易斯和 L. 尼米罗提出的质疑是,在玛丽被放出来时,她没有学到某种东西或没有获得有关意义上的知识。她被放出来时所获得的是某种表征的或想象的能力;它是关于怎样的知识,而不是关于那的知识。因此,物理主义者可以承认,玛丽获得了具有知识种类的意义的东西——这一点几乎不能被否认——当然他们不承认:这表明了她先前的实际知识是有缺陷的。以前,她知道关于他人经验的所有应知道的那,但是在被放出来之前,她一直缺乏一种能力。^③

玛丽在被放出来之后,肯定会获得种种能力。比如,她能够想

102 第二部分

象:观看红色可能是什么样子,还能记住它可能是什么,还能够理解:为什么她的朋友认为她被剥夺了(某种东西,直到她被放出来时,它一直在蒙骗她)。但是有理由说,这就是她将获得的一切吗?假设在她被禁闭的时候接受了一个关于他心的怀疑主义的讲座。当她被放出来时,她在正常条件下,看到了一个成熟的西红柿,进而她有了一个关于红色的感觉。她的第一个反应就是说,她现在对别人看成熟的西红柿时所具有的经验知道得更多了。后来她记起了这个讲座,开始担心了。关于他们的经验可能是什么,她是知道了更多的东西呢,还是沉溺于根据一种情况所作的疯狂的泛化呢?最后,她断定她确有所知,并认为怀疑主义是错误的(即使像我们许多人一样,她不能确定怎样去证明它的错误)。她犹疑不定的是什么呢——她的能力吗?肯定不是;她的表征能力众所周知,是始终恒定的。那么她苦苦寻思的,除了她是否已获得她人的实际知识之外,还有什么呢?如果她被放出来时所获得的只不过是能力,那么就没有什么需要苦苦思索的。

我承认,我没有证据证明玛丽在被放出来时,除了能力之外,还获得了关于他人经验的实际知识——而且不仅仅是因为我没有反驳怀疑主义的证据。我的主张是,知识论证是一种有效的论证,它从合理的、即使是公认不可证明的前提出发,得出了物理主义是虚假的结论。这就是对人们在哲学的这个领域中所能期望得到的一个足够好的驳难所要说的话。

(二)丘奇兰德的第二个责难(24/5)是,这个论证肯定存在着某种错误,因为它要证明的太多了。假设玛丽通过她的黑白电视机接受了一个彻头彻尾的二元论者所作的系列讲座,他解释说,

“规律”控制着“灵的外质”^{*}的行为,还跟她讲到了感受性质。这不妨碍主张她被放出来时学到了某种东西的合理性。所以如果这个论证是反对物理主义的,那么它也得反对二元论。

我的回答是,黑白电视机上关于感受性质的讲座并没有把有关感受性质的、应知的一切东西告诉玛丽。他们可能会告诉她关于感受性质的某些东西,比如说,它们没有出现于物理主义的描述中,我们用“黄色的”表示的感受性质与我们用“蓝色的”表示的感受性质极不相同,就像黑不同于白一样。但是为什么要假设他们把关于感受性质的一切告诉了她了呢?另一方面,说黑白电视中的讲座可能在原则上把物理主义描述中的一切东西告诉了玛丽,似乎是合理的。你学习物理学或功能主义的心理学,不一定要用彩色电视机。为了得到反二元论的有力的论证(属性二元论;灵的外质显得有点好笑),知识论证中的这个前提,即玛丽在被放出来之前就有在物理主义看来是全面的描述,必须为这样的一个前提所取代,即她有在二元论看来是全面的描述。前者是合理的,后者则不然。因此,对于使用知识论证的二元论者来说,不存在“理由相同”的麻烦。

(三)丘奇兰德的第三个责难是,知识论证断言,“玛丽甚至不能想象有关的经验可能是什么,虽然她有详尽的神经科学知识,因此一定还是遗漏了某些关键的信息”(25),这是他要进一步驳斥的一个主张。

但是,正如我们先前所强调的,知识论证断言的是:玛丽不知

* 据说是灵魂在降神的恍惚状态中发出的一种黏性体外物质。

104 第二部分

道有关的经验可能是什么样子。她能够想象什么,这是另一个问题。如果她的知识是不完全的,尽管在物理主义看来是所有应知的知识,那么物理主义就是虚假的,无论她的想象力有多大。

注 释

① 这里的主张不是说,如果物理主义是真的,那么唯一能用简明的物理语言表述的东西就是知识的项目;而是强调:如果物理主义是真的,那么只要你知道用简明的物理语言表述和被表述的一切东西,你就知道一切东西。T. 霍根有不同的看法,见“杰克逊论物理信息和感受性质,”《哲学季刊》,XXXIV,135(1984,14):147-152。

② 就是说,它在我的“副现象的感受性质”一文中,出处同上,XXX II,127(1984,4):127-136。也可看 T. 内格尔,“成为一只蝙蝠可能是什么样子?”《哲学评论》,LXXXIII,4(1974,10):435-450(以上两篇均收入本中文选辑,可参阅。——译者),以及 H. 鲁滨逊:《物质和感觉》,纽约:剑桥,1982。

③ 参见 L. 尼米罗、T. 内格尔的评论:《致命的问题》,《哲学评论》,LXXXIX,3(1980,7):473-477,以及 D. 刘易斯:“疯子的疼痛和火星人的疼痛续编”,《哲学论文集》,第一卷,纽约:牛津,1983。丘奇兰德提到了尼米罗和刘易斯两人,可能是他打算使他的驳难在本质上成为我刚刚给出的那一个,但是他很明确地说(p.23底),他的驳难用不着对有关知识作出“能力”分析。

成为一只蝙蝠可能是什么样子*

托马斯·内格尔

意识是使身心问题难以真正得到解决的东西。这大概就是为什么目前关于这个问题的讨论要么几乎不理睬意识,要么明显错误地加以理解的原因。最近,还原主义的欣快症冲击波已对心理现象和旨在解释某些唯物主义、心理物理同一或者还原的可能性的心理概念作出了分析。^①但是要解决的问题是这种类型的还原和其他类型的还原所共通的问题,而且使身心问题具有独特性的东西、使之不同于水-H₂O问题或图灵机-IBM机问题或闪电-放电问题或基因-DNA问题或者橡树-烃问题的东西,则被忽视了。

每个还原主义者可根据现代科学作出自己所偏爱的类推。成功还原的任何不相干的例子对于心与脑的关系的说明几乎不可能提供什么帮助。但是,哲学家们具有人类共同的弱点,喜欢用一些

* 译自《哲学评论》,LXXXIII,4(1974年12月)。

“What it is like to be……”在当今心灵哲学中是一极常见的表述,表达的是一个已受到广泛争论的问题。在中文中,有的将其翻译为:“成为……像什么。”这是不妥的。因为作者明确表示,它的意思“不是像或类似什么”,而是指“它对于主体自身是怎样的”(见本文注⑥)。根据这一说明,结合表述的实际形式,本中译一律译为“成为……是什么样子”。这当然也未完全传达作者的意思。现在看来,任何中译都难以做到这一点。最好的办法是记住作者自己的解释。

106 第二部分

适合于表示尽管是完全不同的、但却是为人们所熟悉和易于理解的东西的术语去解释那些无法理解的东西。这就导致人们接受那些关于心理现象的不合理的说明,原因主要在于,这些说明承认了那些司空见惯的还原。我想解释的是,常见的例子为什么无助于理解心身关系——我们对什么是关于心理现象的物理本性的解释在目前为什么的确没有什么概念。如果没有意识的话,心身问题将会是索然无味的;但如果有意识,心身问题又好像无望解决。对有意识心理现象的最重要和最独特的特征,人们所知甚少。大多数还原主义理论甚至也不想予以解释。细心的考察将会发现:目前根本没有适用于它的、行之有效的还原概念。为了这一目的,人们也许会想出一种新的理论形式,但是这种办法如果有的话,也只会存在于遥远的、智力发达的未来。

有意识的经验是一种很广泛的现象。它发生于动物生活的许多层面上,尽管我们不能肯定它存在于更简单的有机体中,而且也很难一般地说:什么能证明它。(一些极端主义者甚至还准备否认人以外的哺乳动物有有意识的经验。)毫无疑问,它以无数的、我们完全无法想象的形式,发生在整个宇宙的其他太阳系的其他行星上。但无论形式如何不同,有机体有有意识经验这一事实的确意味着:存在着某种东西,它是成为那个有机体可能具有的东西。经验的形式可能有更进一步的意蕴;有机体的行为甚至也可能有其意蕴(尽管我持怀疑态度)。但是从根本上说,有机体有有意识的心理状态,当且仅当存在着成为那种有机体可能具有的某东西——对那有机体来说,它是可能的某东西。

我们可以把这称作经验的主观特征。对心理的任何司空见惯

的、以及最近设想出来的还原分析都不能把握这个特征,因为所有这些分析从逻辑上说只能符合于这个特征的不存在。它也不能根据关于功能状态或者意识状态的任何解释系统去分析,因为这些状态也可以归之于机器人或者自动机,尽管它们没有任何经验,但它们却能像人那样行动。^②它也不能按照与典型的人类行为有关的经验的因果作用去加以分析——理由同上。^③我不否认,行为是由有意识的心理状态和事件引起的;我也不否认,对它们可以作功能描述。我只是否认,他们的分析能详尽无遗地用之于这类事情。任何还原主义的程序都必须建立在分析什么应被还原这样的基础之上。如果这种分析遗漏了某种东西,那么就会错误地提出问题。把对唯物主义的辩护建立在对心理现象的分析的基础之上,是一钱不值的。因为这种分析并未清楚地说明这些现象的主观特征。当无意说明意识时,还原似乎是合理的,但假设还原能推广到包括意识在内的东西之上,则没有任何道理。因此,如果对经验的主观特征是什么没有某种观点,那么我们就不能知道什么是物理主义理论所必需的东西。

对心灵的物理基础的说明一定能解释许多事物,而这似乎是最困难的。从还原中排除经验的现象学特征是不可能的,就像人们从它的物理或化学还原中排除普通物质的现象学特征——即把它们解释成对人类观察者心灵的作用——是不可能的一样。^④如果物理主义应予辩护,那么就必须对现象学特征本身作出物理学说明。但当我们去考察它们的主观特性时,这样一个结果似乎是不可能的。这是因为,每种主观现象从根本上说都与一种独特的观点相联系,客观的、物理的理论似乎不可避免地会放弃那个观

108 第二部分

点。

我首先想用比通过指出主观和客观或自为与自在的关系更全面的方式陈述这个问题。当然这绝非举手之劳。关于成为一个X可能是什么样子的事实是非常特别的,以至于有些人可能会怀疑它们的真实性,或怀疑关于这些事实的论断的意义。为了说明主观性和观点之间的联系,为了突出主观特征的重要性,相关于清楚地显示了主观和客观两种类型的概念之间差异的事例来考虑这个问题,是大有裨益的。

我假设,我们都相信蝙蝠有经验。毕竟,它们是哺乳动物,而且没人怀疑老鼠、鲸鱼或鸽子具有经验,同样关于蝙蝠具有经验就更不必怀疑了。我之所以选择了蝙蝠而不是黄蜂或比目鱼,是因为,如果某物沿种系发生之树走得很远很远,那么人们就会逐渐抛弃经验的确存在于那里这一信念。蝙蝠虽然和我们的关系比其他物种更加密切,但它们表现出的大量活动和感觉装置与我们的如此不同,以致我想提出的这个问题就是异常鲜明的(虽然对其他物种也能提出这个问题)。一个人即使没受过哲学反思的训练,但只要他和一个活跃的蝙蝠在一个封闭的地方呆一段时间,便会了解到:遭遇一个完全异类的生命形式是什么情形。

我已说过,相信蝙蝠有经验这一信念的实质就是:成为一只蝙蝠就好像有某种东西。现在我们知道,大多数蝙蝠(精确地讲,是微翼手目动物)首先是通过声纳或回声,分辨来自一定范围内的对象的反射,亦即它们自己的迅速、巧妙调节的高频率的尖叫声的反射,从而感知外部世界的。它们的大脑注定使向外的冲力与随后的回声联系起来,进而如此获得的信息使蝙蝠能够精确地分辨出

距离、大小、形状、运动和构造特征。这可与我们用视觉所作的分辨相媲美。很明显,即使蝙蝠的声纳有知觉的形式,但是它的作用与我们所具有的任何感觉是不可同日而语的,没有理由假设它在主观上像我们能够经验或想象的任何东西。这似乎为关于成为蝙蝠可能是什么的概念提出了种种难题。我们必须考察的是,是否有什么方法让我们根据自身的情况去推知蝙蝠的内在生活。^⑤如果没有的话,理解那个概念还有什么别的替换的方法吗?

我们自己的经验为我们的想象提供了基本材料,而想象的范围由此又受到了限制。试想某人手臂上有网状物,这使他能在黄昏和早上来回地飞,用嘴去捕捉昆虫;试想某人视力很差,只能用反射高频率声音信号的系统去感知周围世界;试想某人整天在顶楼里用脚倒钩悬吊着,这些想象都是无济于事的。就我能想象这一点而言(它并不太离奇),它告诉我的仅仅是我按照蝙蝠的行为方式行动可能是什么样子。但是问题并不在这里。我想知道的是,蝙蝠之成为蝙蝠可能是什么样子。然而,如果我尽力这样去想象,那么我严格地受到我自己的心灵资源的限制,这些资源和这个任务是不相称的。不管是通过想象增加我现有的经验,还是通过想象从中逐渐减去一些断片,或是通过想象把增加、减去和改造混合起来,我是没法完成这个任务的。

在一定程度上,我可以不改变我的基本结构,而像黄蜂或蝙蝠一样看和行动,但是我的经验则全然不像这些动物的经验。另一方面,断言我将拥有蝙蝠的内在神经生理学构造这一推测有什么意义,这则是可疑的。即使我会慢慢地变成一只蝙蝠,但我现有的构造中没有什么能使我想象:我如此变形后的将来的经验可能

是什么。如果我们只知道它们可能是什么的话,那么从蝙蝠的经验中就可得到最好的证据。

所以,如果根据我们自己的情况所作的推论涉及到成为一只蝙蝠可能是什么样子这一观念,那么这个推论就必定是不完善的。关于它可能是什么,我们不可能形成一个以上的图式化概念。比如说,我们可以根据那个动物的结构和行为对经验的一般类型作出归因。这样,我们就可把蝙蝠的声纳看作是三维前向知觉的一种形式;我们相信,蝙蝠可感觉到某些形式的疼痛、恐惧、饥饿和欲望,而且除了声纳之外,它们还有其他更常见的知觉类型。但是我们相信,在每一种情况下,这些经验也有独特的主观特征,对之进行设想超出了我们的能力范围。如果在宇宙的其他地方,存在着有意识的生命,那么其中的一些即使是用我们唾手可得的最一般的经验术语也不可能予以描述。^⑥(但这个问题并不局限于异常的情况,因为它还存在于一个人和另一个人之间。比如,一个天生就既聋又瞎的人,对于他的经验的主观特征我是不可能理解的,我的经验特征对于他也可能是一样。这并不妨碍我们每个人都相信:别人的经验有这样的主观的特征。)

如果有人想否认:我们能够相信存在着这样的事实,其实存在的本质是我们所不可能设想的,那么他应当想到,在我们思考蝙蝠的时候,我们在很大程度上将像有理智的蝙蝠或火星人的那样,^⑦陷入它们试图形成关于成为我们可能是什么样子的概念时所陷入的处境。它们自己的心灵结构可能使它们不能如愿以偿,但我们知道,如果它们得出结论说,不存在什么精确的东西,它是成为我们像什么的东西,只有某些能归之于我们的一般类型的心理状态,那么它

我们就错了(知觉或欲望也许是我们两者都共有的概念,但也许不是)。我们知道,它们得出这样一种怀疑论结论是错误的,因为我们知道成为我们可能是什么样子。而且我们还知道,当它包涵的内容极为广泛和复杂的时候,当我们没有足够的词汇去恰当地描述它的时候,它的主观特征就是异常独特的,在某些方面只能用我们这样的造物所能理解的术语加以描述。我们永远不可能奢望用我们的语言提供对火星人或蝙蝠的现象学的详尽描述,这样的事实不应当使我们将下述主张作为无意义的东西而予以抛弃,即蝙蝠和火星人有完全能与我们相媲美的丰富而细腻的经验。如果有人阐发了能促使我们思考这些事情的理论和概念,那就再好不过了;但这样一种理解可能由于我们有限的本性而永远为我们所拒斥。否认我们不能描述或理解的东西的实在性或逻辑意义是认知的不和谐性的最天然的形式。

这便把我们带到了需要比我这里所能给予的讨论更多的讨论的主题面前,即一方面的事实与另一方面的概念图式或表征系统之间的关系问题。我关于以各种形式表现出来的主观领域的实在论,隐含着对于人类概念所及之外的事实之存在的信念。当然,人类可能相信存在着人类永远无法用所需的概念予以表征或理解的事实。假如整个人类的期望是有限的,那么怀疑这一点就是愚蠢的。即使每个人在圣咏队领唱发现他们之前,都死于黑死病,但毕竟存在着超限数。不过,人们也可能相信,存在着永远不会为人类表征或理解的事实,即使该物种永远延续——这仅仅是因为我们的结构不允许我们以所需的概念类型运作。这种不可能性甚至可能被其他生物观察到,但尚不清楚的是,这生物的存在或者它们存

112 第二部分

在的可能性是存在着人类不可接近的事实这一假说的意义的先决条件。(毕竟,有办法理解人类所不能理解的事实的生物的本性本身大概就是一种人类不可理解的事实)因此,对于成为一只蝙蝠可能是什么样子的思考似乎把我们引向了这样一个结论,即有这样的事实,它们不存在于能用人类语言表达的命题的真理之中。我们即使不能陈述或理解这些事实,也能被迫认识它们的存在。

但是,我不想纠缠这个问题。它对我们面前的课题(即身心问题)的影响是,它使我们能够对经验的主观特征作出总的观察。关于成为一个人,或一只蝙蝠,或一个火星人可能是什么样子的事实,无论有什么样的重要地位,它们似乎是包含着一种独特观点的事实。

我这里不是要对把经验的私人性归之于其所有者评头品足,这里所说的观点并不是只能为单独的个体所理解的观点。毋宁说它是一种类型。采取某人自己以外的观点常常是可能的,因此对于这类事实的理解并不局限于某个人自己。有这样的一种看法,据此现象学的事实完全是客观的:某人可以知道或者说出别人的经验的质是什么。然而在下述意义之下,它们是主观的,即只是对于完全相似于能够采取他的观点——比如说同时理解第一人称和第三人称情况下的归因——的归因客体的人,这种对经验的客观归因才是可能的。别人的经验越是与他自己的不同,他的这个艰巨而冒险的计划之成功的希望就越小。就我们自己来说,我们具有相关的观点,但是如果我们从别人的观点出发去探讨我们自己的经验,那么我们在准确地予以理解时便会碰到极大的困难,就像我们不采取别的物种的观点去试图理解它们的经验一样困难。^⑧

这与身心问题有直接的关系。因为如果经验的事实——关于成为经验着的有机体可能是什么的事实——只能从一种观点出发才能得到理解,那么经验的真实特征怎么可能在那有机体的物理运作中显现出来,就是个难解之谜。后者纯然是一个客观事实的领域——一种能够被具有不同知觉系统的个体、从许多观点出发去观察或理解的领域。人类科学家在获取关于蝙蝠的神经学知识时的障碍是无可比拟,难以想象的,而且有智力的蝙蝠或火星人对人类大脑所知的可能比我们想知的更多。

这本身不是一个反还原的论证。一个不懂得视知觉的火星科学家可能把彩虹、闪电或云彩理解为物理现象,虽然他从不理解人类关于彩虹、闪电或云彩或这些事物在我们的现象世界中所处的位置的概念。这些概念所把握的这些事物的客观本质能够被他所理解,因为这些概念本身即使与一种独特的观点、一种特殊的视觉现象学相联系,但是从那种观点来理解的事物却不是下述情况,即它们可从那种观点来观察,但外在于它;因此它们也能够从其他的观点予以理解,要么是被相同的有机体,要么是被其他的东西。闪电有一个客观的、并没有被它的视觉现象充分表现出来的特征,而且这能够为没有视觉的火星科学家研究。更确切地说,它有比在它的视觉现象里显现出来的特征更客观的特征。在谈到从主观描述到客观描述的转化时,对一种极端的观点之存在,即事物的完全客观的、人们可能或不可能接近的内在本质,我想保持沉默。将客观性看作是理智前进过程中所瞄准的方向可能更准确。在理解像闪电这样的现象时,尽可能远离严格的人类观点是有其道理的。^⑨

另一方面,就经验而言,与独特观点的联系似乎更加密切。撇

114 第二部分

开主体借以理解经验的客观特征的独特观点,就难以理解经验的客观特征可能具有什么意思。毕竟,如果蝙蝠自己抛弃了蝙蝠的观点,那么成为一只蝙蝠可能所是的东西中还会留下什么呢?如果除了主观的特征之外,经验不具有能用许多不同的观点加以理解的客观本质,那么怎么能设想:正在研究我们的大脑的火星星人,只能从一种不同的观点观察作为我的心理过程的物理过程(就像他观察作为闪电之突发的物理过程)呢?^⑩对于这个问题,人类生理学家怎么可能从另外的观点来观察它呢?

我们似乎面临着心理生理还原的一般性难题。在其他领域中,还原的过程是一种向更大的客观性的转化,即指向关于事物的真正本质的更精确观点。通过把我们对于个体的或者物种的特定观点的依赖性还原于所考虑的对象,就可完成上述转化。我们不是依照它加之于我们的感官的印象,而是依照它的更一般的结果和可由人类感官之外的别的途径识别的属性而予以描述的。它依赖的特殊的人类观点越少,我们的描述就越客观。沿着这条路走下去是可能的,因为我们在思考外部世界时使用的概念和观念起初即使是根据与我们的知觉器官有关的观点而加以应用的,但我们常用它们来表示它们自身之外的事物——对于它们,我们具有的是现象的观点。因此我们能将它放弃而采用另外一个,并能一如既往地思考相同的事物。

不过经验本身好像不符合这个模式。从外表到真实转化这一观念在这里似乎没有什么意义。如果是这样,通过放弃关于它们的最初的主观观点,而采用另一个虽涉及同一的事物但却更加客观的观点,来实现对同一现象的更加客观的理解,与此类似的是什

么呢？诚然，抛开人类观点的独特性，用不能想象成为我们可能是什么样子的生物可理解的术语尽力去进行描述，我们似乎不可能由此接近人类经验的真实本质。如果经验的主观特征只能从一种观点出发才能完全予以理解，那么向更大客观性的任何转换——即更少地系缚于一种独特的观点——都不会使我们更加接近于现象的真实本质：它将让我们离它更远。

在某种意义上，这种反对经验的还原性的种子在还原的成功案例中已初露端倪了；因为在揭示声音实际上是空气或其他媒体中的一种波现象时，我们遗漏了一种观点而采纳了另一种观点，而且我们所遗漏的听觉的、人类或动物的观点仍未被还原。极其不同的物种的成员可能用客观的术语理解同一的物理事件，这并不要求它们理解这些现象的形式，即以这样的形式，那些事件呈现给其他物种成员的感官。因此它们指称一共同实在的一个条件是，它们的更加独特的观点不是它们两者都理解的共同实在的组成部分。只有在物种的独特观点从将要被还原的东西中省略掉，这种还原才算大功告成。

在寻求对外部世界更完满的理解的过程中，尽管我们有权撇开这个观点，但我们不能永远无视它，因为它不只是关于它的一个观点，而同时是内部世界的实质。新近哲学心理学中的大多数新行为主义根源于这样的尝试，即用客观的心灵概念代替真实事物，以铲除不能被还原的东西。如果我们承认，关于心灵的物理理论必须说明经验的主观特征，那么我们必须承认，目前尚没有现成的概念能向我们暗示：怎样才能予以说明。这个问题是独一无二的。如果心理过程真的是物理过程，那么就可能存在着内在地^①经历

116 第二部分

某些物理过程的某事物。说这样的事物就是那个的事实,这究竟是什么意思,仍是一个谜。

从这些反思中应该得出什么样的启示呢?我们下一步应该怎么办呢?断言物理主义肯定错了,这是一个谬论。物理主义假说对心灵采取了错误的客观分析,这一缺陷什么也说明不了。更准确地说,物理主义是我们没法理解的观点,因为目前我们没有关于它如何可能为真的任何概念。或许,要求这样一个作为理解的条件概念会被认为是不合理的。毕竟可以说,物理主义的意思是十分清楚的:心理状态是身体的状态;心理事件是物理事件。我们不知道,它们是哪一些物理状态和事件,但这不妨碍我们理解这个假说。有什么词语比“is”和“are”(是)更清楚明白的呢?

但是我相信,单词“是”的这种显而易见的清晰性显然具有欺骗性。通常,当我们被告知: x 是 y 时,我们知道它怎样被假定为真。但是,这依赖于概念的或理论的背景,而不是仅由“是”这个词孤立地表达的。我们知道“ x ”和“ y ”怎样指称,也知道它们所指的事物的类别,而且关于两种指称路径怎样汇合于一个事物之上,我们有大致的观念,如果它是一个对象、一个人、一个过程、一个事件或诸如此类的东西的话。但是当这两个认同术语是根本不同的时候,它怎么可能为真就可能是不清楚的。对于这两条指称路径怎样会合,或它们会合于什么样的事物上,我们甚至可能没有大致的观念,因而必须提供一种理论构架,以便使我们能够理解这一点。没有这个构架,神秘主义的气息就会笼罩这种认同。

这解释了基本科学发现的常见陈述的神秘风格,它们常作为人们在真正理解时必须予以赞同的命题公布出来。比如,人

们在很小的时候就被告知,所有的物质其实是能量。但他们事实上虽然知道“是”的意思是什么,但大多数人对什么使这种说法成真并没有形成概念,因为他们缺少这样的理论背景。

此刻,物理主义的地位类似于物质是能量这一假说为前苏格拉底哲学家说出时所享有的地位。关于它怎么可能为真,我们没有形成概念的基础。为了理解心理事件是物理事件这一假说,我们不只是需要对“是”一词的理解。关于心理术语和物理术语怎么可能指称相同的事物的观念尚属阙如,而且与其他领域中的理论同一性的常见类比尚未提供这种观念。之所以没有提供,是因为,如果我们根据通常的模式解释心理术语对物理事件的指称,那么我们得到的要么是分离的主观事件的再现,它们是对物理事件的心理指称借以实现的作用,要么是一个关于心理术语如何指称的错误说明(如一个因果行为主义说明)。

奇怪的是,我们可以证明我们不能真正理解的某物之真实性。假如一个蠋被一个不熟悉昆虫变态的人锁在无菌的菜橱之内,几周后,这个菜橱被重新打开,里面便出现了一只蝴蝶。如果这个人知道菜橱一直是关着的,那么他有理由相信,这个蝴蝶现在是或者过去曾经是一个蠋,即使对它在什么意义上可能如此没有任何观念。(一种可能性是:蠋包含有一个微小的有翅的寄生虫,它吞食了这个蠋,然后变成了一只蝴蝶。)

可以想象,对于物理主义我们就处在这样的地位上。戴维森曾论证过,如果心理事件有物理的原因和结果,那么它们必定有物理的描述。他认为,即使我们没有——而且实际上不可能有——一种一般的心理物理理论,我们也有理由相信这一点。^⑩他的论证

118 第二部分

适用于有意识的心理事件,但是我想我们也有一些理由相信,感觉是物理的过程,虽然没有条件理解这是如何发生的。戴维森的观点是,某些物理事件有不可还原的心理属性,而且能用这种方法描述的某个观点也许是正确的。但是我们现在能对之形成概念的任何东西与它风马牛不相及;能使我们设想它的理论究竟是什么样子的,我们对此没有任何观念。^⑬

对是否能阐明经验有客观特征的意义这一基本问题(在这里,大脑完全可以不予提及)几乎看不到什么探讨。换言之,问我的经验的确像什么,以对立它们怎样呈现于我,这样问有意义吗?除非我们理解了这个更根本的观念,即它们有客观的本质(或者说客观的过程可能有主观的本质),否则我们就不能真正理解它们的本质可由物理描述加以把握这一假说。^⑭

我想提出一种猜测性的建议,以此来结束本文。从另外的方向来探讨主客之间的鸿沟问题是可能的。暂时撇开心灵和大脑的关系问题,我们可以按其自身的方式寻求对心理的东西的更加客观的理解。在目前,倘若我们不依赖于想象——不采取经验主体的观点的话,我们就完全没有条件去思考经验的主观特征。形成新的概念,设计一种新的方法——一种客观的、不依赖于移情作用或者想象的现象学,这应当被认为是一种挑战。虽然它什么也不能抓住,但它的目的应当是:以一种能为不具有这些经验的生物所理解的形式描述至少部分描述经验的主观特征。

我们必须阐发这样一种现象学以描述蝙蝠的声纳经验;但是从人类开始也是可能的。比如说,人们可能想阐发这样一个概念,它可以用来向生来就盲的人解释:观看可能是什么。人们最终大

概会头撞南墙,但是,也应有这样的可能,即设计一种用比我们目前更多的、更精确的客观术语去予以表述的方法。这种不严格的、多种方式联用的类推是没有用的——如“红色就像喇叭的声音一样”——它常出现在关于这一课题的讨论中。这对于任何一个既听到过喇叭声又看到过红色的人来说是一清二楚的。但是,知觉的结构特征可能更容易进入客观的描述,即使会有某种遗漏。用来置换我们所熟悉的、以第一人称形式出现的那些概念的概念,可以让我们得到一种对我们自己的经验的理解,而轻描淡写的描述和主观概念所造成的无间距性则否认我们有这种理解。

撇开现象学自己的兴趣不说,在这个意义上是客观的现象学也许会承认经验的物理^①基础的种种问题具有更好理解的形式。接受了这种客观描述的主观经验方面可能是那种更熟悉的客观解释的更好的候选者。但无论这种猜测是否正确,只有当对主观与客观的一般性问题作出了更多的思考时,关于心灵的物理理论似乎才有可能被周密地考虑。否则,即使我们没有规避心身问题,我们甚至也不能把它提出来。

注 释

① 例子有:J. J. 斯马特:《哲学和科学实在论》(伦敦 1963);D. K 刘易斯:“对同一理论的论证”,《哲学杂志》,LXIII(1966);H. 普特南:“心理学谓词”,载卡皮坦和梅里尔:《艺术、心灵和宗教》(匹兹堡,1967);D. M. 阿姆斯特朗:《唯物主义的心灵理论》(伦敦,1968);还可参阅 S. 克里普克:“命名与必然性”,载戴维森和哈曼:《自然语言的语义学》(多德雷赫特,1972),尤其是 pp. 334-342;丹尼特:《内容和意识》(伦敦,1969)。我已在“阿姆斯特朗论心灵”一文中阐述了更早的怀疑,见《哲学评论》,LXXIX(1970),394-403。

② 也许实际上没有这样的机器人。复杂到像人一样行动的任何东西也

120 第二部分

许有经验。但如果的确是这样的话,那么它就是一个不能仅仅通过分析经验概念就能发现的事实。

③ 它不能与我们对之无法矫正的东西相提并论,因为对经验我们不是没法矫正的,而且经验也存在于缺乏语言和思想的动物中,而它们对它们的经验完全没有信念。

④ 试比较 R. 罗蒂:“心身同一性、隐私和范畴”,《形而上学评论》,XIX,尤其是 pp. 37-38。

⑤ 我所谓“我们自己的情况”,不仅意指“我自己的情况”,而且意指我们没有疑义地用之于我们自己和别人的心理主义观念。

⑥ 因此,英语短语“它可能是什么”的类似的形式易引起误解,它不是指“它类似于(在我们自己的经验中的)什么”,而是指“对于主体自身来说它是怎样的”。

⑦ 地球以外的任何有智力的生物与我们完全不一样。

⑧ 这比我借助想象的帮助,设想出超越物种间的障碍可能要容易得多。例如,盲人能够通过声纳形式或通过拐杖的咔哒声或敲打声探知附近的物体。也许一个人如果知道那可能是什么,就能够通过外推,大致地想象出:拥有蝙蝠的、精致得多的声纳可能是什么。某人自己与其他人、其他物种之间的距离可能一以贯之地存在于任何地方。甚至就其他人来说,对成为他们可能是什么的理解,仅仅是局部的,当一个人转向一个和他自己很不相同的物种的时候,仍可以得到一个更低程度的局部理解。想象显然是非常灵活的。但是,我的观点不是说我们不能知道成为一只蝙蝠可能是什么。我没有提出这个认识论上的问题。更确切地说,我的观点是:即使对成为一只蝙蝠可能是什么要形成一个概念(更不必说知道成为一只蝙蝠可能是什么),人们也必须采取蝙蝠的观点。如果他大致地或部分地予以采取,那么他的概念也会是大致的或部分的。它好像就存在于我们现在的理解状态中。

⑨ 因此,即使只有在一个更广泛的人类观点之内才能对更加主观和更加客观的描述或观点本身作出区分,我想提出的问题也是可以被提出来的。我不赞成这种概念相对主义,但没有必要拒绝提出这样的观点:从其他情况中所获悉的从主观到客观的模式不可能容纳心理物理的还原。

⑩ 问题不仅仅是这样,当我看“蒙罗丽莎”的时候,我的视觉经验有某种

质,它的踪迹不可能被任何观察我的大脑的人发现。即使他的确观察到有关于“蒙罗丽莎”的微弱映象,他也没有理由把它和经验同一起来。

⑪ 因此,这种关系不会是一种偶然的的关系,就像原因与它的不同结果的关系那样。一定的物理状态以一定方式被感知,这是千真万确的。S. 克里普克(同上)论证说,对心理的东西的因果行为主义及相关的分析之所以失败,是因为他们仅仅把“疼痛”解释为疼痛的偶然的名称。经验的主观特征(克里普克称之为“它的直接的现象学的质”(p. 340))就是这样的分析所遗漏的基本属性,而且它也是这样的属性,即根据它,它必然成了它所是的经验。我的观点与他的观点有密切的关系。像克里普克一样,我发现,如果没有进一步的解释,某种大脑状态一定有某种主观特征这一假设就是不可理解的。在把心灵和大脑的关系视作偶然的理论中,没有出现这样的解释,但是也许还有其他的选择,不过尚未被发现。

一种解释心灵和大脑的关系是如何必要的理论仍然会把克里普克的问题即解释它为什么是偶然的留给我们。在我看来,这个困难用下面的方法是可以克服的。我们可以从知觉、感应或者象征方面,通过把某物呈现给我们自己的方式去想象它。我并不要说,象征式的想象如何起作用,但是在另外两种情况下所发生的情况部分地是这样的。为了从知觉上想象某事物,我们应把我们自己放在一种类似于我们在知觉它时会进入的有意识的状态中。为了感应式地想象某事物,我们应把我们自己放在这个事物本身的有意识的状态中。(这个方法只能用来想象我自己或其他人的心理事件和状态)当我们试图去想象在没有它的相关联的大脑状态的情况下也发生的心理状态时,我们首先得感应式地想象这个心理状态的发生:即是说我们把我们自己放在一种在心理上类似于它的状态中。同时,我们试图去知觉地想象这个有关联的物理状态的未发生,方式是把我们自己放在与第一个状态不相关的另一状态中:它类似于这样的状态,即如果我们知觉到这个物理状态未发生,我们就会处在其中。在物理特征的想象是知觉式的且心理特征的想象是感应式的地方,据我们看来,我们能想象:即使没有它的相关联的大脑状态,任何经验也会发生,反之亦然。它们之间的关系会是偶然的,即使它是必要的。因为类型根本相异的想象是各自独立的。

(如果人们错误地理解感应式的想象,好像它是像知觉式的想象那样起

122 第二部分

作用的,那么偶尔会有唯我论发生:因此想象任何不属于某人自己的任何经验似乎是不可能的。)

⑫ 参阅“心理事件”,载福斯特和斯旺森编:《经验和理论》(安伯斯特,1970);虽然我不懂反心理物理规律论证。

⑬ 类似的评论适用于我的论文“物理主义”,见《哲学评论》,LXXIV (1965),339-356。

⑭ 这个问题也存在于其他心灵问题的核心里,它和心身问题的密切联系经常被忽视了。如果一个人理解了主观的经验怎么可能有客观的本质,那么他就会理解除了自己之外还有主体存在。

⑮ 我没有给“物理的”这个术语下定义。很明显,它不仅仅适用于能为当代物理学的概念所描述的东西,因为我们预期有更进一步的发展。一些人可能认为,没有什么能阻止心理现象最终被认作是有其自身地位的物理的东西。但是不管物理的东西有别的什么,它必须是客观的。所以,如果我们关于物理的东西的观念得到拓展以致包括心理现象,那么它将不得不赋予心理现象以客观的特征——不管是不是通过根据别的、已被当作是物理的东西对它们加以分析做了这样的工作。但对我来说,更有可能的是心理-物理关系最终会在这样的理论中得到表述,它的基本术语显然不属于两者中的任何范畴。

感受性质与物理主义、取消主义*

金 在 权

感受性质问题

通常所谓的“感受性质问题”实即给予感受性质以符合于物理主义基本框架的说明问题,即是说对感受性质作出可为物理主义接受的说明的问题。^①我们已讨论了主张感受性质并不随附于大脑中发生的物理或生物过程的理由。但是如果感受性质在物理上不是随附的,那么就必须把它们当作是这个物理世界的项目之外的项目,进而我们就必须考虑感受性质是超物理的事实这一事实。这一来,非随附性的感受性质之存在便对物理主义提出了一种直接的挑战。

因此,物理主义者在这一点上有两种选择:(1)要么为一种较弱形式的随附性命题作论证,(2)要么像通常所设想的那样,论证感受性质并不存在——根本就不存在现象的、质的状态或具有现象特征的经验。下面我们来看看每种选择是怎样展开的。在这一节的剩余部分和下一节,我们将讨论第一种选择,在后面,我们再

* 节译自作者的新著《心灵哲学》,第7章,韦斯特维耶出版社,1996。标题为译者据内容所加。作者的英文名字为 Jaegwon Kim,一般译为耶格旺·金。作者在给译者的信中署名为“金在权”,现照此译出。在文中,如遇名字缩写,则简译为J. 金。

124 第二部分

讨论第二种选择。

我们的物理主义者可能会承认:感受性质并不以形而上学的必然性随附于物理或生物过程,但能以较弱形式的必然性即是说法则学上的必然性随附于后者。他承认:“行尸世界”在形而上学上是可能的,而且完全可以设想:你的物理的复制品有一种视觉光谱,它相对于你自己的而言是颠倒的。然而他又断言:这些事物在完全相似于我们世界的世界里是不可能的,因为它们是由遵循着同样的基本物理规律的相同种类的基本物质粒子所构成的。他大概会说,在谈到感受性质随附性时,对于包含有非物理精神的或由非常不同的物理规律在其中起着主导作用的真正超自然的物理世界,我们用不着忧心忡忡。相反,我们应记住这些世界是非常相似于这个世界,即“法则学上可能的世界”的。^②

这自然弱化了物理主义。它只是提出了关于现实世界和相关的、类似的世界的断言,认为在类似于我们的世界的世界,两个事物如果没有物理上的区别,那么根本就不能从心理上——尤其是在现象方面——相互区别开来。也许这就是十足的物理主义;^③我们大概可以称之为“法则学物理主义”,以把它与“形而上学物理主义”区别开来。人们可以沿着下述路线对之作出辩护,即把热还原于分子运动,把基因还原于 DNA 分子,而且科学中的类似的还原只出现于完全相似于我们世界的世界,就基本物质材料与根本的物理规律组成的自然界来说更是如此。在离我们的世界非常遥远的、完全不同的规律在其中起着作用的世界里,大概不是 DNA 分子而是某种别的类型的分子或机制对基因信息的传递发挥着原因的作用。因果力依赖于起主导作用的规律,假如基因概念根据

因果作用而定义(详后),那么很显然,基因还原于 DNA 分子只会出现在法则学上相似于我们世界的世界中。只要 DNA 在我们的世界和非常接近于我们世界的世界起着那种作用,那么就足以看作是基因对 DNA 分子的还原。这说明的是:物理主义、随附性、还原诸如此类的问题只需要相对于法则学上可能的世界,即基本物理规律和物质构成相似于我们世界的世界,加以考虑,对于超现实的、异在的世界,我们可以坦然地置之不理。

让我们以下述形式表述关于感受性质的法则学随附性的主张:

对于每一现象属性 P 而言,每当某有机体 x 例示 P 时,就存在着物理或生物过程 B ,以至于 x 在那时例示 B ,进而它就成了这样的自然法则(即作为法则学的必然性而起作用):每当任何有机体 y 例示 B ,那么 y 在那时就例示 P 。

因此,假如你在一特定时间体验到了头疼,根据随附性命题,你就处在特定的物理或生物状态(大概是你的大脑的某种中枢状态)——不妨称之为“ N ”——于是下述普遍原则就可作为规律而起作用:

(S)每当 N 在人类身上得到例示,他们就体验到了一种头疼
(简言之, $N \rightarrow I$)

我们无须假定: N 对于头疼是唯一的随附性基础;在别的物种中,别的中枢状态可能会引起头疼感。(S)所述的不过是:在人类身上,每当 N 得到了实现,那么一种头疼就会被体验到。

物理主义者所面临的问题是:法则学随附性足以涵盖物理世

126 第二部分

界——至少是充分地相似于我们世界的物理世界——的感受性质吗？有理由认为：需要做更进一步的解释工作。为了说明这一点，让我们假设发痒的感觉随附于某种中枢状态 M ；即是说，下述规律是适用的：

$$(S^*)M \rightarrow T$$

像 (S) 和 (S^*) 这样的规则适用于我们的世界和别的相同的世界。但是为什么呢？这能根据这些世界的物质构成(更具体地说，人的物理的或生物的构成)和基本的物理规律加以解释吗？如果 (S) 与 (S^*) 之类的相互关系只就基本规律和它们所包含的物质材料而言存在于相似于我们的世界，那么我们一定有望根据这些规律及其所控制的物质材料的属性对这些相互关系作出解释。但是这样一种解释的前景似乎极其渺茫。“疼痛”、“痒”之类的术语根本就不出现在物理或生物科学中，更不用说在基本物理学之中；我们的基本物理科学并不论及像经验头疼、痒、感觉到绿等等之类的属性。我们怎么能够从物理的或生物的前提中推论出或以某种方式引申出关于感觉的感受性质的事实，这是很难弄清楚的。如果是这样，那么这个世界上就存在着不可能根据物理事实和规律解释的事实，亦即关于现象经验发生的事实。世界的现象超越于物理主义的解释力之外。

如果我们把 (S) 与 (S^*) 这样的随附性规律作为解释的前提，那么我们自然就能解释头疼感觉或痒的感觉的特殊例子为什么发生在人类身上：她头疼，因为她处在中枢状态 N 中，她痒，因为她处在 M 状态中。但很明显，必须予以解释的是： (S) 与 (S^*) 这样的随附性关系为什么优先起作用。为什么正好当我们处在中枢状

态 N 时才经验到头疼呢？我们为什么没有经验到痒？能说明头疼起源于 N 、痒感起源于 M 而不是别的方式这样的事实的 N 状态 and M 状态又该如何解释呢？任何现象意识为什么会产生出来？在完全是按照物理规律运动的物质粒子构成的世界中，为什么存在着像感受性质这样的东西？这些是物理主义者必须妥善解决的问题。

感受性质能从物理上加以解释吗？

试考察一下温度和平均动能。我们似乎可以沿着下述路线很圆满地解释气体的温度为什么随附于大气分子的平均动能——气体的温度为什么事实上只不过是它的平均的分子运动能量。首先说明温度概念是一种根据把温度与其他属性关联起来的一组规律而得到定义的功能概念。温度是这样一种对象的物理量，即在该对象以更高的量与另一对象发生联系时，其量会得到提高；即是说，当它足够高时，便使蜡块近乎熔化；当它非常高时，它还能引起铁块发热，直至变成熔液；当它足够低时，还能使钢变碎；如此等等。一种对象的温度就是具有那些因果作用的对象的那种属性，因此可以得出结论说：在这个世界以及在基本规律和物质构成相似于这个世界的世界中，平均分子能量就是发挥这些作用的东西。而且，考虑到世界的基本的物质构成和物理学规律，我们便能解释平均分子能量为什么具有这类因果力量。与这些因果力量相符合的一切属性都是温度。这也就解释了：温度为什么随附于更基本的微观物理事实。对于基因和 DNA 分子也能如是说：基因就是

128 第二部分

按照某些有规律的模式在代与代之间起着传递表现型特征的因果作用的生物学因素或机制。而且我们还发现 DNA 分子可以充任这种因果作用——事实上它就是在我们的世界和别的相似于我们世界的世界行使着这种因果作用的东西。而且我们至少粗略地、概要地但令人信服地知道:DNA 分子是怎样完成这个任务的。不管怎样,这些根据起基础作用的物理机制对高水平现象的解释,似乎例示了科学中的微观还原解释模式。我们为什么不能期待科学(大概是神经生物学)基于同样的模式对感受性质作出物理解释呢?

我们已考察过的那类还原解释是这样开始的,即用现象的因果功能作用来解释现象的被还原。因此像我们已看到的那样,温度被解释为这样一种对象的属性,它与如此这样的现象具有如此这般的因果的或法则学的关系。同样,基因被认为是这样的生物学机制,不管它最终被证明是什么,从因果上说它是与传递着遗传特征有关的如此这般的现象的原因。这意思是说,被还原的那些现象是依照它们与别的现象的因果或法则学关系从关系上或外在方面被概念化的,而不是依照它们的内在的质的特征或构成性结构而内在地被概念化的。

因此等待着接受还原解释的属性首先必须进行这样的“准备”即被关系化,即是说被重新解释为外在的关系属性。这对我们所探寻的、并在许多例示中已找到了微观物理解释的许多宏观物理属性也是有意义的。因为某东西能溶解于水也就是当它浸在水中,它有一种能引起它溶解的属性,某物是透明的,也就是它有让光穿透的那种结构,如此等等。但是对于现象特征来说,我们最终

直接面对的似乎是内在属性的范例性的例示。疼痛的伤痛性、硫磺的刺激性的气味、菠萝的味道和气味——如果有的话,这些东西就是内在的性质。像我们刚做过的那样,我们经常用外在的摹状词来指谓这些感受性质(如说“哪种气味一般是硫磺在人身上引起的”),但是这些外在的摹状词所述说的并不是外在的性质,而是内在的性质。我们之所以要求助于外在的关系性的摹状词,以使它们可传达于他人,这显然是因为所涉及到的性质是内在的(也许是主观的)。这一过程并非感受性质所独有;它也适用于基本的物理量值。一根棍子为一米长,就是当把标准米尺两端放在头尾时,它正好与之相符(或者说它就是镉或别的什么多次发射出来的光的波长)。对于质量我们也可以说同样的话:一千克就是这样的物体的质量,即在如此这般的体积静力加之于它时它以如此这般的速率作加速运动。但是我们只是一般地把质量、长度、电极等当作是物质事物的、基本的、内在的物理属性。

如果感受性质是内在的属性,那么根据我已描述的标准模型,对它们作出还原解释就几乎是无望的。如果这是正确的,那么N-头疼、M-发痒以及其他如此之类的联系就一定会被认为是无法接受进一步解释的“不合理”的事实。我们将完全无法理解:为什么头疼产生于N,发痒产生于M;我们所知道的不过是它们的所作所为。正如倏忽进化论者所论证的,我们也许不得不以“自然的虔诚”承认这些倏忽进化或随附性关系。

那些为了说明意识、将其希望寄托在神经生物学研究上的人,要求我们把神经生物学的某些最新进展当作是在寻求关于现象意识的物理解释过程中的最终成功的吉兆。例如近来脑科学家中有

130 第二部分

这样的推测:大脑中某些神经元以 35-75Hz(有时被称之为“卡玛振荡”或“40-Hz 振荡”)的频率同步激活对于解决下述问题具有至关重要的作用,这个问题就是所谓的结合问题:由不同感觉中枢通道加工的彼此分立的信息(颜色、形状、大小、运动、滋味、声音、结构等)怎样结合在一起,进而形成成为一种连贯的、有意识的认识,比如说把运动着的汽车当作统一的事物。^④某些人似乎认为:如果这种假说最终是真实的,那么它就是对“意识问题”的一种解答,或者说至少是朝着解决迈出了重要的一步。^⑤

这个假说如果是真的,那么它也许是对意识的科学问题的一种部分的解答。这其实是对作为意识状态和过程之随附基础的中枢机制的认同问题。^⑥假设在未来的某一时间点,我们对与不同中枢状态(比如说具有 40-Hz 的振荡以用作“意识统一性”的中枢基质)有关的各种感受性质有一幅详尽的图画。这可能表达了对于意识的一种根本科学的理解。但是它解决了关于意识的哲学问题吗?或者说它推进了这一问题的解决吗?^⑦

不管我们关于这种中枢机制的知识多么详尽,多么激动人心,但是也难以说明:意识的中枢机制之发现怎么可能对驱散笼罩在意识的现象特征上的神秘迷雾发挥作用。像我们已注意到的那样,这里的高深莫测的神秘性就是特殊的现象特征为什么与它们事实上由以产生的中枢基质相关联,或者说前者为什么从后者产生出来。神经科学家在某一天也许会向我们提供关于感受性质与大脑状态相互关联的详尽的一览表,这可能会丰富我们关于大脑的知识、关于我们有意识的生活依赖于大脑中所发生的东西的特殊方式的知识。但是这些相互关联实际上正好是引起哲学困惑的

东西;我们需要解释的是,这种特殊的关联系统,且不说其他无数的可能的关联,为什么在我们的世界中起作用。^⑧而且还有这样一个更深层次的问题,即在本质上是物理的世界中为什么存在着现象意识这样的事情?

谁需要感受性质

我们发现我们自己陷入了二难境地:如果感受性质随附于物理或生物过程,那么它们为什么随附——为什么它们随附于或来自于它们事实上所出自的特定的中枢基质——就仍是一个秘密,即是某种好像完全不能用物理的观点予以说明的东西。(这并不是说,从非物理主义的观点去解释就更容易。)然而如果它们不随附,那么就一定得把它们当作是完全外在于物理领域的现象。在这点上,某些哲学家可能会感觉到:对心性的基本的物理主义探讨已是穷途末路,已经到了寻求非物理主义的替代方案的时候了。但是这又会作为一种凄凉的前景刺伤某些哲学家,而且某些人已选择了否定感受性质存在的英雄的(也许是堂·吉珂德式的)路线。我们大概可以称之为“感受性质虚无主义”(或“感受性质取消主义”)。

感受性质虚无主义以两种变种出现。第一种是主张:不管感受性质是否真的存在,它们在心理学科学中是没有地位的,也没有什么作用。在这种科学中,心理学被理解为这样的科学,其目的是对受到广泛分析的人类行为作出以规律为基础的解释和预言。这种形式的感受性质虚无主义应当与行为主义区别开来;不像后者

132 第二部分

那样,它在构建关于行为的理论时,不反对假设内部状态,不管它们是生理的还是心理的。它没有必要害怕述及意识:它能在觉知的意义上承认意识具有能动的心理作用。它只是说:我们的意识的纯粹质的方面在解释和预言行为时没有任何作用。我们不妨称这种理论为“理论的感受性质虚无主义。”

你可以沿着下述考察路线而到达理论的感受性质虚无主义。感受性质是主观的,因而不是主体间可理解的,这就使对它们的直接的科学研究没有可能性。而且不像物理科学的不可观察的理论假设那样,感受性质的内在特征似乎没有任何可从观察上检验的结果;因为从行为方面看,主体经验到什么感受性质似乎没有任何差异。有差异的地方不过是,主体能把红与绿区别开来——例如她能在交通指示灯显示“走”或“停”时作出可靠的分辨。^⑨根据这样的分辨行为,我们甚至可以推断说:那个主体以某种直接或间接的方式意识到红与绿之间的质的差异,这种分辨能力对解释她的行为(如她的驾驭行为)可能有其作用。然而她的红的经验的内在的质的特征是什么,红对她来说是怎样的,或者说她的红的感受性质像这而不像那,似乎没有什么重要意义——这一切对解释下述行为毫无用处:当红灯亮时,她为什么刹车,在变为绿灯时,她为什么驱车前进。在这里,至关重要的仅仅在于:她能把她红的感受性质与绿的感受性质区别开来。因此有行为结果、进而在主体间可得到理解的东西是主体分辨不同感受性质的能力,而不是她的经验的内在性质。内在性质存在于那里,并起着一种作用,但有作用的东西是关系性的,而且正是感受性质中的差异才携带着信息。换言之,如果没有内在的性质,那么就没有什么需要分辨!因此在

这个意义上,必定存在着感受性质,但是一种感受性质像这,另一种像那,这样的事实是无关紧要的;就分辨的目的而言,可能正好相反,而就心理学而言,情况则依然如故。不过,我们应该注意到:这些考虑并没有为一种整体的感受性质虚无主义提供支持;充其量它们只是表明:感受性质的内在性质并无因果的、理论的作用,因为它们承认感受性质的关系的、结构的属性在我们的分辨行为中具有不可或缺的作用。

现在让我们转到我们称之为“哲学的感受性质虚无主义”这个题目上来。这种观点的支持者论证说:“的确”不存在感受性质这样的东西,对感受性质这一概念的直接分析将表明感受性质只不过是一种哲学的发明。对这种感受性质虚无主义的论证一般从列举与感受性质有经常联系的属性开始,诸如可靠的、不可变更的第一人称可理解性,不可言喻性,第三人称的不可企及性以及内在性。接着,这些论证试图表明:这些概念不是自相矛盾的就是完全晦暗不明的,要么就是空洞的,不适用于我们心理生活中任何可认识的东西。^⑩这些论证中有这样一个问题,即它们倾向于夸大某些哲学家在赞成感受性质时所断言的认识的和别的属性,以至于带来下述不足为奇的结论:没有什么东西有资格作为感受性质。例如,要相信感受性质,就没有必要坚持绝对的第一人称不可错性或第三人称的不可企及性。这一来就有憎恶感受性质的人^⑪所提出的完全不费力气的类比,即认为对感受性质的信念类似于对巫婆、燃素、磁性以太的难以令人信服的信念。总而言之,当神经科学发展到不求助于内在心理生活也能解释人类行为的地步时,对感受性质的信念将被证明是不可信的,正像对燃素的信念在关于燃烧

134 第二部分

的氧化理论为人接受时就被抛弃一样。但是这样的论证依赖于下述前提,即感受性质只是为了解释人类的行为而作出的理论构造,只是一种将为那些严肃对待感受性质的人拒斥的假设。

撇开对这些论证的评论,让我们再来想想维特根斯坦的“盒子中的甲虫”的比喻:

假设每个人都有个盒子,在里面有某东西;我们将其称为“甲虫”。每个人都不能窥探别人的盒子,而每个人都认为,他只要看看他的甲虫就知道甲虫是什么。——在这里,每个人都在他的盒子中看到了某种不同的东西是完全可能的。^⑫

正如维特根斯坦所述,盒子甚至有可能是空的,而且又不清楚:“甲虫”一词在公共语言中传递主体间信息时怎么可能有作用。这意思当然是说:我们对于存在于我们的私人的小盒(我们称它们为“心灵”)中的私人的、直接可理解的感受性质的谈论正像这种关于甲虫的谈论。

要弄清楚怎样反驳诸如此类的关于感受性质、关于感受性质话语之合理性的否定论证,那是很困难的——除了说当我向内看我的盒子时,我发现那里有具有这些特征的某东西,没有别的办法。我还能添加进一步的解释:当某人踩了我的脚趾或一根针刺了我的手而不是某人按摩我的背时,我常能发现具有相同特性的某种东西。如果这仍然于事无补,那么也许就没有什么办法了。我们怎么可能用表示感受性质的语词进行主体间的交流——至少是那些我们通常假定为主体间交流的东西——的确是一个秘密。有这样一个故事,即当阿姆斯特朗被要求解释爵士音乐是什么时,

他回答说:“如果你能够提问,那么你不会不知道的。”^③如果有人想严肃地、诚恳地探讨感受性质是什么,那么它将表明:这些讨厌的东西不仅在形而上学上是可能的,而且其中的一些的确存在于这个世界!

注 释

① 某些作者说的是关于感受性质的“自然主义的”而非物理主义的说明。不过,称之为“自然主义的”远没有称之为“物理主义的”清楚,而且感受性质为什么被“自然化”,或者说那意味的除了“物理化”还有别的什么意思,这都不明确。感受性质为什么不能没有疑义地成为自然界的组成部分呢?只有当引入物理主义的思考时,感受性质才开始带来种种问题。

② 关于这个概念的进一步讨论,请参阅 T. 霍根:“随附性与微观物理学”,《太平洋哲学季刊》,63(1982):29-43。

③ 关于对立的观点,可参阅查默斯:《有意识的心灵》,牛津大学出版社,即出。

④ 关于这假说的流行的分析,请参阅 F. 克里克:《惊人的假说:关于灵魂的科学探讨》,纽约,1994,第 17 章。

⑤ O. 弗拉纳根在他的《再论意识》(麻省理工学院出版社 1992, pp. 59-60)一书中似乎倡导这类观点。

⑥ 克里克本人说:“考克和我把这种观点作为下述主张的一个步骤,这主张是:这种对于或近于卡玛振荡的冲击(在 35-75 赫兹的范围内)的同步激活可能是视觉认识的中枢关联物。”见《惊人的假说》,p. 245。

⑦ 《纽约时报》(1992. 12. 27, p. 10)援引了加利福尼亚技术研究中心一位科学家的话:“我们理解了基因的分子基础之日,就是遗传学的进步到来之时……同样,要理解意识,你就得进到神经细胞、电脉冲密码和脑网络的水平。”

⑧ 关于这些争论,请参阅 J. 莱文:“唯物主义和感受性质:解释的真空”,《太平洋哲学季刊》,64(1983):pp. 354-361;“抛弃‘可能是什么’”,载 M. 戴维斯等编:《意识》,牛津:布莱克韦尔,1993。

136 第二部分

⑨ 即使当绿、黄、红光的方位是随机的时,也是如此(我认为:色盲是把这些光的相对方位作为线索的)。

⑩ 关于按这些路线所作的论证,可参阅 D. 丹尼特:“取消感受性质”,载 A. 马塞尔等编:《当代科学中的意识》,牛津大学出版社 1988;G. 哈曼:“经验的内在的性质”,《哲学前瞻》,4(1990):31-52。

⑪ 相信感受性质的人即所谓的“感受性质怪人”。

⑫ L. 维特根斯坦:《哲学研究》,牛津:布莱克韦尔,1953,§ 293。

⑬ 这一轶闻出自 N. 布洛克,他在“功能主义的麻烦”一文中利用它提出了同样的观点,载 C. 萨维奇编:《明尼苏达科学哲学研究》。第 9 卷,明尼苏达大学出版社 1978,p. 281。

功能主义与感受性质*

悉尼·休梅克

I

在《心理状态不是什么》这篇新作中,N. 布洛克和 J. 福多对“功能状态同一论”(FSIT)提出了许多反对意见。FSIT 主张:“对于符合心理谓词要求的任何有机体来说,存在着独一无二的最好描述,以至于该有机体的每种心理状态都同一于与该描述有关的、它的机械状态中的一种。”^①FSIT 是功能主义的一种变种,而他们把它看作是这样一种更一般的学说,即认为“心理状态的类型—同一条件指称的不过是它们与输入、输出的关系以及相互之间的关系。”^②布洛克和福多认为,他们所提出的大多数反对意见只是反对 FSIT,并不针对更广泛地被阐释的功能主义。这里我对这些反对意见不感兴趣。但是他们提出了一个反对意见,他们说:它“大概可以用来说明:心理状态绝不能从功能上加以定义,它们也不可

* 该文原发表于《哲学研究》XXV II, 5(1975, 5): 292 - 315。后经修改重印于 N. 布洛克选编的《心理学哲学文选》(哈佛大学出版社 1980)。D. 罗森塔尔在编辑心灵哲学原著选辑《心灵的本质》(牛津大学出版社 1991)时,收录了修改过的该文。本中译据后者译出。

138 第二部分

能与任何能在抽象的自动机上被定义的属性相一致。”^③简言之，该反对意见就是：FSIT 以及一般的功能主义所假定的“类型上认同”心理状态的方式“未能涵盖这类状态的这样一个特征，即对决定它们的类型即它们的‘质的’特征至关重要的特征。”^④

布洛克和福多只用了两页的篇幅来讨论这一反对意见，并用一种极富试探性的方式予以表达；因此极有这样的可能，即我在这里讨论它所用的篇幅与他们赋予它的重要性是不相称的。但是他们对一种反对意见给予了明确而生动的系统阐述，这种反对意见在谈话中比用文字表达更经常被碰到，被听得出来，在我看来，它提出了根本性的问题。其他哲学家也提出了极其相似的反对意见，他们说：功能主义（或行为主义，或唯物主义，或关于心灵的“因果”理论——该反对意见旨在反对所有这一切）不能说明心理状态的“原感觉”构成成分，或它们的“内在的”或“现象学的”特征。我这里首先关心的不是：这种反对意见对 FSIT 是否是致命的；如果我对这一理论的理解没有错的话，那么它足以为布洛克和福多所提出的、反对它的其他意见所驳倒。但是由于他们是“从广义上”描述功能主义的，因此粗略地说，许多哲学家包括我自己在内都觉得它是富有吸引力的一种观点；在我看来，很有必要考察的是：它能否在这种反对意见面前得到辩护。

我将像布洛克和福多一样，把心理状态（或毋宁说某些心理状态）当作有“质的特征”或“质的内容”的东西。我希望在接下来的讨论中出现这样的局面，即这并没有把我与“私人对象”或“私人语言”的头脑清醒的反对者觉得可予责难的东西扯在一起。

II

布洛克和福多分两步阐发他们的意见。他们把第一步称之为“颠倒感受性质论证”，第二步即所谓的“缺席感受性质论证”。

因为布洛克和福多觉得，证实主义者反对“颠倒光谱假说”的概念融贯性的常见论证不可信，因此他们倾向于认为：“颠倒感受性质”这样的事情是可能的。他们对它的理解是：“无论在什么机器中，当工作台的状态据说同一于疼痛时，如果每个人真的有略微不同的感受性质（或极不相同的感受性质）”，那么就可能有感受性质的颠倒（大概是它的一个极端的例子）。^⑤这种可能性与功能主义的下述似乎合理的假定势不两立，即“没有什么东西是‘疼痛’状态类型的一个标记，除非它感觉到好像是一种疼痛……即使它以疼痛所表现的一切方式关联于有机体的其他所有心理状态。”^⑥

布洛克和福多并不认为，感受性质颠倒的可能性本身就构成了对功能主义的判决性驳难，因为他们认为，否定疼痛一定在质上相同这种表面上似乎合理的假定（大概还有这样一种有关的假定，即任何在质上同一于疼痛的东西本身就是疼痛），可能正对功能主义者的胃口。^⑦如果感受性质颠倒实际上在疼痛过程中发生了（即如果在功能上同一于疼痛的状态在质的特征方面不同于它），那么他们会说：“有机体的感受性质的特征与是否处在疼痛中无关，或者（同等地）说疼痛对不同的有机体来说，感觉是极不相同的，这样说可能是合理的。”^⑧这样一种观点事实上未曾听说过。在唐·洛克先生看来，“一种处在疼痛感觉中的感觉并不是它怎样感知的问

140 第二部分

题,而是它所属的种类的问题,这个种类由身体的伤痛所引起,进而又引发行为”。^⑨A. 多纳根认为,维特根斯坦有这样的看法,即“如果你我碰到了某种令人恐怖的事情,并且我们都通过咬、磨擦上下颚以及某种怪相等自然地将其表现出来了,那么我们就可正确地说,我们有相同的感觉,比如说牙痛。这些方式所表现的事情的内在特征对你、对我是否相同,这与‘牙痛’一词的意义无关。”^⑩

但是布洛克和福多并没有把这种对颠倒感受性质论证的答复当作明显的错误加以抛弃,而是认为,它可能为通向更严厉地打击功能主义的论证即缺席感受性质论证打开方便之门。他们的想法也许是:一旦承认了一特定的功能状态没有特定的质的内容也能存在,那么就很难否认这样的可能性,即它完全没有质的内容(或特征)也能存在。不管怎样,他们进一步指出:

就我们所知的而言,在法则学上有这样的可能性,即两种心理状态在功能上可以是同一的(即同等地关联于输入、输出和后续状态),即使其中仅有一种状态有质的内容。如果是这样,FSIT 就会要求我们说:有机体即使什么也没有感觉到,仍可处在疼痛中,这个结论似乎是完全不可接受的。^⑪

如果“缺席感受性质”的事例是可能的,即如果一种状态在功能上同一于一种有质的特征的状态,而其自身并没有质的内容,那么不仅 FSIT,而且还有广义意义上的功能主义似乎都是站不住脚的。

III

如果心理状态在“质的特征”方面可以是一样的或有区别的,

那么我们就应当能够说到一类状态,不妨称之为“质的状态”,它们的“类型—同一条件”可以用质的(或现象学上的)相似性概念而加以具体说明。对于一状态所具有的、每一起决定作用的质的特征而言,存在着(即我们能够定义的)一种起决定作用的质的状态,它是一个人在得到了明显有那种质的特征的状态的条件下正好具有的状态。例如有这样一种质的状态,某人具有它的唯一条件是,他有一种感觉,其感受就像我最近的头痛感受那样。这一来,质的状态本身就是“心理的”或“心理学的”状态。因而这就带来了对布洛克和福多下述主张的非议,即功能主义者要对付“颠倒感受性质论证”,就得坚持“有机体的感受性质的特征与它是否处在疼痛中无关”。如果心理状态包括了质的状态,那么功能主义者关于疼痛所说的就不可能适用于一般的心理状态,因为说有机体的感受性质的特征与它具有什么质的状态无关就是自相矛盾的。当然,如果质的状态本身能从功能上加以定义,那么感受性质颠倒的可能性对功能主义就没有提出什么难题,功能主义就没有必要对下述问题作出反直觉的否定回答,即有机体的感受性质的特征与它是否处在疼痛中有关。但是如果像布洛克和福多那样明确地假定,质的状态不能从功能上定义,那么就存在着一类心理状态,即质的状态本身,它们不能从功能上加以定义。

这引出了这样一些问题,我在后面将回过头来讨论,即(a)在什么意义上,质的状态不能从功能上定义(或者说如果感受性质颠倒是可能的,那么在什么意义上,质的状态不能从功能上定义), (b)它们不能从功能上加以定义(无论在什么意义上),这对功能主义构成了严重威胁吗?正如我们在本节余下部分将看到的那样,

142 第二部分

在思考所谓的“感受性质缺席”之可能性时也提出了这个问题。

只要我们能够说明:如果一种状态在功能上同一于有质的内容状态,那么它本身必定就有质的内容,基于此我们就能断定“缺席感受性质”这种情况是不可能的。人们试图用下述方式阐释功能同一这一概念以做到这一点,即质的状态包含在“别的心理状态”中,根据与后一类状态的关系,以及与输入、输出的关系,特定心理状态的“类型-同一”就得到了定义。因此人们可能会论证说:如果一特定心理状态有某种质的特征,那么这便与它所处的、与某种特殊的质的状态(即一个人在他处在具有质的特征的状态时所处的质的状态)的某种决定性的关系有关,而且还可以论证说,任何在功能上同一于它的状态一定处在与那质的状态的同样的关系中,因此一定有相同的质的特征。^⑩但这个论证是难以令人信服的。一种驳难就是:既然质的状态本身不能从功能上加以定义(假定颠倒感受性质的可能性),那么把它们包括在其他心理状态由以从功能上得到定义的、或者说“功能同一性”据以得到定义的心理状态中就是不合逻辑的。后面我会再回过头来讨论这种驳难,因为它乍一看也是一种对我下面要表述的更合理的论证的驳难。另一驳难是:一种状态与一质的状态所具有的关系,在具有与那质的状态一致的质的特征时,绝不是因果关系,因此也不是心理状态据以从功能上得到定义的那种关系。但是我下面要表述的论证是不会受到这种驳难的,在我看来,那论证表明的是:根据对功能同一概念的任何合理的阐释,一种状态如果其本身没有质的特征,那么绝不可能在功能上同一于具有质的特征的状态。

疼痛关联于别的心理状态的一种重要方式就是:疼痛在适当

的条件下引起了内省性觉知,使其知道有某种质的特征,即是说感觉到了某些方面。我要假定的是:这里的含义可以部分解释如下,即由于有适当的环境,处在疼痛中常常引发了我将称之为“质的信念”的东西,即大意是相信某人有某种感觉的信念(或者更抽象地说,相信某人处在一种有一定质的特征的状态中,或再换种方式说,相信某人有某种质的状态)。在功能上同一于一种疼痛状态的任何状态都会具有那种疼痛状态,不仅是(1)它倾向于以某些方式影响外在的行为,(2)它倾向于在那个人身上产生这样的信念,即原来他有某种器质性的不适(例如他被划破了或烧伤了),而且(3)它倾向于在那人身上产生质的信念,即让他想到:他有一种充满着某种质的特征(他不喜欢的特征)的疼痛。然而根据“缺席感受性质论证”,这种状态可能缺乏质的特征,因此不是疼痛。下面让我们看看,这样说是否有道理。

假设“缺席感受性质”这种情况是可能的,但如果它发生了,我们怎样去分辨呢?我们每一个人有什么权利拒斥这样的主张,即他自己也许就是这个样子,他本人就没有充满着质的特征的状态?更有甚者,我们有什么权利拒斥这样的主张,即也许根本就没有人有过那种感觉(或其他有质的特征的状态)?行为为一个人的心理状态有什么质的特征(如果有的话)提供了无判决效力的证据,这自然是众所周知的观点。但通常作为其基础的则是这样的观点,即人本身有到达这种质的特征的通道即内省,它比行为可能提供的通道更“直接”。不管内省还会是别的什么,但它肯定是人的心理状态与他对这些状态的信念(或他的知识或他的觉知)之间的桥梁。因此提出我们的问题的一种方式就是追问:假如根据某人现

144 第二部分

在所处的状态加上适用于他的心理学规律(它描述了他的状态的相互关系,以及与输入、输出的关系),可得出结论说:全部可能的行为资料以及全部可能的内省资料都清楚地证明了他处在疼痛之中,那么是否有证据(对任何人来说)证明:他并不处在疼痛之中?我不清楚这是怎样可能的。(也许)我们确实可以想象,“脑检镜”显示:那个人不处在我们疼痛(因此这样认为)时常常进入的那种神经生理状态之中。但这显然提出了这样的问题,即我们根据什么说,我们真的有疼痛(即有充满着质的特征、并在它与输入、输出和别的心理状态的关系中起着适当功能作用的状态)?在这里,如果行为和内省的根据还不够,那似乎没有什么是不够的。但是如果它们对我们来说是足够的,那么它们对我们假设的人也是如此。无论如何,如果我们得知:一个人的状态在功能上同一于在我们身上是疼痛的状态,那么难以说明的是:他与我们之间的心理学上的差异怎么就能证明:他的状态缺乏质的特征;因为如果有材料能为我们证明他的心理状态,那么证明他的状态在功能上等同于我们的状态的根据事实上也能证明:我们与他之间的任何生理学差异与疼痛状态是否在他身上实现是无关的,尽管与怎样在他身上实现不无关系。

坚持有这样的逻辑上的(或更严格地说,法则学上的)可能性,即一种缺乏质的特征的状态从功能上可能同一于有质的特征的状态,就是使质的特征既与我们在知道别人的心理状态时让我们自己知道的东西无关,又与我们在知道我们自己的心理状态时让我们自己知道的东西无关。任何状态(根据这种观点)大概没有可能的物理结果,从这种状态出发,我们借助“对最好的解释的推论”,

能够证明它有质的特征;因为如果有,那么我们就通过说明它在如此这般的环境中倾向于引起那些物理结果,而对享有质的特征至少作出部分的功能描述,进而我们就不能承认:一种缺乏质的特征的状态在功能上可以同一于有此特征的状态。基于已述的理由,如果“缺席感受性质”的案例是可能的,那么质的特征一定不可能到达内省。如果质的特征是以这种方式与心灵的一切知识包括自知和他知无关的某种东西,那么否认疼痛有质的特征就不是完全“不能接受的”,而恰恰是正确的判断。但是当普通人谈论他们感觉怎样、或谈论事物对他们来说看上去、听起来等像什么时,假定他们谈论的是任何人根本不可知的东西,这则又是荒谬的。(而且,正如关于知识的因果理论所暗示的:独立于它们所描述的事物之因果力的状态或特征原则上是不可知的,指称的因果理论也暗示说:这种状态和特征原则上是不可名状的、不可指称的。)言归正传,如果我们把质的特征当作是某种可知的东西,就像我们把人类的感觉当作是可知的一样(在最低限度上,如果它能为内省所知),那么缺乏质的特征的状态就不可能、甚至在逻辑上也不可能从功能上同一于有此特征的状态。

这不是一种“证实主义的”论证。它并未假定有意义与可证实性(或可知性)之间的任何一般联系。它假定的是:如果应当有什么理由假定(像“缺席感受性质论证”所做的那样):它们有“质的特征”对于疼痛和别的心理状态来说是必不可少的,那么我们就一定得承认:“质的特征”指称的是某种不可知的东西,至少像在某些方面我们把疼痛(我们自己的和别人的)当作是不可知的一样。它还假定:如果某心理状态有一种特征,它完全独立于那状态的因果力

146 第二部分

量(即是这样的,其在场与不在场对那状态倾向于引起别的状态等的产生没有任何影响),因此与它的“功能同一性”无关,那么这种特征就完全是不可知的(可以说,这承认了知识的因果理论)。

针对这个论证,就像针对先前的论证一样,可能有这样的反对意见,即别的心理状态——由于与它(以及输入和输出)的关系,一特定的心理状态便从功能上得到了定义——一定不包含本身不能从功能上加以定义的状态。因为可以说,我称之为“质的信念”的状态据说不能从功能上定义,就像质的状态本身不能被如此定义一样。这些状态与其他状态的最重要的关系似乎是它们与独特地引起它们的质的状态的关系,然而(该论证继续论证说)后者不能从功能上加以定义,因此就不能合法地在关于前者的功能定义中被述及。而且(请记住,颠倒感受性质的案例之可能性在这里是没有疑义的)似乎可以合理地假定:如果两个人在他们疼痛的质的特征上彼此有别,但是这种区别在任何可能的行为中没有暴露出来,那么他们在质的信念方面也彼此有别,而且这后一方面的差异也是这样,其存在在任何可能的行为中没有显露出来。如果这是可能的,那么似乎就有足够的理由否认:质的信念能有功能定义,就像应否认质的状态能有功能定义一样。

这种驳难并未触及到我的论证所隐含的实质,即是说由于我们并未承诺关于他人疼痛的极端的怀疑主义,我们也就不能否认:某人感觉到剧痛之说出就足以证明他有某种或别的质的状态,之所以如此,是因为某人这样说通常是他有充满着质的特征的状态的一种结果——这本身再清楚不过地表明:如果一个人的心理状态有质的特征,如果另一个人的别的相同的状态缺乏质的特征,那

么这两种状态就其有助于影响行为(输出)而言是彼此不同的,从而在功能上也相互有别。而且,“颠倒的感受性质”之可能性似乎意味着:质的状态从而质的信念不可能从功能上加以定义。为了说明这是否与功能主义相容,它是否会摧毁上述论证,我们有必要考察:在什么意义上,说质的状态(以及质的信念)不能从功能上加以定义是真的,在使不能从功能上定义的心理状态的指云得以进入其他心理状态的功能定义的过程中,将有什么样的限制。

为了考察这些问题,我想改变例子,把我们从对疼痛的考察转向对视觉经验的考察。这种转换是有好处的,理由有二。第一,在我看来,“光谱颠倒”之可能性(一个人的颜色经验在其质的或现象学特征上完全不同于另一个人关于同一颜色的经验),其疑义比疼痛案例中“感受性质颠倒”(疼痛在不同的人身上感觉起来极其不同)的可能性要少得多。第二个与此有关的理由是,把(例如)观看蓝颜色与其质的特征区别开来比把疼痛与其质的特征区别开来要容易得多,因而考察对质的状态的指云怎么可能进入对观看颜色的功能说明比考察对这种状态的指云怎么可能进入对疼痛的功能说明要容易得多。

IV

如果我观看某事物,那么它就以某种方式显现给我,它显现出来的方式在变化程度和各个侧面上,与在另外的情况下别的事物显现给我或曾显现给我的方式既相似又有区别。我们之所以能看到相似与差异由以被发现的这些对象及其属性,是因为这些“显现

148 第二部分

的方式”之间的相似和差异以整体的方式与被看的对象之间的相似和差异相互关联着。^⑬当事物以某种方式被显现时,例如事物呈现给某人的方式就像我现在注视窗外时事物呈现给我的方式,我便认为他处在一种质的状态之中。因此观看必定涉及到质的状态的出现。而且对这些质的状态的指云便进入了看上去颇似关于观看的功能说明之中。因为说某人看到某物是蓝颜色,其意思是说:

S 看到某物是蓝色,当且仅当(1)S 有质的状态的贮存,其中包括一组状态 K,它们与对象的颜色相联系,因此(a)具有某种颜色的对象在“标准条件”下形成的视觉刺激在那个人身上产生有关的质的状态,(b)K 中的状态之间的“质的”或“现象学的”相似性的程度对应于有关颜色之间的相似性程度,(2)那个人 S(a)现在处在与蓝色有联系的质的状态中,(b)这是蓝色的某物的视觉刺激的结果,(c)因为(a)和(b),他便相信他面前有蓝色的某物。^⑭

现在我必须限定下述断言:以某种方式“被显现”就是一种质的状态。如果要一个人描述他怎样被显现,或更自然地说,事物怎样显现给他,那么他首先就可能说:某对象在他看来是蓝色的,或者说它显现给他的好像是他曾看到过的某种蓝色的东西,或者说(如果他是一个操“显现语言”的哲学家)他“被显现了蓝色”。接着自然会让某人处在那时在他身上与蓝色东西的视觉刺激有联系的质的状态这一点成为他被显现了蓝色的条件;即是说自然要对“S 被显现了蓝色”作出分析,它与上述对“S 看到某物是蓝色”的分析是一样的,只不过排除了条件(2)中的子句(b)和(c)。但是如果我

们这样做了,那么“被显现了蓝色”本身就不是质的状态。或者说,如果光谱颠倒是可能的,那么无论如何将是如此。我们可能会这样概括那个情境,比如说:基于预计的分析,被显现为蓝色是一种功能状态,对它的功能描述总是要求它有某种(或别的)质的特征,但是在不同的人身上(假定主体间颠倒光谱的可能性),或在不同时间的同一个人身上(假定主体内光谱颠倒的可能性),并不要求它有相同的质的特征。不过这再次引出了这样的问题,即质的状态本身能否从功能上被定义,如果不能,那么在对其他心理状态的描述中,它们能否合理地被指云。

我想,“被显现为蓝色”这一表达式可能有一种用法,在运用中,它表示的是质的状态。我可以确定这个表达式的指云,方式是规定它指称现在(1974年4月)在我身上与看蓝色的东西有联系的质的状态(或者说,既然它是谓项而不是单个的术语,那么它就是对那状态的断定或归因。)^⑤以这种方式去理解该表达式,如果我明天经历了光谱颠倒,那么就不再有这样的情况发生,即当我看到蓝的东西时,我被正常地显现了蓝色,而将是这样的,即在此情形下,我被正常地显现了黄色。^⑥(相比较而言,在上述“被显现为蓝色”的“功能”意义上,下述说法在主体内的光谱颠倒之前和之后可能是真的,即当我看到蓝色的事物时,我正常地被显现了蓝色,即使被显现为蓝色在后来自然有这样的质的状态,它是另一视觉状态比如说在先前被显现了黄色时具有的那种状态。)我认为,在所拥有的以这种方式作为视觉感受性质的“固定的指示词”(或“固定的谓项”)的表达式中,并不存在什么重要的有用的东西。另一方面,对于我们为什么不能拥有它们,我根本不明白有什么理由。但是如果我们真的拥有

150 第二部分

它们,它们又不能从功能上被定义。这类术语不得不为克里普克的“指称确定”或显定义(是这种情况的特例)所采用。可以肯定,通过运用同类的其他表达式对这些表达式之一给予语词定义是有理论上的可能性的;正像我利用“不是黄或红或绿等的颜色”这类形式的摹状词可以定义“蓝色”一样,我可以把“被显现为蓝色”定义为下述形式的摹状词的同义词,即“既不是被显现为黄、又不是被显现为红、绿等的那种颜色感受性质。”但是这完全于事无补,因为关于所有视觉感受性质的名称(或谓词)显然不可能以此方式没有循环地被定义。因此如果假定定义功能状态的话语等同于定义表示质的状态的名称或“固定的指示词”的话语,那么似乎在一定的意义上,质的状态就不能从功能上加以定义。

但是把这个结论强加于我们的东西是光谱颠倒之表面上的可能性。我认为(如果稍有区别的话):强迫我们承认光谱颠倒的可能性的东西是主体内的光谱颠倒的表面上的可设想性和可分辨性。如果我们反思一下后者,那么我相信我们将看到:当我们不能从功能上定义特殊的质的状态时,在一定的意义上,我们就能够从功能上定义质的状态的类别——我们能够从功能上定义这个类别的成员之同一性条件,因为我们能从功能上定义质的(现象学上的)相似性与差异性的关系。这就是我下一节要论证的。

V

开门见山地说,光谱颠倒是可能的这一主张隐含着这样一种我认为在经验上是虚假的主张,即它认为,存在着一种把颜色的确

定样式映射到颜色的下述确定样式的方法,(1)每一确定样式(包括“模糊的”、暗淡的颜色和纯光谱颜色)被映射到某种确定样式上,(2)某些样式至少映射到自身以外的别的样式上,(3)对于任何视觉正常的人来说,映射保留了颜色之间的“距离”和“中间状态”这样的关系(因此如果样式 a、b、c 被分别映射到样式 d、e、f 上,那么视觉正常的人对于与 b、c 有关系的 a 大致相似于与 e、f 有关系的 d 就会作出相同的判断),(4)除了为关于色素等的混合属性的知识在经验上所影响的直觉之外,映射保留了我们的所有直觉。如关于哪一些样式是“纯”颜色,哪一些在其中包含了别的颜色(例如,如果样式 a、b 分别映射到橙色和红色上,那么我们就倾向于说:a 没有 b 纯,而且也许其中有 b)。但是即使我们的颜色经验事实上不是这样,以致这类映射是可能的,那么在我看来它大概是可以设想的——而且这就是与我们现在的哲学目的有关的问题。^⑩例如我认为,我们清楚地知道:不从彩色的角度,如用黑、白以及各种形式的灰的观点去看世界会是什么样子——因为我们常常在相片、动画片和电视中以这种方式去看它。而且显然存在着符合颠倒所需的条件的、非彩色样式之间的相互映射。在下面的讨论中,为了方便起见,我将假定:这种映射对于一切颜色都是可能的——但我不认为:任何重要东西都取决于这一假定是否正确。

假如有这样的一种映射(为方便起见,再进一步假设:只有一种映射),我们不妨称之为每一样式都把那样式的“颠倒”映射于其上的样式。如果每一颜色样式向一个人显现的方式就是它的颠倒样式向另一个人显现的方式,或者换句话说,如果对于每种颜色样式而言,在一个人身上与看到那样式相联系的质的状态就是在另

152 第二部分

一个人身上与看到这种样式的颠倒有联系的质的状态,那么我们将有主体间的光谱颠倒。如果颜色的各种样式对某人显现的方式发生了变化,每一个看上去就像先前被看到的它的颠倒,那么我们就有主体内的光谱颠倒。

关于光谱颠倒,最使我们感到困惑的是:如果它能在主体间出现,那么似乎就没有办法分辨两个人的颜色经验是否是一样的,或者说他们的颜色的光谱相对于对方来说是否是颠倒的。主体间光谱颠倒所构成的经验之间的整体的差异性对任何人的内省来说自然是不开放的。这些差异在行为中似乎没有表现自身的方式——但这样的假设即你的光谱相对于我的而言是颠倒的,以及这样的假设即我们的颜色经验是相同的,似乎引起了关于我们行为的相同的预言。这里我们自然不会忘记下面假设的情况,在其中,不同的颜色对一个人来说看上去总是一个样子,而对另一个人则是另外的样子。当我们考察主体内的颠倒光谱的事例时,情形似乎完全不同了。首先,一种变化会显现在经历了那种变化的人的内省面前,或内省加记忆面前。但是如果是这样,那么别的人就会经过那个人的报告知道那种变化。而且,几乎不为人注意的是,存在着非言语行为,像言语行为一样,它可以说明这种变化。如果对一动物进行训练,让其以特定方式对有某种颜色的对象作出反应,然后再让它用那些方式、自发地对具有颠倒颜色的事物作出反应,如果它令人惊异地表明:它作出反应不再需要以习惯的方式加以奖赏,那么这肯定能证明它经历了光谱颠倒。就人而言,我们可能同时提供这类证据和人的证言证据。^⑨

如果我们认为,我们不可能有关于主体内光谱颠倒的证据,那

么我想我们就完全没有理由主张:任何类型的光谱颠倒,不管是主体间的,还是主体内的,在逻辑上都是可能的。断言光谱颠倒可能但又认为在主体内,它是不可分辨的,这割断了我们所假定的、存在于质的状态与对它们的内省认识之间(在它们与由它们所引起的质的信念之间)的联系,割断了它们与关于世界的知觉信念的联系以及通过信念与行为的联系。毋庸置疑,人们能如此定义“质的状态”这一术语,以至于使它们具有这些联系对于质的状态来说是多余的。但是这样一来,事物在人们看来是相似的、不同的,不是由于“质的状态”(在那个被定义的意义)之间的相似性和差异性。而且假设人们在看到具有不同颜色的事物时所得到的质的状态极其不同,是没有什么哲学价值的,其实,它也不是人们通常所理解的那种“颠倒光谱假说”。诚然,假设主体内光谱颠倒发生了,但不可分辨,这就完全像“缺席感受性质假说”一样是自相矛盾的,后者假定,“从功能上同一于”有质的内容的状态的状态本身可能缺乏质的内容。这样假设是没有意义的,除非它们内部的关键概念含蓄地被定义或重新定义,而这样做结果又使那假设成为空洞无物或索然无味的。

但是在假定个人内部的颠倒光谱是可分辨的时,关于质的状态、关于这些状态之间的质的或现象学的相似性与差异性关系,我们又该作何假定呢?接下来,我将把标记质的状态当作“经验”,并认为:如果经验能同时为人所意识到,即当他正确地记住它,当他实际上正拥有它时,经验就可算为他意识到了,那么我将说:经验是共意识的。如果我们以我已说明的方式把主体内的光谱颠倒理解为可分辨的,那么我们假定的一个东西就是:当经验是共意识的

154 第二部分

时,它们之间的相似性就倾向于引起这样的信念,即相信物理世界存在着客观相似性,亦即经验出现于其知觉中的对象之间的相似性,而它们之间的差异性则倾向于引起对世界上存在的客观差异性的信念。进而,这些信念(与人的需要和别的心理状态一道)又引起与之相应的外部行为。这说明了怎么可能存在着作为光谱颠倒之证据的非言语行为;该行为是关于下述事物的错误信念的表现形式,这些事物来自于这样的事实,即假如在主体内发生了颠倒光谱,那么具有相同颜色的事物在颠倒之后会产生质上不同的经验,它们不像颠倒之前产生的经验,而具有每一种颜色的事物在颠倒之后又会产生这样的经验,它们在质上酷似颠倒之前、由具有不同颜色的事物所产生的那些经验。

但是由于某种理由,颠倒光谱的受害者即使不会以这种方式被引导得到关于客观相似性与不相似性的错误信念,不会据以行动,但我们仍有证据证明:他的光谱颠倒了——因为他能告诉我们它颠倒了。如果假定在这种情形下他能知道有光谱颠倒,因而有条件告诉有关它的信息,那么我们对质的相似性与差异性的关系就进一步假定了某种东西,即是说当他们处在共意识经验之间的关系时,这就倾向于引起对处于这些关系的内省认识,即它倾向于引起适当的正确的“质的信念”,意即关于这些关系存在的信念。

那些谈论心理状态有行为标准的哲学家有时说:经验相似的标准就是其主体诚实地报告或倾向于报告经验是相似的。^⑩如果我们用功能主义的术语重新表述这一观点,那么它就成了这样的观点:使经验成为质上相似的东西至少部分是:它们引起或倾向于

引起它们的主体有质的信念,意即关于这种相似性存在着的信念,由于这个信念,就有作出类似的语言报告的倾向。但是作为对质的相似性的功能定义,这自然是循环的。如果我们想解释:说经验是相似的意味着什么,那么我们就不能把关于相信经验是相似的这一概念当作是已得到充分理解的东西,不能把它作为我们解释中有用的东西。

但是从功能上、根据我所述及的第一种关系定义质的相似性和差异性概念并不会陷入这种循环,所谓第一种关系,即下面两方面的关系,一方面,一个人的经验在某些方面是质上相似或相异的,另一方面,他相信世界上客观存在着某种相似性或差异性,进而最终以某些方式行动。我相信能够想出一个案例,即使我这里不打算这样做,因为说感觉经验倾向于引起对自身、对它们的相似性与差异性的内省认识,对于具有人类的概念能力的生物而言,就是它们倾向于引起对世界中的对象以及它们之间的相似性和差异性的知觉认识的一个不可避免的副产品。我的主张在于:使经验之间的关系成为具有质的(或现象学的)相似性的关系的东西显然是,这种关系在对客观相似性的知觉认识中所起的某种“功能”作用,即它倾向于产生关于这种相似性存在的知觉信念。同样,使经验之间的关系成为具有质的差异性的关系的东西是它在对客观差异性的知觉认识中所起的相应的作用。^②

这个主张自然是模糊的、概要性的。但我这里必须予以述及的是:我们按照这些方法可以对质的相似性和差异性作出明确说明这一主张是合理的,就像主张信念、愿望这类心理状态能从功能上定义一样。因为我的目的不是这样一个狂妄的目的,即说明功

能主义提供了完全令人满意的心灵哲学;它只是一个非常诚实的目的,即说明某些心理状态有“质的特征”这一事实对功能主义者并没有提出什么特殊的难题。说明后者的重要一步就是说明:关于质的相似性和差异性的概念像其他心理概念一样,是功能主义定义的合理的候选者。我早就承认:有这样一个意见,即认为特定的质的状态不能从功能上加以定义。但是也得记住:把质的状态与其他类型的心理状态区别开来的东西是:它们的“类型-同一条件”应当根据质的相似性概念而被给予。在我们开始讨论时,用这些术语详述同一条件与用功能术语予以详述似乎有鲜明的差别。但像我已表明的那样,如果质的相似性概念本身能用功能术语加以定义,那么这种差别就变模糊了。如果后者是这样,进而质的状态的同一条件能用功能术语加以说明,那么下述说法正如我前面说过的那样就不是不恰当的,这说法是:即使特定的质的状态不能从功能上加以定义,但质的状态的类别是能从功能上去定义的。

VI

现在我们开始讨论这样的问题:在为其他种类的心理状态下功能定义时参照质的状态是否合理。

根据一种阐释,心灵哲学中的功能主义是这样的学说,即主张心理的或心理学的术语原则上是可以用某种方式加以取消的。为了简化问题起见,如果我们把我们的心理词汇当作是由表示心理状态及关系的名词(而不是对这些状态和关系作出归因的谓词)构成的,那么其主张就将是,这些名词能被看作是与限定摹状词同义

的,即使不用任何心理词汇,这样的每一摹状词在原则上也是可以公式化的。心理状态的确可以用这些限定摹状词从量上加以限定,而且在某些情况下可以同等地得到指称;但一旦如此,它们便被描述了和被等同了,当然不是明确地用心理主义术语,而是用关于它们的相互的、以及与物理输入、输出的因果关系及别的“局部中枢”关系的术语。^②

我已说的这些所隐含的意思是:质的状态(如果我们有的话)的名称不能像这类限定摹状词那样被定义——这自然是根据“感受性质颠倒”是可能的这一假定。如果一特定质的状态(与其他心理状态一道)在沟通输入与输出的联系中所起的因果作用也可能为其他质的状态所承担,如果那质的状态能够发挥不同的作用,那么它发挥那种因果作用对那个状态来说就是不必要的,进而它不是这样一种术语的意思或意义的组成部分,这术语固定指示的是:如此被指示的状态就是那起着这样一种因果作用的状态。再者,既然这种术语不能以这种方式、在肯定限定摹状词时被取消,那么它就不能出现在从功能上定义关于某种别的心理状态的名称的限定摹状词之内——假如这种功能主义定义的目的是取消心理术语,而肯定物理的和局部中枢的术语。

但是在这里没有这样的含义:质的状态不存在于用定义其他种类的心理状态的限定摹状词以量词限定的状态中。它似乎是用量词对这类状态的限定,而不是对这类状态的特殊状态的指称,因而它在定义别的心理状态时就是必要的。如果光谱颠倒是可能的,那么我们就不要让任何质的状态的出现成为看到(或好像看到)某蓝色的东西的必要条件,但是我们必不可少的是:在人的历

158 第二部分

史的任何特定时刻,存在着某种或别的质的状态,它(在那时)一般与他看到(或好像看到)蓝的事物有关联。毫无疑问,对看蓝色事物时的质的状态的作用的说明利用了质的相似性与差异性的概念;但是像我所建议的那样,如果这些概念本身能从功能上加以定义,那么这对功能主义者并没有带来什么麻烦。

然而,就这里考虑的那种强意义来说,即就定义中不存在对心理术语的必不可少的(不可排除的)运用而言,似乎有某些心理状态(除质的状态之外)是不能从功能上加以定义的。试考察我称之为“质的信念”的状态。所谓质的信念即关于质的状态的信念,尤其是对人们(自己)处在特定质的状态中的信念。质的信念可以分为两大类,一是在其命题内容中,存在着对特定质的状态的指云,二是在其命题内容中存在着对质的状态的量化,但没有指云特定的质的状态。就我的理解而言,第二类质的信念对功能主义者并没有提出什么特殊的难题;如果其他类型的信念能从功能上加以定义,那么这些质的信念也是如此。但第一类质的信念似乎拒绝功能定义。试考察我在下述条件下所表达的信念,即如果我说“我处在被显现了蓝色的状态中”,这里所用的短语“被显现了蓝色的状态”固定指示的是特定的质的状态。如果我想从功能上即从它与其他心理状态、与输入和输出的关系上描述这种相信的状态,那么我们似乎不得不在描述中提到那信念所关于的质的状态——我们不得不说:相信某人被显现了蓝色这一状态一般是被显现了蓝色这一状态的结果。如果是这样,那么没有心理术语的必要运用,就不可能定义这种状态(第一类质的信念)。

但是这对我们从功能上定义其他类型的心理状态并不构成任

何障碍。因为当我们想在我们对某些类型的心理状态的功能描述中放进它们引起了第一类质的信念(即这样的信念,在其命题内容中,存在着对特定质的状态的指云)时,这就用不着我们在我们的功能描述中同等地述及这类信念;有必要涉及到的不过是对这种信念的量化。例如我们可以在我们关于疼痛的功能描述中加入这样的内容,即处在疼痛中常常导致某种质的信念,意即相信某人具有某种特殊的质的状态,而用不着述及任何质的状态,说处在疼痛中倾向于引起对它的信念。如果对质的状态进行量化容许作出功能定义,那么我不明白为什么就不允许量化功能信念。

现在让我们回过头来简要地谈谈第3节的反“缺席感受性质”可能性的论证。在该论证中,我曾说过:疼痛的独特之处就是让它们内省式地认识到:自己有特殊的质的特征,进而引起“质的信念”,据此我论证说:任何状态如果在功能上同一于有质的特征的状态(如疼痛),那么其本身必定有质的特征。针对这一论证,有责难说:既然质的信念像质的状态一样不能从功能上加以定义,那么它们就不能合乎逻辑地进入对别的心理状态的“类型-同一条件”的功能说明之中。对此责难,我们现在可以给予答复了。不错,疼痛引起了那类(因此我承认)不能从功能上加以定义的质的信念,即关于某人有某种特定的质的状态的信念。但是它们也引起了关于某人处在疼痛中的信念——(正像“缺席感受性质论证”所明确假定的那样)如果疼痛必然是一种有质的特征的状态,那么相信某人处在疼痛中这一信念大概涉及到(至少就喜欢反省的人而言)这样的信念,即某人处在一种有某种或别的质的特征的状态中。当最后一信念是一质的信念时,它的命题内容从量上限定的就是质的

160 第二部分

状态,而不涉及到对特殊质的状态的指云。为什么不能认为这类质的信念可从功能上加以定义,是说不出任何理由的。而且如果它们能从功能上加以定义,那么没有任何理由说明:其他状态倾向于引起这种信念为什么就不应是构成那些别的状态的功能同一性的东西的组成部分。第3节的论证需要说的就这么多。

VII

在过去十几年中,心灵哲学中的大量争论都卷入两类似乎对立的直觉之间的斗争。一方面,有这样一种直觉,即心理状态不知什么原因在逻辑上或概念上与物理的事态、尤其是与被认为表现它们的行为联系在一起。这种直觉体现在一系列不同的哲学理论中——逻辑行为主义、由维特根斯坦所激发的“准则学的”观点以及最近的对心理状态的功能或因果分析(这些常常与某种形式的唯物主义或物理主义结合在一起)。^②另一方面,有这样一种直觉,它认为:心理状态与行为之间的联系从根本上说是偶然的;在对心理状态的大多数“内在的”描述之下,它们按照它们与行为、与别的物理状态的关系被陈述,这只是一个偶然的事实。这种观点的一种常见的表达式是这样的断言:光谱颠倒和其他类型的“感受性质颠倒”在逻辑上是可能的;因为说这些在逻辑上是可能的,显然就是说:这些心理状态有什么固有的、内在的特征,它们的“质的内容”在逻辑上与它们按照它们对它们的身体原因和行为表现的关系被陈述没有关系。我承认这种观点包含有基本的真理要素。因为我认为:光谱颠倒是可能的,而且还承认:这意味着至少有一些

质的状态(以及质的信念)不能从功能上加以定义。但我又相信,另一观点也包含有基本的真理要素。我认为,另一观点——即主张心理状态在“逻辑上”或“概念上”与行为有联系——有其最大的合理性的地方就在于,它适用于信念和愿望这样的状态。而且我认为,即使这些状态有时可能为质的状态所伴随,但从与我们这里有关的观点看,它们是没有“质的特征”的。不过正如我一直试图表明的那样:质的状态甚至可以在对心理状态的功能或因果分析的构架之内得到认可。在不能用一般的术语、或用不包含有质的状态的名称的一般术语去述说一个人处在特定的质的状态中是什么样子这个意义上,它们是“不可言喻的”,这可能是质的状态的本质,但是这并不妨碍我们对下述问题作出功能说明,即一种状态成为一种质的状态意味着什么,质的状态之同一性条件是什么。因此把这些关于心理状态和物理世界之间的偶然的或别的什么关系的、根深蒂固的、似乎有冲突的直觉调和起来大概是可能的。

在宣布这种尝试性的调解取得成功之前,还有很多问题必须加以研究。我所倡导的对质的差异性和相似性的说明是针对知觉经验的事例而提出的,因而必须考察的是,它能否合理地应用于疼痛这类感觉。它应用于疼痛实例时必需的东西是:把关于疼痛的观点当作肉体的感觉印象即躯体伤害之类的印象等(没有必要是诚实的)来对待。^②这种对质的相似性与差异性的说明也是针对下述事例而提出的,即在那里,被比较的经验就是同一个人的经验,因而必须考察的是:它是否能阐明或恰到好处地阐明主体间的经验的比较。这首先牵涉到这样的考察,即不同的人的经验在先前定义的意义上下是否有可能“是共意识的”;我认为这可归结为这样

162 第二部分

的问题,即是否存在着近来的人格同一性的讨论中所想象的那类人与人“合并”的可能性,也就是说,是否能把两个人结合成一个人(或意识的单个主体),他记住了两个人合并之前所有的经验,并能作出比较。(值得注意的是,如果合并是可能的,那么根本就不会有这样的情况,即任何可能的行为都不会暴露这样的信息,即两个人的颜色经验是否相同,或者说他们的颜色光谱相对于对方而言是不是颠倒的;因为如果两人合并了,那么新产生出来的人的行为大概能解决这个问题。)但这些都是复杂的争论,我不打算在此予以讨论。

注 释

① N.J. 布洛克与 J.A. 福多(1972):“心理状态不是什么”,载《哲学评论》,81, p. 165。

② 同上, p. 173。

③ 同上, pp. 173 - 174。

④ 同上, p. 172。

⑤ 同上, p. 173。

⑥ 同上, p. 172。值得注意的是,这个假定,或极相似的假定在克里普克近来反心理物理同一论的论证中起着关键性的作用;他是这样表述的:疼痛“并不是借它的偶然属性中的某一个而被分辨出来的;而是通过成为疼痛本身这样的属性、通过它的直接的现象学的质而被分辨出来的。因此疼痛……不能为‘疼痛’固定地指示,但是指示词的指称是由所指的一个基本属性决定的。”见“命名与必然性”,载 D. 戴维森和 H. 哈曼(编):《自然语言的语义学》,D. 里德尔出版公司,1972, p. 340。

⑦ 除了文中述及的那个方式之外,布洛克和福多还提到了另一种方式,功能主义者大概正是想以此对付颠倒感受性质论证;他可能会坚持说:“如果颠倒感受性质发生了,即使它们对他的理论提出了反例,但作为一个法则学的事实,功能上同一的心理状态也不可能在质上不同。”(p. 172)这里的思想

一定是:感受性质颠倒的单纯逻辑的或概念的可能性与功能主义不是不可调和的。

⑧ 布洛克和福多,同上,p.173。

⑨ 唐·洛克(1968):《自我与他人》,牛津,p.101。

⑩ A.多纳根:《维特根斯坦论感觉》,载G.皮彻(编):《维特根斯坦:哲学研究》,纽约,1966。

⑪ 布洛克与福多,引文同上。

⑫ 一种状态要有一致于该状态的质的特征,就必须有质的状态,这是什么关系?在与我们有关的事例中,它不可能是同一的关系(同一关系只允许质的状态有质的特征,而不允许我们说到下述状态的质的特征,它们的“类型-同一”条件是用功能术语所提供的)。据推测,这一定是比“被伴随”或“一同被例示”的关系更强的关系。我所能说的充其量是:状态S的特定标记有符合于质的状态Q的质的特征,如果在那种情况下S的标记(例示)一定与Q的标记(例示)有关的话。我不敢肯定这一点,但我能够予以强调,并用下述方式使之不那么含糊,比如说,在这种情况下,S的标记就是Q的标记。

⑬ “被显现”这一术语,我是从R.齐硕姆那里借用的;见他的“‘显现’、‘接受’与‘证明’”,载R.J.斯沃茨(编):《知觉、感觉与知识》,纽约,1965,尤其是p.480,脚注6。在诚实的知觉中,在幻觉和错觉中,人都可以“被显现”,在所有这类情况下,他能以相同的方式被显现。文中所用的“被显现为蓝色”这一技术措辞,是作为“看到或似乎看到某蓝色的东西”这一表达方式的缩略语而用的(根据对那表达方式的非认知性理解)。

⑭ 作为一种分析,这是远远不够的。我能看到某事物是蓝的,即使它看上去像绿的(也就是说,即使我的视觉的质的状态是与绿有联系的状态),如果我被“暗示”:在这些条件下,蓝色事物看上去像绿的。

⑮ 我从S.克里普克那里借用了“指称确定”和“固定指示词”两概念;参阅他的“命名与必然性”,pp.269-275及其他地方。用限定摹状词“x以致 Fx ”来“确定术语T的指称”,与把T定义为意义相同于限定摹状词的词,即定义为限定摹状词的缩略语,这两者形成了鲜明对比;在前一情况下,陈述“如果T存在,那么T是x以致 Fx ”将是偶然而非必然地为真。如果一种表达式在所有可能世界指示同一对象,那么它就是一个固定指示词。在克里普

164 第二部分

克看来,通名是固定指示词,而许多限定摹状词则不是。当一限定摹状词用来引出一名称(进而引出一固定指示词)时,它就是用来“确定指称”,而不是定义它或给予它以意义。

⑩ 我区分了两种意义,一是“被显现为蓝色”的“功能”意义,一是这样的(可能的)意义,在其之下,它固定地指示(或更准确地说固定地表示)一质的状态。我的这种区分相似于齐硕姆在“看上去像蓝色”这样的表达式的“比较”的意义与“非比较的”意义之间的区分。参看他的《知觉:哲学研究》,伊萨卡,1957,第4章。

⑪ 在写成本文之后,《形式与内容》(牛津,巴兹尔·布莱克韦尔,1973)一书引起了我的注意。在这本书中,B. 哈里森对我所谓的光谱颠倒的可能性提供了否定性的经验证据。可参阅。

⑫ 有时,也有人主张,如果某人报告经历过光谱颠倒,那么对我们作出结论最合理的东西就是某种东西随着他对颜色词汇的把握而消失了。

⑬ 例如参阅 C. 吉内特:“单词怎样意指各种感觉”,见《哲学评论》,77,1 (1968,1),p. 9。

⑭ 对此观点的进一步论证见我的“现象相似性”一文,载《批评家》,7,20 (1975,10),pp. 2-34。

⑮ 对功能定义是什么的说明以及接下来对它的阐释,是以 D. 刘易斯在“心理物理与理论的同一”一文中的说明为基础的,参看《澳大利亚哲学杂志》,50(1972,12),pp. 249-258。

⑯ 关于心灵的因果或功能理论的某些倡导者尤其是那些不接受第6节和脚注^⑩对功能主义的描述的人,不赞成归入这一类之中。但其他的人显然又把他们的说明当作是对行为主义和准则学观点中正确的东西的综合,或作为对使行为主义和准则学观点具有合理性的直觉的解释。例如可参阅 D. 刘易斯:同前,p. 257, D. 阿姆斯特朗:《唯物主义的心灵理论》,伦敦,1968,p. 92, A. 戈德曼:《人类行动理论》,英格伍德,克立弗,新泽西,1970,p. 112。

⑰ 这种观点事实上得到了 D. 阿姆斯特朗和 G. 皮彻的阐发。参阅阿姆斯特朗,同上,p. 313ff. 皮彻:“疼痛知觉”,《哲学评论》,79(1970),pp. 368-393。

感受性质、表征论与自然主义*

弗雷德·德雷特斯克

对于唯物主义者来说,不存在只能为一个人所理解的事实。可能有这样的时间与空间条件,它们能使一个人知道另一个人不知道的东西,还可能有使一个人成为权威的时间和空间条件,但是不存在个人所独有的事实,即只有一个人才能知道的事实。如果另一生物的主观生活即成为那生物可能是什么样子,是不可理解的,那么这必定是因为,我们在谈论它的主观状态时,我们未能理解我们所谈论的。如果 S 以某种方式感觉,而且 S 以某种方式感觉是 S 的一种物质状态,那么我们怎么不可能知道 S 怎样感觉呢?

表征论把经验的质——感受性质——与对象被表征_s 具有的属性等同起来。S 表征_s 事物所具有的属性在原则上是可为他人知道的。即使我们每一个人都有关于我们自己的经验(参看第 2 章)的直接信息,但是也不存在特许的通道。如果你知道在哪里去看,那么你也能够得到我所具有的关于我的经验的特征的同样的信息。这就是用自然主义术语对心灵进行思考的结果。主观性成了客观秩序的组成部分。对于唯物主义者来说,这是理所当然的。

* 译自 F. 德雷特斯克:《论心灵的自然化》(麻省理工学院出版社,1994,“让·尼科德讲座”讲稿)一书第 3 章,标题略有变动。

1. 法国的长卷毛狗和法国的酒

苏珊,一个视力和智力均正常的小孩,从未见过狗。她不知道狗是什么。她看到的第一只狗是一个法国长卷毛狗。对她来说,它看上去像长卷毛狗吗?苏珊不会说或不会认为它是(或看上去像是)一只长卷毛狗。她不会把它作为长卷毛狗看待,或不会认为它是一只长卷毛狗。这不是她对她所见的东西进行概念化的方式。尽管如此,对她来说,它看上去像长卷毛狗吗?

阿瑟是一年幼的、有正常的癞蛤蟆视力和智力的癞蛤蟆。他从未见过狗。他所看到的第一只狗碰巧也是长卷毛狗——苏珊看到的那一只。它看上去像长卷毛狗吗?同苏珊一样,阿瑟不会说或不会认为它是(或看上去像是)一只长卷毛狗。他不会把它当作或认为它是一只长卷毛狗。这不是他对他所见的东西进行概念化的方式。尽管如此,对他来说,它看上去像长卷毛狗吗?

基于我们关于人类和癞蛤蟆的视力所知道的东西,似乎很清楚,那只长卷毛狗在苏珊和阿瑟看来是不一样的。狗在苏珊看起来就像长卷毛狗在你与我看起来那样,这与叭儿狗、梗和牧羊犬看上去的样子是不一样的。你我都能描绘狗看上去的样子(“像一只长卷毛狗”),而苏珊不能,但这对她怎样看待狗肯定没有任何影响。阿瑟则不同。即使那只狗在阿瑟看起来,就像长卷毛狗在正常的癞蛤蟆看起来那样,而且阿瑟是一正常的癞蛤蟆,但是那只狗对阿瑟来说,看上去则不像长卷毛狗。尽管有这样的事实,即它们看上去的样子就像长卷毛狗在正常情况下、在半盲人看上去的样

子,但长卷毛狗对半盲的人来说看上去并不像长卷毛狗。长卷毛狗对半盲的人来说像是模糊的斑点——与看上去有这种大小的一切别的东西是一样的。癞蛤蟆有半盲人那样的视觉敏锐度。因此长卷毛狗对癞蛤蟆显现的样子就像它们对半盲人显现的样子。

在说狗对苏珊来说看上去像长卷毛狗时,我并不是说苏珊关于长卷毛狗的视觉经验把它表征为一只长卷毛狗,因此如果它并不真是长卷毛狗,那么在某方面便错误地表征了它。不,我这里讲的是关于长卷毛狗的感性表征而非概念表征。苏珊关于狗的经验把那只狗表征为一只长卷毛狗,其意思是说,它把狗表征为具有麦金(1982,p. 40,还可参阅米勒,1991,p. 42)所描述的长卷毛狗的表面属性,也就是使长卷毛狗看上去如此有别于别的狗(更不用说自行车之类)的那些属性。各种非长卷毛狗也有长卷毛狗的表面属性:长卷毛狗的逼真的画像,伪装的梗,长卷毛狗机器等。这些对象看上去也像长卷毛狗。它们也在苏珊身上产生关于长卷毛狗的感受性质。即使她会错误地相信(即在概念上表征)精巧的赝品是长卷毛狗,但当她用与在视觉上表征一真实的长卷毛狗相同的方式表征一假冒的长卷毛狗时,她是不会错误地表征(从感官上错误表征)它的。精巧的赝品据说引起了与原型同类的经验。

对没有关于长卷毛狗概念的人来说,把任何东西说成看上去像长卷毛狗,会使某些人觉得荒诞不经。例如,米勒(1991,p. 32)说:一个人可能具有与我看到南瓜时实际相同的南瓜经验,但如果那人没有关于南瓜的概念,不知道南瓜是什么,那么那人“似乎”不会认为那里有南瓜。我都不想搞语词争论。“看上去”、“显现”等词也许有这样的意义(尤其是在用了宾语补足语从句时),在此意

168 第二部分

义上,米勒是正确的。为了直截了当,我将称这些用法为“看上去”(“显现”“看来好像”等)的信仰(=信念)意义,并有根据地用下标予以标示(杰克逊[1977,p.30]称之为“看上去”的认识用法)。说一只狗对S来说看上去_d像一只长卷毛狗,就等于说:在没有起抵消作用的思想的情况下,这就是S所认识的狗,就是S关于狗的知觉(通常)促使S相信的东西(参看米勒,1991,p.19)。对S来说,把狗描述为看上去_d像长卷毛狗,意味着S有关于长卷毛狗的概念,理解了长卷毛狗是什么,能对她以这种方式看到的東西作出划分或分辨(或在没有起抵消作用的思想的情况下会如此去做)。在信念的意义上,苏珊所看到的狗对她来说看上去_d并不像长卷毛狗。那不是她关于狗的知觉引起她相信的东西。

即使那些词有另一种意义,在此意义下,如果那只狗对她来说与对我来说,看上去是相同的,而且在我看来,它像一只长卷毛狗,那么对苏珊来说,它看上去必定像长卷毛狗,不管她是否理解了长卷毛狗是什么,不管她是否有关于长卷毛狗的概念。按照一种悠久的传统,我将称之为“看上去”(look_p)的现象意义。对S而言,说狗看上去——从现象上——像长卷毛狗(=look_p),就等于说:(1)那只狗在S看来就像长卷毛狗在通常情况下由S所看上去的样子;(2)那只狗对S来说看上去不同于别的狗(叭儿狗、等)。^①第二个子句——被称之为区别对待子句——是防止让狗对癞蛤蟆(或半盲人)来说看上去像长卷毛狗所必需的,因为它们来说,所有的狗看上去都是一样的。如果你是色盲,如果你不能把一种颜色与另一种区别开来,那么一对象对你来说,看上去(look_p)就不像红的,这只是因为它看上去与你正常情况下看到的红的对象是

相同的。你并不满足那个分辨条件。在这里,分辨条件被故意弄得模糊不清了,以反映这些判断的情景敏感性和环境相对性(对你来说,K 在环境 C 而不是在 C' 中看上去像一只长卷毛狗)。即使存在着你不能把它与法国长卷毛狗区分开来的稀奇古怪的狗类,但狗“对你来说看上去($look_p$)可能像一只法国长卷毛狗”,只要被隐含的那一系列比较不包括那古怪的种类。这就是狗对某人为什么看上去($look_p$)像一只法国长卷毛狗的理由,即使某人不能把它与一非常精巧的赝品区分开来。假冒的长卷毛狗不在——在正常情况下,无论如何不在——那种类别比较中。^②因此尽管有区别对待的子句,但为了使一只狗对你来说看上去($look_p$)像一法国长卷毛狗,人们不一定非得要把法国的长卷毛狗与非常逼真的赝品区分开来。^③

应当注意到,这仅仅是因为,一对象对两个人看上去($look_p$)像 F,并不意味着它对他们看上去($look_p$)是相同的。原来, S_1 和 S_2 同样善于把 F_S 与别的对象区分开来,一特定的对象 K 对 S_1 看上去的样子就像 F_S 对 S_1 通常看上去的样子,它对 S_2 看上去的样子也像 F_S 通常对 S_2 看上去的样子。这并不意味着,那对象对 S_1 看上去的样子就像它对 S_2 看上去的样子。它也许是,也许不是。对苏珊和另一个人来说,一只狗看上去像($look_p$)长卷毛狗,这一事实并不意味着它对他们看上去是相同的。这些界定并不妨碍颠倒感受性质的可能性。关于这一点,等一会再详述。

有了这些术语上的约定,我们就可以说,那只狗对苏珊看上去像($look_p$)长卷毛狗,但对阿瑟则不然(对任何哪一个,它看上去 [$look_p$] 自然不像长卷毛狗)。癞蛤蟆的视觉系统注定不能用来分

170 第二部分

辨中等大小的对象,而人类的视觉系统则相反。其根据就在于,癞蛤蟆没有在知觉上分辨中等大小对象的形状的能力。长卷毛狗对癞蛤蟆看上去像什么部分取决于癞蛤蟆的分辨力。如果像我们所假定的那样(基于神经生理学和行为资料^④),正在运动的中等大小的动物(癞蛤蟆似乎不能看到静止的动物)对癞蛤蟆来说看上去是相同的,那么癞蛤蟆就不能看到正常的人所看到的差异,不能看到形式和细节上使长卷毛狗对我们来说看上去不同于叭儿狗的差异。癞蛤蟆的视觉系统不能像我们表征_s长卷毛狗那样表征_s它们。因此我推测,只有当我们在微弱的灯光下,且又没有戴眼镜去看长卷毛狗时,长卷毛狗对癞蛤蟆看上去($look_p$)的样子才像它们对我们来说看上去($look_p$)的样子——极像叭儿狗和狸看上去的样子。

在描述某些(别的)令人迷惑的情境时,我们也处在更加有利的地位。假设所有红葡萄酒在我尝起来都是一样的。这意味着我正在品尝的那种勃艮第红葡萄酒对我来说尝起来($tastes_p$)像勃艮第红葡萄酒吗?它尝起来的味道毕竟就像勃艮第红葡萄酒在正常情况对我所是的味道。不,如果我不能把它与西昂蒂葡萄酒区别开来,那么就不是这样。^⑤如果在我看来,勃艮第红葡萄酒尝起来与西昂蒂葡萄酒一样,那么勃艮第红葡萄酒和西昂蒂对我的品尝来说就都没有机会像勃艮第红葡萄酒。两者很有可能尝起来像普通的“大罐”红酒。也许不是。它取决于我能否把大罐“葡萄酒”与可口可乐区别开来。如果我(在味觉上)固着的是葡萄酒,就像癞蛤蟆在视觉上固着的是狗那样,那么勃艮第红葡萄酒和西昂蒂葡萄酒对我来说就是长卷毛狗、叭儿狗对于癞蛤蟆所是的那种东

西——模糊斑点的味觉上的等价物。

运用相同的标准来判断事物对一系统“看上去是”或“看起来像”怎样,这是不言而喻的,就像我们思考和谈论工具那样。如果一测量仪器不适于在 7.00 与 7.01 之间作出分别,如果小数点之后的数字对它是没有意义的数字,那么该仪器就不能把任何东西表征为 7.00。对于一个没有表示 6.99 与 7.01 之间的值的功能的内在状态的仪器来说,任何东西都不能“看起来”像 7.00。从根本上来说,这就是任何东西对阿瑟来说“看起来”都不像是长卷毛狗的理由。阿瑟的视觉系统没有必要的“辨析”力。在阿瑟身上,任何东西都没有这样的功能,即显示远处的对象有这个(长卷毛狗)形状而无那个(叭儿狗)形状的功能。而苏珊的视觉系统则是更灵敏的工具。^⑥她似乎经验到了 7.00 的状态,而阿瑟则没有。

不过,我们在利用分辨材料作出关于显现(*appearances_p*)——关于某人的经验的质——的推理时得小心谨慎。这种推理从关于动物能和不能以特定的感觉样式分辨什么的前提出发,过渡到这样的结论,即某物怎样以那种方式显现 *p* 给它们。这是靠不住的——至少是可疑的——从考察关于酒的例子就可看出这一点。即使所有的红葡萄酒在我尝起来都是一样的,但它们尝起来(*taste p*)也许像精美的勃艮第红葡萄酒,正如我假定的,而不像一般的红酒。果真如此,那么你我在分辨力上就是一样的(我们两人在品酒舞会上都将陷入尴尬),即使所有的酒在我们品尝起来是不一样的。所有的酒由我尝出的滋味就像精美的勃艮第红葡萄酒由行家尝出的滋味。所有的酒由你尝出的滋味就像廉价的西昂蒂葡萄酒由行家所尝出的滋味。这种味觉上的异常是很难(或不可能?)觉

172 第二部分

察的。每次,你在呷了一口昂贵的勃艮第红葡萄酒时便说:这酒喝起来像那(用手势指向廉价的西昂蒂葡萄酒),我赞成。我品尝它实际上也是这个样子。

如果酒的例子有点牵强,那么克拉克(1993, p. 167)则提供了更具挑战性的例子:

如果你闭上一只眼睛,注视某种鲜艳的颜色 30 秒钟,接着连续地眨眼睛,你就会发现大多数事物表面的色彩的更替。两只眼睛之间的适应差异迅速清除后,那种后果也会很快消失。不过当它们持续时,这些适应后果在许多方面与人中间的所谓差异是相似的。用适应了的眼睛所作的分辨将一致于用不适应的眼睛所作的分辨。任何两个对不适应的眼睛表现不同的色彩的项目将对适应的眼睛表现出不同的色彩。一致的项目还将继续一致。但是每一事物显而易见的色彩则是变化的。

克拉克得出结论说:这类人际间差异的可能性对于根据能被分辨的东西定义感受性质的方案似乎是致命的一击。休梅克(1975)提出了同样的看法。

再者,我们的分辨行为,可能是没有什么区别的,然而事物在我们尝起来的滋味则似乎是不同的——我们从感性上表征我们感知到的对象是不同的。这自然是颠倒光谱问题。^⑦该“问题”就是对于那些人例如行为主义者和功能主义者^⑧的问题,他们认为:经验的质无论如何必须用行为的或功能的术语去定义。

表征论是一种能避免这个难题的自然主义理论。^⑨它承认,知

觉经验的质的特征不可能从功能上去定义。不过,它可从物理上去定义。关于心灵的表征方案通过把感受性质与那种经验表征_S事物所具有的属性等同起来,做了两件事情:(1)它尊重人们(甚至功能主义者^⑩)广泛享有的下述直觉,即经验的质的方面是主观的或私人的:它们并不必然在它们存在于其中的系统的行为(或行为倾向)中表现自身;(2)它解释了使经验的质的方面可从客观上加以确定的感觉经验。经验的表征方案,在把感受性质与被经验的属性,被经验的属性与被表征_S的属性,被表征_S的属性与这样的属性即感官有提供关于它们的信息的自然功能的那些属性等同起来时,使感受性质像躯体器官的生物学功能一样可从客观上加以确定。要发现某种状态的功能_S是什么,这可能是很难的——有时(从实践的观点看)是不可能的,但不存在关于功能的、必然属于私人的或专属第一人称的东西。

一位名叫 E. 比撒克(1992)的神经心理学家对于发现研究感受性质的客观方法深感失望。他说,“关于意识的自然科学没有与感受性质打交道的任何办法”(p. 115)。在本章其余部分,我的目的就是表明这种悲观主义是可予避免的。对经验的表征说明不仅为感受性质腾出了地盘,而且为研究它们提供了客观方法。

2. 作为表征属性的感受性质

对感受性质,很难找到两个(更不用说所有)哲学家都接受的描述,但在开头似乎有足够的把握说,在感觉样式 M(对 S 而言)中的感受性质就是对象在 M 中从现象上显现给 S 或在 S 看起来

所是的方式。根据表征论,我继续把感受性质等同于现象属性——(根据该理论)对象在感觉中被表征(被表征_s)具有的那些属性。这意味着,关于另一个人的(或动物的)感受性质的问题就是关于人(或动物)的表征_s状态的问题,关于这些状态有指示什么属性的功能_s的问题。

正如我们在第2章第2节中所看到的那样,经验表征_s对象的方式就是这样的方式,即表征系统正常运作时,那对象所表现的方式,即它被设想起作用的方式,它被预想起作用的方式。在常规表征的情况下,我们关心的是设计者和构建者的目的与意向。如果你想知道:事物在响着的门铃“看来似乎是”怎样的,门铃通常是怎样表征门前的情况,你只须问:如果那装置以被设想的方式运作时,实际上会怎样呢?那将是事物在门铃“看来似乎是”怎样的。用这来检验,当门铃响了,那系统“似乎认为”,一定有某人站在门口。不管门口有人还是没有人,这就是铃响所意指的东西。它并不意味着,议案收集者在门口,即使事实上门口的人就是他。假如这个系统在运作,但在最理想的条件下,门铃也不能在议案收集者与造访的亲戚之间作出区分。这就是我们为什么说:门铃意味着某人在门口——不是议案收集者,不是造访的亲戚,而是某人。听到铃声的人大概会把它表征为(或相信是)议案收集者,但那不是门铃所意指的,不是它所说的或所表征的。表征系统的状态意味着它有指示任何东西的功能。这就是为什么门铃对谁在门口不会“撒谎”的理由。但它们绝不能说出那是谁。

阿瑟的视觉系统不能把某物表征为长卷毛狗,同理,门铃系统也不能把门口的某人表征为议案收集者。门铃似乎不可能“认识

到”议案收集者站在门口,其理由同阿瑟不可能认识到附近好像有一只长卷毛狗一样。这不是——我要强调的是不——因为阿瑟不能把长卷毛狗与叭儿狗区分开来。我们可以想象,苏珊的视力如此糟糕,以致她也不能把一个与另一个区分开来。不过,即使苏珊由于受伤而不能把长卷毛狗与叭儿狗区分开来,但我们仍能设想她有关于长卷毛狗的感受性质。我们能设想她梦见过长卷毛狗。^⑩我们能设想她在幻觉中看到了长卷毛狗。我们甚至设想,苏珊具有如此被损伤的视力,以致不能把长卷毛狗与叭儿狗区分开来,当她看狗时,则生起了关于长卷毛狗的幻觉——于是当她看长卷毛狗时,便有关于长卷毛狗的感受性质。这正是酒的例子所要说明的东西。我们能够设想,完全没有鉴别酒的能力的人在饮酒时,他或她经验到了一种精美的勃艮第红葡萄酒的滋味。另一个人体验的则总是廉价的西昂蒂酒的滋味。不,苏珊和阿瑟之间的关键差异不是他们的分辨力方面的差异。我们能设想他们的分辨力是一样的,不过他们有相当不同的经验。关键的差别也不在于他们的视觉系统提供了什么样的信息(这可能相同),而在于他们的视觉系统具有提供什么样的信息的功能。^⑪苏珊处在这样的知觉状态中,以至于不管它们能不能执行它们的功能,不管它们是否能使她把长卷毛狗与叭儿狗区别开来,但她都有把中等大小对象(如长卷毛狗和叭儿狗)相互区别开来的功能(即功能_S)。阿瑟没有进入这样的状态。这就是他为什么不能把长卷毛狗经验为长卷毛狗的原因,就像标准论题中的门铃不能把议案收集者表征为议案收集者一样。

试想有两个工具,J与K。J是一种用来测量每小时百英里速

176 第二部分

度的精密的装置。当它记下 78.00 时,这意味着(正如小数点后的数字所提示的)不是 77.99(及以下),也不是 78.01(及以上)。J 有分辨速度的“鉴别力”。而工具 K 是不太贵重的装置,旨在用来提供关于速度的大概的信息。它所记录的 78 意旨的不是 77(及以下),也不是 79(及以上)。K 不存在这样的状态(像 J 所具有的),即它有(根据设计)指示 77.99 与 78.01 之间的速度的作用。既然状态没有这个功能,K 的状态就不能把速度表征为 78.00。K 对速度的“鉴别力”没有 J 那么强。J 是速度的鉴别专家。它能把 77.92 与 77.98 的速度区别开来。而 K“感觉到”所有这些速度都是一样的。J 有 K 从未“经验过”的速度“感受性质”。

J 是苏珊的工具上的类似物,K 是阿瑟的类似物。在把速度分别表征为每小时 78.00 和 78 英里时,J 和 K 反映的是相同的客观条件,正如苏珊与阿瑟对同一长卷毛狗作出的反应一样。但苏珊是以一种更精确的分辨方式表征该对象的。形态上的微小变化对苏珊都有影响。正像在速度从 78.00 变为 78.05 时,J 的记录也发生变化(从“78.00”到“78.05”)那样,如果对象的形态从长卷毛狗变为叭儿狗,那么苏珊的表征“指针”也将移动。正像 J 的构造对速度上的那些差异很敏感一样,而且原来设计就是如此,苏珊的视觉系统对形态上的那些差异也很敏感,而且(我们假定)原来注定就是如此。这不适合于 K 和阿瑟。它们不会记录这些差异,它们内部在设计时就没有东西能记录这些差异。阿瑟把长卷毛狗表征为一个斑点,就像他把叭儿狗表征为一个斑点——同样,K 把速度 78.00 表征为 78,就像它把 78.05 表征为 78 一样。

我们可以想象,两个工具在分辨行为及能力方面没有区别,而

它们“经验到了不同的感受性质”(如不同地表征速度)。假设计速器 J 由于损伤或磨损而丧失了敏感性。它对速度就像 K 一样变得不敏感了。受损伤前,在 J“看来好像”(即被表征为)不同的速度,现在则似乎是相同的。在 J 的信息分辨系统出现损伤之后,两个工具对速度提供的是相同的信息。当速度在 77.84 与 78.23 之间变化时,J 的指针(正像 K 的指针)没有移动(也许有助于把这些作为计数装置)。当速度从 77 跳到 78 时,J(像 K)以大约 77.50 的数字迟钝地反映从 77.00 到 78.00 之间的变化。两个工具因此在功能上变得没有区别了。尽管有这种等同,J 上所记录的 78.00 意指的是不同于 K 上记录的“78”的某东西。J 不再传递作为其功能应提供的信息这一事实并不意味着它丧失了提供信息的功能。只要它保留了这个功能,J 上所记录的“78.00”意指的就是不同于 K 上的“78.00”的某东西,即使两种反应携带的是同样的信息。在受损伤(受损、老化及别的什么)之后,速度在 J“看来”仍像每小时 78.00 英里。这就是 78.00 的读数所具有(和保留)的功能要指示的东西。J 不能再(真实地)“知觉”每小时 78.00 英里的速度,但仍能(幻想或梦见)到这个速度。而 K 则不能。K 从未经历过每小时 78.00 英里这样的感受性质,现在依然如故。他不能梦见到每小时 78.00 英里的速度。K 中不存在具有这种表征内容的状态。K 与 J 表征上的差异不在于它们做了什么,甚至不在于能做什么,而在于它们的不同的状态有做什么的功能。

J 与 K 之间的差异是,J 有、而 K 没有注定要指示速度 78.00 这样的内在状态,这是一种具有这种功能的内在状态。因此 J 能把那速度表征为每小时 78.00 英里。而 K 不能。当情况异常时,

178 第二部分

两装置的内在状态指示同一的东西,这一事实并不意味着它们意指同一的东西。损伤改变了一状态所携带的(它与之关联的)信息,但没有改变它意指的是什么(它有与之关联的功能)。基于此,我们便能料想:苏珊尽管有视力障碍,把每只狗看成了长卷毛狗,但她仍不是阿瑟。她有意指这的感觉状态,即使这些状态不再携带这种信息。

因此,我赞成休梅克的观点(1991, p. 508),他与 N. 布洛克、J. 福多(1972)一样认为:感受性质是不能从功能上加以定义的。但这并不意味着,感受性质不能由目前这种表征说明来把握。因为两种表征装置在分辨力和能力方面可以是等值的(因而在功能上是等值的),然而又可处在不同的表征状态中。于是经验可以是不同的,即使这种差异不可能再在分辨行为中“表现”自身。即使这意味着感受性质不能从功能上加以定义,但不意味着它们不能从物理上加以定义。只要存在着一种用物理术语作出的、关于系统在其之下具有携带信息功能的状况的描述,那么它们就可从物理上加以定义。只要我们有关于指示功能的自然主义说明,我们就有关于表征进而有关于感受性质的自然主义理论。

3. 观点

T. 内格尔(1974)说:“每种主观现象必然与一独特的观点相联系,进而不可避免的似乎是,客观的、物理的理论将放弃那种观点。”对主观现象的表征说明构成了一种关于心灵的客观的、自然主义理论,这种理论是不会放弃主观的观点的。它不放弃观点,而

且还会对之作出解释。

如果我们比较两个表征了同一可测定因素(速度、质量、温度等)集合的表征系统,那么它们的观点就由它们所表征的对象以及它们表征这些对象所具有的那些可测定属性的限定性形式确定了。 S_1 和 S_2 都表征了 O , 但 S_1 把它表征为蓝色, 而 S_2 把它表征为青色(深蓝色)。或者两者都把某物表征为青色, 但是它们结果又把不同的对象——或同一对象的不同部分——表征为青色。无论如何, 它们有一种不同的观点, 即由表征事实所决定的观点。

奥赫尔(芝加哥)和拉瓜第亚(纽约)的雷达表征了它们附近的飞机的位置和运动。这些工具有不同的观点。它们表征世界上的不同对象。人们不能仅仅借助观看雷达屏幕而判断它们有不同的观点。因为那屏幕看上去毕竟是相同的。它们的观点上的差异不是由它们怎样表征它们的天空部分而决定的, 而是由它们表征哪一部分天空所决定的。给予两个雷达系统以关于世界的不同图景的东西是它们与世界发生联系的不同方式(我先前用“C”标示的关系), 亦即是使它们对世界的不同部分敏感进而能够予以表征的联系。不过, 观点上的这种差异不是关于那些系统的一个表征的(因而就自然系统而言不是心理的)事实。经验之所以在主观上不同, 正是因为它们是关于不同对象的经验, 进而构成了不同的观点。

如果我们假设任何对象在世界上必然占有不同的位置——进而让表征不同对象的任意两个系统有关于世界的不同观点(即关于不同位置的观点)——那么由不同对象正被表征这样的事实所引起的“观点”上的差异就是比罗(1991, 1992)所谓的固定观点

180 第二部分

——能够通过相互改变位置而交换的观点——中的那种差异。奥赫尔的雷达能搬到拉瓜第亚——因而有同拉瓜第亚的雷达一样的(固定的)观点。比罗指出:固定的观点没有任何心理的东西——说“可能”(在内格尔的意义上)享有或改变固定的观点,这是胡说。不同的观点在经验上可能是相同的。奥赫尔和拉瓜第亚的屏幕即使有不同的观点,但事实上仍可能处于相同的表征状态。经验状态中的差异不是观点上的差异的结果,而是根源于观点由以被观察的方式上的差异。如果我们看到了不同的对象,但它们是孪生兄弟,我们便拥有不同的(固定的)观点,但我们的经验是相同的。我们能交换我们的固定的观点,而在我们的有意识生活中不起任何波浪。在经验的质中,使“有一种经验可能是什么”表现出差异的东西不是那种经验所关于的对象——因此不是我们的固定的观点——而是这些对象从那种观点出发得到表征的方式。

甚至是当多种 F- 表征系统在表征同一对象时,其实也可以说它们能以不同程度的放大或超出不同的范围表征它。因此它们能包含关于这对象——当我们用无遮蔽的眼睛和通过放大镜观看一小点时我们所经验到的那同一种差异——的不同的“知觉”。在第 2 节中,我们已明白了一仪器怎么可能提供关于同一量值的“近似”观点。一测速器把一速度表征为每小时 78.00 英里,而另一个把它表征为每小时 78 英里。第一个有“近似”的观点,它能随着速度的微小变化而变化。对于一灵敏的仪器来说,78.05 的速度“看起来不同”于 78.00 的速度。对于一个不太灵敏的仪器来说,这些速度“看来好像”是一样的。因此这些仪器即使真实地“知觉到了”(表征着)同一个对象,但它们是不同地加以看待的。一个仪器有

更好的视觉,或者也可以说它“更接近”。它有关于对象的“更好的观点”。观点上的这种区别有助于定义具有那观点可能是什么样子,但是在表征习语的范围内就可以轻而易举地抓住这些区别。

4. 经验到电子场可能是什么样子?

有一源于F. 杰克逊(1986)的论证,这就是,如果你没有亲自经验过颜色(色彩),那么某一方面的实际知识是不可能为你所得到的。你不知道经验红、蓝和别的颜色可能是什么样子。该论证很容易加以泛化。没有嗅觉的人怎么可能知道丁香花或烤焦的面包闻起来有什么气味呢?如果你没有某些经验,那么似乎就有你不知道的事实。但是如果是这样,那么某种知识,即关于这些经验可能是什么的知识就只能暴露给对它们有经验的人。进而,这就似乎导致了这样的结果,好像存在着主观性的方面——经验F可能是什么样子——这是不可能得到客观的裁决的。如果你为了知道经验F可能是什么样子,而不得不经验F,如果有理由认为别的人——肯定是别的动物——经验了你没有经验过的事情,那么就有理由认为,关于别人的经验,存在着不能为你所知道的事实。

他心知问题的这个翻版使许多人深信不疑。它还使人想到:对心灵——至少是留给主观经验的那部分心灵——的客观研究即对心灵科学的期盼是一个虚幻的目的。我们怎么可能知道周围的人的经验可能是什么样子呢——且不说异邦人、动物的经验——除非我们能进入他们的头脑,经验他们正经验的东西?

表征论暴露了这类推论的错误之所在。知道蝙蝠、鱼和周围

182 第二部分

的人的经验是什么,从根本上说不同于知道事物对测量工具“似乎是”怎样的。在两种情况下,它是一个决定一系统怎样表征世界的问题。即使这很困难——有时从实践的观点看是不可能的——但并不意味着,进入另一存在的头脑内部在概念上是不可能的。

我借用杰克逊的主人公玛丽来说明这一点。玛丽是一位精于电磁现象的专家。她精通磁场、电子学、麦克斯韦方程、量子电子动力学等。但是她从未经验过电子场。她曾听说鲨鱼和角鲨有电磁感觉,这使它们能通过感觉地球电子场中的畸变(常常是由被捕食者引起的)而分辨出被捕食者。她惊奇不解的是:作为一个角鲨,经验电子场会是什么样子呢?

玛丽没有必要成为一头角鲨。在想知道成为一头角鲨可能是什么样子时,玛丽想知道的不过是:当角鲨借助它们的电磁感受器分辨被捕食者时,事物对角鲨似乎是怎样的。什么样的感受性质是由分辨磁场中的变形的感觉系统产生的?

既然角鲨是相当复杂的动物,因此如果我们从更简单的东西——只有单一表征的寄生虫——开始,那将是有裨益的。过一会我再来讨论角鲨。^⑩试想我们的寄生虫依附于寄主,当且仅当那寄主处在一定的温度:18℃中。既然寄主的温度对那寄生虫的存在是至关重要的,那么它便进化出了一种敏锐的热感官。它能在它与之有联系的一个亲等的躯体的部分之内记录下温度。^⑪寄生虫把一潜在的寄主感觉为18℃——理想的温度——并依附于它。寄主不一定非要是18℃。使其依附于寄主的必不可少的东西是那寄生虫把它感觉为、表征为18℃。问题是;当寄生虫感觉到接纳的寄主时,成为这个寄生虫可能是什么样子呢?

如果你知道达到 18°C 意味着什么,那么你就知道寄主在寄生虫“感觉起来”是怎样的。你知道:当寄生虫“感觉”到寄主时,它的经验像什么。如果知道成为这样的寄生虫像什么就等于知道事物对它来说似乎是怎样的,它怎样表征它所知觉的对象,那么为了知道成为它可能是什么,你就不必要成为一个寄生虫。你必须知道的是:温度是什么。如果你知道了,以致足以知道处在 18°C 的温度中是什么样子,那么关于寄生虫的经验^{的质}你就知道了应知道的一切。要知道经验的质对这寄生虫看上去像什么,应注意的不是寄生虫,而是那寄生虫正“注视”的东西——寄主。因为如果情况正常,那么寄主所是的东西—— 18°C ——就是事物在那寄生虫看来似乎是怎样的。因此如果你想知道事物在那寄生虫看来似乎是怎样的,关注那寄主就行了。

这在某些人看来肯定是荒诞不经的。确切知道温度是什么并不能告诉人们:寄生虫(甚或另一人类)感觉到这种温度会是什么样子(如果它像什么东西的话)。聋子可能知道声波是什么,但不知道听到声波可能是什么样子。我无意于否定这一点。事实上我也没有否定。在本节其余部分尤其是下一节,我将更充分地说明我为什么没有否定这一点。而此刻,我请读者宽恕。我只概述一下关于下述几乎为人们普遍承认的事实——完全是本讲座中所辩护的表征观点之外的事实——的结论。第一个事实是:感受性质被假定为事物在这里所讨论的感觉形式中看上去或显现出的方式。因此例如,如果一西红柿对 S 来说,看上去是红的、圆的,那么红和圆就是 S 对西红柿的视觉经验的感受性质。如果是这样,那么(第二个事实):倘若事物是它们看来所是的方式,那么由此可以

184 第二部分

说,感受性质,即定义有那经验可能是什么的属性,实际上就是当知觉是真实的时候被知觉的对象所具有的属性。在寄生虫的例子中,既然对象(寄主)在知觉真实时所具有的属性是 18°C 的属性,那么它就必然是寄生虫关于寄主的经验的感受性质,不管那寄生虫是否真的知觉到寄主(不管寄主是否实在地具有这种属性)。因此任何知道 18°C 是什么的存在,都知道这属性是什么,知道寄生虫的经验有什么感受性质。就这单一的感受性质来说,他们知道成为那个寄生虫可能是什么样子。如果这个结论是荒谬的,那么导致这一结论——不是表征论——的两个事实中的一个就应予以摘。

再来看角鲨。玛丽对电子场和电子场中的无规律性无所不知。关于电子场,她知道应该知道的一切,就像你对以每小时 4 英里的速度的运动知道了应知道的一切一样(没有更多的)。如果角鲨的电磁感觉正常地起作用,那么它表征的就是电子场中的模式。该场在这里是正常的,它在那里则“收缩”了,它“膨胀”在这里与那里之间。因此在其几何学是可描述的场中存在着一种无规律性。我们可以描述场的构型。我刚才就是如此做的。玛丽既然知道电子场的一切,知道鱼、岩石和植物怎样使之改变,那么她就能勾画出场的实际的图景。她关于电子场所勾画的(所描述、表征或知道的)就是那条鱼对电子场所感觉到的,正是在此场中,它发现了自己。在勾画、描述、表征和知道场的几何学时,玛丽勾画、描述、表征和知道了:成为一个(真实地)感觉那种类型的场的角鲨可能是什么样子。正像你——知道有四条边可能是什么——知道(从现象上)看四条边可能是什么一样,玛丽在知道场的形态是什么时,

便知道关于具有这种形态的场的经验像什么。^⑤在知道关于这类形态的场的经验像什么时,就此经验的这个单一的维度来说,她知道有这类经验可能是什么样子。她知道角鲨的这种经验的感受性质。关于那种形态的电子场在角鲨看来似乎是怎样的,再没有什么要知道了。根据关于经验的任何理论——确切地说,根据表征理论——在经验蓝时,人的经验不可能有更多的质要知道,就像没有更多的蓝颜色要知道一样,因为蓝颜色就是人们所经验的颜色。正是那种质把人的蓝色经验与其他颜色经验区别开来。同样,对 T 类电子场,也没有什么更多的要经验,就像对成为 T 类电子场没有什么要经验一样,因为 T 实际上就是使这电子场经验不同于其他的电子场经验的东西。T 就是这经验的感受性质。如果玛丽知道 T 类电子场是什么,那么关于这类经验的质,她便知道了应该知道的一切。

知道了经验的质是什么,在你独自经验它时,并不必然就能认识到某物具有那种质。我有足够的音乐理论,以致知道音调的变化是什么。如果我知道你听到了一种音调的变化,那么我实际上也知道你在这方面的经验像什么。你的经验有那种音调的变化的质,而且我知道此质是什么。不过,尽管知道关于你的经验的这个事实,但当我独自听它时,我本人并没有技术或能力认识音调的变化。正如刘易斯(1983)、尼米罗(1980)和其他人所指出的那样,存在着一种感觉知道“像什么”的能力,它是我听音调的变化像什么(在事实的意义上)时用不着具有的。如果我知道音调的变化是什么,那么在事实的意义上,我就知道听音调的变化像什么。它像“那”,在这里,“那”指的是我知道的、音调的变化所是的东西。这

186 第二部分

并不是说,我通过听就能分辨某物就是那。那就是这样的能力,即当你的经验有音调变化的质时,你知道你的经验有什么质用不着这种能力。这对玛丽也是适用的。知道角鲨的经验的质并不要求玛丽有这样的能力,即能通过任何感觉形式分辨出角鲨借它的电子感觉所分辨的东西。

这就是说,角鲨并未把那种电子场表征为一种电子场——更不用说具有如此这般形式的电子场,理解这一点极为重要。它是一个有这种状态的电子场,这一事实并不是被表征的东西。那条鱼关于电子场所表征的就是它的构型,它的几何学,它的状态。关于它是有那种状态的别的什么的理论构架,又是另一个问题。如果我们假设,气体的温度是气体的分子的能量,那么温度计在把一种气体的温度表征为 78°C 时,它表征的实则是那些分子有 78°C 的能量。但在把那些分子表征为有 78°C 的能量时,温度计并没有把那气体表征为具有分子构成的东西——更不用说有具有 78°C 能量的分子构成。温度计把分子构成的能量表征为 78°C (不同于 77°C 或 79°C),而不是表征为具有分子构成的能量。为了吃食,角鲨不得不知道电子场的构型。它们没有必要知道它是有这种构型的电子场。这对玛丽也是适用的。为了知道那条鱼怎样表征那个场,她必须知道它们把它表征成什么状态。但她没有必要知道把某物表征为电子场可能会像什么。她这样做,并不是因为角鲨没有把电子场表征_s为电子场。它们为什么是这样呢?

因此,在对角鲨的重复分辨测验中,玛丽要确定的是:角鲨对哪一种电子场变形敏感,对哪一种“看不见”。玛丽由此实实在在地弄清了(角鲨)感觉到电子场可能是什么样子。在用图表示角鲨对电

子场的分辨力时(假设这些鱼的电子感觉具有向它们传递关于这种场的状态的信息的功能),玛丽要确定的是鱼的知觉经验的特征,即它们的经验怎样表征电子环境。一当知道这一点——而且没有什么阻止玛丽知道关于它的一切——那么关于角鲨对电子场的了解,就再没有什么要知道了。^⑩

有一件事是角鲨能做而玛丽不能做的。假设一条鱼经验到一种电子场模式 P。如果这些鱼能思考,那么这条鱼就能想到这(指被经验的电子场)有模式 P,而玛丽则不能。由于她不知道任何电子场,因此她对能用这种形式表现出来的东西一无所知。对玛丽来说,(关于这)作为具有 P 模式的电子场,没有玛丽所知晓的对象。知道了她关于电子场所知的东西,也许就实在地知道了那条鱼经验到了什么,知道了那条鱼用指示代词“这”指的是什么,以及鱼经验到它所具有的是什么状态。尽管如此,玛丽本人并不能像鱼经验它那样描述她所知道的、鱼经验了什么。但是可以肯定,这并不能说明角鲨经验到了玛丽所不知道的某东西。你我都知道我的钥匙在桌子上,但只有我才能用这样的词语表述我们所知道的:“我的钥匙在桌子上。”你不得不用不同的词语指称你所知道的、在桌子上的东西(那是你知道我所知道的东西)——例如,“你的钥匙在桌子上。”我们不能以相同的方式指称我们所知道的、在桌子上的东西,这一事实并不意味着存在着我们所知道的、不同的东西。玛丽不具有与电子场的这种关系,有了这种关系,她才能像角鲨经验它那样(像一个这)指称它,这一事实也不表明,关于角鲨的经验有某种不为玛丽所知的东西。

当我经验到蓝色时,我所经验的东西是我能用下述语言描述

188 第二部分

的某东西,即“这是蓝颜色”。你可能知道我正经验到蓝色——因而实际上知道我所知的——不过,因为你处在不同的位置,因而不能用我的表述方式表达你所知的。根本不存在这样的东西,你把它说成是这,而它正好是我碰巧经验到的颜色的质。这并不表明:你不知道我的经验的质是什么。玛丽没有经验到电子场这一事实并不表明:她实际上不可能知道角鲨的经验像什么。^⑩

在整个讨论中,我一直假定,关于被表征的可测定的东西并不存在什么特殊的难题,而只有这样的问题,即什么决定了那系统在特定时间所表征的那个可测定的东西的形式。这一假定隐匿了一个重要难题。该难题在颜色(以及其他的所谓第二性质)的知觉实例中已被证明是特别令人困惑的。我关于红的经验是一种什么属性的经验(表征)呢?不错,是红,但这是什么样的客观属性呢?如果颜色感受性质同一于颜色经验所表征 s 的属性,而且该经验表征 s 这样的客观属性,即它们有功能 s 予以指示的属性,那么颜色经验有指示什么样的客观属性的功能 s ? 它是(网膜上)入射的光的波长吗?它是整个光学配置的整体属性吗(参看兰德,1977)?它是把光反射到眼睛上的对象的表面属性(反射)吗?就颜色来说,在发现视觉系统对之作出反应的可能的客观属性时的困难使许多哲学家变成了关于颜色的主观主义(哈丁[1986]就是新近的一个)。主观主义者认为,颜色不是客观的属性——从而不是视觉的功能能传递其信息的客观属性。即便在理想的条件下,我们的颜色经验也没有客观条件向我们传达关于它们的信息。正是基于种种条件,不同波长的光才唤起相同的颜色经验。因此颜色经验并没有把那种波长“告诉”我们。根本就不存在这样的特殊的光谱

反射,其表面非得要被看作是绿色的。由于照明,许多不同的表面性质就可能引起相同的颜色经验。这一普遍现象——即这样的事实,各种各样的(事实上是无限的)客观环境可以引起相同的颜色经验——就是所谓的条件配色(metameric matching)。如果色觉的功能就是指示颜色,颜色与对象的某种客观属性(如表面的光谱反射)是同一的,那么要么人的色觉干了名副其实的坏事,要么颜色是客观条件即所有引起相同颜色经验的东西的复杂析取。如果是这样,我们便面临着这样的问题,即一系统(如视觉)为什么被选择来指示远侧状况的这种高析取(完全是杂色)集合的出现?事情的真相难道不是这样吗,即在一个类别中,集合了我们称之为红色的各种形形色色的对象的唯一的東西就是它们都引起同一种颜色经验——红的经验这一事实?但是如果是这样,那么我们就不能按照一种经验有功能 s 予以指示(进而表征 s)的某种客观属性定义红这种感受性质。因为我们是通过这些客观状况所产生的主观感受性质(红的经验)说明这样的客观状况(表面光谱反射的高析取集合),即那经验有功能 s 予以指示的状况。

这是一个复杂的问题,我不可能在几页的篇幅里妥善地予以解决。^⑩但是与常规系统的再一次类比可能有助于对该问题作出正确回答。在前面(第1章,§3),我描述过一种仪器——计速器——它用关于车轴旋转的速率的信息提供关于速度的信息。制造这样一种装置比制造自动地补偿(通过顾及路面上的车轴高度)车胎大小的装置要便宜得多。既然造它比较便宜,且每一正常地装有该仪器的汽车都装配了标准大小的车胎,因而就没有必要增加额外的花费了。设计者之所以挑选安装在汽车中的廉价的装置,

190 第二部分

是因为它能出色地履行自己的职责——足以证明它优于那具有更多用途、更可靠但在根本上更浪费的装置。基于任何合理的选择程序,这一装置(而不是更复杂的那种)就被选中了。假如它被选中了,那么这装置的表征状态(指针位置)因而就有指示速度而不是车轴旋转的整体功能,即使该系统用了关于车轴旋转所做功的信息。

一当我们拥有装配了这种仪器的媒介物,我们就可以把条件配色与速度进行类比。能引起每小时 78 英里的表征的速度多种多样(事实上无限):对正常车胎来说,每小时的速度为 78 英里,对小车胎来说,每小时的速度为 62 英里,大车胎为每小时 93 英里等等。如果人们开始用不同大小的车胎,那么这个装置就不再能做属于其功能应做的事——指示速度。它之所以被选择去完成某一工作,是因为它原先做过那事情——指示速度——在某些已被说明过的条件下(正常大小的车胎)。在另外的条件下,它不能完成那项任务这一事实并不意味着它没有指示速度的那种功能 s 。它意味的不过是:人们要确定一系统的功能 s 是什么,有时不得不确定该系统在其之下被选择做那事的条件。因为它被选择做什么——因此它有做什么的功能——进而它表征什么——这都取决于选择发生的时刻所存在的(可能的)特定的条件。

就工具而言,我们(设计者和使用者)决定了工具的行为所依赖的量中的哪一些是它所表征的量。我们授予该工具以功能。我们说:它就是计速器。我们可以通过让它完成那项工作而使同样的装置成为车轴的每分钟转速计量器。如果我们让它干那项工作,那么车胎的大小就与它的功能的完成没有关系,与它的表征的

准确性也没有关系。但是我们不能那样做。我们首先感兴趣的是车辆的速度,而不是车轴的旋转(除了作为速度的指示器)。因此那工具就被确定做一项工作——指示汽车之速度——这使它成了环境的“人质”。除非这些环境符合要求,否则那系统就不能履行其职责,即使它尽其所能,在这个意义上,工具就是这些环境的“人质”。通过给予我们的车轴每分钟转速指示器以指示车辆每小时车速的功能,我们就提高了它不因功能故障而出差错的方式之数量。我们增加了马瑟恩(1988)所说的“正常的错误表征”的数量。我们扩大了条件配色的数量和种类——能产生相同表征状态的不同客观条件的数量和种类。从而我们提高了一系统在理想地运转时可能出故障的方式的数量。通过向外扩展——到更“远”的条件——到那系统有能力指示的东西,我们就使它越来越难借助仔细观察那仪器来分辨它的表征状态是什么。因为它的状态有能力指示的东西愈益不取决于该系统在正常地发挥作用时所传递的信息,而更多地受制于那功能在特殊环境下的表现。

关于实际历史条件在评估系统的功能以及(根据目前对表征的阐释)它的表征成果中的意义,许多作者(哈特菲尔德,1991,1992;谢泼德,1992)已提出了相同的看法。哈特菲尔德(1992)指出:如果色觉的功能就是提高绿色植物与泥土、岩石等的可分辨性,“……如果存在着物理上可能的、而非实在的(非植物的)对于那系统不能分辨的绿色植物的条件配色,那么它就没有任何用处。”对功能的判断依赖于一系统被设计要做的事情,而且你并不能通过考察一系统在现在条件下实际的所作所为,一如既往地分辨一系统注定要做什么——它有做什么的功能——因此它表征什

192 第二部分

么,即使你知道它现在所做的就是它被设想要做的。此外,它还依赖于在相似于下述条件的条件下它是否做了那件事,在此条件下,它被选择要做那件事。当上面描述的计速器错误地把一(具有略小的轮胎的)汽车的速度表征为每小时跑 62 英里时,在一定意义上,它实际上做了被设计要做的事,被设想应做的事。它正是利用关于车轴旋转的信息,依据固定的规则系统,去分辨那汽车跑得多快。这里的麻烦不在于那工具,而在于它应用于其中的生态学。这些不是该工具被选择在其中起作用的条件。¹⁹它用来计算速度的规则系统不再适用了——不具有它被选择的那种可靠性。仅仅因为它正常工作时,它实际上以同一种方式对无限多样的各别速度作出反应,就推断它没有指示一种简单的客观属性(速度)的功能,这是十分错误的。事实的确如此,但这并不表明:它没有以独一无二的方式表征无限多样的条件。

这种看法同样适用于色觉。如果我沿着这一思路进到它的逻辑结论,那么我们就到达了希尔伯特(1992, p. 36)的观点:颜色就是色觉功能所觉察到一切属性。如此众多的条件引起我们经验红色这一事实并不表明:当我们经验红色时,我们所经验到的东西不是客观的属性。它表明的不过是:它所是的那种属性也许不再明显有别于引起我们对它的经验的各种条件。

5. 成为一个经验到电子场的人 可能是什么样子?

但是这真的意味着我们能客观地判定成为一角鲨可能是什么

样子吗？或者说成为一只蝙蝠可能是什么样子吗？事情真的是如此简单吗？如果一蝙蝠的声纳系统把蛾表征为存在于那里，而且我又正好理解存在于哪里（以及存在什么）——由此实际上知道蝙蝠似乎看到的蛾在哪里——那么我真的理解成为蝙蝠可能是什么样子吗？

事情自然并非如此简单，那是因为我们一直在考察：做角鲨和蝙蝠所做的一件事情——感觉电子场的几何形状或蛾的空间位置——可能是什么样子，而不是考察成为一角鲨或蝙蝠可能是什么。但是这肯定不是角鲨和蝙蝠经验的全部。一当我们通过表征它们的环境知道了它们所干的一切时，我们对它们怎样经验这个世界才知道了应知的一切。因此我们不会知道成为角鲨或蝙蝠可能是什么样子。

换句话说，在知道感觉 F(S 所做的)是什么样子与知道成为 S (感觉 F 的东西)是什么样子之间存在着差异。因为要知道成为 S 是什么样子，我们就必须知道(关于感觉的感受性质)S 所感觉的一切属性。在实际上知道角鲨感知电子场变形时所经验到的构型的质时，玛丽对于那条鱼的现象世界还存在什么，它经验到了它的环境或电子场(除了其几何形状)的其他什么属性，可能仍然一无所知。那条鱼不仅感觉到电子场的空间结构，而且还感觉到了其剧烈的变化吗？它还对诱发的磁场敏感吗？对相位关系怎样？别的人基于他们知道了观看对象时关于其位置、大小、形状、方位应知道的一切——即知道了我们观看对象时所经验到(以及他们知道我们过去所经验到)的感受性质，便断言他们知道了观看对象可能是什么样子，这是错误的。如果他们不理解颜色，以及不懂得

194 第二部分

我们在观看对象时也表征了他们的反射属性(如果那肯定是颜色所是的东西),那么我们关于对象的经验有一个方面就是他们没有理解的。他们没有理解:观看对象可能是什么样子。

盲人也许知道:借视觉去经验运动可能是什么样子。如果他知道运动是什么,那就够了。关于运动的经验——不管是视觉的,还是触觉或运动感觉的——都有其质的特征,这特征可根据它是关于什么的经验(表征)而加以定义,如果这些经验都有相同的属性,那么它们在主观上对于这一单一的属性来说就都是同一类的经验。^④不过盲人在知道从视觉上经验运动是什么样子时,他并没有必要知道从视觉上去经验一正在运动的对象会是什么样子。因为看到一对象在运动所涉及到的比经验到一对象的运动所涉及到的更复杂——更更复杂。人们还能经验到对象的形状、大小、颜色、运动方向以及其他大量的属性。这就是观看运动和感知运动为什么如此不同的原因,即使同一事物(运动)以两种方式被表征。甚至当感觉在它们的表征活动中重迭时——如它们在感觉空间属性时那样——它们也还是表征到了不同范围的可判定的属性(麦金,1991,p.35)。

在描述蝙蝠的声纳知觉的假设的胶卷(视觉表征)时,埃金斯(1993,p.264)提出了相近的观点——在蝙蝠扑向山洞时,当我们观察他(借助其声纳装置)所感觉到的东西时,我们并没有发觉与这些运动相应的额外的“惹人喜爱的感觉”。因此在只知道他的声纳系统怎样表征他的环境(胶卷可能向我们显现的)时,我们并没有必要知道成为一只蝙蝠可能是什么样子。需要知道的是:蝙蝠怎样表征他的环境和他自己的内部状态。我不知道,你观看愤怒

的狮子向前冲时,成为你可能是什样子,那恰恰是因为我知道愤怒的狮子看上去像什么。我还得知道:感觉到肾上腺素激增可能是什样子。

注 释

① 这个定义并不是循环定义。原始术语是看上去相同和看上去不同。看上去($look_p$)像 F 被定义为看上去相同于 Fs 在正常情况下看上去的样子(不同于一系列的非 Fs)。

② 这不是说它们从来不是。人们能想象着设法把法国长卷毛狗与仿制品区分开来。在这个意义上,说一只狗看上去像一法国长卷毛狗隐含的意思是:人们能把它与仿制品区分开来。

③ 我认为,我在显现 p 和显现_d之间所作的区分,接近于米勒(1991)在 F-型经验(现象的)与 F 存在于那里的表象(信念上的)之间所作的区分。也相似于皮科克 1983 年的区分。丘奇兰德(1979, p. 14)在主观与客观的意向性之间的区分也是如此。

④ 参阅劳埃德 1989, 第 5 章,该章简述了从莱特文等(1959)到尤尔特(1987)的有关材料。

⑤ 注意,如果我很糊涂,那么我就可能被促使相信:这是一种勃艮第红葡萄酒。那酒可能是,即是在我看来,尝起来_d像勃艮第红葡萄酒。

⑥ 即使撇开马上要述及的那些问题,但在把知觉分辨力当作感觉系统表征了什么属性的标准时仍存在着技术上的困难。克拉克(1993, p. 59)指出:两种刺激的不可分辨性并不足以表明:它们授予了观察者以相同的感受性质。如果有某种 Q_i ,它与 Q_1 是不可区分的,但与 Q_2 则相反,那么尽管 Q_1 与 Q_2 是不可区分的,但它们一定是不同地“被给予的”(即被表征的)。因此,一系统不能从位置上区分两种状态还不足以表明:该系统把两种状态表征为相同的。必不可少的东西是克拉克所谓的“整体”不可区分性。这里,我将略去这些复杂的细节。它们并不影响我要阐述的观点。

⑦ 我对“颠倒光谱问题”略微作了这样的新的篡改,以避免对标准阐释的责难。例如,哈丁(1986)和范·久里克(1993, pp. 144 - 145)坚决主张:哲学

196 第二部分

家们所想象的大多数颠倒瓦解了具有高级的、有结构的质的空间的组织,从而是有害的后果的(即造成了分辨力上的差异)。就我的想象来说,几乎没有什么质的空间可被瓦解。

⑧ 雷伊(1992, pp. 59-61)和利康(1987, p. 297)论证说:颠倒感受性质问题之出现于行为主义者和功能主义者面前是有根本的差异的。

⑨ 表征理论的这个优点得到了哈曼(1990)、泰伊(1991, 1992, 1994)、戴维斯(1991)、利康(1987, 1990)和麦金(1991)的赞赏。即使我把所有这些哲学家归结为“表征主义者”(休梅克(1990b)称他们为“意向主义者”),但许多人不赞成我关于表征主义的自然主义观点,几乎没有人同意说:所有可由内省把握的质是意向的(即被表征的属性)。参看休梅克 1994, p. 22。

⑩ 例如,休梅克(1975, 1991)。但参阅利康(1987, p. 60):“相对于输入-输出关系的被颠倒光谱的可能性是有牢固基础的,因而值得重视,即使我假设了可宣布无效的模态直觉。相对于输入-输出关系的颠倒光谱可能性,以及倡导者觉得其合理的、以无论多么低的抽象水平表现的内在的功能组织,完全是晦暗不明的,并有悖于某些直觉上合理的随附性观点。”在第5章中,我再讨论随附性问题。

⑪ 达马苏等人(1993, p. 57)指出:大脑某些部位的损伤,不仅会使颜色知觉丧失(色盲),而且会使想象颜色的能力消失。大脑其他部位的损伤,以及视觉加工的早期损伤则不会导致这类问题。即使一个人不能看颜色,但还能想象它们。他们断言,在某种意义上,颜色的概念(我认为是表征颜色的能力)依赖于大脑的这个区域。我设想苏珊的大脑的这一部分就应做手术。受损伤的是负责把关于颜色的信息传递到大脑的这一部分的视觉系统中的某个部分。

⑫ 值得记住的是:一系统可能表征(即有提供有关信息的功能)原本对那系统没有任何特别“利益”的属性。

⑬ 既然这些就是例证,因此这里我并不担心角鲨和寄生虫是否有经验。我早就说过(第1章, § 3),一种表征 s 要成为一种经验,它必须服务于这样的表征 a (概念)系统,即能经过学习而校准它的指示器状态的信息功能的系统。我不担心这一条件在角鲨和寄生虫的情况下能否得到满足。

⑭ 根据我所选择的例子,值得注意的是:蚊子(格里尔 1984, p. 447)据报

告对 0.02°C 这么低的温度差异敏感,而鱼只对 0.2°C 这样低的温度差异敏感。而且某些动物对这些量有“绝对的倾斜度”。啮齿动物、蜜蜂和鱼经过训练可以选择一种特定的温度,它不依赖于与先前温度有关的相对差异。

⑮ 这与知道看一正方形可能是什么样子不可同日而语。在看一正方形时,四边之外的许多属性也被表征了,而人们并不可能知道这些属性。在 § 5 中,我再回头来讨论这一点。

⑯ 克拉克(1993, p. 206)描述了一个类似的分辨力测试模型,(就蝙蝠而言)把它作为对成为一蝙蝠可能是什么样子这一问题的(“关于类别的”)回答。他说,我们得到了关于成为一只蝙蝠可能是什么样子的“看起来”“客观的”描述。克拉克与我唯一不同的地方在于:关于表征的目的论观点使修饰词(“关于类别”和“什么看起来是”)成为不必要的。

⑰ 在回应杰克逊的论证时,许多哲学家已提出了这一观点,即不存在玛丽不知道(关于其他造物的经验)的事实,除非事实如此严格地被个体化了,以至于我关于这是绿的(我所看到的唯一的事情)的知识与你关于我所看到的是绿的知识被当作是关于不同事实的知识。请参阅戴维斯等 1993, pp. 16-17;埃文斯 1984;霍根 1984;洛尔 1990;利康 1990;休梅克 1991, p. 508;泰伊,1986;以及范·久里克 1993, p. 142。

⑱ 参阅汤普森等,1992,其中有对流行的观点的有价值的考察。

⑲ 试比较 C. 麦金(1989, pp. 77-78)关于模式认知装置的例子,该装置始终如一地“错误地解释”新环境中的刺激。

⑳ 有价值的是,梅尔佐夫等(1979)证明了不满一月的婴儿身上的多感觉系统的等值性。他们让一个月的婴儿熟悉放在他们嘴里的两个橡皮奶头中的一个。其中一个有平滑的乳头罩,而另一个粗糙,其上有隆起物。熟悉之后,向婴儿显示两个橡皮奶头的视觉模拟物,他们对他们先前用口所知觉的那一个表现出稳固的视觉优先性。

参考文献

C. 麦金(1982):《心灵的特征》,牛津大学出版社。

——(1991):《意识问题》,牛津:布莱克韦尔。

A. 米勒(1991):《理性与经验》,牛津:克拉伦登出版社。

198 第二部分

F. 杰克逊(1977):《知觉》,剑桥大学出版社。

A. 克拉克(1993):《感觉的质》,牛津:克拉伦登出版社。

S. 休梅克(1975):《功能主义与感受性质》,载《哲学研究》,27:292-315。

——(1991):“感受性质与意识”,载《心灵》,100,no 4,507-524。

——(1994):“自我知识与内感觉”,《心理学与现象学研究》,52,2:249-314。

E. 比撒克(1992):“理解意识”,载米尔纳等编:《意识的神经心理学》,伦敦:学术出版社1992。

N. 布洛克与 J. 福多(1972):“心理状态不是什么”,《哲学评论》,81:159-181。

T. 内格尔(1974):“成为一只蝙蝠可能是什么样子?”《哲学评论》,83,4:435-450。

J. 比罗(1991):“意识与主观性”,载戴维斯等编:《意识》,牛津:布莱克韦尔,1993。

E. 兰德(1977):“色觉的视网膜理论”,《科学美国人》,237,6:108-128。

G. 哈特菲尔德(1991):“知觉与认知中的表征”,载 W. 拉姆齐等编:《哲学与联结主义理论》,希尔斯代尔,新泽西:劳伦斯·埃劳恩。

——(1992):“颜色知觉与中枢编码”,载 M. 福布斯等编:《PSA》,1992两卷本。

R. 谢泼德(1992a):“论颜色的物理基础、语言表征和有意识经验”,载 G. 哈曼编:《心灵的概念》。

——(1992b):“彩色空间的结构是由世界上的什么决定的?”《行为与脑科学》,15,1:50-51。

K. 埃金(1993):“无质的蝙蝠”,载戴维斯等编:《意识》,1993。

C. 皮科克(1983):《感觉与内容》,牛津:克拉伦登出版社。

P.M. 丘奇兰德(1979):《科学实在论与心灵的弹性》,剑桥大学出版社。

D. 劳埃德(1989):《简单的心灵》,麻省理工学院出版社。

J. 莱特文(1959):“蜜蜂的眼睛把什么告诉给它的大脑”,《无线电工程协会论文汇编》,47:1940-1951。

- J. 尤尔特(1987):“放松机制的神经行为学”,《行为与脑科学》,10:337-368。
- C. 哈丁(1988):《对哲学家而言的颜色》,印第安纳波利斯。
- 范·久里克(1993):“理解现象性的心灵”,载戴维斯等编:《意识》。
- G. 雷伊(1992):“被调换的感觉句子”,《哲学研究》,68:289-319。
- W. 利康(1987):《意识》,麻省理工学院出版社。
- (1990):“什么是心理的主观性”,载 J. 汤柏林:《哲学展望》,之 4,“行动理论与心灵哲学”,1990。
- G. 哈曼(1990):“经验的内在的质”,《哲学展望》,1990。
- M. 泰伊(1991):《想象的争论》,麻省理工学院出版社。
- (1992):“视觉感受性质与视觉内容”,载 T. 克兰编:《经验的内容》,剑桥大学出版社 1992, pp. 158-176。
- (1994):“感受性质、内容与颠倒光谱”,《心灵》。
- M. 戴维斯(1991):“个体主义与知觉内容”,《心灵》,100,4:461-484。
- A. 达马苏(1993):“大脑与语言”,载《心灵与大脑》,纽约。
- J. 格里尔(1984):《动物行为的生物学》,St. Louis, MO: Times Mirror/Mosby。

第三部分

随 附 性

随附性的种种概念*

金 在 权

1. 引言

我们认为,环绕我们的世界并非孤立的物体、事件和事实的简单集合,而是组成为一个系统,即显示出结构的某种东西,其组成要素以多种多样的方式相互联系着。这种世界观对我们关于事物的图式来说似乎是极为重要的;它反映在下述常识性的假设之中:在一个地方发生的事物能够影响另一个地方发生的事物,在某种程度上,这使我们能够根据一个事物而说明另一个事物,通过关于一个事物的信息而推知关于另一事物的信息,或者通过影响一个事物来影响另一事物。这一关于事物相互联系的观点的核心就是依赖性(或其逆概念:决定性)的观念:即事物之所以相互联系,是因为一个事物是否存在,或它具有何种属性,依赖于或决定于其他某种事物的存在以及这些事物是何种类型。正是由于这些依赖或决定的关系,世界才能为人所理解;而且我们还可以借助对它们的

* 译自《哲学和现象学研究》,1984年12月,第155卷,第2期。其雏形是呈交给1983年在马萨诸塞大学和史密斯学院举办的“赫伯特·海德尔伯格纪念研讨会”的一篇文章。

204 第三部分

利用来干预事件的进程,并改变它们以符合于我们的愿望。解释、预测和控制之类的活动对于一个缺乏这些联系的世界是没有意义的。关于存在着“真实的联系”的观点与关于世界可理解并可控制的观点有理由认为是等价的观点。

因果关系是我正在谈论的决定或依赖关系的一个极好的例子;撇开那些有逻辑上的根据的东西比如衍推不说,它是唯一被明确认识和广泛讨论的那种关系。原因决定了它们的结果,结果在其存在和属性上依赖于它们的原因。休谟将因果关系称作“宇宙的黏合剂”是不无道理的;因果关系是宇宙的胶水,它将孤立的物体和事件结合起来,使它们彼此具有意义——即使在休谟的原子论世界里也是如此——从而有助于为预测和控制自然现象提供一个必要的根据。以为世界的因果关系网络支撑和支持着我们关于它的知识中再现出来的解释性以及其它认识关系网络,这种观点与我们大多数人所赞成的广义的实在论世界观是不谋而合的。

部分—整体的关系也是重要的;然而,其重要性在很大程度上似乎来自于这一信念,即一个整体的一些至关重要的方面,包括其存在与本质,依赖于其各个部分的那些方面。即是说,部分学的关系之所以重要,是因为部分学的决定性或“部分学的随附性”^①是或被认为是一个普遍的事实。

近来人们对随附性概念的兴趣与日俱增,这主要是因为它可应用于心—身问题、微观还原和物理主义之上。把随附性归入具有重要哲学意义的、包括因果关系在内的那种关系,是大有裨益的,因为它们描述了物体、属性、事实、事件等由以进入相互依赖关系的方式,以至于产生了一个赋予世界以及我们关于它的经验以

结构的相互关联的系统。依赖性或决定性的样式在许多方面可能相互有别;如果随附性被看作是一种样式,那么实际上就会产生这样一些问题:它究竟如何区别于其他样式;它是一个单纯同质的关系还是实际上代表着两个或多个可辨别的关系;随附的决定性是否在哲学上提出了一种取代其他决定关系的重要的方案。

随附性的思想似乎起源于道德哲学。在下面的众所周知的段落中,G.E. 摩尔描述了后来被称作“随附性”的道德与非道德属性之间的一个特定的依赖关系。

……如果一特定的事物在某种程度上具有任何类型的内在价值,那么该事物不仅在所有情况下,在同等程度上一定具有这种价值,而且任何实际上相似于它的事物在所有情况下,在极为类似的程度上也一定具有这种价值。^②

然而,摩尔本人似乎没有使用过“随附性”这个术语;我认为,正是 R.M. 黑尔本人,及他多年后的作品才使它像现在一样在哲学中流行开来。下面出自黑尔的引文,现在一般认为是形成这一概念的雏形的经典来源:

首先,让我们以“善良”的特征为例,这特征被称之为善良的随附性。假使我们说,“圣·弗朗西斯是一个善良的人。”那么,在逻辑上就不可能既主张这一观点,同时又坚持认为,可能还存在着另一个人,他处在和圣·弗朗西斯完全相同的环境中,而且以完全一样的方式行动,但仅仅在下述方面与圣·弗朗西斯不同,即他不是一个人。^③

在这里,黑尔把随附性说成是“善良”这个术语的“特征”。然而,如

206 第三部分

果把它解释成一种关系,即“善良”和表示行为方式及品质特征之类的东西的术语之间的关系,那显然更有用些。黑尔的意思是说,对于两个人来说,他们在后面这些方面完全相同,而在“是一个善良的人”方面却不相同,这是“逻辑上不可能的”。同样明显的是,最好不要把随附性看作是各个单个考虑的属性或术语之间的关系,而应该看作是它们的集合或族之间的关系。这样,我们就可以说,所有价值评价的属性(即所有价值评价的属性集合)随附于所有自然的或描述的属性集合。在本文中,我们将主要讨论属性的而非谓词的随附性;这样做在一定程度上是无关紧要的,但是我们将看到,并非完全如此。一个人也可以在句子、事实、事件、命题和语言方面谈论随附性;而我在下面要论证的是,事实的随附性可以根据属性的随附性来理解。我认为,下述说法似乎是合理的,即属性的随附性是基础性的,其他大部分实在的随附性能够根据它来解释。

正是这种超出伦理学领域的显而易见的泛化才使得随附性成为一个值得密切关注的、富有吸引力且充满希望的概念。或许因为这一点,现在才可看到“随附性”这个术语在各种领域中日益广泛的运用,这也说明其中存在着共同的、基本的直观内容。如此说来,一件艺术作品的美学属性已被认为随附于其物理属性。^④有些哲学家已发现,心理物理随附性有望取代还原论者的物理主义;有人认为,随附性的论点承认了物理的首要地位,而又没有把关于物理的还原主义的更强的主张强加给我们。^⑤那种认为价值评价的术语通常随附于非价值评价术语的观点已经扩展到了认知术语,即用来作出认知评价的术语诸如“明显的”、“确实的”、“可证明

的”。认知证实的标准必须以非认知的术语来陈述,这一观点可以看作是对下述论点的一种表述,即认知属性随附于非认知的特征与关系(比如,因果属性与逻辑关系)。莱布尼兹关于关系判断可有可无的晦涩学说也许可以解释为这样的论点,即关系随附于属性。奎因关于翻译的不确定性的观点一般被解释为对意义随附于全部物理事实的主张的否定。^⑥部分学的随附性也已得到了论述。还有其他一些有趣的问题,我们也可能根据随附性来予以阐释:因果关系是否随附于非因果事实的特殊事态?规律是否随附于它们的例示?理论是否随附于论据?那种相信在一个给定的领域内部,比如在相对于物理的心理领域,存在着一种随附性关系的信念往往会形成这样一个具有重大意义的、含蓄的前提,它促进关于那一领域的特殊理论的形成,并决定其发展方向。接受还是拒绝心理对物理的随附性会导致有关心-身关系的理论之间最基本的分野:接受心理物理随附性的理论基本上是唯物主义的,而拒绝它的那些理论则基本上是反唯物主义的。这一区别在哲学上似乎比通常将心-身理论划分为“一元论者”或“二元论者”的分类更基本并更有意义。

本文打算对作为依赖或决定关系的随附性作一般性的讨论。我的主张是,存在着两个可区分的随附性概念,其中一个强于另一个,而且通常在哲学讨论中提及的是两个中较弱的一个,而所需要的则是较强的一个。我还要论证的是,较强的关系等价于“全体随附性”,这是为某些作者所喜好的一个替换概念。重点将是这样一个争论,即两个领域间的什么样的随附性可衍推出它们之间的类别-类别联系的存在,对于作为这两个领域之间的可定义性与可

还原性这样的关系来说,这意味着什么。

2. 弱随附性

前面出自黑尔的引文提出了关于随附性关系的这一最初概念:道德的东西随附于自然的东西,意思是说,如果两个对象(人、行为、事态等)在所有自然方面相同,那么它们必然在所有道德方面相同。这就是说,事物不可能在某些道德特征方面不同,除非它们在某些自然方面存在着不同。差不多同样的观点也出现在 D. 戴维森关于心理物理随附性的阐述中:

尽管我所述的观点否认存在心理物理学规律,但它与下述观点是一致的,即心理特征在某种程度上依赖于或随附于物理的特征。这种随附性可能意味着,不可能有在所有物理方面相似但在某些心理方面却不同的两个事件,或者说没有某些物理方面的变化,一个对象不可能在某些心理方面发生变化。^⑦

这里,戴维森对随附性作出了两种解释,第一是说明事件,而第二是说明对象。我并不关心这样的事实,即一个是关于事件的,而另一个是关于对象的;我只是对这两种解释的一般形式感兴趣,而且我想指出的是这样的事实,即第一个符合黑尔的陈述中所指出的模式:心理特征随附于物理特征,因为没有两个事物(物体、事件等)能够在某些心理特征方面不同,除非它们在某些物理特征方面也不相同——就是说物理上的一致可衍推出心理上的一致。如果

我们制作一个关于你的精确的物理复制品,那么它和你应该在心理上是不可分辨的。(戴维森的第二个解释,正如我后面要提及的,表示的是一种较强的随附性关系。)

下面将直接展开对包含着这些观念的随附性的一般性分析。假设 A 和 B 为两个非空属性族(为了简明,我们将关系排除在外),并从属于通常的布尔属性-形成运算、互补、合取以及析取(或许还有其他的,如无限合取与无限析取)。那么,这就是“弱随附性”的一个定义(称之为“弱”的理由将在下面说明):

A 弱随附于 B , 当且仅当必然地:对于任意的 x 和 y , 如果 x 和 y 共有 B 中的所有属性,那么 x 和 y 共有 A 中的所有属性——那就是说,关于 B 的不可分辨性可衍推出关于 A 的不可分辨性。

我们将称 A 为随附族而 B 为随附的基础(族); A 中的属性是随附的属性,而 B 中的则是基础属性。

例如,试考察集合 A , 它包含这样的属性,即“是一个善良的人”(G),并具有布尔闭合属性;假设 B 是包含“是勇敢的”(C)、“是诚实的”(V)、“是仁慈的”(H) 的属性的集合,并从属于布尔运算。 A 除了一个同义反复的($G \vee \neg G$)和一个不可能的($G \wedge \neg G$)以外,只包含两个属性, G 和 $\neg G$ 。假设 A 弱随附于 B 。这意味着,如果两个人共有 B 中的相同属性,比如说,都是诚实的和仁慈的但缺乏勇敢(这将确保他们共有 B 中的所有其他属性),那么他们必然都是善良的人或者都不是善良的人(他们在同义反复的或不可能的属性方面自然不可能有差异)。或者同样地,如果一个人是

210 第三部分

善良的而另一个不是,那么 B 中必然有某一属性,在这方面,他们相互有别(比如,第一个是勇敢的而第二个相反)。A 中的任何差异一定得由 B 中的某种差异来说明。

为了进一步明确这一点,试考察我们称之为“B-极大属性”的概念:它们是 B 中可构想的、最强的、一致的属性,并且对我们当前的例子来说,有 8 种这样的属性: $C \wedge V \wedge H$, $C \wedge V \wedge -H$, $C \wedge -V \wedge H$, \dots , $-C \wedge -V \wedge -H$ 。这些属性是相互排斥的,每个对象必然只具有其中的一种。显然,只有当两个对象具有相同的 B-极大属性,它们在 B 中才是不可分辨的。A 对 B 的弱随附性因此可归结为:任何具有相同的 B-极大属性的两个对象必然有 A 中的相同的属性——它们都是 G,或都是 -G。或用“可能世界”的术语,我们可以说:没有这样的可能世界,在它里面,两个对象具有相同的 B-极大属性,但在 G 方面却不同。

假设 A 弱随附于 B,那么,在每一个可能世界里,下述形式的泛化将是成立的:

$$(1) (\forall x)[(B_i(x) \rightarrow G^\star(x))],$$

在这里,对每一个 i 来说, B_i 是一个 B-极大属性,而 G^\star 要么是 G 要么是 -G。无论 G 还是 -G 都应与一个给定的 B-极大属性相联系,这是特定的可能世界的一个特征;但是在每一个世界里,一定有下列毫无例外的普遍条件式,即存在于“是一个善良的人”的属性与勇敢、仁慈以及诚实的美德之间的条件式。实际上,在每一个世界里,下面的双条件式是成立的:

$$(2) (\forall x)[B^\star(x) \leftrightarrow G(x)]$$

$$(\forall x)[B^\#(x) \leftrightarrow G(x)]$$

这里, B^\star 或 $B^\#$ 每一个都是 B -极大属性的一个析取。

一般说来, 当 B 有限时, 所有这些观点对于弱随附性来说都是有效的; 如果 B 不是有限的, 这些结果将取决于 B -极大属性的可成形性, 而这些可成形性又离不开无限的合取以及它们的无限析取。稍后我将证明这些运算对于属性(区别于谓词的)来说是可以接受的, 并且它们在下面将以某些形式论证的形式表现出来(其实, 在它们得到运用的地方就清楚了)。

我详述这些细节的目的是要证明: 尽管“弱随附性”的定义非常接近于文献中对随附性的标准解释, 出自黑尔和戴维森的引文就是很好的证明, 但它所定义的关系比人们所期望的要弱得多——实际上, 对于它的一些典型的预期的应用来说太弱了。理解这一点的关键在于, 上面公式(1)的泛化对于基础族中的每一极大属性来说关联于一种随附属性, 在这种形式的泛化中, G^\star 是 G 还是 $\neg G$, 取决于所考察的特定世界, 而且这并不是跨可能世界的一个不变的特征。这意味着, A 对 B 的弱随附性(就我们的例子来说)有这样一些可能性:

(a) 在这个世界, 任何勇敢、仁慈并且诚实的人都是一个善良的人, 但在另一可能世界, 这样的人不是善良的; 实际上, 在这另一世界, 每一个这样的人都是邪恶的。

(b) 其次, 在这个世界, 任何具有勇敢、仁慈和诚实特征的人都是善良的; 在美德的分布方面与这个世界极为类似的另一世界里, 没有人是善良的。

(c) 就人有或没有这些品质特征而言, 另一可能世界正好相似于这个世界, 在那里, 每一个人都是善良的。

212 第三部分

很显然,弱随附性容许这些可能性,因为它只是要求:在任何可能世界内部,不存在这样的两个事物,它们在 B 方面一致,但在 A 方面却不同;并且这个条件适合于这些例子中的每一种情况。它并不要求:如果一个对象在另一世界拥有与它在这个世界所拥有的相同的 B-属性,那么它必然也拥有与它在这个世界所拥有的相同的 A-属性。将一个给定世界里的 A-属性与 B-属性之间的特定联系转移到其他世界是不可靠的。

因此,弱随附性不符合下面的条件:确定了一个对象的基础属性就确定了它的随附的属性。这个条件说出了随附性说明中假设为必需的东西:基础属性一定决定了随附的属性,在这种意义上,一旦前者对于一对象而言得到了确定,那么它的后者就没有变动的自由了。弱随附性对这种决定观很有帮助:如果你以相同的方式确定一个给定世界里的两个对象的基础属性,那么你必须以相同的方式确定它们在那一世界里的随附的属性。但是在弱随附性下,基础属性甚至限制着随附的属性的归属。不管用什么方法都可以看到,这还不是我们对决定或依赖关系所预期的东西。决定或依赖自然被认为携带着一定的模态力量:如果“是一个善良的人”依赖或决定于某种品质特征,那么具备这些特征一定确保或保证“是一个善良的人”(或者说缺乏其中的某些特征一定决定了某人不是一个善良的人)。这些特征与“是一个善良的人”之间的联系必定不只是随着世界变化而变化的一个事实上的巧合。我们应该可以说:尽管查尔斯现在不是一个善良的人,但只要他是诚实的和勇敢的并且在他的本性中具有某种仁慈,那么他就会是一个善良的人。我们还应该可以说:任何具备这三种美德的人都将是一

个善良的人,尽管不幸的是无人具备所有这些。这些断言似乎就是我们说到“使之成为善的特征”时所固有的意思:任何“使之成为X的特征”一定是这样的,即任何事物如果具有它,那么它必然具有X(至少,它必然正相关于它具有X)。道德属性之于非道德属性的弱随附性并不能衍推出:道德属性有非道德的“条件”或“标准”。

与这种随附性观点常常联系在一起的另一观点是:如果道德的东西随附于非道德的东西,那么任何两个在所有非道德方面极为类似的世界必定在所有道德方面类似(实际上,它们必定是同一个世界)。但正如我们已知的那样,这在弱随附性之下是不存在的。关于心理物理随附性可得出类似的结论:心理物理弱随附性与这样一个世界的存在是一致的,该世界在每一个物理细节方面都与现实世界非常类似,但被证明无智能、无意识,并且除了一种低级的痛苦渗透于每个角落的每个对象之外,其余方面完全相同于我们的世界。因此如果我们像戴维森那样,指望随附性成为一种心理对于物理的依赖关系,那么我们很可能会发现,它并不存在于弱随附性里面。

我觉得,摩尔的下述评论既富有启发意义又令人惊叹不已:

我从未想到要提出这样的主张:善良是“非自然的”,除非我已经假定,它在下述意义上是“衍生的”,即无论何时一个对象是善良的(在这里所说的意义上),其善良(用布罗德先生的话说)“依赖于某些非伦理特征的存在”,这些特征为刚才所说的事物所具有:我一直在下述意义上假定它是如此的“依赖”,即如果一个对象是善良的(按我的理解),那么它之所以如此

214 第三部分

是出自于这样一个事实：它具有某些自然的内在的属性，这些属性是这样的，从它是善良的这个事实不能反过来得出它具有那些属性这样的结论。

为了弄清楚其力量超出了弱随附性的范围，我们用不着确切地知道摩尔在这里所用的术语“出自”或“依赖”的含义。因为弱随附性，如我们所看到的，只要求任何具有相同自然属性的两个事物必定都是善良的，或者都不是善良的。这确实不足以说，一事物之“是善良的”“出自”于它正表现它所拥有的自然属性；因此，弱随附性不能阐明摩尔心里所想到的“依赖”这一概念。

那么，弱随附性还有有价值的哲学用途吗？尽管它明显地不够强，因而不能看作是对依赖或决定的全部关系的分析，但我仍然相信：它表明了一个有趣而且有意义的部分的依赖或决定关系。还是来看道德随附性的例子：也许黑尔所要的就是弱随附性。^⑧在弱随附性下，某人褒扬一个对象（说它是善良的），但没有褒扬在所有描述性细节方面是或被认为是与它完全相似的另一个对象，这里可能有一种不一致性；然而，不褒扬任何一个，而同时承认两个对象有相同的描述性属性，则不存在前后矛盾或不一致。从根本上说，这就是伦理学背景下“病症相似则治疗相似”的处方。所以，弱随附性为我们提供了受到大量讨论的伦理判断的普适性原则，它常被理解为一致性要件。^⑨但是，伦理判断的普适性已在一种更强的意义上被理解：每一单个的伦理判断一定能得到一个完全的普遍适用的原则的支持。这一较强的必要条件超出了弱随附性的范围，而更吻合于将在下一节解释的“强随附性”概念。普适性原则的这两个变体与本文所区分的随附性的两个概念原来正好相

符,这充分说明两个概念的自然本性及其哲学意义。关于普适性的两个原则的这些看法显然适合于其他情况,包括价值评价判断(比如,在美学和认识论中)。

戴维森把真理的语义学概念与句法学概念之间的关系比作心理物理随附性:尽管有这样的事实,即真理不可能根据句法加以定义或还原,但在一定的意义之下,一个句子的真理仍然依赖于其句法属性。^⑩我认为,这可以被看作类似于弱随附性的某种东西:任何句法上不可分辨的两个句子实际上是相同的句子,因此必定具有相同的真值。但很显然,一个句子的真值一般是不能指望在各个世界都保持不变的。戴维森在用这个例子解释随附性时,说明了这样的可能性,即弱随附性也就是他说心理物理随附性时心中所想到的东西。这种解释与戴维森关于心理物理的异态一元论的学说是完全一致的,其大意是说,心理的与物理的类别之间不存在似规律的联系。似规律的联系在所有可能世界(至少相对于某种可及性条件而言)都一定是稳定的,而且心理与物理之间的这种联系恰好是心理物理弱随附性所不需要的。另一方面,这种解释有一个缺点:我认为,任何坚定的唯物主义观点都会断言,物质的东西决定那个世界中存在的一切,^⑪而这种弱随附性不可能提出这种主张。尽管我不清楚,戴维森是否会接受我所理解的唯物主义者的全部主张,但他想要的似乎不只是弱随附性。

虽然还未达到完全成熟的唯物主义,但心理物理弱随附性也许是一个值得深思的可能的论点:例如有人可能会论证说,关于一个有机体,不管是其行为还是其生理机能,没有什么物理事实迫使我们将其某种特定的心理状态,或任何的心理状态完全归之于它,但

216 第三部分

一致性仍然会要求,如果两个有机体表现出了相同的行为和生理机能,那么相同的心理状态必然可归之于其中的每一个,这是对心理状态归因的唯一的约束。我认为,这种观点可能会受到这样一些人的欢迎,他们认为心理状态的归因正好是推断理论上的解释状态的(比如说,与行为相关的)另一个例子,而且他们严肃地对待“颠倒光谱”的可能性。

另一个相关例子是:即使如一些哲学家们所相信的,理论是由所有可能的材料“不充分地决定的”,它们可能在下述意义上弱随附于材料:尽管没有任何材料迫使人们选择一个特定的解释理论,但“相当类似的材料”必须以“相当类似的理论”来解释。因此这里所用的弱随附性便产生了关于理论建构的一致性要求,就像弱道德随附性产生了对伦理判断的一致性要求、从而澄清了材料由以约束理论的一种明确的方式一样。弱随附性可能还有其他许多有意义的应用;不过,我希望我们所知的东西已足以使我们相信,它作为一个哲学概念具有潜在的意义。

3. 强随附性

当我们考察下面对弱随附性的等价阐述时,就会看到一条以适当方式强化弱随附性以获取一种更强的关系的线索:

A 弱随附于 B,当且仅当必然地,对于 A 中的任意属性 F 来说,如果一个对象 x 具有 F,那么在 B 中就有一个属性 G,使得 x 具有 G,并且如果任意的 y 具有 G,则它具有 F。

让我们先来看这两个定义是如何等价的。首先,我们要说明,前一个定义所给定的弱随附性可衍推出刚才所定义的随附性。假设对 A 中的某个 F 来说, x 具有 F 。我们需要证明,对于 B 中的某个 G 而言, x 具有 G ,并且任意的事物 y ,如果它具有 G ,则它具有 F 。假设 G 是 x (在所考察的任意给定世界里)的 B -极大属性。因此一般来说, x 具有 G 。要证明具有 G 的任意事物 y 具有 F ,就得假设某个 y 具有 G 。既然 x 和 y 都具有 G ,并且 G 是一个 B -极大属性,因此 x 和 y 共有 B 中的所有属性。根据最初定义的弱随附性, x 和 y 必定共有 A 中的所有属性。但是 F 在 A 中,并且 x 具有 F ,所以 y 也必然具有 F 。

其次,要证明第二个定义可衍推出第一个:假定 x 和 y 共有 B 中的所有属性,并且假设它们不共有 A 中的所有属性——即,对于 A 中的某个 F , x 具有 F 而 y 不具有。既然 x 具有 F ,因此第二个定义中所定义的弱随附性便可衍推出:对于 B 中的某种 G , x 具有 G ,并且具有 G 的任何事物都具有 F 。根据假设, x 和 y 共有 B 中的所有属性;所以, y 也具有 G ,由此 y 具有 F ,这便陷入了矛盾。

第二个定义的关键方面是其最后的子句,即具有 G 的任何对象也具有 F 这一规定。该句的确切意义是:在每一世界之内,这个 $G-F$ 泛化一定是成立的;它并不要求 $G-F$ 联系在各个世界中都是固定的。这意思是说,为了获得一种更强的随附性关系,它将确保随附的属性和它们的基础属性之间联系的稳定性,我们应该设法在该子句前面加上一个合适的模态算子。结果,这便产生了我们所需要的东西。

在前面引证的那一段中,戴维森对随附性所作的第二种解释

218 第三部分

也提到过这一方法;此处将予以重复,这个解释是:“没有某些物理方面的变化,一个对象不可能在某些心理方面发生变化。”“不可能”的模态力量以及对心理和物理“方面”的指称强有力地表明:理解戴维森这里所想说的意思的正确途径就是诉诸于心理特征与物理特征之间的一种联系,这种联系在可能世界间是恒常的。上面出自摩尔关于善良对自然属性的“依赖性”的引文也表达了同样的取向。

所以,和前面一样,假设 A 和 B 是属于布尔运算下的属性族。

A 强随附于 B,只有在这样的条件下,并且必然地,对于每一个 x 和 A 中的每一个属性 F 而言,如果 x 具有 F,那么在 B 中有一个属性 G,使得 x 具有 G,并且必然地,如果任何 y 具有 G,则它具有 F。

要说明这点,让我们回到“是一个善良的人”以及勇敢、仁慈和诚实这三个品质特征的例子上来。强随附性的观念可以这样概括:如果圣·弗朗西斯是一个善良的人,那么必定存在这些美德的某种组合(比如,诚实与仁慈),组合如此形成,因而圣·弗朗西斯具有它,并且具有它的任何人必定是一个善良的人。然而,这些特征的这个特定组合在能够“证实”“是一个善良的人”的基础族中不必是唯一的组合;苏格拉底也是一个善良的人,但他所具有的美德是勇敢与诚实而非诚实与仁慈。苏格拉底因为勇敢与诚实而成为一个善良的人,圣·弗朗西斯则因为诚实与仁慈而成为一个善良的人。一般说来,一个随附的属性通常具有多种可供选择的随附性基础——即基础属性,它们中的每一个对于随附的属性来说都是

充分的。如果 A 强随附于 B, 一个对象的 B-极大属性就是这个对象所具有的每一个 A-属性的随附性基础。但是, 一个 B-极大属性要作为一个给定的 A-属性之基础, 常常要比所需要的更强, 并且在任何比它弱的属性都不是随附性基础的意义上, 有价值的可能是一个极小基础。(相反, B-极大属性则可称作“极大的基础”。) 如果“是一个善良的人”强随附于自然属性, 那么任何善良的人的极大自然属性(或许是其所有自然属性的一个长合取)就会是“是一个善良的人”的随附性基础; 但是, 这显然超出了我们所需要的(它可能包括这个人的身高、体重、出生日期等等)范围, 而且也不够清楚明白。另一方面, 诚实和仁慈的合取属性可以构成一个极小基础——这是一个实际上包含更多信息且更有用的概念, 它将证明我们的下述说法是正确的: 这个人是善良的, 因为其诚实与仁慈; 他之善良在于他具有这些品质特征; 或者说他是善良的, 因为他是诚实与仁慈的。^⑫

模态术语“必然地”在强随附性的定义中出现了两次。在这里, 事先具体说明应怎样理解必然性, 既不可能, 也不值得; 一种恰到好处的好处具体说明一定有赖于所思考的特定的随附性命题, 而且对“必然地”的不同解读将引出有待思考的不同的随附性命题。例如, 如果某人对道德之于自然特征的随附性感兴趣, 那么这个术语的两次出现最好都应理解为信号逻辑的或形而上学的必然性。对于心理物理随附性而言, 有可能把第一次出现解释为形而上学的必然性, 而把第二次解释为法则学的必然性; 或者还有可能把两者都解释为形而上学的, 或都解释为法则学的。就部分学随附性而言, 最合理的解释可能是这样的, 即第一次出现表示形而上学的必

220 第三部分

然性,而第二次出现表示法则学或物理学的必然性。要点在于,对模态术语的不同解读将产生不同的随附性命题,而且这种灵活性是所述定义的一个值得注意的特征。所以,我们应该把对“必然地”的精确解释作为参量,让其相对于所用的特定例子来加以确定。

随附性的两个概念之间的下述关系是显而易见的:

(3) 强随附性可衍推弱随附性;而弱随附性则不能衍推强随附性。

下面一点也是一目了然的:

(4) 两种随附性关系都是可递的、反射性的,并且既非对称,也非不对称的。

在大部分重要事例中,随附性实际上似乎是不对称的;例如,尽管许多人断言价值评价对于非价值评价属性的随附性,但很显然,反过来并不成立。同样地,尽管心理物理随附性是一个有待论证的观点,但要认为物理的随附于心理的则显然是令人难以置信的。随附性的这种不对称性恰好是不对称性依赖观点的核心,一见到这种依赖性,我们就联想到了随附性关系。因为,当我们考察那个定义中已说明的一个强随附的属性与其基础属性之间的关系时,我们所有的一切就是基础属性可衍推出随附的属性。仅此还不能保证我们有理由说:随附的属性依赖于或决定于基础属性,或者说一个对象具有随附的属性,是因为它具有基础属性。后面这些关系强烈地暗示着一种不对称关系。从对因果关系和因果模态逻辑的研究中,我们已经得到了这样一种严厉的教训,即关于因果依赖性或决定性的观点是不能如此轻而易举地或直接地从简单的

模态概念就能够获得的;这很可能也适用于随附的决定与依赖的观点。关于依赖性与决定性的观点,不管是因果的、随附的,还是其他类型的,根本不可能用更简单的或更明晰的术语予以把握。我想提出的、唯一可能有助益的主张是:随附的属性对其基础属性的不对称依赖性完全有理由来自于下述不对称的依赖性,即一个综合性的、完整的属性系统——随附的属性是该系统的一个要素——对一个同样综合性的、系统的基础属性族的依赖性。这样,比如说,作为一个心理事件的痛苦对神经系统电化学过程的假定的依赖性,完全有可能根源于心理现象的整个家族对物理过程的不对称的、随附的依赖性。根据刚才的说明,这后一种不对称性不过是这样的事实(如果它是一个事实的话),即心理的东西在这里所定义的意义上强随附于物理的东西而非相反。因此,我所倡导的是一种整体主义:个体的依赖性根源于系统之间的依赖性,而不是相反。

4. 全体随附性与类-类联系

我们现在转到分析随附性的另一种方案,即为某些致力于心理物理随附性以及唯物主义研究的作者所津津乐道的方案。^③这种替代方案从全体的角度谈论“世界”和“语言”,并提出了可称之为“全体随附性”或“世界随附性”的概念。^④这一来,通过说明物理上不可分辨的世界在心理上也是不可分辨的(实际上,这样的世界是同一个世界),就可对心理物理随附性作出解释。道德的东西对非道德的东西的随附性也能以类似的方式予以解释:不可能存在

222 第三部分

两个这样的世界,它们在每一个非道德细节上不可分别,但在某个道德方面却相互有别。你可能还记得,在对弱随附性的缺点的讨论中,我们曾作过这样的阐述。有人可能觉得,这种全体的方案比我们自己的更可取,他们认为,我们的定义由于明确利用了随附族与其基础之间的属性-属性联系(如在我们关于弱随附性的第二个定义和关于强随附性的定义中所说的那样),因此是以一个重要的、未经证实的假设去反对这样一些人,他们显然是因为随附性有望成为一种依赖关系,而又不想承认隐含着不可信的、各种各样的还原主义的属性-属性联系,因此才求助于随附性。

这样一来,两个世界在心理上(或物理上等)是可分辨的还是不可分辨的,基本上就是关于心理属性如何分布于两个世界的各个个体之中的一个问题。如果各个世界在某个普遍的心理事实方面各不相同,那么这必然会反映在它们所包含的特殊的心理事实中的某种差异之上。因此,说两个世界在心理上是可分辨的就等于是说,对于某个心理属性 P 和一个个体 x 来说, x 在一个世界具有 P 但在另一世界不具有;说两个世界在心理上是不可分辨的,就等于是说,对于每一个心理属性 P 和每一个个体 x 而言, x 在一个世界具有 P ,仅仅是当, x 在另一世界具有 P 。

与前面一样,假设 A 和 B 是两属性集,试考察:

A 全体随附于 B ,仅仅是当,在 B 方面不可分辨的(简称“ B -不可分辨的”)诸世界也是 A -不可分辨的。

第二节中的讨论已经表明:全体随附性比弱随附性更强。这里的论证旨在表明:全体随附性等价于强随附性。为了证明强随

附性可衍推出全体随附性:假设 w_1 和 w_2 是 B -不可分辨、 A -可分辨的。这一来,对于 A 中的某个 F 以及对于某个 x 而言, $F(x)$ 在 w_1 中而 $\neg F(x)$ 在 w_2 中。假设 B^\star 是 x 在 w_1 中的 B -极大属性;于是,根据 A 强随附于 B ,就有 $\Box(\forall y)[B^\star(y) \rightarrow F(y)]$ 。既然 w_2 与 w_1 是 B -不可分辨的, $B^\star(x)$ 就在 w_2 中。因此,说 $F(x)$ 在 w_2 中,就陷入了矛盾。接下来要证明的是逆命题:假设强随附性不能衍推全体随附性,那么,对于某个对象 x 和 A 中的属性 F 而言,就有 $F(x)$,如果任意的 G 在 B 中,并且 x 具有 G ,那么 G 不能衍推 F 。(“ G 可衍推出 F ”是“必然地,具有 G 的任何对象具有 F ”的简称。)这就等于是说,对于该 x 和 F 来说, x 的 B -极大属性不能衍推 F 。假设 w^\star 表示现实世界:在 w^\star 中,我们具有 $F(x)$ 和 $B^\star(x)$ 。试考察另一世界 $w^\#$,它在 B -属性分布于个体方面与 B^\star 正好相似;尤其是, $B^\star(x)$ 在 $w^\#$ 中。然而,由于 B^\star 不能衍推 F ,因此我们可以始终假定 $\neg F(x)$ 在 $w^\#$ 中。这一来, w^\star 和 $w^\#$ 是 B -不可分辨的,但 A -可分辨的;即是说, A 不全体随附于 B 。证明结束。

因此,全体随附性不过是强随附性而已。这两个概念的等义性有一个相互补充的作用:两个独立构想出来的概念最终成为等义的这一事实说明了它们具有自然性,并包含有直觉的哲学内容。而且它表明,根据对类-类相互关系的承诺,在全体随附性和强随附性之间根本不存在任何区别;以为全体随附性没有这样的承诺的想法是错误的。

如果我们用“事实”定义全体随附性,又将会是什么样子呢?我们可能会有这样的某种东西:“ P 类事实随附于 Q 类事实,仅仅

224 第三部分

是当,在 Q 类事实方面相等同的世界在 P 类事实方面也是等同的。”这个阐述显然没有论及不论是待分析对象还是供分析对象中的个体的属性;然而,它与我们上述根据属性所作的阐述似乎基本相同。说两个世界“在 P 类事实方面等同”,这意味着什么? 意味着,把诸世界看成是某些极大的事实类别;因此说两个世界在 P 类事实方面等同就是说它们含有相同的 P 类事实。诸世界作为事实的类别的最大化条件可能意味着,两个世界含有相同的 P 类事实,当且仅当,它们含有 P 类的相同的单个的事实。依我看来,一个单个的事实就是像“a 是 F”形式的某种东西,在这里 a 是一个个体,而 F 是一个属性;并且说 a 是 F 这样的事实是 P 类的一个事实(比如说,一个心理事实),大概就等于说 F 是 P 类的一个属性(比如说,一个心理属性)。因此结论是,说两个世界在 P 类事实方面是同一的,就等于说:对于 P 类的任何属性 F 以及对于任何 x 来说, x 在一世界中具有 F,当且仅当, x 在另一世界中具有 F。这样,就 P 类事实而言的世界同一性概念就可归结为这样的概念,即先前所解释过的、就 P 类的一个属性集合而言的世界不可分辨性概念。而且,至此可以说真相大白了,根据现在对“事实”和一个事实“属于 P 类”意味着什么的解释,随附性讨论中关于“P 类属性”的话语一般可取代“P 类事实”的话语,尤其是,事实的随附性可以还原为属性的随附性。^⑤

对于两个族中的属性之间相互关系的存在来说, A 对 B 的随附性意味着什么呢? 根据弱随附性和强随附性的定义,部分答案已经清楚了:

(4) 如果 A 弱随附于 B,那么对于 A 中的每一个 F 来说,在 B

中就有了一个属性 G , 从而就有 $(\forall x)[G(x) \rightarrow F(x)]$ 。

(5) 如果 A 强随附于 B , 那么对于 A 中的每一个 F 来说, 在 B 中就有了一个属性 G , 从而就有 $\Box(\forall x)[G(x) \rightarrow F(x)]$ 。

我们先前的讨论表明, 如果假定无穷的合取与析取, 那么(4)可以强化为:

(4a) 如果 A 弱随附于 B , 那么对于 A 中的每一个属性 F 来说, 在 B 中就有了一个属性 G , 从而就有 $(\forall x)[G(x) \leftrightarrow F(x)]$, 即是说, 每一个 A -属性在 B 中都有一个共外延。在同样的假设下, 一个伴随的结果也会出现在(5)中:

(5a) 如果 A 强随附于 B , 那么对于 A 中的每一个属性 F 来说, 在 B 中就有了一个属性 G , 从而就有 $\Box(\forall x)[G(x) \leftrightarrow F(x)]$, 即是说, 每一个 A -属性在 B 中都有一个必然的共外延。

(5a)可以这样来予以证明: 假设 F 是 A 中的一个属性。我们可以假设 F 是偶然的; 这就是说, 某个 x 在某个可能世界 w 中具有 F 。(一般来说, (5a)对于非偶然的 F 来说为真。)根据强随附性的定义, 在 B 中有一个属性 G , 从而 x 具有 G (在 w 中), 并且 $\Box(\forall y)[G(y) \rightarrow F(y)]$ 。假设 $B_{x,w}$ 是 x 在 w 中所具有的 B -极大属性。那么我们就有:

$$\Box(\forall y)[B_{x,w}(y) \rightarrow G(y)],$$

由此:

$$\Box(\forall y)[B_{x,w}(y) \rightarrow F(y)].$$

而且对于在一个世界 u 中具有 F 的每一个 v 来说, 我们将有:

$$\Box(\forall y)[B_{v,u}(y) \rightarrow F(y)].$$

假设 B^* 是这些 B -极大属性的无穷析取; 那么

$$\Box(\forall y)[B^\star(y) \rightarrow F(y)].$$

不难发现,我们也能得出逆命题:

$$\Box(\forall y)[F(y) \rightarrow B^\star(y)].$$

假设不是这样;那么在某个世界 $w^\#$, 就会有一个对象 x , 从而有 $F(x)$, 而不是 $B^\star(x)$ 。但是根据强随附性, 在 B 中有某个属性 K , 从而 $K(x)$ 在 $w^\#$ 中, 并且 $\Box(\forall y)[K(y) \rightarrow F(y)]$ 。假定 $B^\#$ 是 x 在 $w^\#$ 中的 B -极大属性, 那么和前面一样, $\Box(\forall y)[B^\#(y) \rightarrow F(y)]$, 并且可得出: $B^\#$ 是 B^\star 中的选言支之一。因此, x 必定具有 B^\star , 这便陷入了矛盾。这一来, 我们便有:

$$\Box(\forall y)[B^\star(y)F(y)].$$

请注意, 这个双条件句中的“ \Box ”的实质就是在强随附性定义中内层模态术语(即“ \Box ”的第二次出现)的实质。就特定的随附性而言, 由我们赋予这个术语究竟是逻辑的、形而上学的必然性还是法则学的必然性所决定, 在必然性的那种意义上, 对于每一个随附的属性, 我们在基础属性族里面就有一个必然的共外延属性。就法则学必然性而言, 有些人可能会怀疑这一点; 我们将在下一节讨论这一问题。

5. 一些哲学上的思考

前面一节的主要结论首先是, 强随附性承认, 对于每一个随附的属性而言, 在基础族里都有一个必然的共外延的存在; 其次, 如果接受全体随附性, 这个承诺就是不可能避免的。因为两种随附性关系实际上就是同一种关系。对于一些用随附性来阐述某些哲

学主张的哲学家们来说,这乍看起来是令人困惑的。如前所述,弱随附性不可能抓住依赖性的全部意义;强随附性或全体随附性是必要的。但是要得到这种程度的依赖性,就事实本身而言,必须承认一个具有必然的属性-属性可衍推性的普遍系统的存在,根据强随附性的定义,这是显而易见的。而且,(5a)加强了这一点:每一个随附的属性在基础族里有一个必然的共外延的属性。这可能出乎某些哲学家们的意料之外。

我们已注意到一个可能的两难推理,戴维森可能会发现自己深陷其中:心理物理弱随附性似乎太弱,因此不可能导致唯物主义,而强随附性似乎又太强,它衍推出了一个具有心理物理等价性的普遍系统的存在。但戴维森倡导心理物理随附性的最初动机恰好是要承认心理对于物理的依赖性,而同时又否认存在着把心理属性和物理属性联系起来的规律。我们的结论要表明的似乎是这一点:如果你要承认心理物理的依赖性,那么你最好要准备好心理物理的规律——或者,至少是必要的心理物理的可衍推性。有的人可能会对这种思维方式提出质疑,理由是:“法则学的属性”,即规律所容许的那些属性不可能从属于布尔运算——那就是说,这些运算,当运用于这样的属性时,并不总是产生适于在那些规律中出现的属性。^⑩B-极大属性及其无限析取等“太复杂了”、“太人工化、太不自然了”、“太异质性了”,因此有理由认为不是“自然的类别”。

这引出了有关规律、还原、定义等的本体论和认识论的各种复杂的问题。不过,我这里只能述及我认为我们应该采纳的总的方案。首先,我们有必要对谓词与属性之间的差异保持敏感,并且要

228 第三部分

认识到,归之于谓词(或一般的语言构造)的复杂性或人工性不必归之于它们所表达的属性。通常,一个长布尔谓词组合作为谓词可能是复杂的;另一方面,它所表达的属性不必继承那种复杂性(布尔表达式可能等同于一个简短的表达式)。本文中的定义是为属性而不是为谓词所构想的;诸如无穷合取与无穷析取这样的运算对谓词而言可能是极不可靠的,但对于属性则不一定——无穷并集与交集对于类别也不一定。“不到一米长”这样的属性可看作一个无穷析取(比如,在那种形式的所有属性中,对于每个自然数 n 来说,不到 $n/n+1$ 米长)。实际上,我们有属性集合就足够了,而毋需无穷合取与析取。要点是,从属性的结构细节到它们的复杂性与人工性,不存在直接的推论,不管这些东西对属性来说意味着什么。

当我们说到规律时,我们可能想到的,要么是句子,要么是属性之间的某些非语言的、非概念的、客观的联系。如果把规律当作句子,那么我们的结论表明的就是,心理物理随附性并不能衍推双条件式规律的存在。因为我们并没有理由保证:存在着表达了被构造的属性的谓词,尤其是适当地简单而又明确的谓词。根据谓词而非属性对我们的基本定义进行再阐述是没有意义的;因为那将使非限定的程序极为可疑,甚或不可接受。而且,由心理和物理谓词所陈述的心理物理强随附性似乎远不如由属性陈述时那样合理:因为它断言的是,对于每一个心理谓词来说,都有一个(在逻辑上或法则学上可衍推出它的)物理谓词。比如说,可衍推出“厌烦”的物理谓词是什么呢?要作出回答,似乎至少要求助于“理想的物理语言”之类,而这可能恰好又把我们带回到属性的话语中去了。

另一方面,如果规律被解释为能被句子或陈述(“法则学的”或“似规律的陈述”)所表达的属性之间的客观联系,那么就必须承认,(5a)所陈述的是:在随附属性与基础属性之间存在着双条件式的规律——如果诸规律“仅仅”在法则学上是必要的,那么存在的就是等价性,这种等价性至少与规律一样强。但这样说的意思是什么呢?它意味着随附属性(或根据它们所阐述的理论)必然地“可还原”为其基础吗?意味着道德的属性可根据自然主义的术语予以定义,心理学可还原为物理理论吗?如此等等。看来这些双条件式的规律,或必要的等价性,提供了为理论间还原的经典概念所必需的“桥梁规律”。但这个结论可能是不成熟的。还原、解释等都是认识活动,而且仅仅根据这些等价性或双条件式的“存在”这类事实,还不能证明它们对还原或解释的运用是或永远是有效的。我认为,这里的“有效性”最好是根据一个得到很好证实的解释理论中的表征来加以理解,而且作为解释,就我们觉得它们令人欣慰和满意来说,这又特别有赖于我们自己的认知能力、我们的倾向和特性。所以,每一个心理属性的一个必要的物理的共外延的存在与我们完成心理学的物理还原的能力并无直接关系。

我们能够更有理由地预期的东西是:随着科学的进步,人们将成功地识别心理属性的越来越多的局部的物理的共外延,即是说,局限于特定领域的物理的共外延(比如说,特定的生物物种);而且这种局部共外延的一个足够广泛的系统可能成为心理学理论“局部还原”的基础。如许多人所指出的那样,任何给定的心理状态在不同的物理结构或生物物种中很可能有“多种多样的物理实现”;然而,对于任何给定的物种或结构类型来说,完全可能有一个不变

230 第三部分

的基础,而且比如说,如果这种基础的一个广泛的列阵相对于人类心理状态而得到确认,那么人类心理学就能够“局部地还原”为物理学理论。^⑦如果火星人的心理状态由于不同的火星人的解剖结构和生理机能而具有不同的物理基础,那么火星人的心理特点就得接受一种不同的局部物理还原,即使火星人与人类例示的是同样的心理特点。

摩尔坚持这样的观点,即善良是一种“简单的非自然的”属性,这里所谓“简单的”,他指的是不可定义性,而“非自然的”指的则是一般感性经验的不可接近性。假如这样,他完全不为善良对自然属性的随附性所动就值得注意了。事实上,在上面引证的最后一段中,他说道:除非他认为一个对象的善良“来自于”它的自然属性,否则他“绝不会想到要主张,善良是‘非自然的’”。如果一个对象的善良“来自于”它的那些我们假定可通过一般的感性经验接近的自然属性,那么善良本身为什么不是可接近的呢?也许有人回答说,这种意义上的可接近性并非争论的焦点;关键是善良在感性经验中没有直接的或即时的“呈现”。但如果善良确实来自于自然属性,那么这为什么不足以成为善良的自然主义认识论的一个基础呢,为什么不能使直观主义的道德认识论完全失去其价值呢?

关注一下摩尔的意见是饶有兴味的,他认为,尽管一个事物的善良来自于它的某些自然属性,来自于它是善良的这个事实,但是并不能由此得出结论说,它具有这些自然属性。我们的(5a)难道使摩尔陷入窘境了吗?可能没有,因为“来自于”一词正如摩尔、也许还有一般的哲学用法所使用的那样,似乎有一种明确无误的认识论维度:如果善良“来自于”某些自然属性,那么我们通过获悉一

个事物具有这些自然属性,就能够“得知”或“推断”它是善良的。善良的必然的自然主义共外延,就本文的论证来看,并不具有这样的认识论地位:我们知道它必然存在,如果强随附性得到认可的话,但可能永远不知道“它是什么”。不可能指望这样的一个共外延为“善良”这个术语提供定义的依据;实际上,其存在甚至不足以证明用自然主义术语对“善良”的“原则上”的可定义性。因为对定义的看法包含着某些语义学的和认识论的联想,即使我们能够识别善良的作为基础的自然主义共外延,我们也不能指望这些联想能支持它。

我们似乎已得到了这样的结论,即随附性关系本身对于诸如可定义性和可还原性这样的关系,并没有直接提出什么意见,如果“可定义性”或“可还原性”所涉及到的可能性在一种相当强的和实在论的意义上来理解的话。然而,如我在后面要论证的那样,在接受关于两个领域的随附性命题与拒绝这两个领域间的所有重要的认识或概念的关系之间,存在着一种至关重要的张力。但我们还是先来讨论意志自由这个问题。

对于与其基础有关的随附的东西的意志自由来说,随附性意味着什么呢?尽管一个详尽的讨论要求对意志自由的相关概念有一个更准确的理解,但弱随附性似乎能够与意志自由完全一致;事实上,那可能是它的主要的吸引力之一。然而,强随附性则不同了:在强随附性之下,基础完全决定了随附的属性。如果心理物理强随附性成立,那么心灵王国所发生的事情在每一个细节上都决定于物理王国所发生的事情。这种决定关系是一个客观事实;它不以人们是否知道有关它的什么为转移,或者说不以人们用何种

232 第三部分

表达式来谈论心与物为转移。不同于还原和定义,认识论上的考察在这里鞭长莫及。这或许就是全体随附性为什么常常用来陈述唯物主义学说的原因。同样地,道德对自然属性的强随附性可能标志着一种形式的“道德自然主义”——并非定义上的伦理自然主义观点,而是承认自然对于道德的本体论上的优先性的一个形而上学观点。

我深信,对因果关系决定性的思考将为我们思考随附的决定性提供一个有价值的类比基点。如果因果决定论(“每一事件都有一个原因”)成立,那么每一事件就都有一个时间上的在前的决定条件。然而,对于我们将如何成功地识别原因并构建因果解释,它无可奉告;对于我们将如何成功地发现因果规律,它也是缄默不语。解释是一种认识论上的事情,而且那种认为所有事件可从原因上予以解释的主张是一个认识论观点,或方法论学说,并不是孤立地从普遍因果关系的形而上学观点中衍推出来的(除非前者可明确地理解为意指后者)。与此相仿,一个给定领域随附于另一个领域的观点是关于两个领域之间的一种客观存在的依赖关系的形而上学观点;它不可能说明:依赖关系的细节将是否或如何成为可知的,以至于使我们能够对解释、还原或定义作出阐释。

在将因果决定论的形而上学观点与同它联系在一起的、关于因果解释可能性的认识论观点区分开之后,我们现在就有条件协调它们的关系,并正确评价它们的相互关联。首先,有这样一种非常直接的关系:在没有因果关系的地方,就不可能有正确的因果解释。当一事件引发另一事件时,就造成了使一个相应的因果解释“正确”或“真”的客观事实。更一般地说,我们可以认为,因果决定

论的命题为要求我们依据其因果前提来寻求对自然事件的解释的方法论的战略或原则提供了一个形而上学的基础,而且还为这个战略为什么能像通常那样发挥作用提供了一个解释。所以,接受因果决定论就可视之为赞成作为认识论战略的因果解释方法的一种体现。反过来,为我们关于因果决定论的信念建立必要的基础,这则是在发现因果联系并阐述因果解释过程中的成就,尽管这可能是有限的。

同样,相信两个领域之间存在着一种随附性关系,能够推动我们对属性-属性联系的探索,而根据这些联系,有利于开阔视野的还原和富有启发意义的定义就可能得到系统的阐述。在强随附性存在的地方,(5a)使我们坚信:这些采取必然等价性形式的联系必将被发现,但又自然未进一步保证:我们将成功地发现它们或者说它们将会再现于一种解释理论之中。一个恰到好处的例子是部分学随附性,即有形事物的宏观属性随附于它们的微观属性的学说。似乎正是这种形而上学的原子论学说支撑和支持着现代理论科学中对微观还原的庞大而卓有成效的研究战略。反过来,这种研究战略的成功又增强了我们对部分学随附性的信念。或许,类似的评论也适用于道德的随附性:对道德的东西随附于非道德的东西的信念可能已经引导着道德哲学的一些主要的假定和工作,诸如对伦理学术语的自然主义定义的探讨、相信道德的归因一定有非道德的“标准”、相信存在着诸如“行善”或“主持公正的特性”之类的东西,还有用自然主义和描述性术语来陈述行为合理性或制度公正性的条件的持续尝试。

我认为,这些评论阐明了(5a)的要点:它有助于我们认清关于

234 第三部分

两个领域的随附性命题与我们对它们可能采取的某个认识论战略之间的联系。它解释了:对随附性命题的信念如何能够导致、反过来又为其支持的下述预期:一个领域能够根据另一个领域、通过发现(5a)向我们保证存在的必然等价性而得到理解——即得到还原、定义、解释等。在几段以前,我所述及的张力之所以出现,显然是因为这种联系在赞成随附性、同时又拒绝所有重要的概念的或认识的关系时被否认了。这就是说,可能性有这样一种意义,尽管是弱的,在此意义下,(5a)表明:强随附性衍涵着有关领域之间的还原或定义的可能性。

注 释

① 金在权:“随附性和法则学上不可公度的量”,《美国哲学季刊》,15(1978):149-156。

② 摩尔(1922):《哲学研究》,伦敦,p.261。

③ 黑尔(1952):《道德语言》,伦敦,p.145。

④ F. 西布利:“美学概念”,《哲学评论》,68(1959):421-450;J. 莱文森:“美学随附性”,《南方哲学杂志》,22 增刊(1984):93-110。

⑤ 见注①;另见金:“心理物理随附性”,《哲学研究》,41(1982):51-70,以及“作为心-身理论的心理物理随附性”,《脑与认知理论》,5(1982);J. 豪格兰德:“弱随附性”,《美国哲学季刊》,19(1982):93-103。

⑥ 详见奎因给乔姆斯基的回复:《语词与异义》,D. 戴维森和 J. 辛迪卡主编(多德雷切,1969),p.302。

⑦ 戴维森(1979):“心理事件”,见 L. 福斯特和 J. W. 斯旺森主编:《经验与理论》(阿姆斯特),p.88。

⑧ 根据豪格兰德与黑尔谈话的记录(豪格兰德,引文同前),黑尔心中所想的似乎就是我所说的弱随附性。黑尔在谈话中所解释的“随附性”以及“衍涵”的概念已证明分别和我所说的弱随附性与强随附性大致吻合。

⑨ 有关普适性要件的两个变体的区分源自 J. 霍华德·索贝尔未刊论文

“依赖的属性”。

⑩ 戴维森：“心理事件”，p. 88。

⑪ 对唯物主义的阐述可见 T. 霍根：“随附性和微观物理学”，《太平洋哲学季刊》，63(1982)：29-43；霍根：“随附性与宇宙注释学”，《南方哲学杂志》，22 增刊(1984)：19-38。D. 刘易斯：“全称命题理论的新研究”，《澳洲哲学杂志》，将刊。

⑫ 这种关于一个因果条件的概念会引出类似的问题；有意义的通常是条件这样一种极小集合，它对结果是充分的，而不是任意的条件集合。

⑬ 详见霍根、豪格兰德和刘易斯等人的论文，同前。

⑭ “全体随附性”一词，我是从 P. 特勒的论文“关系整体论和量子力学”（即将发表）中借用过来的。

⑮ 状态与事件的随附性也可还原为属性的随附性，如果它们被解释为属性-例证的话；关于状态与事件的概念，可见金：“作为属性例证的事件”，《行为理论》，M. 布兰德和 D. 沃尔顿主编（多德雷切，1976）。如果与戴维森的名字联在一起的、关于事件的多种可供选择的观念（比如，可参考“事件的个体”，戴维森：《行动与事件论文集》，牛津，1980）被采纳，那么诸事件就可简单地被看作个体，即是说，在随附性的定义中，作为变量“x”、“y”等等的值，又形成了关于事件随附性无必要性的一个特殊的概念。

⑯ 关于按这些思路所作的有趣的思考，可见 P. 特勒的“对 J. 金论文的评价”，《南方哲学杂志》，22(1984)：57-61。我这里的看法，包括对特勒的某些观点的回应，表达了我对先前所辩护的观点的修正与增补，尤其是在“随附性与法则学上不可公度的量”一文中是如此，同前。

⑰ 关于“局部还原”，可见金：“作为心-身理论的心理物理随附性”，同前；亦见 R. 理查森：“功能主义与还原论”，《科学哲学》，46(1979)：533-558。

本体论的随附性*

约翰·豪格兰德

1. 背景

两年前,我写成并发表了名为《弱随附性》(1982)的论文**。但我孤陋寡闻,不知该文的几个重要论点早在几年前就已为 G. 赫尔曼和 F. 汤普森在两篇合作论文中明确地阐述过(1975, 1977)。迟做总比不做好,在这里,我愿承认他们的优先性,并为没有及时这样做而致歉。既然我已知道了他们的开拓性工作,因此我想明确地探讨他们所提出的别的一些问题。

赫尔曼和汤普森(以及我)都试图以不太含蓄的方式明确阐述物理主义唯物主义的基本直观;尤其是我们都想避免任何可衍推出传统还原主义的表述。他们主张把两个不同的论点结合起来,这两个论点他们分别称之为物理决定性原则和物理穷尽性原则。

* 选自作者的《思:心灵的形而上学论文集》,哈佛大学出版社 1998。该文原为在 1983 年举行的“孟菲斯州随附性斯宾德大会”上的讲演稿,后发表于《南方哲学杂志》(增刊)22(1984):1-12。

** 在“弱随附性”一文(见他的上述论文集《思》,pp. 89-107)中,他提出了物理主义一元论的一种变种,即他所谓的“弱随附性”,认为它既不同于标记同一论,又有别于二元论、类型同一论。他说的随附性如此之弱,以致“它不蕴涵任何种类的同论,然而又坚持物理的东西的适度的优先性。”

前者与我(后面)所说的“弱随附性”相当接近;尽管有许多不同,但我是非常赞同的(一点也不奇怪)。^①另一方面,物理穷尽性原则在我的论文中找不到对应物;而且我是很不赞成的。因此这正是我要商榷的。

先追溯一下历史。近来有许多作者拒绝所谓的“类型-类型”同一论(以及直接相关的还原主义),根据有两方面,即它强得太离谱,而且对唯物主义又是不必要的。他们说,所需要的不过是一种更弱的“标记-标记”同一论。其基本观点是:即使心理(或无论什么)类型(种类,属性)不能同一于(或从法则学上关联于)物理类型,但个体的心理标记(特殊的事件,状态等)则仍可以在一一对应的基础上同一于个体的物理标记;这足以满足物理主义的形而上学要求。

现在的问题在于,标记同一论是否还会太强,对唯物主义是不是多余的。我的看法是肯定的,决定性/随附性原则本身就足够了。对于该原则,可以作出如下初步的解释:

如果世界在某种严格意义上的物理方面——即在某些可用规范的物理学语言加以描述的方面——没有什么变化,那么世界在任何方面都不可能变化。

值得注意的是:这并不涉及任何特殊的个体(除了整个宇宙);因此与原先的定义不同,^②它并未预先明确地设定标记同一性。我的论文的主要意图是论证:随附性和同一论事实上是各自独立的,后者大概是错误的。赫尔曼和汤普森赞同独立性,^③而且对同一论的通常阐释也不满意。但是他们提倡一种新的、更富思辨性的理论,而不是拒绝同一性。

2. 赫尔曼和汤普森的物理主义

许多哲学家包括我本人一直乐意认为:某种“适当的”物理谓项所适用的任何实在是物理的——至于这些谓项是什么,则另当别论。赫尔曼和汤普森有一种相同的直觉,不过又有更好的策略。他们由此出发的仅只是“基本的”物理谓项——诸如“ x 是中微子”,“ x 是电磁域”,“ x 能吸引 y ”等。他们假定:基本谓项的肯定的事例无疑是物理的,这些东西合在一起构成了整个物理宇宙。我们尚未被告知;哪些谓项才是“基本的”;但它们一般比“适当的”谓项的可疑性显然要轻得多。事实上,我的理解几乎是,根据当代物理学的看法,单个的谓项如“ x 有不等于零的物质-能量”就足够了。

无论如何,必要的改革在于:运用对个体的演算(即部分的或整体的)从基本谓项的肯定例示中把所有物理实在建构出来。首先,物理宇宙被规定为所有那些事例的合成(即总合成时空集合);进而全部物理实在且只有物理实在才变成那个宇宙的部分(时空上的份额)。合成宇宙,进而再分割宇宙,其根本点在于,新的部分在“界限”上是没有限制的。它们可能是任意(或许无比)复杂和异乎寻常的;在任何特定的语言中,它们没有必要予以定义,不管是明确的还是不明确的。

对于赫尔曼和汤普森的同一论来说,物理宇宙的那些任意部分就是物理的“标记”。在任何(有限论者的)语言中,既然这种标记之存在比谓词之存在确实要多,因此与以前相比,他们不仅提供

了得到更明确说明的物理基础,还提供了更广阔的基础。不过物理穷尽性原则与标记-同一论并非一模一样,因为它也划定了抽象实在的界限——即是说,限定在我将称之为“有物理基础的”集合上。这些正好是出现在集合论的等级系统中的集合,它们全部的原始(ur-)元素(基本个体)都是物理标记。^④因此在赫尔曼和汤普森看来,所有真实存在的东西都是物理标记和有物理基础的集合。

这显然是对同一论的一个绝妙的阐述,较之于它的先驱来说既强硬又温和;它为决定性/随附性原则制造了一个健美的伴侣。但问题依然如故:那个原则想要一个伴侣吗?即是说在我们努力去表述唯物主义的有点不成熟的观点时,有没有论证把物理穷尽性与物理决定性连在一起的理由?我认为没有——即使我承认,我尚不完全清楚怎样充分地论证这一点。

3. 成为唯物主义者的理由

一个似乎合理的出发点就是先去追问:我们为什么要做一个唯物主义者。然后我们再考察哪一个候选原则(如果有的话)能满足那些原始的动机。原来,唯物主义的通常理由居然是一令人瞠目的大杂烩;这一事实唯一的作用就是授予它们的庄重的一览表以价值。我已区分开了六个松散的范畴,我先将它们罗列出来,然后简要地予以评述。

(1)“一”的呼唤。某些人严肃地认为,科学具有一种所谓的统一性。如果这种统一性被设想为一种需要解释的、确定的事实,那

240 第三部分

么我简直有如坠五里云雾之感。另一方面,如果它被设想为一种迫切希望得到的东西或目的,那么我们但愿它的寻求者都有好运气,如愿以偿;但是他们最美好的希望几乎都不能算是接受唯物主义的理由。事实上,我认为,只有那些已成了唯物主义者的人才真的倾向于期盼科学的“统一性”;他们用这个词的真正意思不过是说:他们的唯物主义将会得到证明。^⑤

然而一个迥然有别的统一性问题更加令人有趣。如果不拘泥于细节,那么我们大致可以从三个方面规定实在:(1)(几乎是)每个人都同意的那些东西是物理的;(2)那些有争议的东西也是如此;(3)(几乎是)每个人都同意的那些东西不是物理的。第三类范畴包括数量、集合、柏拉图式的形式等——即是说内在的甚或必然的实在,在此意义下,它意指的不是易于产生、变化或衰变的东西。对比之下,前两类范畴包括暂时的或偶然的東西——即我们一般地称之为易于产生、变化、衰变的“偶发事件”。从形而上学来说,这种暂时/非暂时的对比是相当重要的;它把所有有争论的问题放在与显而易见的物质的东西相同的那一边了。举一个恰当的例子,信念和情感就其是暂时的、偶然的偶发事件而言,像高山、分子一样(不像数量和集合);这也适用于风俗习惯、道德行动以及你所拥有的任何东西。

这似乎有可能是一个有待解释的事实;唯物主义可能已作为那种解释崭露头角。什么是暂时的/偶然的東西的统一性原则?物质!机智地说,如果我们与康德一起问,为什么时间是内、外感觉的形式,那么我们能够回答说,与笛卡尔相反(更不用说康德),时间(而不是空间)是物质的本质——而且一切(可感的)都是物质

的。

(2)富有诱惑力的类比。一篇小说甚或一幅绘画都可充满勇气、冲突和悲怆;当它启动新的艺术闸门时,激情和美就会跃然纸上。然而在某种意义上,每一本书不过是墨水和纸张;把那些纸张蒸发掉,你就等于消灭了那份原稿——剥夺了它的存在。消灭了所有的原稿,(无论如何)你就消灭了著作。人(文化以及别的一切)不都是这样吗?——把他们的物质实体蒸发掉,你就完全剥夺了它们的存在。在我看来,就是那个样子。

(3)事物的发展趋势(“进步”)。从否定方面看,唯物主义的敌人似乎像殊死搏斗的战士一样,不顾一切地固守着某些击中了形而上学滑轮的陈旧体系。更明确地说,科学在凯歌行进。他们首先肯定:星球就是(普通的)物质;接着就进到尿素的合成、进化论、双螺旋结构,……现在又想标新立异地谈论他们的计算机。似乎运动的物质已一再被证明取得了决定性胜利——别的一切纯属多余。自然而然地,与其说我们是殊死搏斗的战士,不如说是获胜者。

(4)对隐秘的恐惧。有些人说,有心理的自然化动物是不清楚的、幽灵般的、难以捉摸的,一般来说也是不可思议的——因此在本体论上是可疑的。即使不对该前提提出争辩,我们仍可以探讨这个推理:幽暗不明的东西为什么就不可能是真实的呢?噢,存在的就是合理的,^⑥这正好是形而上学的古老原则(即使是古怪的,且总没有得到承认)。难以查明的是,那东西由什么构成。从表面上看(在没有理智的创造的情况下),这似乎有点武断,但我仍然确信:它的根深深扎在丰润的哲学沃土之中。

尽管如此,但从仅仅没有幽灵到对唯物主义的论证仍可能采取多种途径。一种方式是说,那种理解是解释的产物,而我们所能得到的唯一真正清楚的解释存在于物理科学之中。因此除非别的科学刚开始萌芽,否则它们的领域最好是同一于物理科学,不然的话就会成为令人沮丧的非存在。而不同的方案则强调:合理性是以认知的有意义为前提条件的,后者进而又总是基于客观的观察。因此假如只有物质对象(状态、事件等)才能客观地予以观察,那么只有它们才是合理的,从而是真实的。

(5)吝啬鬼的直觉。那个可怜的孩子,为了化解形而上学的说明不知背负了多少债务,乃至倾家荡产?从根本上说,奥康的著名剃刀“如无必要毋增实在”自然是从比特尔·贝利的更锐利的“如无必要什么也不要做”中得到其锋芒的。但是在讲求实际的新英格兰,警告更加锐利经济:“在一切承诺中尽可能节约——不,要吝啬。如有必要就买基本的物理学,但要避免一切不必要的浪费。千万千万不要把你珍贵的液体浪费在过度奢侈的假说中!”

我有时不解的是,什么使本体论如此昂贵,或者说什么限制了我们的预算,或者说我们实际上花费了什么。由于在西南比在东北度过了更多的夏季,因此我就不太渴望沙漠的景色。坦率地说,奥康的剃刀在那里几乎没有留下任何印记。不过,我也许可以遵守精明的加利福尼亚人的折衷方案:我们不要提前假定任何实在。

(6)规律与秩序。物理学力求(并允诺)严密;即是说,力求提供一种关于毫无例外的规律的封闭的、广泛的系统。^⑦换言之,撇开对其他领域的任何暗示,物理学总是非常恰到好处地关注自身。物理现象可以在相应的描述层次上得到物理上的充分解释。这种

允诺结果引起了相互作用论的标准批评：如果物理现象能用它们自己的术语完全予以解释，那么非物理的东西（例如任何心理的东西）对它们又会有什么影响呢？似乎只有违反物理学规律才能如此。而且心理的东西最好能够影响物理的东西（例如在深思熟虑的行动中），否则它就是没有任何用处的。因此不幸的是，使心理在既是非物理的、又是有作用的这一价格标签似乎就是物理学本身的内在严格性。它绝不会在波斯顿出售。

4. 这些为什么不是同一论的理由

关于唯物主义的小节目我就说这么多；不用说，某些考察给我的印象比其他的更深。现在我回到对这个问题的阐述上来。尤其是，这些动机除了支持物理决定性原则之外还支持同一论吗？按顺序再审视一番那个目录正好是举手之劳的事情。

(1)“一”的呼唤。对暂时的/偶然的東西的统一性的解释是一种引人入胜的形而上学的珍品，是唯物主义的一种巨大的附加物。但是在唯物主义中，实际上是什么允诺给予我们这一珍品呢？在我看来，物理决定性原则清楚地说明了这一点。因为该原则明确指出：无论什么都可能有差别，换言之必定是暂时的/偶然的。该原则用来约束物理的东西的——从而借助对物质的常见的随附性（正如规范物理学中所描述的）对范畴的统一性作出阐释——正好是这个范畴。

我不明白：物理穷尽性对前述的东西究竟增加了什么，但它可能提供了一种替换说明。因为它可衍推出：每一实在不是一个物

244 第三部分

理的标记就是一个集合,那种区分可能与暂时的/偶然的区分相匹配。说暂时的偶发现象不过是物理的标记,其实既解释了那范畴的统一性,又解释了它与时间的联系(因为物理标记是宇宙的时空要素)。不过,这一战略怎样包含暂时的事件、过程和状态尚不明朗。它们不得不同一于物理标记(而不是集合),不然的话就会瓦解那一观点;而这似乎又可能引出种种麻烦。

顺便值得一提的是,我要对一种颇有诱惑力的简单化提出警告。按照可能的变化,重新阐述随附性是很诱人的:

如果没有某种(可从法则上加以描述的)物理方面的变化,世界在任何方面都不会有变化(变成不同的东西)。^⑧

不幸的是,存在着明显的反例。例如,当午夜钟声敲响时,一个人可能从已婚变为单身汉,或者说一笔财产由你的变成了我的,因为某种更重要的合法的法令在那时生效了——即是说,除了钟的滴答声之外,那时并没有任何物理变化。唯物主义者想坚持的自然那是法令颁布时后面所发生的物理变化(以及在婚姻的、财产的制度等的演变期间,许多别的变化)。因此必须在“全世界”的一切时间及每一当下作出对随附性的一般陈述,进而根据世界(就其全部而言)怎么可能发生变化来作出更自由的系统的阐述。

(2)富有诱惑力的类比。众所周知,绘画/故事类比易受拒斥策略的攻击;在那个故事中,其实并不存在怜悯——那个故事正好引起了人的怜悯之情。而且,其实并没有真实的故事,而只有似故事的铭文;在某些似哲学家的有机体看来,“真实的”心理实在可能同样要被排除。不过我看到的全是拱和孔。

因此假设我们一开始就承认这个类比的价值：它真正暗示的是什么呢？噢，从根本上说，如果你使足够多的物质结构解体了，那么你也同时消灭了其他许多实在。但是哪些实在会遭受这种毁灭呢？可以推测，数量、集合和柏拉图式的形式不会；事实上，潜在的受害者似乎只限于暂时/偶然的東西——即是说，只限于随附于物理的东西的那些实在。其实，没有物理王国，就不可能有任何暂时的王国，这正好是表达物理决定性原则后面的直觉的一种更粗糙的方式。

(3)事物发展的趋势。一当我们对科学成功意味着什么作出了独立的论证，按科学的趋势前进在这里才有助益。合成尿素和统计热力学之类的成就如果被视之为归纳的基础——因为它们还是还原，而且我们一直放弃了将其泛化的尝试——那么它们就太令人尴尬了。另一方面，像进化论和功能心理学这样的事例（如果可以算作成功的话）从形而上学上说是有不同的解释的。它们充其量只是说明了：仅仅是“因为”物质像其所是的方式被配列，因此才有生命形式、心灵之类的存在。换言之，如果某些物质结构消失了或瓦解了，那么各种生命和思维也会随之消失。但是这并不意味着，每一特定的心理的（或生物学的）偶发现象同一于某种特定的物质结构。相反，结论不过是：瓦解了足够的物质结构就会毁灭许多别的实在——这就是我们刚讨论过的东西。

(4)对隐秘的恐惧。试考察具有隐秘性的自然化动物 D，他正在奋争的恐暗症者中寻找唯物主义的避难所。古典 D-类类型——还原让他们都充满着物理学的防御之光。但是唯物主义者不再有提供还原的全权；而某些祈求者一定会大失所望。在物理的决定

246 第三部分

性中有安全保障吗？噢，它迫使 D 直接地开辟物理路线；因为如果物理世界没有变化，那么 D 也一样不可能有变化——而且这自然适用于 D 中的每一可能的变化。我以为，这足以排除精灵、鬼怪、离体精神、有魔力的幽灵以及各种庸俗的虚幻物——而不一定排除了思维、情感、习惯、风俗等。另一方面，必须承认，决定性对其所省去的实在并没有带来特殊的福音；随附性可以淡化非物质的阵营，但它并没有为剩余的东西提供解释。

物理的穷尽性又如何呢？与物理宇宙中的某种任意的碎片的同一不会使 D 之作为 D 有丝毫的合理性。须知，穷尽性原则并不意味着：D 除了作为 D 还有任何别的同一描述；尤其是，即便 D 充斥的全是物理的东西，他也不一定非要用物理学术语予以描述。换句话说，唯物主义的子句（若无还原论）根本不可能担保所有那些东西具有物理上的可理解性。如果是这样，同一论就没有填补随附性所留下的空白。

(5) 吝啬鬼。比特尔·贝利的剃刀开辟了两条路线。可以肯定，有这样一个必然的结论：“如无必要毋增实在。”而另一强劲的、重要的结论则说：“如无必要毋增原则。”这两条路线的难题在于搞清楚什么是“如无必要”。奢侈的要求是吝啬鬼的天敌。我一直主张：同一论对直觉上令人满意的唯物主义来说是多余无用的。我的主张之一是：在对实在的说明中的纯粹的吝啬本身并不是强制性建立与物理的东西的婚姻的理由。

(6) 规律与秩序。这个论证一定得谨慎从之。说某一分子的运动“完全”可从物理上加以解释就等于说，加之于它的所有物理力量完全决定了它的运动；其他的力量是多余的或不允许的。但

是这并不排斥别的解释。同一个分子可能寄居在我的拇指甲中,进而可能提高 3 英尺,因为我决定举起我的手。物理学的严格性并不阻止这些解释;它阻止的只是:任何抉择为任何分子提供了一种额外的力量。换言之,心理学解释不允许干涉或限制物理学解释——而只是通过提供另一种观点对之作出补充。

试以同样的方式考察一下打在这个句子标记末尾的标点符号。它是一点儿墨水;在手边的纸纤维上,把力量加于其上,这就足以解释它现在的空间位置(在某种指示图式中)。人们也可以通过说:它与一陈述性文字在末尾并置,而解释:那时它为什么在它所在的位置(这反映了我把它放在那儿的理由)。这两种解释都是完全正确的;它们在这里几乎是等值的;它们不是势不两立的。严格的完善性只意味着:物理学绝不容许对它自己的地盘入侵;它绝没有这样的意思:物理学是都市里的唯一的解释游戏(或唯一好的游戏,或我们确实需要的唯一的游戏)。

这向我们说了关于心理的东西的什么呢?如果物理学就其自身的术语而言是完善的,那么当物理的东西本身(在粒子、场、力以及难以归类的东西的层次上,足以解释一切的东西)没有伴随的变化时,心理的东西(能对物理的东西产生影响)也不可能有任何变化。但是这正好是物理决定性原则(根据严格的物理学)的一个特例;物理穷尽性与此无关。

不过,人们可能坚持一种更强的解释。造成相同的变化——引起相同的结果——不仅是因为必须有某种或别的“伴随的”物理变化;而且必须有某种物理变化。然后再增加一些故弄玄虚的言辞,如因果作用是自然的最亲近的个体化原则,进而标记同一就会

248 第三部分

脱颖而出。但暂时不要表态。

对于用物理学语言明确描述的每一物理结果(例如每一分子运动)而言,一定有用同一种语言表述的严格的物理学说明。但是行动的抉择(比如说举起我的手)“涉及到”我身体内外的不可名状的无数分子运动。物理学的严格性绝不要求把这些运动归入如下种类:

A. 有效的原因,作为整体单元(复合的物理标记,同一于我的抉择);

B. 有关的结果,作为整体单元(复合的物理标记,同一于我的行动);以及

C. 单纯的旁观者(如我的血液循环,我的拇指甲中的热运动,我的皮质中的蛋白质合成等)。

不过,这些等级中的每一个对标记同一论都是必要的:它们实际是关于它的一种陈述,因为事实就是如此。

这里的观点不是说:抉择和行动的“界限”从微观物理上说是模糊的、不可分辨的甚或可互换的。毋宁说,根本就没有这种界限:抉择不可能(甚至在原则上)在微观物理学的层面上区分为孤立的个体。或者说,物理学的严格性无论如何无法衍推出它们有如此的可分性。因此我再一次断言:关于唯物主义的众所周知的理由支持决定性/随附性原则,而不赞成同一论或物理穷尽性。

5. 其他有诱惑力的解释意图

当然,可能有支持同一论(除了随附性之外)的独立的论证,且

不说有利于唯物主义的一般理由。我将考察其中两个论证。第一,人们可能设想,物理穷尽性原则是必要的,它可解释心理的东西(以及别的东西)为什么由物理的东西所决定。但是这样一种解释绝不会出现,正如赫尔曼和汤普森所说的那样,两个原则是各自独立的。随附性是两组真理或事实之间的一种关系;就此而言,它对“被随附的”——假设是理论物理学的——语言的表现力提出了强烈的要求。另一方面,物理穷尽性原则对物理学只提出了最低限度的要求:少数“基本”谓词是必要的,而于此,有关的理论就可能是非常弱的,且没有表现力的。

同时,仅仅为了解释一个已确定的原则便宣布一全新的原则,这是令人费解的。换言之,如果有关于同一论的更进一步的根据,那当然是求之不得的(让它们去证明它)。但是如果没有,那么随附性就肯定是我们所需要的唯物主义形而上学。它本身完全有可能是一个“原则”:穷尽性不再像原来那样是根本性的了,而且再也不会得到有效的支持。

第二个论证更不可靠。物理的决定性(甚至与严格的物理学相一致)与平行论、副现象论以及过去司空见惯的哲学幽灵似乎完全可以和睦相处。它不是认为同一论把那些东西抛回到坟墓里去了吗?我自然表示怀疑。试考察另外两个假说:

A. 每过一千年,宇宙便停顿——一动不动地停着——约 10 分钟,接着恢复它的运动,实际上如过去一样;这最后一次正好发生在威廉·奥康出生之前。

B. 一亚 - 原子粒子实际上有自己丰富多彩、独一无二的个性,这一点表现在它与邻居的深层的、充满激情的谈话之中。不幸

250 第三部分

的是,再好的翻译也无济于事,因为那些粒子的量子力学行为没有受到什么影响——除非是在统计非决定性中。

这些假说与平行论、副现象论何其相似;尤其是它们没有科学上的动机,魔法似地难以捉摸,完全是稀奇古怪的。依我看来,这足以摧毁它们。但是我们可以作出原则性泛化的样子,方法是对贝利的下述剃刀阐述一种新的结论(我认为迄今未受到注意的),这剃刀是:“如无必要,毋故弄悬殊。”不用说,这个结论(如前此的奥康的)必定是一种辩证法的托词,主要用于把证明的担子转嫁到别的人身上。

乍一看,同一论像似一种更巧妙的、更自信的回答。由于它,平行论和副现象论最终不仅仅是古怪的,不可证明的,而且它们与物理穷尽性原则是势不两立的——真正被禁止干的事情。不过,表象可能是有欺骗性的。因为假如我们把例子变换成(粗略地说)斯宾诺莎式的一元论:

实际上只存在一个实体,即神,但它有无限多的属性,其中每一个又是完善的、无限的,与别的每一个完全相符;神的属性中有两个,一个是心灵,一个是物质。

这一来,当我们讲到物理标记时,我们便说到了神——尤其是它的物质属性之下的样式。同样当我们提到别的标记时,我们说的也是神——还是同一个神。换言之,不管我们谈论什么,从本体论上说,它与我们谈论物理标记时所谈论的东西正好是同一的(等同的)。与赫尔曼、汤普森的系统不同,斯宾诺莎的系统自然是对称的:它也蕴涵着心理的穷尽性。

关键在于,极其乖僻的形而上学并不知道什么界限。(如果17世纪不足以使你相信,那么想想19世纪。)我认为:任何确定的原则也不足以排除那些正令人困惑的体系。换言之,我预期,我们将有对新的结论的需要,以掌握那些能与随附性、物理穷尽性以及其它你想要的一切相容的任意多的、特殊的图式。如果是那样,那么铲除关于它们的一种稀奇古怪的虚构似乎只是同一论的脆弱的根据。

6. 作为形式质料说的物理穷尽性

此刻,我想转换思路,更具体地讨论物理穷尽性原则。从根本上说,它依赖于两种“构成样式”:时空合成和集合抽象。它们是从你已有的一切实在中产生新的实在的两种方式;因此对于许多事物的组合来说,你可以形成它们的总合,或形成把它们作为成员包括在内的集合。而且,在赫尔曼和汤普森看来,这些是唯一的、有本体论意义的构成样式。而其他所谓的复合要么并不真实存在着,要么等值于其中的一个(即是说,一个物理标记或集合)。

每个人都有权质问:合成和集合抽象怎样得到这种排他的权利。我猜想,这又回到了古希腊人所钟情的持久性以及变化问题的判决性的探讨。形式(或本体)处于支配地位:永恒的,因而是不可毁灭的。然而普通的事物显然是要消亡的;因此它们都不是纯形式。而消亡就一定有另一因素即质料的介入,它因形式而成为暂时的实在。这样的质料也是持久的,因为它存留着;但是无论是保存还是消失,它都没有可理解的特征。毋宁说,就质料和形式

252 第三部分

结合在一起而言,所有变化都发生在特定的地点——哪种质料有哪种形式,这取决于时间。

哲学家的毛病就是进行泛化。我们知道,除了普通事物,可能还有非质料化的形式(例如鹰头马身有翅怪兽);但是不可能有未被形式规定的质料。因此……每一事物要么是纯形式,要么是某种被形式规定的质料。我认为,赫尔曼和汤普森的物理穷尽性原则正好就是这种披着现代(数学)外衣的、被泛化的形式质料说。集合是无时间的形式的(更广泛、更明确的)替代品;在物理上具有时空的任意部分都是被形式规定的质料的(更广泛和更明确的)替代品。(质量-能量甚至可以代替“原始的”质料。)

对形式质料说的致命的打击是:它几乎没有为划分留下什么空间:作为具体的、个别的、暂时的(偶然的)和质料的东西都搅混在一起。即是说,所有的物质实在都是特殊的、暂时的具体——别的一切都是抽象的、永恒的类别。而且在两个方面之间只有一种可能的关系:内在(例示,集合的全体成员)。最后,构成的样式是非对称性的:你可以沿水平方向前进或向上,但不能向下。即是说,你可以得到任意一束物质个体,并把它们组合成一个新的个体;或者,你能对它们所有的物质性作出抽象——提升,以得到一个纯形式(集合);用现代的话说,你能从任意的集合中抽象出一个“更高”的集合。你没法做到的是:沿着另一方向,使质料由形式构成——当成为永恒的,便永远是永恒的。结果便有一严格的等级,所有暂时的个体实际上一无例外地处于其底层。

试考察一篇小说及其各种“标记”。所写的标记是特定的铭文——暂时的个体,可合理地等同于一系列墨迹。(我这里不管

口头标记、口头叙述；它们只会是使问题更难处理。)但是那小说，那些铭文是它的标记，本身又怎么样呢？如果我们把它称之为“类型”，那么我们不得不解释那意味着什么。在某种意义上，一种小说—类型是由它的标记组成或构成的：它让自己内在于、贯穿于它们之中——没有它们，它就根本不可能存在。但是它是通过合成还是通过集合抽象而形成的呢？我可以说，都不是。

先看合成。小说本身是物理宇宙的一种更大、更零散的虚构——更大的一片墨迹。由此大概可以说，小说类型有部分和卷次，它令人惊奇，但也许并不令人讨厌。更为重要的是，把标记/类型关系降至为部分/整体关系：小说的标记(例如抄本)显然就变成了这种关系的部分。比如说，这就使它们可比作章节，(根据这种观点)章节也就是整体的时空部分。但显而易见的是，章节/小说的关系完全不同于抄本/小说的关系；甚至它们之间没有相似性。这似乎还不够，假如一小说只有一个抄本未佚失，那么类型与它的标记就是同一的——这肯定是不能令人信服的。

因此再看看集合抽象。它使小说—类型成了无时间性的种类，这乍一听好像是合意的。但是稍作思考，我就觉得，断言小说是暂时的、偶然的个体是不可证实的。首先它们在某一时刻开始存在，它们演化，它们变迁，有时它们也消失(可能留给后代)：这并不是无时间性的种类的特性。

其次，小说是作为个体而不是作为亚变种关联于文学作品的明显的种与属。因此在“菲多，长毛垂耳狗，狗，哺乳动物……”这个排列中，我们认为第一个有别于其他的。菲多不像长毛垂耳狗那样是一种狭义类别，而是一种被划分的项目。同样，在“《酸葡

254 第三部分

萄》，传说，寓言，小说……”这个排列中，《酸葡萄》不像其他的项目那样是一更狭窄的类别，而是一种被划分的项目——即是说是一个体。因此把标记/类型降格为例示/种类或分子/集合，就像把标记/类型降格为部分/整体那样，几乎是没有什么前途的，这就是我为什么要说几乎没有可能作出区分的理由。

再看看另一简单的例子。从历史的长河中随便挑出 12 个人来。于是就有一种独一无二的实在，它是他们的合成(或者不管怎么说是他们的身体的合成)；而且还有一种独一无二的无时间性的实在，它是实际上包含着它们的集合。但是假设 12 个人是具有共同业余爱好的朋友，他们打算成立一个俱乐部。在某种意义上，一个俱乐部是由它的成员组成或构成的。但是我视之为自明而不作进一步论证的东西是：俱乐部既不同一于其成员的集合，又不同一于他们的躯体的合成。很显然，要构成新的实在，我们除了需要物理穷尽性原则所允许的两点之外，还需要别的方法。

“很明显”，我不愿作出明确的表态，因为众所周知，集合论对于那些想对其他日常术语作出惊世骇人的解释的人来说，是一种无疆界的游戏场，甚至整数对于没有成见的人来说也是无疆界的游戏场。谁知道什么样的花样翻新的集合(或数字)同一于(在某人看来)《酸葡萄》或《匹兹堡缝被人的三角尺》？谁知道“作为暂时的个体”这类谓词怎样被阐释才能应用于它们？有谁在意呢？如果保护物理穷尽性原则是这种训练课程的首要动机，那么我没有感觉到这种冲动。而且，如果是这样，那么我不得不沉痛地复述我们对贝利的新的推论：“如无必要，毋故弄悬殊！”

注 释

① 赫尔曼和汤普森实际上区分了决定性原则的三种变体;弱随附性类似于他们所说的“真值的决定性”。这是我在本文中关注的唯一的观点,显然他们最令人满意的也是这一点。

② 例如戴维森 1980,金 1978 和 1979。

③ 参阅他们对平行论的评述(1975,561)。

④ 这包括“纯”集合论(以及经典数学,它是由此重构出来的),因为空集合是物理标记集合的一种子集合。

⑤ 例如可参看他们(1975)的最后一句。

⑥ 通常不是一种双条件句——但可能有时如此(梅农?)。

⑦ 这些短语均出自戴维森 1970 或 1980。

⑧ 参阅我论及戴维森(见我的 1982,p. 99)的开头一段;并可与奎因关于“世界的差异”和变化(1978,162)的论述作一比较。

参考文献

G. 赫尔曼和 F. 汤普森(1975):“物理主义:本体论,决定论与还原”,《哲学杂志》,72:551-564。

——(1977):“物理主义的唯物主义”,《心灵》,II:309-345。

D. 戴维森(1970/1980):“心理事件”,见戴维森:《论行动和事件》,牛津,克拉伦登出版社。

金在权(1978):“随附性和法则学上不可公度的量”,《美国哲学季刊》,15:149-56。

——(1979):“心身问题中的因果性、同一性和随附性”,《中西部哲学研究》,4:31-49。

J. 豪格兰德(1982):“海德格尔论人的存在”,《心灵》,16:613-19。

W. 奎因(1978):“关于物质的事实”,《西南部哲学杂志》,9(第二卷):155-169。

上帝心灵中的还原*

丹尼尔·博恩瓦克

伯特兰·罗素在他的逻辑原子论的心脏置入了他所谓的“科学哲学化中的最高原理”：“只要有可能,就要用已知实在中的构造取代关于未知实在的推论”(1924, p. 326)。鲁道夫·卡尔纳普在阐明世界的逻辑结构时把这个原理当成了自己的格言(1967, p. 65)。20 世纪的形而上学家们在探究还原战略时也心照不宣地采纳了罗素的原理。

然而,在名目繁多、与日俱增的领域里,已形成了这样的共识,即这种还原战略面临着暗淡的前景。在神经科学中,我们依然还是新手,但心灵哲学家们对于把心理语言还原为物理语言已不抱太大的希望。物理学迄今为止尚未找到一种关于基本物理现象的统一理论,更不用说找到一种把它们与关于世界的明白无误的印象统一起来的理论。而心灵哲学家们对于把我们关于周围世界的日常话语最终还原为一种理想的物理学语言丧失了信心。

对还原丧失信心并未导致对二元论信念的恢复,因为它假定了一个非物质的心理实在的王国,这些实在具有相互影响以及影响物质世界的因果力量,这似乎是不科学的和缺乏解释力的。对

* 译自 E.E. 萨维洛斯等编:《随附性:新论文集》,剑桥大学出版社 1995。

许多人来说,物理主义依然是关于心身问题的唯一值得敬重的态度。当务之急是找到这样一种非还原的物理主义,它在抵制对非物质实体的任何承诺的同时,又承认心理学和其他更高层次的理论有独立于物理学的自主性。

随附性似乎是一种既能表达物理主义的中心思想又不使自己去做不可能或任何程度上不大可能的事情的方法。然而金在权(1989、1990、1992)近来论证说,非还原的物理主义是不严谨的。他指出:还原的物理主义和二元论之间的两难困境依然充满着强大的生命力。他声称,非还原的物理主义者碰到了突现论者如 S. 亚历山大(1920)、C. 劳埃德·摩根(1923)和 C.D. 布罗德(1925)等人在本世纪初所碰到的同样的难题。他们要受到“上向的决定性”即心理为物理所决定和“下向的因果性”即心理有影响物理的能力的双重约束。前者是随附性的要求;后者则是源于对心理学的自主性的捍卫。而金在权对它们的谐和一致则提出了异议。

在本文中,我将为非还原的物理主义的可能性进行辩护。我认为,随附性和还原之间的关系通常被误解了,并且被完全看成是认识论上的东西。有时被等同于还原的强随附性在认识论的许多重要方面事实上是更弱的。但是从实在论的本体论视角来看,这种不同仅仅是认识论上的不同;强随附性显然具有还原的本体论意蕴。相反,对于一个反实在论者而言,随附性和还原之间的差异则可能具有重大的本体论意义。因此,随附性是否足以为物理主义作出辩护还取决于远远超出心身问题的广泛的形而上学问题。而且下向的因果关系和心理学的自主性是否见容于强随附性则取决于一种关于因果关系和规律的理论。

下面,我描述的是这样一种非还原物理主义,它证明,只要按照某些通常的理解方法来加以解释,上向的决定性和下向的因果性是可以相容的。这里的观点并没有什么新颖之处,它就是非还原的物理主义的一种强随附性变种,我要辩护的正是它的一致性。我的辩护只是在关于本体论承诺和因果关系的某些概念上取得了成功。但是,至少在我看来,这些概念是合理的。否定它们将使人陷入二元论和还原论之间的两难困境,这也许应看作是对它们的支持。

1. 强随附性和还原

随附性通常被理解成属性的种类之间的一种关系;而还原则被理解为理论之间的一种关系。它们的前景是不一样的,但幸运的是,至少根据关于属性的某些理论,它们是可以轻易地予以相互转译的。我将把一种属性描述为一组关于某种语言的开公式:即对象所满足的那些公式,当且仅当对象具有这里所说的那类相应的属性。我还将经常把那个种类等同于一组谓词,开公式正是借助标准的逻辑运算而从这些谓词中产生出来的。这对于我来说是有益无害的,尽管它预先假定:在(有穷)合取、(有穷)析取以及大多数或许有争议的否定词的情况下,那些种类的属性是封闭性的。(参见培根 1986、1990;范·克利夫 1990)。

同心身问题关系最为密切的还原是微观的还原,亦就是 R. 考西(1972a、b, 1977)和胡克(1981)所讨论过的那种还原。贝克尔曼(1992b)运用微观还原论证说,强随附性比这种与心身问题有关

的还原更弱,这在某种程度上是和金在权(1978、1983、1987)、培根(1986)的观点相对立的。然而我感兴趣的是随附性和还原之间的另外一种差异,亦即是在关联于作为解释的更为传统的还原观点时突出地表现出来的那种差异。它对解决本体论争论特别有价值,同时又没有否认:微观还原在心身关系问题上同更为传统的观点相比,是一个更好的候选者。我主要关心的是后者。这还有一个额外的好处,即它可以把演算技术和其他得到恰当理解的理论作为标准的事实。

假定 L 和 L' 是语言,暂时没有函项符号,而 T 和 T' 则是理论,即分别是这些语言中的句子集合,它们在逻辑推论之下是闭合的。假定 D 是 L' 中的一个一元谓词,再假定 h 是从 L 的 n 元谓词到具有 n 个自由变量的 L' 的开公式的一种函项,同时,也是从 L 的常项到 L' 的个体常项的一种函项。这一来, $* = \langle h, D \rangle$ 解释了 L' 中的 L , 当且仅当

$$(1) c^* = h(c)$$

$$(2) v^* = v$$

$$(3) R(t_1 \cdots, t_n)^* = h(R)(t_1^*, \cdots, t_n^*)$$

$$(4) (-A)^* = -A^*$$

$$(5) (A \& B)^* = A^* \& B^*$$

$$(6) (A \vee B)^* = A^* \vee B^*$$

$$(7) (A \rightarrow B)^* = (A^* \rightarrow B^*)$$

$$(8) ((x)A)^* = (x)(D_x \rightarrow A^*)$$

$$(9) ((\exists x)A)^* = (\exists x)(D_x \& A^*)$$

这个解释 $*$ 也说明了 T' 中的 T , 当且仅当它把 T 中的定理映

260 第三部分

射到 T' 中的定理上,也就是说,当且仅当 T^* ,即在 $*$ 之下 T 的映射是 T' 的一个子集。 T 可还原于 T' ,当且仅当在 T' 中存在 T 的一个解释。

一般来说,在 L 和 L' 包含着函项符号的情况下, h 将把 L 中的 n 元函项符号 f 映射到 $n+1$ 的自由变项中的 L' 的开公式 $h(f)$ 上,这样一来, T' 就必然包含: $(x_1) \cdots (x_n)((D(x_1) \& \cdots \& D(x_n)) \rightarrow (\exists y)(A(y) \& (z)(h(f)(x_1, \cdots, x_n, z) \leftrightarrow z = y)))$ 。(因此这个定义不仅将一种语言 L 和另一种语言 L' 联系起来了,而且将它和一种理论 T' 联系起来了。)在这样的情况下,常项应该被看作 0 元函项符号。(3)必须按照函项符号的号数加以变化,以循环地定义原子公式的翻译。假如没有变化,那么 $(R_{x_1}, \cdots, x_n) * = h(R)(x_1, \cdots, x_n)$ 。否则,就得把公式 A 中最右边的函项符号 f 替换成一个新的变项,并在那变项前加一个量词,进而由那函项符号构成的前件与公式 $h(f)$ 相联系,最终就有: $(z)(h(f)(x_1, \cdots, x_n, z) \rightarrow A[f(x_1, \dots, x_n)/z])$ 。

这个还原概念具有这样一些属性,它们使该概念在形式语境中极其有用。如是理解的还原就包含着相对的连贯性,几乎也包含着相对的可判定性和不可判定性。如果 T 还原为 T' ,那么 T' 的连贯性就意味着 T 的连贯性; T' 的前后一贯的可判定性就意味着 T 的前后一贯的外延的可判定性;而 T 的连贯性和必不可少的不可判定性就意味着 T' 的不可判定性。这些属性引起了 20 世纪对可解释性的大多数数学应用;例如,非欧几里德几何学对相对连贯性的证明、歌德尔和科恩对选择原理的连贯性和独立性以及对广义连续统一体假说的证明等。它们也引发了对可解释性的大多

数逻辑研究(例如,塔尔斯基、莫斯托斯基和鲁滨逊 1953)。

然而,我刚才描述的还原概念早在 19 世纪就由于某些不同的原因而产生了。大约 1800 年左右,许多数学领域正处在混乱状态,其中最值得注意的或许就是分析。微商和积分的概念根源于对无穷小的直观讨论。关于连续的函数以及关于函数本身的概念依然是直观性的。级数的收敛还没有得到很好的理解:对无限性的论证常常是混乱不堪的,负数和复数的本质与合法性则聚讼纷纭,而实数的基本属性也未引起重视。分析与代数受到关于特殊定理和基本原理的争论的困扰(参见克莱因 1972)。考奇、狄德金、维尔斯特拉斯、坎托和皮亚诺等人阐发了关于函数、有限、实数和无限的现代概念。在怀特海和罗素的《数学原理》(1910-1913)中,数学的精确性达到了顶点。而更精确的工作则内在地依赖于我所描述的还原概念。更进一步地说,它的结果显然是本体论的。例如,维尔斯特拉斯关于连续性的定义表明:这一观念并不依赖于对无穷小的任何承诺。考奇关于函数的定义表明:它并不是什么神秘的东西,而可以被等同于一类 n - 个元。精确性实际上导致了关于负数和复数的理论向没有这些承诺的理论的还原,加快了这些数在数学应用中的采纳,并使作为它们的属性的这种实践之基础的意见趋于一致。

因此完全可以说,在 19 世纪,还原的主要诱因至少就部分而言是本体论方面的。而在 20 世纪的还原论哲学家中间,本体论的动机无疑已经跃居于主导地位。例如,罗素在他的逻辑主义纲领中使用了这种还原概念,以消除对任何关于数学实在的康德式的直觉的需要;他主张,数学并未承诺逻辑的承诺之外的任何东西。

262 第三部分

这种概念正是逻辑原子论的基础,也是卡尔纳普根据现象学的东西构造物理的东西的基础。

与随附性的比较对我在本文中的目的来说极为重要,因此对还原的本体论意义的更深入的考察就是值得的。按照罗素、卡尔纳普以及后来的哲学家如奎因等的观点,还原具有本体论的力量,因为它表明,某些话语方式是可以排除的。基于这样或那样的理由,对一种令人困惑的理论的承诺是可以避免的,方式就是说明这种理论即使没有同一种承诺也能翻译成另一理论。因此在对世界作出一种完全的描述时,这些令人困惑的对象就不一定要被提及(罗素 1918);没有必要把它们算作是宇宙根本构成因素中的东西。但这并不意味着还原说明了某些对象是不存在的,或者必须被等同于其他某些对象。它表明的不过是,它们可以被看作是不存在的,或者常常被等同于还原理论的领域中的其他某些对象。因此还原使人们可以避免对于作为独立存在的对象的承诺。

乍一看,强随附性和还原似乎完全不同。当且仅当两个有着相同的 K 属性的对象也拥有同样的 J 属性, J 类属性才会强随附于 K 类属性——即使它们可能处于不同的可能世界之中。用有点接近于早期表示还原的术语来说,就是把一类属性等同于来自于一组谓词的一组开公式,或更直截了当地说,把它等同于一种语言。因此当且仅当在语言 L' 中难以分辨的任何两个对象在语言 L 中也难以分辨时,语言 L 才会强随附于语言 L', 即使它们处于不同的可能世界之中。不像下述别的随附性概念,这些概念——比如说,弱的随附性、全体的随附性等——未能从量上限定对象和世界,而这个概念则是强的,足以支持独立性和决定性观念似乎必不

可少的反事实。

然而正如金在权(1978、1983、1987、1989)和培根(1986)所论证的那样,强随附性与还原并不像似乎所是的那样相去甚远。假如属性的类别在有限的合取、析取和否定的情况下是封闭性的——而且我的语言表征又假定它们是这样——那么,如果 L' 有着有限数量的谓词而没有函数符号,这一来强随附性与可还原性就是等同的。鉴于关于 L' 的这些假设,如果 L 强随附于 L' ,那么 L 的每一个 n 元谓词便等同于 L' 的、具有 n 个自由变项的一个开公式。而且这足以使 L 在 L' 中成为可解释的。

为了更直接地说明与可还原性的联系,应予以注意的是,这个定义与更传统的表述是等值的,根据这一公式, T 在 T' 中是可以解释的,当且仅当存在着可能定义或桥梁规律的一种递归集合,即必要的普遍的双条件句,它们把 L 的谓词与 L' 的开公式关联起来,后者与 T' 结合在一起就蕴涵着 T 。如果 L' 是有限的,且缺乏函项符号, L 强随附于 L' ,那么 L 的真值在这一严格意义上便还原为 L' 的真值。

这个结论并未粉碎对以强随附性概念为基础的非还原的物理主义的全部希望,因为任何有趣的语言都有函项符号;物理学的语言肯定也是这样,因为它包含着数学,而且带有如此附加成分的任何东西都包含着函项符号。任何具有从谓词中形成谓词的算子的语言——例如英语——也超出了这一结论的狭隘的规定。因此我们必须在一个更具普遍性的关系中去探求强随附性的特性。

正如金在权所指出的,如果 L 强随附于 L' ,那么 L 的每一个 n 元谓词都等同于 L' 中的具有非限定外延的一个开公式,而 L' 是

264 第三部分

一种具有允许无穷合取与析取的 L' 语汇的语言。一般来说,不可能找到 L' 自身的等值的开公式。这一问题导致了根据可多样实现性或者构成的可塑性所作的论证;详细描述一种既定的心理状态的各种可能的物理实现也许是不可能的。可还原性可衍推出强随附性,而不是相反。

当务之急是对强随附性作出更直接的考察,以弄清它与可还原性究竟具有哪些相同的和不同的方面。用翻译的术语来说,强随附性可衍推出: L 的那一组真句子可翻译成 L' 的具有非限定外延的一组真句子。然而由于翻译不是循环性的,所以翻译的概念在这里被弱化了。没有把握找到一种循环的方法,它能找出一特定 L 句子的翻译,或能决定 L' 的非限定描述中的一个特定语句是不是那种翻译。由此可以说,尽管强随附性仍然可以用来建立相对的前后一致性,但是关于相对的可判定性或不可判定性它却说明不了什么。毫无疑问,关于某些种类的话语,进而关于本体论的承诺的原则上的可排除性,强随附性也是缄默不语,因为从 L 的视角看,想用非限定的 L' 说我们想说的东西也许是不可能的。不管说的是不是很糟,我们要永远继续谈论下去,这是不合适的。我们可能永远无法将我们原本想说的东西翻译出来,因为在一个定义中所使用的那组选言支不可能循环地列举出来。

还是来看心身问题。一个非还原物理主义者可以一以贯之地坚持说,心理语言强随附于物理语言,而无须承认心理事实还原为物理事实。金(1983)断言:“ A 对 B 的强随附性表明了 A 还原为 B 的可能性”(p. 50)。金(1989)重申:“这种强度的随附性可衍推出把随附的东西还原为被随附的东西的可能性”(p. 46)。然而,孤立

地看,这则是错误的。金(1983)还说,“如果还原要提供解释性的理解,那么可还原性将极大地依赖于用 A 对属性 B 中的、作为基础的共外延的可清晰描述性。仅靠随附性还不能证实:一种将提供这种描述的理论已经出现了或将要出现”(p. 50)。要不然,我就会补充说,它甚至有可能出现。强随附性原则上并不蕴涵着可还原性,因为问题是围绕着一个人在实践中能否通过设计合适的翻译来完成还原而出现的。人类认知代理程序能够使用的一种翻译,从根本上说也许是不可能的。贝克尔曼(1992b)的下述断言:“强随附性和反还原主义是不相容的”(p. 97),肯定是错误的。

例如,我们可以考察一下用一种有 0、'(后继符)、+ 和 \times 的语言所进行的一阶演算。不妨将该语言称为 L。而 L' 通过在 L 中增加一个真值谓词扩展了 L。塔尔斯基的定理表明, L' 的真值不能还原为 L 的真值;算术真值是不可能从算术上加以定义的。(假如能的话,那么在算术中就可能出现说谎者悖论。)但是 L' 中的真值却强随附于 L 中的真值。L 的赋值明显地决定了 L' 的赋值。这一结论并不仅仅限于算术:它可以扩展到有办法谈及它自己的句子的任何理论(博恩瓦克 1991)。金的论证表明,对算术真值的非限定定义事实上是可能的,但这没有什么大惊小怪的:该定义仅仅是 $(x)(\text{真}(x) \leftrightarrow x = n_1 \vee \dots \vee x = n_m \vee \dots)$, 在这里 n_1, \dots, n_m, \dots 是算术的真句子中的哥德尔数字。显然,这并不是一个有用的定义,我们没有办法形成这样的定义;集合 $\{n_1, \dots, n_m, \dots\}$ 是不可能递归地予以列举的。因此算术真值随附于算术,但不能从算术上加以定义,因而在原则上就不可能还原为算术。

因此,有一点非常清楚,那就是,金提出的强随附性“为我们提

供了依赖性,但是威胁了自主性”(1983,p. 50)这一结论是错误的。例如,如果人类行动者不可能完成把心理语言还原为物理语言,那么心理语言就是不可排除的,且是自主的。心理语言中的真值可能依赖于作为基础的物理语言中的真值,但是这种联系或许太复杂了,以至于它们不可能为人类所理解。

2. 依赖性和本体论

对于非还原物理主义的强随附性阐释,批评者已提出了两类驳难。首先,根据贝克尔曼(1992a)的复述,格兰姆斯(1988)和金(1990)认为强随附性太弱了,以至于没有抓住依赖性观念的实质。许多作者(如博伊德 1980、福多 1981、莱波尔和洛伊尔 1989、贝克尔曼 1992a、b)都论证说,实现的概念更为恰当:在心理状态由物理状态所实现的意义上,心理状态依赖于物理状态。正如其倡导者所理解的那样,实现比强随附性更强。因此实现的说明就是一种强随附性的说明。就此而言,这种观点对我在本文中的论点并无威胁。我并不反对实现理论。然而,对于金旨在削弱表达了我们关于依赖性的直觉观点的强随附性所使用的某些论证,我的确是不敢苟同的。

格兰姆斯和金实质上都认为,强随附性表达的只是属性的协变,而不是依赖性。例如,两种属性——或者用我们在这里所使用的术语,两种语言——相互强随附于对方,这是可能的。 L 可能强随附于 L' ,这并不是因为 L 依赖于 L' ,而是因为两者都反过来随附于 L' 。相应地,心理强随附于物理就可以与二元论、副现象论

以及其他非物理主义的心灵理论协调一致。贝克尔曼(1992a)在详述金的论点时断言,“如果人们想系统地表述关于非还原物理主义的一种切实可行的观点,那么强随附性概念是没有多大用处的,因为物理主义所要求的比因果的或法则学的依赖性要多得多”(p. 14)。

毫无疑问,强随附性本身并没有完全抓住人们在想到依赖性时心里所想说的“因为”或“由于”的要津。只有在某物因为或由于自己的物理状态才具有心理状态这种意义上,物理主义者才会认为心理依赖于物理。但是这些论证给人的明显印象则是,它们与还原可能是背道而驰的,而且一直如此。这就是说,如果强随附性由于这些原因不能满足物理主义的需要,那么还原也是如此。同强随附性一样,可还原性是反射性的、可转换的。两个理论的相互还原是可能的;事实上,卡尔纳普试图表明的是,物理的东西和现象的东西是可以相互还原的,一个理论也可以凭借与另一个理论一道还原为第三个理论的方式而还原为另一个理论。从物理主义者的观点来看,这大概就是社会学和心理学或者生物学和化学常有的情况。而且可还原性与被还原的理论和还原的理论两者的对象的独立存在是可以相容的。就像罗素曾经小心翼翼地指出的那样,一种还原(或者用他的术语——一种分析)并不能证明任何东西的不存在:“我要澄明的是,我不是在否认任何东西的存在;我仅仅只是拒绝证实它”(1918, p. 273)。例如,罗素有时把没有属类的理论当作这样的迹象,它表明:集合“仅仅只是方便的工具,而不是真正的对象”,是与“当今法国国王”一样的“虚假的抽象”(1910, p. 72; 1906, p. 166)。然而,他用一种更深思熟虑的语气强调说,它允

许我们对属类的存在持不可知论态度,比如说,他用拉普拉斯的话说:“我从来不需要这个假设”(1919,p.184)。

因此,可还原性并不能衍推被还原对象的不存在或者被还原对象与还原对象的同一性;它承认同一,但并不以之为条件。同样,它也没有肯定依赖性或者决定性,因为理论是可以相互还原的。从心理到物理的还原与二元论、副现象论是完全可以相容的。然而,由于心灵独立于物体而存在这一假设不再是必要的,所以这些观点也就失去了意义;它将不具有解释的价值,而且似乎呼吁人们到奥康式的沙龙去做客。

我认为,上述推论也适用于强随附性。强随附性并不能衍推随附对象的不存在,或者与它们随附于其上的对象的同一性。它本身并不能衍推依赖性或决定性。同还原一样,随附性允许依赖性甚或同一性的假设,但并不需要它。尽管如此,同还原一样,随附性可以具有本体论的意义,因为它使独立王国的假定成为多余的。金正确地指出,强随附性本身并没有建立本体论的结论,或者说没有完全抓住我们关于依赖性的直观观念的实质。但是要完成这些工作,必不可少的是得到奥康剃刀的帮助。

这便把我带到第二类反对意见上面了,它改进了第一类反对意见。随附性也许无望衍推本体论的结论。但是它甚至不能与本体论的结论相容吗?正如罗素、卡尔纳普和奎因等人所设想的那样,还原在一定的方面完善了本体论。因为它表明,给它们带来本体论承诺的某些言语方式是可以避免的。如果心理可以还原为物理,那么我们就无须在世界的全部编目表上提及心理,因为在为物理的东西登记编目时,我们已对实在作了完全的描述。我们想用

心理语言说的每一件事只用物理语言就可以说出来,而且经常这样说了。显然,这个论证在强随附性的情况下是不适用的。正如我们所知的那样,强随附性隐含着具有必要的双条件句形式的随附语言与基础语言之间的联系,但这些联系可能具有高度的复杂性;它们也许不能用任何现成的语言或人类事实上能学会的任何语言加以表达。这也就是随附性为何要承认随附王国的自主性的原因。例如,根据我已经为之辩护过的那种非还原物理主义的观点,关于心理状态的语言是自主的、不可避免的,因为那种语言和物理学的语言之间的联系可能是人类无法理解的。

因此我们可以用这种方式表述第二种反对意见:随附性不同于还原,它没有肯定随附语言是可以排除的。因此它也没有本体论的意义。随附性对于物理主义的建立是没有任何作用的。

在博恩瓦克(1988,1991)的两篇文章中,我阐发了旨在克服这个难题的随附性观。随附性自身不能保证随附语言的可排除性。然而从背景语言的角度来看,我们或许能够发现,随附语言所能做的每一件事,基础语言也能作。那就是说,退一步思考就可以发现,随附语言所起的作用可以为基础语言所顶替,尽管前者向后者的直接还原也许是不可能的。假设 T 是语言 L 的真句子,而 T' 是 L' 的真句子,那么我们就可以说, L 在本体论上随附于与背景理论 T'' 有关的 L' ,当且仅当存在着解释 $*$ 和解释 $**$,以至于 $*$ 说明 T 在 T'' 中, $**$ 说明 T' 在 T'' 中,而且 $T*$ 是 $T'*$ 的一个子集。即是说,只有当 L 中的真值和 L' 中的真值都还原为 T'' ,而且 L' 中的还原真值所需要的 T'' 的部分包括了 L 中的还原真值所需要的部分, L 才会在本体论上随附于与背景理论 T'' 有关的 L' 。在这样的条

270 第三部分

件下,不是从 L' 的观点来看,而是从背景理论 T' 的观点来看, L 是可予排除的。

从本体论的观点来看,本体论随附性的最重要的方面就在于,如果 T 在本体论上随附于与 T' 相关的 T' ,那么 T' 在 T' 中的映像的每一个原型就是 T 在 T' 中的映像的一个原型。因此根据这种随附性概念, L' 的事实不仅仅只是决定了 L 的事实; L 的论域也是 L' 的论域的子集。这一来,本体论随附性不仅与本体论的结论是相容的,而且直截了当地说,前者似乎还证明了后者。

本体论的随附性似乎只有微不足道的作用,因为它需要一种背景语言,而且离不开向背景理论的还原,而这种还原像一种理论向另一种理论的还原一样正好是成问题的还原。假如我们对心理学向物理学的还原失望了,那么我们为什么还要期盼两者都还原为使它们两者的联系清楚明白的另一种理论呢?更有甚者,假如我们认为适用于心身问题的严格的背景理论直接就是基本的物理学,那么本体论的随附性与还原相比就没有任何进步,因为当且仅当 L 中的真值还原为 L' 中的真值, L 中的真值才在本体论上随附于 L' 中的、与 L' 的真值有关的真值。

在某些情况下,背景理论和还原都是有用的。例如,具有附加的真值谓词的演算从本体论上说就随附于演算;塔尔斯基或者克里普克或者 G. 赫茨伯格或者孔斯的真值理论正好提供了所需要的背景理论。再来看本体论上一个更有趣的例子,在具有抽象名称 λ 的语言中的真值在本体论上随附于缺乏这种抽象名称的语言的部分中的真值(博恩瓦克,1991)。背景理论再一次变成了语义学理论。内涵的罗素悖论表明,含有抽象名称的语言不会还原为

它的具体部分,但是整个语言和它的具体部分都可以通过解释了它们的真值子句而还原为一个背景理论。在那种语义学理论中,不难说明,全部语言都在本体论上随附于它的具体部分。在这类情形下的本体论随附性并不隐含着还原,但是它的确与还原的一种非限定的普遍原则有关;作为整体的语言的真值可以根据超穷递归而加以定义。

因此本体论随附性的意思是说,如果我们轻而易举地得到了一种正确的背景理论,那么我们就有可能发现,一种语言是可以排除的。试想一下这样的责难:人们可以摒除本体论承诺,唯一的办法就是表明,作为这些承诺的那类的话语是可以摒除的。而强随附性无意于说明任何种类的话语是可以摒除的。因此强随附性不可能摒除本体论的承诺。本体论随附性战胜了上述驳难,因为它表明:从某种背景理论的视角来看,随附的话语不仅是可摒除的;基础理论已经承诺的对象也是如此。

因此,本体论的随附性战胜了这种反对意见,并且可以建立本体论的结论。但不幸的是,尽管它在语义学和其他形式领域内有用,可是几乎无望澄清心身问题,因为背景理论似乎是不可能得到的。我们不可能像在从事语义学和数学工作时那样,仅仅只用集合论。因此我们似乎陷入了两难困境。本体论的随附性可以确立本体论的结论,但似乎不适合建立物理主义。而强随附性似乎有助于建立物理主义,但似乎又没法建立任何本体论的东西。

非还原的物理主义者指出了一条出路:否认第一个前提。摒除使本体论承诺陷入困境的话语就是摒除本体论承诺的一个途径,但它不是唯一的途径。如果非还原物理主义者直截了当地说,

“强随附性是另一种途径”，那么除了特别的申辩，我们再也没有别的办法了。但是没有必要为此而修正本体论的原则。

罗素－卡尔纳普－奎因的本体论观点很显然是语言学上的观点，它曾经是这种二难推理以及 20 世纪大部分哲学的基石。例如，奎因关于承诺的那个著名的标准，除了详述了一个理论所说出的存在着的東西之外，并没有提到存在着什么问题。在考察前面提到的二难推理时，一个理论是什么，这样的问题显得至关重要。不难说明的是，一个理论就是在逻辑推理之下的一组封闭的句子。到目前为止，一切都不错。但是那是什么类型的语言中的句子呢？一种回答是把我們限制在人們使用而且能够使用的那类语言之中。从某些方面说，这似乎是人为的限定，因为人们的记忆和运算能力等有许多限制，这些限制又极其严格地限定了人类可以使用的语言的类型。另一种回答则试图消除这种限制，它把我们限定在那种人类在原则上可以学习和使用的语言中。如果人类的语言能力实质上是计算的，那么这就等于把我们限制在计算机所能学习和使用的语言中了。这种语言的句子一定是有限的；其语法也一定是根据一个有限的基础从一组有限的规则中循环地产生出来的；语义学也一定是计算的；如此等等。按照这种思想，我们可能会说，我们能根据一种本体论的观点考察的任何理论一定是人类和计算机在原则上可以学习和使用的理论。这无异于把我们限制在这样一种语言的、可循环地列举的句子集合之中。

基于这一方案，先前的第一个命题可表述如下：一个人可以作出一种本体论承诺，方法是假定一种语言中存在着可循环地列举的一种理论。这种语言是在有限的基础之上由有限的规则所循环

地产生的。此人可以通过收回那个假定或者表明它是可予剔除的从而剔除这个承诺。在此意义上,这个假定等同于假定另一个可循环地列举的理论或等同于别的没有作出这种承诺的假定。强随附性则没有表明这种类型的可剔除性,它能够表明的是:最初的假定等值于没有任何承诺的关于另一个理论的假定。但这个理论在计算上也许是难以处理的,它不可能循环地予以列举。它的句子在长度上可能是无限的;它可能没有有限的语法或有限的基础。

这个反对意见要求我们得出这样的结论,即强随附性没有本体论的意义。但是似乎可以合理地说,对我们的二难推理所依赖的语言和理论的那些限定是不妥当的。很明显,它们并非不重要。所讨论的语言和理论是否能为人学会,这个问题至关重要。然而,在一定意义上,这是一个认识论的问题。强随附性承认这样的可能性,即随附的语言与基础语言之间的联系如此复杂,以至于不能为人类所理解。这种不可理解的可能性正是使心理学具有比如说自主性的东西。但是从根本上说,它是认识论的问题,即人类能够或不能够说出和理解什么的问题。很明显,这种认识论问题与本体论是没有关系的。任何类型的实在论者都倾向于认为,存在的东西并不依赖于我们能够说出、知道或理解的东西。如果实在论者是对的,那么我们把自己限制在可计算的语言范围内就无异于掩耳盗铃。

实在论者关于本体论的观点纯属一种古老的观点。事实上,实在论者要求我们用上帝的观点,而不是用可能的人类求知者的观点来考虑本体论问题。我们可以继续谈论语言和理论,但是我们应该想到的是上帝的语言和理论。我们自身认知上的局限性不

274 第三部分

会限制上帝;也不会限制存在着的東西。我們感興趣的不是我們自己对世界的概述——我們不得不把它看作世界的根本要素——而是上帝的概述,即上帝不得不如此加以看待的東西。

因此,實在論者沒有這樣的麻煩,即避免這種二難推理或說明強隨附性為什麼具有本體論的意義。強隨附性簡直就是在上帝心靈中的還原。但要記住,強隨附性意味的是,對於隨附語言中的每一個謂詞而言,在基礎語言或它的非限定的外延中,都有一種等值的開公式,它具有同樣數目的自由變項。正是上述子句,即“或它的非限定的外延”把強隨附性和還原區別開來了。從人的觀點來看,這一區分是極其重要的。但從上帝的觀點來看,它是沒有意義的。這裡不會出現可理解性和可表達性的問題。

這一來,實在論者可能會說,還原和強隨附性除了認識論方面之外是完全一樣的。在本體論上,它們同樣表明:承諾是可以避免的。還原與隨附性相比有著明顯的認識論和解釋上的優點,因為那些王國之間的聯系是可以得到系統闡述的。然而,從本體論上來看,它們扮演著同樣的角色。

3. 因果關係

金(1989,1992)論證說,非還原的物理主義者不可能為心理因果關係提供充分的說明。他的論點簡明扼要:心理屬性是真實存在的。所謂真實存在就是有因果力(亞歷山大的名言)。因此,心理屬性有因果力。任何有原因的物理事件都有一個物理的原因,因此心理原因要么無關緊要,要么同一於物理原因。

“假如任何物理事件都有一个物理原因”,金问道,“心理原因又是如何可能的呢?”(1989, p. 44)把物理原因和心理原因等同起来的确最有意义。但是这又要求把物理属性和心理属性等同起来。还原论者可能承认这一点,但是金认为,反还原论者则可能相反。他主张,非还原的物理主义者走进了死胡同。心理原因要么没有什么意义,永远是多余的——不同于还原主义和取消主义,它似乎是难以辩护的——要么就是和物理原因相等同的,从而还原论者终归是正确的。

金自己描述了解决这一两难困境的方法。我们可以认为,宏观事件中的因果关系随附于微观事件中的因果关系。同样,我们也可以认为,心理事件通过随附于由之引起的物理事件而引起物理事件。所以强随附性的支持者可以说,心理原因随附于物理原因。上一节中的实在论者也可以进一步认为,心理原因纯粹就是物理原因,同时又用不着由此而支持还原主义。

金拒绝这种图式,因为它“包含着心理物理的规律”,进而“足以使任何可能的非还原的物理主义者踌躇不前”(1989, p. 46)。戴维森的异态一元论毕竟是当代哲学中对非还原的物理主义的一种最初尝试,而且它强调不可能得到严格的心理物理规律。

然而,实在论者用不着为金的随附性因果关系说明所要求的心理物理规律而惊慌失措,因为它们只不过是上帝心灵中的心理物理规律;它们也许不能用人类能学会的任何语言予以表达。随附性的因果关系就像通常的强随附性一样,离不开把随附层次的属性和基础层次的属性连接起来的必要的双条件式。用我的话来讲,它们离不开把随附的语言中的谓词和基础语言的非限定外延

276 第三部分

中的表达式连接起来的双条件式。表达这种双条件式也许需要一种具有无限长的句子的语言;在其他问题中所涉及到的无限的析取,由于反映了可多样实现性,甚至不可能循环地加以列举。因此这种战略所提出的心理物理规律不可能用任何有限的、循环地产生的语言予以表达。如果我们把规律看成是属性之间的关系,那么可运用于心身问题中的随附的因果关系和通常情况下的强随附性就将产生心理物理规律。然而,如果我们把规律看成是人类能够学会的语言中的句子,那么它们就不会产生这些规律。金(1984,p.172)认识到了这一点,但是并没有严肃地接受关于规律的语言学观点。

如果我们把随附性看成是在上帝心灵中的还原,就像我曾说过的,实在论者可能这样认为,那么我们就可以将心理状态等同于物理状态,而无须相信:我们终究会掌握甚或可能掌握把心理王国和物理王国连接起来的心理物理规律。我们所获得的规律要么果然是局部性的,从属于一种有限制的情境(即某种被很好地定义了选言支的集合),要么是普通的而不是严格的规律。从上帝的观点来看,心理和物理可以通过严格的双条件规律相互联系,但是这些规律是如此复杂,以至于人类无法理解它们。按照这种观点,心理事件可能有真实的因果力量,但是它们是由于随附于物理事件进而与之同一才具有这些因果力量的。

这是否意味着,心理的东西之存在无关紧要,我们最好是成为取消主义者呢?我认为,回答是否定的。我们肯定不能抛弃关于心理的因果关系的话语,尽管上帝可以。心理学是自主的,因为没有心理学我们就做不到这一点。它所发现的联系——常常是局部

的和普通的而非普遍的和严格的——是不能从物理学中引申出来的,或者说不可能根据物理学而加以预言的。心理学将不会被物理学所取代,甚至在原则上也不行。更进一步地说,心理和物理可能以如此不同的方式获得现实性,以至于某些类型的因果联系与其被理解为作为心理事件基础的物理事件中的联系,倒不如被简单地、直接地理解为心理事件中的联系。也就是说,上帝甚至可能会发现:关于心理因果关系的话语是有用的,因为它可能比物理学语言更为有效地把握某些类型的规则。

参考文献

- S. 亚历山大(1920):《时间、空间和上帝》,伦敦,麦克米兰。
- J. 培根(1986):“随附性、必要的共外延和还原性”,《哲学研究》,49, 163-176。
- A. 贝克尔曼(1992a):“导言:还原的和非还原的物理主义”,见 A. 贝克尔曼等编:《突现论还是还原论——对非还原的物理主义的展望》,pp. 11-21,柏林。
- (1992b):“随附性、突现性和还原”,同上书,p. 94-116。
- D. 博恩瓦克(1988):“随附性和本体论”,《美国哲学季刊》,25,37-47。
- (1991):“语义学和随附性”,《综合》,87,331-361。
- R. 博伊德(1980):“没有还原论的唯物主义:物理主义没有衍涵的东西”,见 N. 布洛克编:《心理学哲学读物》,卷 1,pp. 67-106,哈佛大学出版社。
- C.D. 布罗德(1925):《心灵及其在自然中的地位》,伦敦:劳特利奇与基根·保罗。
- R. 卡尔纳普(1967):《世界的逻辑结构》,伯克利:加利福尼亚大学出版社。
- R. 考西(1972a):“属性同一与微观还原”,《哲学杂志》,69,407-422。
- (1972b):“微观还原”,《综合》,25,176-218。
- J. 福多(1981):《表征》,剑桥,马萨诸塞:麻省理工学院出版社。
- T. 格兰姆斯(1988):“随附性的秘密”,《太平洋哲学季刊》,69,152-160。

278 第三部分

C. 胡克(1981):“关于还原的一般理论”,《对话》,20,38-60,201-236,496-529。

金在权(1978):“随附性和法则学的不可公度的量”,《美国哲学季刊》,15,149-156。

——(1983):“随附性和随附因果关系”,《南方哲学杂志》,22 增刊,45-56。

——(1984):“随附性的概念”,《哲学和现象学研究》,45,153-176。

——(1987):“再论‘强’和‘全体’的随附性”,《哲学和现象学研究》,45,153-176。

——(1989):“非还原的唯物主义之谜”,《美国哲学学会会刊》,63,31-47。

——(1990):“作为一个哲学概念的随附性”,《元哲学》,21,1-27。

——(1992):“在突现论和非还原的物理主义中的‘下向’因果关系”,载 A. 贝克尔曼等编:《突现论还是还原?——对非还原的物理主义的展望》,pp. 119-138,柏林。

M. 克莱因(1972):《古今数学思想》,纽约,牛津大学出版社。

E. 莱波尔和 B. 洛伊尔(1989):“再论使心灵更加物质”,《哲学论坛》,17,175-191。

C.L. 摩根(1923):《突现进化论》,伦敦:威廉斯与诺盖特。

B. 罗素(1906):“关于种类和关系的替代理论”,见 D. 莱克编:《分析论》,175-282 纽约:乔治·巴西勒尔,1973。

——(1918):“逻辑原子论的哲学”,见 R. 马什编:《逻辑与知识》,pp. 175-218,引自普特南的著作,纽约,1956。

——(1919):《数理哲学导论》,伦敦:艾伦与尤昂温。

——(1924):“逻辑原子论”,见 R. 马什编:《逻辑与知识》,pp. 321-343,引自普特南的著作,1956。

A. 塔尔斯基、A. 莫斯托斯基和 J. 鲁滨逊(1953):《非确定的理论》,阿姆斯特丹:北荷兰。

范·克利夫(1990):“随附性与封闭句”,《哲学研究》,58,225-238。

A. 怀特海与 B. 罗素(1910-1913):《数学原理》,剑桥大学出版社。

物理主义、随附性与依赖性*

保罗·K. 莫泽 J.D. 特劳特

当代物理主义者一般拒绝关于物理属性和心理属性关系的两种对立的观点:笛卡尔的二元论和还原物理主义。与此同时,他们拥护某种类型的强调心理属性依赖于但不可还原于物理属性的非还原物理主义。本文将对这种非还原物理主义的有效性作出评判。尤其是,它将考察非还原物理主义在诉诸于某些非还原的随附性关系的过程中能否沿着笛卡尔的二元论与还原物理主义中间的光明大道前进。我们将看到:这条道路阴云密布,因为物理主义对有关依赖性概念的说明面临着种种难题。我们认为,特别是物理主义的心理功能主义基于它所必需的、心理属性之发生的相对位置的(relatively local)因果机制,对非还原的心身随附性关系的流行概念提出了一些严峻的问题。下面我们先概述非还原物理主义者试图回避的两种观点:二元论和还原论。

1. 二元论与独立性

笛卡尔主义的二元论断言:在特定的意义上,心理的东西在本体论上独立于物理的东西。笛卡尔主义的实体二元论说的是:心

* 节译自 E.E. 萨维洛斯等编:《随附性:新的论文集》,剑桥大学出版社 1995。

280 第三部分

理实体(如能思的个体)在本体论上不依赖于物理实体或属性。笛卡尔主义的属性二元论说的是:心理属性——能为个体所例示的心理特征——在本体论上不依赖于物理实体或属性。笛卡尔主义者继承亚里士多德的传统,可能对实体和属性作了这样粗略的区分:属性是事物的可断定的属性,可以以多种方式实现,而实体则是个体,它们不能如此被断言或实现。笛卡尔主义的二元论必须阐明的是,与两类二元论的依赖性有关的那种本体论的依赖性。

下述断言并未揭示某物 x 对于另一物 y 的本体论的依赖性,这断言是, x 就其存在来说依赖于 y 。由于 x 对 y 的因果依赖性,考虑到现实的自然律,因此 x 的存在可能依赖于 y 。倘若如此,根据现实的自然律, y 对于 x 的存在来说就是因果上的先决条件。这种因果依赖性是否定心理在本体论上依赖于物理的笛卡尔主义者能够接受的。尤其是,笛卡尔主义者可以承认,在现实的自然律之下,某些身体事件可以是我们的某些心理事件的因果前提。笛卡尔本人明确主张:心理实体必然缺乏物理特征;但人们没有必要认为,心理对于物理的本体论的独立性可衍推出那个观点。在许多哲学家看来,拒绝那种观点有助于推进对心理与物理之间的因果相互作用的说明。

笛卡尔主义者赞成下述心理对物理的独立性:心理实体、属性、状态或事件能离开物理实体、属性、状态或事件而存在,这是可以一以贯之地予以设想的(参阅笛卡尔 1637, p. 101; 1641a, p. 190)。我们可以假定,一以贯之的可设想性正好是无矛盾的可设想性。如果可以一以贯之地设想:某物 x 没有某物 y 也能存在,那么根据笛卡尔的观点, x 在本体论上就独立于 y 。尤其是,因为如

此一贯地设想:思维没有身体也存在,在笛卡尔看来,思维在本体论上就独立于身体。我们可以用概念表述这种观点:就其语义的确定性而言,关于思维的概念并不依赖于关于身体的概念(参阅笛卡尔 1641c, p. 101)。笛卡尔从这种类型的理性观点中引出了有争议的本体论教条,为他的本体论二元论提供了可设想性论证。(参阅 1637, p. 101; 1641b, p. 23)。^①许多哲学家反对笛卡尔根据心与身之间的可一贯设想的区分去推论它们的现实的本体论区分。

基于熟悉的分类学,我们将继续讨论笛卡尔主义者对本体论依赖性的理解,在那里, x 在本体论上依赖于 y , 当且仅当不能一以贯之地设想 x 离开 y 也能存在。根据这个观点, x 在本体论上独立于 y 容许 x 有对 y 的因果依赖性, 即使可以一以贯之地设想: x 离开了 y 也存在, 但仍有这种可能, 即由于现实的自然律, y 是 x 的因果前件。在承认心灵在本体论上独立于身体时, 笛卡尔坚持认为: 心灵“实际上统一于身体”(1641c, p. 102)。二元论的相互作用论者可能认为, 这种“实际上的统一性”涉及到了心与身之间的(也许是极其)广泛的因果作用。

笛卡尔主义者关于心理在本体论上独立于物理的观点对下述观点构成了威胁, 这观点认为: 物理主义据说衍涵着心身同一论, 因而从概念上或分析上说, 是真实的。笛卡尔主义者关于本体论的独立性的观点对下述关于物理主义的解释也提出了挑战, 这些解释认为: 某些心理概念可根据那些概念所认定的有关心理状态的典型的物理原因来加以定义(即使只是部分的)。(关于这样一些物理主义的变种, 可看刘易斯 1966; 阿姆斯特朗 1968, 不过, 刘易斯和阿姆斯特朗认为他们的心身同一论偶然地为真。)^②例如如

282 第三部分

果我们根据疼痛的典型的身体原因定义“疼痛”，那么我们就不得不否认没有身体的疼痛的一以贯之的可设想性。(参阅刘易斯 1980, 他试图解决这个问题。)然而, 支持物理主义偶然为真论的人也可能容许笛卡尔主义者的那种本体论的独立性, 只要他们承认心理现象离开了物理现象也具有一以贯之的可设想性。

任何类型的物理主义者, 不拘是还原的还是非还原的, 都拒绝笛卡尔主义者关于构成的观点, 即笛卡尔主义的实体二元论的下述结论: 人的心灵(被看作是时间性的个体)是由在类别上不同于身体的非物理的实体合成或构成的。同样, 物理主义者还拒绝笛卡尔主义者的这样的观点, 即某些心理属性实际上是由非物理的实在所例示的, 这些实在在空间上是不能延展的。就他们的所有例示了心理属性的现实实在都是物理的这一观点而言, 物理主义者是彻底的一元论者。不过, 他们在关于心理属性的某些命题包括所有心理属性都等值于或同一于物理属性的某种同时发生这样的命题上存在着分歧。非还原的物理主义者(如博伊德 1980; 波斯特 1987)就拒绝那个论点, 而承认心理属性可以在非物理世界中得到例示。这样的物理主义者还包括下面的功能主义者, 他们坚持认为心理属性由于它们有某些因果的或功能的作用而有别于物理属性。根据许多物理主义者的看法, 决定心理属性的因果或功能作用可由我们最好的心理学理论的分类学而得到具体说明。有关的因果或功能作用依赖于这样一些整体的关系条件, 它们独立于那例示了心理属性的东西所构成的条件。这样一来, 心理属性便能在构成上彼此有别的系统中得到多样的实现。

某些哲学家用下述口号表述了物理主义的随附性: “没有物理

上的变化就没有心理上的变化。”^③不过,由于缺乏具体说明什么构成了一种可容许的变化的原则,因而这一论断的重要性尚不明晰。对这个口号,现有多种阐释,但许多物理主义的随附性论者对之则嗤之以鼻。例如即使许多随附性论者承认某些关系属性是真实的,但大多数当代物理主义者仍坚持认为:世界并非自始至终都是关系性的;他们尤其强调:那些物理学的(至少是某些)属性(特别是涉及到物理世界的根本构成的属性)是非关系的,或者说是内在的。(人们可能会说,一对象的属性是非关系的,当且仅当那属性的同一性并不依赖于它与其他属性或个体的关系。)尽管如此,这里所说的物理主义者仍可能主张:至少某些心理学上的分类属性是关系性的。因此物理主义者可能会拒绝这样的主张,即人的身体的内在属性决定了心理状态的类型-属性。

例如试想,在两个不同的可能世界 W_1 和 W_2 中的相同的标记物理状态。^④如果我们假定:物理状态的同一完全取决于它的内在的属性(以及它的部分内在属性),那么同一的标记-状态就有可能偶然地同一于两种类型-独特的心理状态。如果心理状态在关系上是同一的,那么这就是可能的。如果(但也许不仅仅是如果) W_1 和 W_2 有不同的物理规律,那么不同的心理规律就可能存在于 W_1 中的一些状态的承载者和 W_2 中的物理上同一的状态的承载者中。(在 §3 中,我们再来讨论类似的事例)因此同一的物理状态,由于非关系地(从而非意向性地)被个体化了,就有可能隐含在类型-独特的心理过程之中,只要物理状态偶然地与之同一的心理状态有类型-独特的原因。根据这一方案,“没有物理上的变化就没有心理上的变化”就是错误的,只要这种变化涉及的是身体的

非关系的物理属性。

非还原物理主义者反对那些把心理属性还原于物理属性的老生常谈。我们不妨对之作出简要的考察。在考察之后,我们将提出这样的问题,即如果没有对某种类型的还原的承诺,对莫立于随附性关系之上的非还原物理主义的某些解释能否坚持心理对物理的依赖性。

2. 还原:定义的与法则的

赞成对心身关系作随附性说明的物理主义者总是设法避免关于物理主义的种种还原阐释。不过,对于怎样描述还原关系,以及关于还原的其他基本问题,仍是众说纷纭。对于某些人来说,还原的对象是句法上的(内格尔 1961;卡尔纳普 1966;考西 1977)。就另一些人而言,还原的对象是本体论上的(奥本海姆和普特南 1958;胡克 1981)。根据对还原的传统的形式的说明,理论、规律和术语都能成为还原的对象。关于规律和术语的还原由于涉及到法则的和定义的还原,因此是本节特别要关注的。例如如果作为还原目标的规律在逻辑上可从还原领域的规律中引申出来,那么一个规律就可还原于另一规律。

还原是从理论上加以描述的两个实在领域之间的一种关系,不管这假定的对象是属性、过程、状态、事件、规律还是术语。(一种假定的实在不一定是一种现实的实在。)还原的首要目的就是本体论上的统一,至少与本体论的假定有关。还原在科学史上经历了形式的变迁,由于太复杂,以致这里无法用文字加以证明。而且,成功还

原的严格的标准是什么,即使作了持续的澄清努力,迄今仍不清楚。因为涉及到从理论上不同地加以描述的领域的同一是如此难以确定,以致对还原的程度的判断离不开对理论相似性的判断。而关于相似性的有关维度仍是见仁见智。有关的假设标准是本体论的、解释性的和语义学的。也许满足一种以上的条件是必要的。例如除非还原的理论能够替代——不丧失解释力——被还原的理论的所有解释,包括它的本体论的假定,否则我们就不会把该还原看作是成功的还原。即便如此,本体论、解释和意义的相似性应怎样予以评估呢?最常见的方法就是因果方法,即当高阶领域的假定对象为被设定具有同样因果作用的那些对象所替换时,就可认为本体论、解释和意义得到了成功的还原(一种“有保持力”的还原)。

借助有选择地述及对应规则、桥梁规律、起调节作用的定义的同时性陈述,心理学的本体论可关联于物理的东西的本体论。既然同一是一种对称关系,而还原则不是,那么为了说明依赖性的方向,我们一般得具体说明某些非形式的、理论性的假定。还原可能是旧的理论与新理论之间的一种关系(不妨称之为“旧—新还原”),它还可以是高层次理论如心理学与低层次理论如物理学之间的一种关系(这一直被称之为“层次间的还原”,一种自下而上的还原)。我们关心的主要是自下而上的还原。

就层次间或自下而上的还原来说,桥梁规律或同一性陈述即使不是必然地,但也是经常地把高层次的宏观属性与还原理论内的微观属性关联起来了。一般认为,这种关系具有认识论上的意义,因为宏观属性如温度是可观察的,而还原的属性是理论上的,或者说是不可观察的。无论如何,还原就是从平稳到激进、从本体

论上的保持到本体论上的排除的依次变化。热力学向统计力学的还原被认为是平稳的,而借助对生命力和行动的能力的物理说明去还原关于生命精气的理论则是激进的或排除式的。在正常情况下,还原像所涉及到的理论一样是不均衡的。例如炼金术假定黏合剂具有人格性的、投射性的机制,而这些机制已为现代的说明所排除,尽管如此,它也为下述实践和语汇所承认,这些实践和语汇有时竟分离出了现代的分类学如铅所认同的因素。当一种理论的本体论假定被还原或被排除了时,它的规律便不再被遵循。

法则的还原。心理物理规律表达了支持心理状态和物理状态之间存在着多种关系的反事实。在一种成功的法则的还原中,一种必须予以满足的条件就是最低限度的相关:对心理学规律的遵守伴随着对物理学规律的遵守。因此根据法则的还原,不存在着特定的心理学规律这样的例示,它与特定的物理学规律的例示并不同现(co-occur)。

根据对心理物理规律的相关解释,桥梁陈述把心理学规律的例示与物理学规律的例示关联起来了。事实上,桥梁陈述正好是心理物理规律。不过,这种一般类型的相关可以与心理对物理的独立性相谐一致。可以设想:这里所说的心理学规律可以为非物理种类中的关系即下述例示所例示,这些例示与遵守有关物理学规律的现实的物理-对象关系同现。

心理物理还原据说涉及到的正好是邻近的空间区域中的类型-类型同现,在某种意义上,它离不开心理的东西对物理的东西的依赖性。根据许多物理主义者的观点,只有承认这种依赖性,我们才能坚持物理的东西的本体论上的优先性。不难证明的是,有

关类型的依赖性只能在反事实的形式中得到。沿着这条路线,成功的、法则的还原的第二个必要条件就是具有反事实依赖性的条件:反事实支持的心理普遍原则之所以有效,是因为或由于反事实支持的物理普遍原则。

故名思义,法则还原涉及到规律从一个领域向另一个领域的还原——从热力学的规律到统计力学的规律,从牛顿力学到狭义相对论的还原等。因为法则还原的基本要素是规律,高层次理论中的普遍原则一定是高层次的规律,那些规律一定是根据更基本的还原理论中作为规律的普遍原则而加以描述的。因此一种成功的法则还原一定陈述了核心学科中作为分类学的普遍原则中间的适当的对应关系。如此一来,法则还原据说就拉开了还原的自由战略的序幕,根据这种战略,高层次的规律被假定可还原为大量各不相同的低层次规律的析取。即使基础领域中的每一普遍原则是一种规律,而对应的析取则不是。对于应用这种自由战略以吸收多样实现论证优点的、关于身心关系的说明来说,这里就有一个教训,即向析取的还原并非必然是向理论上的分类学范畴或属性的还原。

当高层次规律双条件式地关联于低层次规律时,我们就有一种类型—类型法则还原。即使桥梁规律被认为陈述了两个类型之间的同一性,但还原几乎不是那么平稳。一般来说,高层次(如宏观)规律可以为低层次规律中的一种有限的类别所还原。例如斯涅耳定律*就能用一种关于电磁辐射——光是其中的一个种类

* 在表述光穿过两种互相接触的媒质的界面时,光所穿行的路程以及两种媒质的折射率之间的关系。为荷兰天文学家 W. 斯涅耳于 1621 年所发现。

——的规律和关于衍射的特征的规律予以解释。还原是有方向的,即从高层次的、宏观的对象向那些更基本的对象的还原。因此仅只是高层次和低层次规律之间的相关还不足以再现还原的方向和结果。为了把握因果依赖性的方向,桥梁陈述必须支持反事实。^⑤法则还原的最终前景将在§4中予以评估。

定义的还原。使有关的同一得以实现的一种方式——无须顾及科学的历史——就是使还原成为高层次理论中的词汇与低层次理论中的词汇之间的一种语义关系。这种尝试在科学哲学中以不同的形式表现出来,开始表现为操作主义运动,后来在卡尔纳普(1966)那里又表现为对一种类型的理论定义的规律-集束(law-cluster)说明。假如有定义还原,那么我们用某种高层次(如宏观)谓词所意指的东西就是用低层次谓词所意指的东西。这样一种定义还原源于下述主张,即“温度”意指的正好是“平均的分子运动”。这种定义被称之为惯例、协调定义、桥梁陈述或意义假定,即使它们在这些各不相同的名义下履行的是同一的功能:它们在语义上把还原理论的理论术语与被还原理论的理论术语关联起来了。

对还原的定义探讨方案提出了两个值得注意的、与我们关于心身关系的流行观点有关的难题。第一,因为被还原的术语(如“温度”)在语义上与还原术语(如“平均的分子运动”)所意味的一切是同一的,因此定义还原可衍推出:儿童、前分子运动理论科学家以及新的初学者一定会被认为意指同一记号的另一方面所包含的一切东西。再则,因为定义还原论与其说是本体论理论,倒不如说是语义学理论,因此定义还原论者就不能在下述主张中得到庇护,这主张(正如中枢状态-类型还原论者所坚持的)就是,他们只

是赞成这样的观点:平均分子运动是温度所是的东西,而不是“温度”所意谓的东西。这个结论似乎违背了把意义与指称分离开来的纯属天真的语义学假定。现在的问题直指的正好是用物理术语定义心理术语的任何企图。

源于定义还原的第二个难题——对于我们现在的目的至关重要的难题——就是:这种还原显然也呼唤非还原的物理主义的随附性。因为定义还原把心理学术语与还原的物理学(如生物学)术语所意谓的一切同一起来了,因此它对心理学术语提供了一种沙文主义的解释。它把有非典型原因或基础的心理状态的可能性作为语义学上语无伦次的东西加以排除,进而把(非还原的)物理主义的功能定义所特别允许的那种心理状态的可多样实现性扫地出门。这样一来,定义还原就排除了关于心理现象的典型的非还原随附性理论所承认的那种可多样实现性。^⑥

3. 随附性和依赖性

非还原的物理主义者拒绝心理属性向物理属性的规则还原和定义还原,但又赞成关于例示了心理属性的东西的一元论。一般来说,关于非还原物理主义的常见的变种是从 D. 戴维森的下述论述中受到启发的:

即使我描述的观点否认有心理物理规律,但它与下述看法即心理特性在某种意义上依赖于或随附于物理特性是一致的。这种随附性的意思可能是:在所有物理方面不可能有两个一模一样的事件,除了在某种心理方面有不同,或者说,一个对

290 第三部分

象如果没有某种物理方面的变化,那么它也不可能有某种心理方面的变化。这种依赖性或随附性并不衍涵借助规律或定义的可还原性。(1970,p.215)

戴维森推荐了物理主义的一个品种,在那里,(a)所有事件包括心理事件都是物理事件,(b)心理事件莫名其妙地随附于物理事件,(c)在心理的东西和物理的东西之间不存在定义的或法则的还原(至少是通过严格的规律)。

戴维森有时这样表述他的随附性论点,认为它指的是谓词而不是事件之间的一种关系:

随附性概念,像我经常使用的那样,最好看作是语言中的一个谓词和一组谓词之间的一种关系:如果对于每两个对象而言,一个谓词P适用于一个,而不适用于另一个,而且一组谓词S中有一个谓词,它适用于一个而不适用于另一个,那么一个谓词P就随附于一组谓词S。所有能用随附性谓词区分的个别的实在都能用被随附的谓词加以区分。(1985,p.242)

戴维森诉诸于随附性,目的就是要避免下述有点麻烦的观点,即心理属性就是物理属性,而且存在着把心理属性与物理属性联系起来的严格的规律。(在戴维森看来[1993,p.8],一种“严格的”规律“用不着为防曲解而作的事先解释,也用不着余者皆同这样的从句”。)他希望随附性能使他避免下述还原的观点,即如果心理事件同一于物理事件,那么心理事件的属性就正好是物理属性,至少可还原于物理属性的某种同现(1985,p.243)。

戴维森用谓词对随附性的描述与他先前关于心理特性“在某

种意义上依赖于或随附于物理特性”的言论不尽一致。试看这样一个事例,在那里,对于每一双对象来说,如果一谓词 P 适用于一个而不适用于另一个,那么一组谓词 S 中的一个谓词就适用于一个而不适用于另一个。这种事例并不必然涉及到 P 所认同或表述的属性依赖于 S 中的谓词所认同或表述的属性。P 所认同或表述的属性可能经常以适当的、符合于休谟主义的方式直接与 S 中的谓词所认同或表述的属性一道发生。因此戴维森 1985 年对随附性的说明未能反映他在“心理事件”一文的下述观点:“心理特性在某种意义上依赖于或随附于物理特性。”^⑦现在我们就来考察对适用于心身随附性的那种依赖性的一些探讨。

戴维森所指的那种心理-物理依赖性对适用于物理主义的任何种类的心理-物理随附性来说似乎是必不可少的。D. 刘易斯解释说:

一个随附性命题就是对独立变化的一种否定……说这件事随附于那件事就等于说:如果没有如此这般的变化,那么就不可能有如此那般的变化。例如在同一或不同的世界中,如果两尊雕像在形状或大小或颜色方面没有变化,它们的美也没有变化,那么雕像的美便随附于它们的形状、大小和颜色。(1983,p. 358)

很显然,刘易斯意在使他的下述观点与依赖性命题保持一致,这观点即:“随附性意味着:没有另一种变化就可能没有这一种变化”,同时他还补充说:“没有模态(‘可能’所指的),我们就没有任何令人感兴趣的东西”(1986,p. 15)。

292 第三部分

基于依赖性命题,我们就可以承认:随附性是对心理事件有相对于物理事件的独立变化的否弃。根据这一方案,随附性必定是一种有依赖性的变化的关系。这一描述与金在权的结论即“在随附性关系中最好把协变性要素与依赖性要素区别开来”(1990, p. 16)也是一致的。我们已说过,心理-物理协变并不能衍推心理-物理依赖性。而且有关类型的依赖性是本体性的,而不(仅仅)是认识上的:随附性命题毕竟不(仅仅)是认识上的。

必然的心理-物理协变又怎样呢?假设必然有这样的**事实**,即对于每两个双对象而言,如果一谓词 P 适用于一个而不适用于另一个,那么一组谓词 S 中的一个谓词也只适用于一个而不适用于另一个。这种事例与由 P 认同或表述的属性对由 S 中的谓词认同或表述的属性的那种相关的依赖性有关吗?在那种情形下,就它们被这里所说的对象例示而言,由 P 认同或表述的属性与由 S 中的谓词认同或表述的属性并不是碰巧结合在一起的。如果承认本体论上的不同领域之间,甚至心理与物理之间存在着一种(不可能得到解释的)协变,那么“必然”这一算子就成了这种碰巧的结合的屏障。

如果只是出于论证的考虑,那么不妨让我们假设:必然的协变可衍推出某种依赖性,即便不是那种适用于非还原随附性的依赖性。如果“必然地”意味着概念的必然性或分析性,那么 P 表示的是对 S 中的谓词的一种语义的(如定义的)依赖关系。我们似乎可以这样描述这种依赖性:人们对 P(或某种同义的谓词)的理解离不开他们对 S 中的谓词(或某种同义的谓词)的理解。相比之下,如果得到的只是逻辑上的必然性,那么 P 便在逻辑上依赖于 S 中

的谓词,即是说,P对某物的适用在逻辑上要求:S中的谓词也适用于它。然而,如果得到的是因果必然性,那么P便从因果上依赖于S中的谓词;换言之,至少是在因果上相似于现实世界的世界中,P对某物的适用要求:S中的谓词也适用于那物。此外,人们还可以引进一种特殊类型的形而上学必然性,它是关于某些类型的对象存在于其中的所有世界的;但我们不要扯远了。

非还原的物理主义者并未根据概念或逻辑的必然协变描述适合于随附性的那种依赖性。那些类型的协变对他们的非还原的目的来说也是必要的。一般说来,非还原的随附性的倡导者寻求的主要是这样一种协变,即它既不要求心理谓词对物理谓词的语义上的依赖性,又不要求逻辑上的依赖性。他们寻求的是一种逻辑上偶然的协变,尽管如此,这种协变仍可衍推出心理对物理的依赖性。我们要考察的是:某种必然的协变对非还原的物理主义的目的来说是否是有帮助的。

戴维森和其他随附论者所承认的心理物理依赖性超出了单纯的协变的那一种依赖性,涉及到了物理主义的一种基本信条,即物理的东西具有本体论上的优先性。根据许多物理主义者的看法,优先性至少可衍推出:物理实在偶然是心理属性的被随附的基础。描述有关的被随附基础的一种典型的方法就是诉诸于有依赖性的、逻辑上偶然的协变。根据这一方案,否认这种依赖性就会削弱物理的东西的本体论上的优先性,从而等于拒绝物理主义。我们主张:与心理功能主义者的物理主义有关的那种随附性-依赖性不仅涉及到必然的心理-物理协变,而且也涉及到与适当的(有相对位置的)因果机制有关的反事实的根据。

294 第三部分

事实上,物理主义者怎么可能得到那种有关的心理-物理随附性呢?非还原的物理主义者独出心裁地赞成戴维森关于心理属性对物理属性的定义还原和法则还原的拒斥。不过,他们对心理物理现象之间的所宠爱的随附性关系又有各种各样的分歧看法。^⑧其中一种一般性的观点颇值得注意。倡导非全体性随附性的人强调的是在唯一的可能世界如现实世界中的个体之间的属性与属性的关系(如与心理的东西和物理的东西有关的关系)。相比较而言,支持全体性随附性的人谈论的则是各种可能世界,以确定心理与物理现象之间的有关随附性;因此他们把整个世界作为具有不可分辨性的、被比较的项目。在具体论述中,他们对不同可能世界中的现象作出比较,以找到心理的东西与物理的东西之间的非还原的、不依赖于特殊的属性对属性的关系的随附性联系。^⑨下面我们将考察在非还原的物理主义者中占有突出地位的一种全体随附性。

这里是J.波斯特所提供的全体随附性(GS)的一种有代表性的原则:

(GS) 已知有两个物理上可能的世界 W_1 和 W_2 ,如果在那里找到了相同的物理条件,那么在那里也可找到相同的非物理条件。(1991, p. 118; 试比较波斯特 1987, p. 185)

波斯特倡导(GS),其根据是认为它能涵盖这样的事例,在那里,某物的非物理属性并不是由那物自己的物理属性和物理关系决定的,甚至不是由那物的附近的时空中的物理现象所决定的。(GS)打算承认下述非物理的真理,它们的成立只是基于(因为它

们关于)物理现象的全体性,而不是基于那种全体性的某种特定的部分(试比较波斯特 1987, p. 188)。

(GS)的一个问题(莫泽 1993, 第 5 章对此作了阐述)就是:它未能说明具有非典型的心理学规律的物理可能世界。试想两个物理上可能的世界 W_1 和 W_2 , 在那里(a)有相同的物理条件, (b) W_1 和 W_2 中的物理条件的最近的非物理结果是相同的, 但是(c)同一性在前一世界的最近的非物理结果中消失了。我们姑且把有关的最近的非物理结果称之为“第一阶心理现象”, 把它们非物理结果称之为“第二阶心理现象”。

W_1 和 W_2 中的物理事实是相同的, 并决定着相同的第一阶心理现象。不过, 考虑到 W_1 和 W_2 中的心理因果关系有不同规律, 那么 W_1 和 W_2 中的同一的、第一阶心理现象就不能决定相同的第二阶心理现象。在多种可能方案的一种中, 普遍承认的心理事件在 W_1 而不是在 W_2 中产生了倾向性的或似习惯的信念-状态与意向-状态, 而根据假说, 相对于普遍同意而言, W_2 不具有 W_1 的心理因果关系的规律。仅就想象力的温和的运作而言, 这似乎是可以想象的。被想象的条件在第一阶从心理学上说是统一的, 而在第二阶, 从心理学上说则是不统一的。然而(GS)意味着: 在得到了相同的物理条件的 W_1 和 W_2 的情况下, 这样的不统一性在物理上则是不可能的。

如上所述, 既然对(GS)的否定在概念上是前后一致的, 因此那些倡导者们就必须解释: 我们有什么根据(如果有的话)认为它适用于具有相同物理条件的所有物理上可能的世界。是什么样的根据——经验的或先验的——使我们能够像在(GS)中那样泛化

296 第三部分

到这样的任何两个世界上？此外，什么样的推理模式能像(GS)那样依据这里所说的证据去保证这样一种普遍性的泛化？除了作出了有强制力的回答之外，(GS)是不会成为一种对心理现象的有正当理由的说明的。

这种关于心理不一致性的想象的事例并不必然就是关于现实的物理和非物理条件的事例。为了说明(GS)为假，前述的例子只需要提供两个物理上可能的世界，它们在相同的物理条件下具有心理上的不一致性；因为(GS)关心的是任何两个具有相同物理条件的、物理上可能的世界。具有不一致性的事例似乎是具有相同物理条件的物理上可能的世界中的一个。它似乎并未违背任何(已知的)物理规律。因此在没有相反根据的情况下，我们就可以继续假定它在物理上是可能的。

赞成(GS)的朋友们可能会断言说：在具有相同的物理条件和物理规律的所有物理上可能的世界中，总是可以发现心理因果关系的相同规律的。这个断言其实是说：心理因果关系的规律随附于有关的物理条件和物理规律。不管这个建议与物理主义者是多么的合拍，但它似乎并未解决这里的真正问题。我们没有任何令人信服的理由主张：在所有具有相同物理条件和物理规律的、物理上可能的世界中，总能发现相同的心理因果关系规律。然而，似乎可以合理地假定：前述第二阶心理不一致性的事例与物理上可能的世界有关联。前面的建议无异于臆断。

考虑到不一致性这样的事实，我们就可以引入随附性的一个原则，它要求心理因果关系的规律像物理因果关系规律一样具有同一性。这样一来，我们便与(GS)倡导者们所欣赏的那种随附性

分道扬镳了。物理主义者将对在被随附的基础中囊括心理学规律忧心忡忡。

(GS)能说明戴维森先前所诉诸的那种依赖性或有依赖性的协变吗?在刘易斯(1983)看来,波斯特试图接受的是:“物理主义者在说非物理的变化依赖于物理的变化时,他们所想到的东西”(1987,p.176)。事实上,他对关于“决定性”的言论的赞赏胜于对关于随附性的言论的赞赏,因为后者至少在日常语言中尚未表达“拟议中的依赖性”(1987,p.182;试比较,p.186)。然而一般地承认决定性或随附性离不开有依赖性的变化,还不能证明(GS)提出了一种适用于物理主义的依赖性。

倾向于关于心理属性的功能主义(正如福多 1981,1987,索伯 1985 在图林机和非图林机文本中所阐发的那样)的物理主义者将会看到:(GS)无法提出一种适当种类的心理-物理依赖性。(附带地说,这大概包括大多数物理主义者。)现在我们可以凑合着用一下 N. 布洛克所提出的下述最低限度的论点:

大致地说,大多数功能主义者可能接受这样一种对功能主义的描述:每种类型的心理状态都是由以一定方式行动和拥有某些心理状态的倾向所构成的一种状态。基于某些感性输入和某些心理状态……功能主义者便想从因果上把心理状态个体化;既然心理状态除了有感性原因和行为结果之外还有心理原因和结果,那么功能主义者就按照与其他心理状态的因果关系部分地把心理状态个体化了。(1978,p.262;试比较福多 1981,p.245)

298 第三部分

如果根据这种普通的心理功能主义,心理过程就是标记的因果接续,并由它们在一种标记系统中的因果作用而个体化,那么(GS)就一定保留了心理状态对相应的因果作用的依赖性。事实果真如此吗?

J. 福多找出了心理功能主义的一种限制,值得注意的是,它与(GS)有联系,其实也与心理-物理依赖性的任何原则有联系。他强调,功能主义对某些哲学家来说似乎太容易了,这是因为:“既然功能主义允许借助状态的因果作用而对状态个体化,那么它似乎容许对任何已观察到的事件 E 的一种平常的解释,即是说它似乎假定了 E 的引起者”(1981, p. 247)。问题在于,这样一种关于 E 的引起者的言论可能是空洞的,否则在解释上就是神秘的,因而不符合心理学解释的要求。福多呼吁功能主义者应避免这个问题,“方法就是承认从功能上定义的理论构造物,只有在那里,才有完成那种功能的机制,只有在那里,他才有关于这种机制可能类似于的东西的概念”(1981, p. 247)。根据这种似乎合理的建议,对心理状态的功能描述离不开对适合于产生这种状态的因果机制的说明。如果没有这种说明,那么心理功能主义提供的就是因果神秘性。在这里,有关的关系就存在于(可能是非时间性的)对象的微观结构和它的宏观结构之间。如果关于“产生”的言论不可避免地意味着作用的超时间的因果传递,那么所能看到的就是“实现”而不是“产生”;因为 X 在 Y 中实现就是 X 和 Y 之间有一种同现的关系。而且有关类型的因果机制本来就是有相对位置的。它们至少比整个世界甚或物理上可能的世界中的物理现象的总体性更具空间性。(注意:承认因果机制的这种相对的位置并不能衍推出心

理个体主义。)

例如试考察与你用理解力阅读这篇论文有关的思维心理过程(包括它们所例示的心理属性)。这其中有你分辨和记住该论文的许多主张的真值条件。根据心理功能主义,你由此例示了一系列的因果属性,在那里,各种因果关系存在于你所处的功能系统的标记之间。在这种涉及到因果上复杂的心理过程的事例中,你的功能系统的标记中的因果关系相应地说也一定是复杂的。根据心理功能主义,这里所说的心理过程其实是通过这里所说的因果关系而个体化的。如果接受福多的建议,我们就一定得承认:有关的因果关系依赖于相应的(有相对位置的)因果机制,而这些机制足以从因果上产生上述复杂的因果关系。尤其是,因果上复杂的心理过程离不开标记中的相应的复杂因果关系,这种关系进而又依赖于足以产生这种因果关系的因果机制。

其标记只有最低限度因果力的因果系统将只会保持这种最低限度的因果关系,而不会享有因果上复杂的心理过程。事实上,根据与福多关于适当的因果机制的必然性的主张相一致的、关于心理功能主义观点的任何解释,这是一个因果上必然的真理。例如仅由两个铜便士所组成的物理系统,对于产生思考这篇论文的心理过程所必需的(根据心理功能主义)心理上复杂的因果关系来说,在因果上就是远远不够的。前一心理过程在因果上依赖于一种对产生思维所必需的那种复杂的因果关系有足够力量的机制(或一组机制)。没有这种类型的因果机制就不可能出现这样一种因果上有效的事态,正是它表现了适合于思维的因果关系。根据心理功能主义,如果没有对这种关系有适当力量的(有相对位置

300 第三部分

的)因果机制,那么那种对思维至关重要的因果关系,即使不是绝对不可理喻的,也会是完全神秘的。这是福多提出关于适当因果机制建议的一个关键动机。

(GS)并不适用于福多的建议;因为它用不着让复杂的心理过程在因果上依赖于有适当力量的(有相对位置的)因果机制。就(GS)而言,一种物理上可能的世界,由于它在因果相互作用中除了包含两个以上的铜币以外,还在物理上与现实世界有类型上的同一,因而才能产生一种心理过程,而正是在此过程中,它的铜币才得以思考这篇论文。(GS)承认这一点,因为它不想让心理过程依赖于有关标记所例示的、有适当力量的(有相对位置的)因果机制,而这些标记就是在可供选择的、物理上可能的世界中的铜币之类。它只是提出了一种关于具有相同物理条件的物理上可能的世界的主张,认为同样的非物理条件也能在它们中找到。具有两种额外的铜币的、物理上可能的世界并不具有与现实世界一样的物理条件;因此(GS)并不否认在那个选择出来的世界中的铜币会思考这篇论文。如此一来,(GS)就没有包含这样一种依赖性,它适用于心理功能主义尤其是符合福多关于适当的因果机制的建议的那种心理功能主义。^⑩

福多的建议与心理属性之多样实现是一致的。它不要求:心理机制由任何特定类型的材料所构成,例如不同于以硅为基础的材料、以碳为基础的材料。就此而言,福多的建议与心理功能主义的关键动机是吻合的。全体随附性的原则把世界比作整体,从而否定对那些世界中的、具有相对位置的因果机制的非难。在这点上,各种有相对位置的因果机制对于心理功能因果关系来说尽

管是至关重要的,但全体随附性对这些机制来说则是不够敏感的。有待证明的是,是否存在着具有两个额外铜币的物理上可能的世界这样的事例。^⑪

因此我们需要一种关于随附性-依赖性的说明,它能揭示心理属性对适当的因果机制的有关的依赖性。尤其是,我们必须回答:非还原物理主义是否有望提供这种说明。如果非还原物理主义与戴维森的前述观点相一致,真的避免了规则还原和定义还原,那么它就必须为避免了两种还原的心理-物理随附性提供说明。

非还原物理主义的前景看起来很暗淡,因为对依赖性的通常理解引出了一个严峻的问题。那种相关于物理主义包括心理功能主义者的物理主义的心理-物理依赖性支持着反事实和虚拟条件句,尤其是与适当的、根本的因果机制有关的那些条件句。例如根据物理主义和心理功能主义,如果一个人对这篇论文的思考依赖于他有一种物理的因果机制,这种机制在功能上类似于大脑,那么在他缺乏这种机制的情况下,他就不能思考这篇论文。除分析功能主义之外,前述条件句的条件后项并不是一种分析上为真的命题,而是其真理为因果上必然的命题,即假如有现实的自然律,其真理就是必然的。在这里,对反事实和虚拟条件句的相关支持本来就是合乎法则的,即使在某些情况下详细说明相应的规律是很困难的。^⑫

否认心理-物理依赖性是合乎法则的,只会带来神秘性这样的后果。在心理属性和物理属性之间缺乏法则的联系时,关于心理依赖于物理的断定究竟涉及到什么,这是完全不清楚的。关键不在于:缺乏法则的联系就妨碍对有依赖性的心理-物理变化作

302 第三部分

出有说服力的解释(即使那一点是相当合理的);而在于:法则的联系这一概念对理解有依赖性的心理-物理协变是至关重要的。关于依赖性的贴切的话语本身就是合乎法则的,因为它支持反事实和虚拟条件句。我们注意到,这样的话语不是关于同时发生的协变的话语(与前面说过的、戴维森 1985 对随附性的描述的含义恰恰相反)。法则的联系使随附性不止是单纯的休谟主义协变;而且与适当的因果机制有关的法则联系,使心理-物理的随附性-依赖性——与物理主义的心理功能主义有关的那种随附性——不只是全体随附性或单纯的必然协变。

人们可能会寻找一种绕过法则联系和适当的因果机制的出路,方法是主张随附性-依赖性自成一类,且不可解释。然而,这对物理主义者来说是一种稀奇古怪的观点,因为对非还原随附性-物理主义的理解就需要对随附性-依赖性作出说明。我们可以把随附性-依赖性当作不可解释的,唯一的代价就是让非还原的物理主义本身成为不可解释的。如果物理主义者要责难笛卡尔主义的二元论增加了神秘性而不是推进了解释,那么他们本身就应提供解释而不是提供神秘性,尤其是当他们关于心理-物理的有依赖性的变化的理论碰到这种情况时更是如此。从根本上来说,物理主义毕竟是一种解释理论;关于同一因果机制的不同解释性假设的趋同,对限制物理主义者的解释并使之在科学上有效的那种综合性和统一性来说,是决定性的。物理主义所承认的因果神秘性的包袱只会妨碍物理主义的重要的、解释方面的权利;因为因果神秘性无法替代对解释性的因果机制的说明。下面我们简要地考察一下关于依赖性的某些有关的讨论。

4. 对依赖性的探讨

我们已说过,非还原的物理主义者在尝试形成关于既享有物理的优先性又享有心理的自主性的依赖性概念时,一定是踏上了叛逆的征程。随附性论者在描述心身关系时,可能利用的依赖性名目繁多。一是决定论的依赖性;另一种是概率论的依赖性,后者至少提出了两种解释——频率和倾向解释。下面我们就来讨论这些方案。

决定论的依赖性。某些规律是决定论的规律,如牛顿的系统中的那些规律。在这些系统中,蕴含着某种过程、状态或事件的普遍原则要表明的是:因果因素 C 是否完全决定了某种因素 E 的出现(或不出现)。不应认为,出现和不出现作为二分法变量这一事实意味着,决定论规律是严格的(参阅亨普尔 1988;试比较金 1993)。关于决定论现象的理论上的普通原则包含有含蓄的假定或“附带条件”,它们具体说明了普遍原则存在于其中的参数。例如磁的相互作用的规律,如亨普尔(1988)所强调的,假定的一般是:有关对象是不受相互吸引的影响的。这些普遍原则的最重要特征并不在于它们是严格的,而在于它们(严格或相反)支持着某种反事实。

关于心身关系的决定论陈述通常表现为这样的形式:

在物理状态 p 的特定集合中,一分子之出现完全决定了心理状态 m 的出现。

304 第三部分

当只考虑关于一个个体 X 的规律时,这种决定论的心理-物理陈述就支持下述形式的反事实:

如果 X 不是物理状态 p 的特定集中的一分子,那么 X 就不在心理状态 m 中。

我们觉得:根据关于心理功能主义的任何合理的物理主义阐释,有关的反事实都与心理现象的适当的因果机制相关联。当心理-物理陈述支持反事实时,在前件和后件之间的关系中所表达的依赖性关系就可以借助物理和心理这两个领域之间的有规律的关系(至少原则上)来加以解释。因此,至少是有适当保留的规则还原充满着这种前景。相应地,金在权注意到:“一当我们谈论特定的心理属性和物理属性之间的相互关系和依赖性时,我们事实上是在谈论心理物理规律,这些规律引出了令人讨厌的物理还原主义幽灵”(1989, p. 42)。因此在反事实依赖性和规则的还原的可能性之间存在着一种直接的关系。即使这种关系在决定论的规律的情形之下尤为明显,但它也出现于随机规律的事例中。

概率依赖性:频率解释。人们可以借助与概率论或统计学的独立性所不能把握的条件的类比,去说明有关的依赖性概念。根据这种概率论的规则,如果两个事件(或过程、属性或状态)的一个(之出现)在概率上的变化对另一事件的概率没有任何影响,不拘是正面的或负面的,那么这两个事件就是相互独立的。更形式化地说,事件 A 独立于事件 B , 如果 $p(A/B) = p(A/B') = p(A)$ 。这种独立性关系是对称的:如果 A 独立于 B , 那么 B 也独立于 A 。

作为随附的心理状态和被随附的物理状态之间的关系,纯概率论的依赖性并未允准物理的东西在本体论上的优先性,至少是

物理主义的心理功能主义所承认的那种优先性。可以设想,处在某种物理状态中会影响某人处在一种特定心理状态中的概率,但这与纯休谟主义的恒常的同时发生也是可以调和的。这种同时发生并不等于本体的依赖性。

在把概率等同于长期频率时,频率解释就会面临两个一般性的困难。第一,它本身不能提供把因果关系与纯粹的(休谟主义的)相互关系区分开来的相关性标准。第二,它不能确定——甚至澄清——所谓的单例的概率。某些事件是独一无二的,如埃利维·普雷斯利之死*,1992年女子100米短跑中的奥林匹克胜利,或第20届美国总统竞选。然而无法弄清的是,这些事件的概率可以由这些事件在那样长的时期以之发生的频率所授予;甚至它意味着什么也说不清楚。(关于有关问题,可参阅厄尔曼与萨蒙1992, pp. 77-81)近年来,倾向解释应运而生,为人们带来了某种慰藉。

概率依赖性:倾向解释。基于单例问题以及关于概率因果关系的广泛接受的观点,一种倾向解释逐渐形成了,根据这个解释,概率就是产生长期频率的趋向。因为频率分布的根源就是倾向或趋向,因此概率论的倾向谓词像其他倾向谓词一样支持着某种反事实和虚拟条件句。这是频率解释和倾向解释之间的一个重要差别。倾向是理论上的,而频率则不是。^⑬

层次间的相互依赖性这一相应的概念一定获得了这样的意义,在此意义下,心理的东西在反事实上相关于物理的东西。在个

* 普雷斯利(1935-1977),美国摇滚乐歌唱家,对改变美国人民的文化面貌起过巨大作用。死后成千上万的人吊唁他。

306 第三部分

体 X 的生命历程中,我们可以设想:

如果 X 不在物理状态 p 的特定集合的一员中,那么 X 不在心理状态 m 中的概率就会超过(比如说).96。

那个概率自然高于或低于.96;它还可能相对于情境或世界的规则特征的变化而变化。某些世界自然在规则上不同于现实世界,因此相应地支持不同的概率。主要之点在于:处于物理状态 p 的集合的一员中就会提高处于心理状态 m 中的概率。 p 的物理特性有一种产生 m 的倾向或趋向。这一论断用不着从时间上去解释。 p 的状态可能与 m 同时发生。有关语言常常用来把握依赖性关系,而不适于把握作用传递中消逝的时间。人们可能会发觉:“产生”的非时间用法是不协调的;因为根据人们关于“产生”的理解,X 不可能产生 m ,除非 X 先于 m 。然而我们试图把握的关系则是对象的微观结构和它的宏观结构之间的(可能是非时间的)关系。微观状态的物理特征可能授予它的(宏观状态)功能实现以因果力。正是在此意义上, m 在因果上依赖于 X。(如果像我们前面提及的那样,产生这一词汇蕴涵着作用在时间中的传递,那么就要用“实现”替换“产生”。)

关于随机条件句,有两点值得评述。第一,随附性论者不应满足于对有关概率的纯认识上的解释;因为如前面所强调的,随附性-依赖性是一种本体关系,而非一种纯认识关系出现的。第二,即使有关的条件句以一定的概率与它们的后件发生联系,但每一种这样的条件句本身都可能有一定程度的确定性或不确定性。

试图包容心理与物理之间的依赖性变化的非还原物理主义者

现在一定得考虑到:所有关于依赖性的相关的探讨至少面临着规则还原的威胁。纯统计的或频率的依赖性对心理-物理依赖性来说是一种不相干的探讨方案,尽管它与心理物理的纯休谟主义的相互关系是一致的。任何关于心理-物理依赖性的相关探讨,不管是决定论的程序或是随机的程序,都要用到规律的概念,并具有反事实和虚拟条件的支撑;尤其是,如果我们承认物理主义的心理功能主义与福多的主张协调一致,那么我们就不得不假定;存在着可衍推出反事实和虚拟条件支持的适当的因果机制。一种规则的还原是否能在心理和物理规律之间实在地实现,这取决于有关的心理规律所依赖的物理规律的多少和种类。

现在,我们不能满足于“倾向”这一未得到阐明的概念,倾向是宇宙的这样一种特征,物理主义者一般没有把它算作宇宙的基本单元。对现在的目的显得更为重要的是,对心身依赖性的倾向说明使我们承诺了理论实在(倾向)的存在,而且还存在着关于这些实在的(心理物理)规律——此即支持着分类学的普遍原则的反事实。要涵盖心理对物理的依赖性,物理主义者就必须严肃地正视规则还原的前途。非还原的物理主义不符合这一训诫;而且只有当这种物理主义企图与符合福多关于适当的因果机制的建议的那种心理功能主义保持一致时,困难才会增加。

5. 构成性与依赖性

既然本体依赖性概念是用于解释物理的东西的优先性的,因此对非还原物理主义作出一种阐释对它从关于依赖性的话语中解

308 第三部分

放出来就是值得的。这种阐释赞成的是关于物质构成的概念。我们将予以简要的考察,以弄清对有依赖性的心理-物理变化的担忧实际上是否击中了要害。

根据 R. 博伊德的构成唯物主义:

唯物主义者应断言的是:心理状态事实上是中枢神经系统的状态,但是它们有中枢神经系统对它们来说并非是必要的。这一说明就像这里被辩护的说明,即心理状态偶然地同一于物理状态。(1980,p.105)

构成唯物主义最独特的特征就是:它反对把某一种类的标记-同一当作是唯物主义的一种过强的变种。根据这种观点,物理的(从而对物理主义者来说的心理的)事件一般来说并不同一于它们的更小的构成特征。在构成唯物主义看来,即使是在单个的物理标记之内也存在着可塑性(或多样实现性),正如在易于为不同物理标记例示的一个类型之内有可塑性一样。

博伊德从我们关于物理对象同一的日常标准中寻找根据:

如果一卡车的发动机(组成部分之一)被替换了,那么该卡车在现实世界还是同一个卡车,因此闹不明白的是:同样的可塑性在诸可能世界为什么就不存在。(1980,p.100)

在博伊德看来,同一个卡车的事故是可能发生的,即使所牵涉到的卡车有不同的发动机。根据这种观点,两个可能世界中的事故的分子构成物中的差异并不妨碍事故的相同性。

如果引进关于同一性的跨世界标准不利于表征不同标记的例示这一目的,那么那卡车出现在不同可能世界中的功能就是不清

楚的;这种看法可能也适用于那卡车所经历的两个不同的阶段。因此,在论证同一标记可以在不同可能世界以多种形式加以构成时,我们就必须明确分类学。如果关于某种标记的分类问题是:作为卡车,它所具有的特定的构成物对它的同一性(作为一个卡车)是否是必不可少的,那么构成唯物主义的观点就好理解了。然而构成唯物主义的宗旨是要阐明这样的看法,据此,心理事件、状态和过程可以认为有一种(也许是偶然的)物理基础。作为物理项目,在不同可能世界被例示的标记卡车是不同的标记,因为它们有不同的物理构成物。

.....

构成唯物主义在恰如其分地强调构成性时,应该使人们认识到:构成上的差异蕴涵着标记上的差异。因此这种唯物主义承认:构成性决定了严格的标记-同一性。根据唯物主义,对构成差异的最好的说明方式就是求助于标记概念,在那里,物理构成的差异造成了物理标记的差异。一种可供选择的观点会冒险阻止物理主义者把构成上的差异解释为物理事物中的差异,在这里,事物是标记而不是类型。物理主义者必须说明,构成上的差异就是物理事物中的差异。如果有关的构成差异并不直接就是类型-差异,那么它们就是标记-差异,即是说是物理标记中的差异。物理的差异不是类型-差异,就是标记-差异,因此当它们不是类型-差异时,它们就是标记-差异。我们关于标记的探讨方案由于对物理构成中的变化敏感,因此能使物理主义者轻而易举地解除这些顾虑。它也允许相同的标记存在于那些可供选择的世界中,只要那标记保持相同的构成物;而且它还允许同一标记在时间中存在,只

310 第三部分

要构成的同一性能持续下去。

除非我们认为：物理事物中的变化伴随物理构成中的任何变化，否则我们会面对这样的观点，即世界的基本物理构成要素本身具有多种多样的例证，至少在可能世界之间是这样。根据这种观点，只有似属性的、可多样例示的实在才能成为物理宇宙中的基本本体论的存货；因此物理宇宙完完全全、彻头彻尾是似属性的。相比较而言，我们关于标记性质的假设概念允许我们重视物理主义者广泛欢迎的下述学说，即物理宇宙的基本构成要素就是物理事物；它是通过把构成物中的任何物理变化当作物理事物中的变化之形成而实现这一点的。

.....

博伊德的唯物主义的显著特征就是它关于各种标记—状态由一组分子运动所实现的言论。如果“被实现”等于“被例示”，那么博伊德的“标记”看起来就很像类型。不过，如果类型是有争议的，那么关于物理构成的言论就是张冠李戴。时空上的具体可以由物理的要素构成或组成，而类型由于是可多样实现的宇宙（具有不同抽象程度的变量），因而严格地说不能看作是由物理要素构成的。

博伊德这样说明他关于转向似类型的“标记”的某些看法：

这些思考带来了这样的结果，即让标记事件、状态和过程不太像老生常谈的“个体”，而更像类型事件、状态和过程——更像“宇宙”——因为例如一种标记事件可能不止一个例证（即使不同的可能世界），它没有必要同于一其中的任何一个。如果不是因为这样的事实，即对时间中的个体（例如物理事物和人）之再认同的思考表明：个体不太像关于个体的方便的哲学

成见那样,那么这就是令人忧虑的。(1980,p. 102)

因此博伊德似乎认为:日常用法把关于个体的人的概念当作是一个标记,不过该标记的同一性允许构成上的变化。

即使根据博伊德的说明,诉诸于“个体”这一术语的日常用法也还不是决定性的。如果是这样,那么在日常心理学的语境或别的情况下,就有该术语的典型的应用,这些应用要求那些标记是个人的时间上的阶段,从而要求个体的人是一种类型而不是一个标记。当我们因一个人的犯罪行为而赞扬、谴责或惩罚他时,一般来说,我们关于责任的评价是以那个人现在的状态与他过去的状态在心理上的相近为转移的,他的犯罪是由他过去的状态一手造成的。这就是罪犯为什么常常通过悔罪(或弃旧),通过转化犯罪特征而得到改造的原因。只有当有关的标记是一个人在时间中表现出的状态时,这样一种关于人能发生根本变化的观点——也许涉及到同一具体的人先前和后来阶段之间的巨大差异,就像不同的具体的人同时的状态之间的常见差异那样——才能得到理论上的说明。如果我们把这些时间上的阶段当作同一标记——人的构成上的不同阶段,那么我们就将得到两个特殊的观点,(a)人作为人的标记在时间中具有明显不同的、与它的类型——同一性无关的内在因果力,(b)内在因果力中的显著变化与标记构成物中的变化之间没有必然的联系。

我们的关键假定是:内在因果力中的有关变化可衍推出标记中的变化,即使关于“个体”的某种日常概念违背了该假定。根据关于物理主义的、解释上充分的任何阐释,内在的因果力中的有关变化与有那种因果力的一切东西的内在的物理特征中的变化有关

312 第三部分

联;否认这种假定就会使内在的因果力神秘莫测;尤其是使在潜在结果中产生变化的东西神秘莫测。关于个体的概念与这种对标记的限制——包括作为博伊德那种意义上的“标记”的人格个体的概念——是有冲突的,因此未能对符合于物理主义要求的构成要素提供精致的说明;尤其是,它们是从物理分类学注定对之敏感的、作为基础的局部因果机制中抽象出来的。这就回复到了我们在第3节末尾所阐述的观点,它论述了物理主义解释的本质,即它对因果神秘性的本能的不可宽容性。

为了解决关于标记性质的争论,诉诸于科学心理学的分类学是有益于物理主义的心理学家解释目的的。但同时也留下了关于产生心理现象的被随附的物理因果因素的无法回答的问题。后一类问题在物理主义对心身问题探讨的全部历史中占据着核心地位,至少在从霍布斯到当代的理论家中是如此。回避关于作为基础的物理因果机制的传统争论,标志着对典型的物理主义方法论的显著背离,这是一种需要专门论证的背叛。这种对作为基础的因果机制的解释重要性的强调并不能衍推还原论,就像强调遗传的发生基础对于进化生物学的解释作用无论如何不能衍推还原论一样。事实上,正是由于进化生物学和遗传学关注实现随附属性的物理因果机制,才使一种解释上的选择优于活力论说明。关于因果机制解释上的重要性的直接类似的观点也适用于物理科学中各种相似的进展。

关于科学的分类学,我们应注意到:心理学解释的某些目的要求我们把单个人的不同时间阶段当作不同的标记。概念获得和知觉学习这类现象应按照儿童现有的信息储存加上它与它的社会

及自然环境的偶然特征的相互作用去加以解释。如果我们把这个儿童当作一种标记,那么我们就不得不承认:尽管该“个体”经历了显著的物理和心理变化——这变化能轻易地影响该个体关于各种特征标度的分类维度的观点——但是这个人仍是一个标记,从不能自制到概念的和运动的复杂性,最后再到不能自制。

因此,博伊德赋予标记的构成上的可塑性实际上是在同一类型的标记中存在的自然变化。例如可以证明:任何事实科学在它独有的范围内并不只有一种类型——一种自然的种类。因此我们关于时间中的人的心理学是一种类型的看法就不一定使我们承认这样的观点,即作为一门关于心理学种类的科学的心理学只能对个体的人作出抽象。在心理学的类型(从而在个人)中,可能存在着允许抽象的相似性,就像在生物学的类型中存在的相似性一样。类似的观点也适用于博伊德关于卡车的例子。

人们可能会建议说:构成关系恰好是一种依赖关系,因此关于心理-物理依赖性,物理主义者完全可以说,每个心理事件在所有法则学上有关的世界中完全是由物理事件构成的。这个建议的主要问题在于它的解释的狭隘性,尤其是当联想到福多关于适当因果机制的详细说明时更是如此。倾向于心理功能主义的物理主义者不仅需要这样的论点,即所有事件都是从物理上构成的,而且还需要这样的解释,即解释心理属性根据什么依赖于或怎样依赖于物理属性——尤其是心理属性之发生怎样依赖于适当的物理因果机制。如果没有这种解释,那么我们对心理与物理属性之间的关系采取过分随意的态度就太冒风险了;这已为前面关于两个铜币的事例所说明。

314 第三部分

在跨世界—同一“标记”中,博伊德的观点所承认的构成上的差异使构成唯物主义不适合成为一种对心理—物理依赖性的说明。根据博伊德的看法,作为“标记”的卡车尽管可以在不同可能世界以不同方式被构成,但严格地说,作为物理具体的、不同地加以构成的标记实际上是不同的标记。如果像我们设想的那样,博伊德错误地把类型例示公认的多样性等同于相同“标记”的跨世界变体的构成可塑性,那么博伊德所描述的现象就真的是类型,本身就不是由物理的东西所构成的具体。严格说来,只有标记才有物理的构成。因此如果构成唯物主义要成为符合于物理主义的、关于构成的理论,那么它必须要么放弃物理的东西的优先性,要么放弃关于状态、事件和过程所固守的分类学。

构成唯物主义反对定义还原和规则还原,但是又没有为心理类型与实现这些类型的物理实在之间的依赖关系提供可供选择的说明。而没有这种对心理—物理依赖性的说明,它就必然要为对物理主义的心理功能主义至关重要的心理物理机制提供神秘的说明。因此对于非还原的物理主义来说,构成唯物主义并未在传统的标记—同一论之上提供什么有价值的东西,而且还为关于标记状态的一个转换概念付出了代价。

我们对特定因果机制的要求可衍推出类型—同一性即心理类型与物理类型的同一吗?我们的要求与标记—同一自然是一致的。而且,我们认为:关于物理主义的任何恰到好处的阐释一定包含这样的观点,即所有标记都是物理的:离开了标记—同一的物理主义就根本不是物理主义。此外,我们的要求能够避免沙文主义的类型—同一性,而对它的拒斥正好是心理功能主义的关键动机。

在心理－物理依赖性中,要求有适当的因果机制并不意味着把心理现象限定在任何特定类型的物理基质之上,比如说与硅基质对立的碳基质。至少从描述上说,适当的因果机制可以出现在物理因果基质的范围之内。在我们看来,物理因果基质对心理现象之发生究竟怎样富有可塑性,这是一个经验问题。因此我们认为:经验的研究一定能回答物理主义的心理功能主义有多少沙文主义或自由主义。沙文主义或自由的程度将取决于能有助于某种心理现象实现的物理因果机制的类别。这完全符合我们的下述观点,即物理主义的心理功能主义是一种逻辑上偶然的理论。

6. 结论

总而言之,非还原的随附性－物理主义的前途似乎是暗淡的。我们已经认识到,心理－物理的有依赖性变化在对心身关系的随附性说明中具有举足轻重的作用,并论证了相关类型的有依赖性的变化在支持反事实时隐藏着对规则还原的真正威胁。我们的论证赞成说,适当的(有相对位置的)因果机制在对心身依赖性的任何充分的心理功能主义的探讨中有着核心的作用。基于此,我们揭露了对全体随附性的标准探讨的问题,强调指出,它们并没有保护有关类型的心身依赖性。由于这个缺陷,因此这些关于随附性的探讨同样没有为物理的东西的本体论的优先性提供地盘。不管非还原物理主义的最终命运是什么,标准的全体随附性和构成性探讨方案都没有抓住公认的、解释所需的心理－物理依赖性。

注 释

① 参阅 M. 胡克(1978)关于笛卡尔的可设想性论证的陈述与评论。R. 斯温伯恩(1986)对笛卡尔为二元论所作的可设想性论证的变种作了辩护。对斯温伯恩的笛卡尔式论证的批评,可参阅莫泽和范德尔·纳特(1993)。

② 刘易斯认为:详述了一种经验的典型原因(如身体原因)的条件有穷集合,“由于分析必然性”而适用于那种经验(1966, p. 165; 试比较 p. 163)。阿姆斯特朗(1977)把有关的因果条件当作来自于“概念分析”的东西。

③ 关于这个口号的讨论,参阅普特南(1975, pp. 223 - 227; 1979, p. 165 - 166; 1988, 第2章),伯奇(1986),福多(1987, 第2章)和波斯特(1987, p. 174 - 180)。

④ 该假定违背了 D. 刘易斯的限定:一个个体不可能在一个以上的可能世界中存在(参阅刘易斯, 1986, p. 213)。不管人们怎样看待刘易斯的限定,博伊德(1980)关于构成唯物主义的观点(见本文第5节)与它不相符合。

⑤ 即便如此,物理主义者仍旗帜鲜明地坚持;关于心身因果系列的表达式应当是延展的,转换性和替代性应该是存在的。

⑥ 在20世纪的科学哲学中,关于与还原主义工作有联系的问题的更详细的讨论,可参阅博伊德等所编的(1991, pp. 387 - 392)、由特劳特所撰的关于还原主义的介绍性论文。

⑦ 金(1990, p. 11)说到了戴维森1985年的描述:“我们可以不费力气地证明:这等于下面的弱协变”,即:如果任何事物必然有 A(属性的随附集合)中的属性 F,在 B(属性的被随附集合)中存在属性 G,以至于那事物有 G,而且有 G 的一切事物都有 F,这是必然的。金的主张是错误的,因为戴维森并没有用模态术语“必然”或任何类似的术语。

⑧ 参看特勒(1984)和金(1984, 1990)的有用的分类。

⑨ 赞成一种类型或另一类型的全体随附性的人有:赫尔曼和汤普森(1975),霍根(1982),刘易斯(1980),波斯特(1987, 1991)等。

⑩ 对全体随附性关于独立于心理功能主义中的因果机制之区域的依赖性的批评,请参阅金(1989, pp. 40 - 42; 1990, pp. 22 - 23),海尔(1992, 第3章)。诉诸于心理过程中的(有相对位置的)因果机制的重要性可能增加了金

对全体随附性的批评的说服力,尽管他没有注意到那种重要性。

⑪ 即使我们承认波斯特的说法:“功能和意向状态在不涉及它们的物理或别的实现的情况下也能得到定义”(1987, p. 161),但我们仍可承认心理过程中有相对位置的因果机制是必不可少的。

⑫ 流行的观点符合于戴维森的下述说明:“类规律的陈述是支持反事实和虚拟论断的、并为它们的例证所支持的一般陈述”(1970, p. 217)。

⑬ 因此倾向解释是一种关于概率计算的不可接受的解释。这是不足为怪的。倾向解释不是作为概率分析、而是为了把握因果论断的依赖性而被引入的;而且它与概率因果性十分吻合。即使倾向解释精彩地说明了符合概率论准则的概率因果性,但它把它的视域限制在复杂的、混合的随机系统上了,而舍弃了频率解释,频率解释强调的是概率计算的理想化假定如独立性假定存在于其中的那些领域。

参考文献

D. M. 阿姆斯特朗(1977):“心灵的因果理论”,载《新哲学杂志》II, 82 - 95。

N. 布洛克(1978):“功能主义的麻烦”,载 C. W. 萨维森编:《明尼苏达科学哲学研究》,第 9 卷。

R. 博伊德(1980):“抛弃还原主义的唯物主义:物理主义不包含什么”,载内德·布洛克编:《心理学哲学读本》,第 1 卷,pp. 67 - 106,哈佛大学出版社。

R. 博伊德等编(1991):《科学哲学》,麻省理工学院出版社。

T. 伯奇(1986):“个体主义与心理学”,《哲学评论》,95, 3 - 45。

R. 卡尔纳普(1966):《物理学的哲学基础》,纽约。

R. 考西(1977):《科学的统一性》,多德雷赫特:里德尔。

D. 戴维森(1970):“心理事件”,载戴维森:《关于行动和事件的论文集》,pp. 207 - 225,牛津大学出版社 1980。

——(1985):“对 H. 刘易斯的答复”,载 B. 弗默真等编:《论戴维森》,pp. 242 - 244,牛津大学出版社。

——(1993):“思维原因”,载 J. 海尔等编:《心理因果性》,pp. 3 - 18,牛

318 第三部分

津大学出版社。

R. 笛卡尔(1637):《方法谈》;(1641a):《第一哲学沉思集》;(1641b):《对第一责难的答复》;(1641c):《对第四责难的答复》。以上各书均载 E.S. 霍丹等编:《笛卡尔哲学著作集》,第 1-2 卷,剑桥大学出版社 1911-1912。

J. 厄尔曼等(1992):“科学假说的证实”,载 M.H. 萨蒙等编:《科学哲学导论》,pp. 42-103,恩格尔伍德·克里弗斯,新泽西:普林泰斯-霍尔。

J. 福多(1981):“心-身问题”,《科学美国人》,244,114-123。

J. 海尔(1992):《真实心灵的本质》,剑桥大学出版社。

G. 赫尔曼与 F.W. 汤普森(1975):“物理主义:本质论,决定和还原”,《哲学杂志》,72,551-564,

C. 亨普尔(1988):“附带条件:关于科学理论的推理功能的问题”,载 A. 格律恩鲍姆等编:《演绎主义的限制》,pp. 19-36,加利福尼亚大学出版社。

C.A. 胡克(1981):“笛卡尔对心身同一的否定”,载 M. 胡克编:《笛卡尔:批评与解释论文集》,pp. 171-185,约翰斯·霍布金斯大学出版社。

T. 霍根(1982):“随附性与微观物理学”,《太平洋哲学季刊》,63:29-43。

金在权(1984):“随附性的种种概念”,《哲学与现象学研究》,45,153-176。

——(1989):“随附性唯物主义的神话”,《美国哲学协会论文和讲演汇编》,63:31-47。

——(1990):“作为哲学概念的随附性”,《形而上学》,21,1-27。

——(1993):“随附性与‘非严格规律’能拯救异态一元论吗?”载 J. 海尔等编:《心理原因》,pp. 19-26,牛津大学出版社。

D. 刘易斯(1966):“同一理论之论证”,《哲学杂志》,63,17-25。重印于 D. 罗森塔尔编:《唯物主义和心身问题》,恩格尔伍德·克里弗斯,新泽西:普伦蒂斯·霍尔出版社,1971。引证出自此处。

——(1980):“疯狂的疼痛和火星人的疼痛”,载 N. 布洛克编:《心理学哲学读本》,第 1 卷,pp. 216-222,哈佛大学出版社。

——(1983):“宇宙论的新探索”,《澳大利亚哲学杂志》,61,343-377。

——(1986):《论多个世界》,牛津:巴兹尔·布莱克韦尔。

- P. 莫泽(1993):《客观性之后的哲学》,牛津大学出版社。
- P. 莫泽与 A. 范德尔·纳特(1993):“幸存的灵魂”,《加拿大哲学杂志》, 23, 101 - 106。
- E. 内格尔(1961):《科学的结构》,纽约。
- J. 波斯特(1987),《存在的外貌》,康乃尔大学出版社,1991。
- H. 普特南(1979):“对古德曼的世界构造方法的反思”,见普特南:《实在论和理由:哲学论文集》,第3卷,pp. 155 - 169,剑桥大学出版社。
- (1988):《表征与实在》,麻省理工学院出版社。
- E. 索伯(1985):“潘格罗式的功能主义和心灵哲学”,《综合》,64, 165 - 193。
- R. 斯温伯恩(1986):《灵魂的进化》,牛津大学出版社。
- P. 特勒(1984):“痴人为随附性和决定所指的路”,《南方哲学杂志》,增刊, 136 - 162。

第四部分

意 识

“意识”一词的两种用法*

诺曼·马尔科姆

在“有意识的”一词的两种用法之间存在着语法上的区别。在一种用法中,该词必须有对象:如说某人意识到某物,或者说意识到如此这般的東西。可以说某人意识到奇怪的味道,令人窒息的热,朋友讥讽的笑脸;还可以说某人意识到房子里有一种奇怪的味道,屋子里热得让人受不了,他的朋友发出了讥讽的笑声。“意识到”这类表达式一般可为“觉察到”或“觉知到”所替换。我将把意识到某物或意识到如此这般的事情称之为“有意识的”一词的及物用法;而且我还会谈到“及物的意识”。

“有意识的”一词还有另一种用法,在这种用法中,它并不带一个对象。如果我们想到一个被撞伤而成为无意识的人又恢复了意识,我们就可以说,“他是有意识的”,用不着加上“关于”或“到”等词。简言之,说某人是有意識的或无意识的这种用法,我将称之为“有意识的”一词的“不及物用法”,而且我会谈到“不及物的意识”。不难看出,当“有意识的”被不及物地使用时,它就不能为“觉知”所替换,因为觉知总是要觉知到什么。

* 节译自他与D.M. 阿姆斯特朗合著的《意识与因果性》(巴兹尔·布莱克韦尔, 1984)一书的第一部分,标题为译者根据内容所加。

及物的意识

及物的意识就是关于什么或关于那的意识：这种意义的意识需要有一个对象。长期争论不休的问题是：意识与其对象的关系是什么？G.E. 摩尔在标题为“驳唯心主义”的早期著名论文中，批判了某些唯心主义者所坚持的这样一种学说，即存在就是被感知。摩尔认为，该观点不过是主张：一切存在的东西，都是被经验。^①这意思明显是说，作为意识对象的任何东西除了作为意识的对象就不可能存在。他的讨论主要集中在他称之为“感觉”或“观念”的东西上。他谈到了“对蓝的感觉”和“对绿的感觉”。他说，即使这些是不同的感觉，但它们之中有某种共同的东西，他将其称为“意识”。至于两种感觉不同的地方，他称之为“意识的对象”。摩尔说：“这一来，我们在每个感觉中有两个不同的因素，其中一个我称之为意识，另一个我称之为意识的对象。”^②接着摩尔又说：

对感觉或观念的真正分析可以这样进行。它们的共同的、我称之为“意识”的因素其实就是意识。感觉事实上就是这样一事实，即“知道”或“觉知到”或“经验到”某物。当我们知道蓝的感觉存在时，那么我们所知道的事实就是存在着对蓝的觉知。^③

摩尔说：“对每一个经验——从最纯粹的感觉到最成熟的知觉或反省——的分析无一例外地要牵涉到关于某物的觉知或意识。”^④进而他作出了一个与我现在的论题有关的断言。他说：这

种觉知或意识“无论如何就具有且必然具有这样一个本质,即我们觉知到它时,它的对象显然就是我们没有觉知(到它)时所是的东西。”^⑤摩尔的论断旨在批判存在就是被感知这样的学说。后一种观点如果意味着:凳子和山峦除了成为意识的对象之外就不或不能存在,那么它肯定越轨了。但是摩尔不是沿着另一方向走得太远了吗?意识的对象包括身体感觉,如疼和痛,因为我们可以说:“当我达到极点时,我才意识到背上的疼痛。”某人背上的疼痛是具有下述本质的某东西吗,即正如摩尔所述,当人们觉知到它时,它显然就是人们没有觉知到它时所是的那种东西?

.....

丹尼特论疼痛概念的不一致性。在哲学中,疼痛是一个聚讼纷纭的课题。例如有这样一些争论,即是否存在着未被感觉到的疼痛,人们是否能感觉到他们没有意识到的疼痛,人们是否错误地认为:他们处在疼痛中,或者说在定位疼痛时是否会犯错误,计算机能否感觉到疼痛,两个人是否有相同的疼痛,疼痛是否是“私人的”感觉等等。疼痛这一课题是一个有助于澄清感觉概念中的许多哲学混乱的好题目。

近来,有人断言,这种混乱并不是根源于我们哲学家带给那个课题的困惑和不成熟的意象,而是由于这样的事实,即“疼痛”这一日常用语的运用是不一致的或不统一的。D. 丹尼特就曾提出了这样的主张。K.V. 威尔克斯说:“丹尼特令人信服地论证说,疼痛的日常概念具有内在的不一致性。”^⑥现在让我们看看他是怎样令人信服地论证的。在《脑猝病》一书中,他说:

“疼痛”一词在各阶层的人们的日常使用[原文如此]中是不一

326 第四部分

致的。一本教科书宣称：一氧化二氮可能使人“感觉不到疼痛”，这是一种极为平常的表达方式，在最敏感的耳朵里也不会引起“异常的”震惊，但它会使人们想到：一氧化二氮绝不会阻止疼痛的出现，只是当疼痛发生时它才会使人们感觉不到疼痛（正像一个视力正常的人在眼睛被蒙住时感觉不到闪闪发亮的灯光一样）。然而同一本书把一氧化二氮归入能防止疼痛的止痛药的类别中（人们可以把它叫止痛药），我们一点也不惊讶。^⑦

丹尼特认为：他已证明了“疼痛”一词的日常用法是不一致的。根据是：那本教科书的大多数读者对说一氧化二氮使人感觉不到疼痛与说它阻止疼痛之间觉察不出意义上的差异。但丹尼特为什么相信意义上有差异呢？显然是因为，他注意的是语词形式，而不是那种语词形式怎样被使用。他说：“不能感觉到疼痛”这个短语使人想到疼痛仍在那里。它使谁想到这一点呢？也许是对意识概念感到迷惑不解的哲学家，而不是想学会怎样减轻病人手术中的疼痛的医学学生。丹尼特的比较（即“某人在眼睛被蒙住时不能感觉到闪闪发亮的灯光”）要说明的是他的假定：“不能感觉到……”这一短语在每种情况下有或者应当有相同的意义。但是那并不是我们语言怎样起作用的问题。如果在某些语境下，而不是在别的语境下，“使某人不能感觉到 X”和“阻止 X”这类表达式被赋予了相同的意义，那么那就是它们怎样被使用的问题。这里对不一致性的责难是缺乏根据的。

.....

不确定性。我想再来考察各别“疼痛阈限”的概念，这不是因

为它为丹尼特的“疼痛”一词的日常用法具有不一致性的观点提供了什么根据,而是因为它为进入一个具有真正意义的课题指明了方向,这个课题就是我们在把感觉和感情归之于人时所说的“不确定性”问题。

假设一女士在一事故中,手臂受了重伤。她仍很镇静,没有大惊小怪,坚决要求先照顾别的受伤的人。她此时的疼痛比通常碰到那类伤害时所感到的疼痛轻一些吗?或者说她很疼痛,但勇敢地克制住了?通常,我们可能不知道是哪一种情况。也许并不总是如此,如果她是我们很熟悉的人,并且我们还知道,平常碰到一点轻微的伤害(撞击、扭伤或轻微烧伤),她都会像大多数人一样叫苦不喋,显出很疼痛的样子,那么我们就不会认为:她有“很高的疼痛阈限”,而会认为在这种情况下,也许是出于对别的受伤的人的关心,尽力不让严重的疼痛表现出来;或者是由于事故的刺激、别人受伤的情境使她的注意力从自身的疼痛中转移出去了。

然而,事实也可能是这样的,即如此看问题是缺乏根据的。某些人认为,她处在极大的痛苦中,但特别勇敢,而别的人则认为:尽管有严重的伤害,但她的疼痛并不严重,还有些人对究竟怎样看拿不定主意。即使她本人说,“并不是那么糟糕”,但这可能还是没有驱散这里的不确定性和不一致性。而且还有别的情况,在那里一个受到轻微伤害的人表现出极其疼痛的样子:某些看到的人就会以为他的疼痛非常剧烈,而另一些人则认为,那行为有点做作,还有些人则无法肯定该作何感想。这种不一致性和不确定性不仅在涉及到身体疼痛的信号时会产生,而且在涉及到情绪和态度的表现如悲伤、懊悔、亲切的感情等时也会出现。

328 第四部分

维持根斯坦在《卡片》中,用几个段落专门讨论过这个课题。他注意的一件事是这样的事实,即大多数人都说:人们在麻木的时候什么也感觉不到;而某人则说,“那可能是某人有某种感觉,但完全忘记了”。^⑧这是意见不一致的一个有趣的例子:每个人都喋喋不休地讲述同样的根据;而有的人有这种怀疑,有的人则没有。

维特根斯坦还谈到了另一个例子,它与存在着不同的疼痛阈限这一观念有关。他说:

我们常以奇怪的方式使用“我不知道”这种习语;例如当我们说我们不知道这个人感觉到的是否真的比那个人多,或者说他是否只是更强烈地表达了他的感情,就是如此。正是在这种情况下,才不清楚什么样的探讨才能解决这个问题。^⑨

在我们的时代,有一种倾向,即假设:如果我们对神经系统中发生的事情有更多的知识,那么这个问题就可能被解决。但是如果我们对在人之内发生的東西有更多的知识,那么按照这词的字面意义,在行为被观察到时,难道就不可能出现同样的不确定性和不一致性吗?维特根斯坦是这样评说的:

假设人们能观察到别人的神经系统的作用。如果是那样的话,那么他们就有一种确定的方法把真诚的感情与模仿的感情区别开来。——或者他们在这里会再度怀疑:当这些标记出现时,别的人是否感觉到了什么东西?——无论如何,确实可以设想:他们在那里所看到的東西决定了他们的态度,他们不会有任何疑虑的。

而且这可能转变为外在的行为。

这种观察完全决定了他们对其他人的态度,而且绝不会产生任何怀疑。^⑩

某人的行为和话语是不是他的疼痛(或悲伤或亲切的感情)的真实表现,或者说这些表现是不是夸张了或部分的伪装,人们的看法有时是有分歧的或不确定的,这是事实。怀疑有时不是没有用的,而能反映在没有热情或缺乏同情心之中。但别的人可能没有一点怀疑,甚至不能理解怎么会有怀疑发生。在承认、评价某人言行举止时的这类差异有时是极其令人困惑的。有的人比别的人似乎有更多的疑义、较少的确信。在特定的情况下,确信的人不可能劝说怀疑的人来相信。

这一事实有助于巩固下述观念,即要发现某人真的感觉到了什么,就必须对他内部发生的東西进行观察。但是如果这个观念就其本来意义而言应当严肃地予以接受,如果神经过程可以被观察到,那么由外部行为引起的怀疑、不一致不是同样会发生吗?不应该假定:对神经过程的观察必然会使不确定性和不一致性寿终正寝。某些观察者难道不会“怀疑:当这些标记出现时别的人感觉到了某种东西”?另一方面,还可以设想,那些观察真的导致了完全的一致,“决定了他们没有任何疑虑的态度”。但同样有根据设想:对人的例如我们现在的行为和环境的观察就不会引起不一致或不确定性。所有博学多闻的观察者会一致主张:“这种观察完全决定了他们对其他人的态度,而且绝不会产生任何怀疑。”

维特根斯坦说:“确实可信的态度、行为难道不是普遍存在于人群之中吗?因此对感情表达的怀疑与他们完全无关吗?”^⑪

可以认为:现存的不一致性应当这样来解释,即是由于缺乏把

330 第四部分

感情归属于他人的明确的规则。的确不存在清楚的规则。但这不是对不一致性的解释。情况恰恰相反。只有当事实上存在着普遍的一致时,才可能有清楚的、明确的规则。关于颜色之间关系的计算、测量判断的确定性正好以此为基础。规则的这种固定性和明晰性就是不可抗拒的一致性的反映。人的感觉和感情是不会出现这种一致性的。

在《哲学研究》中,有这样一种鲜明的对比:

存在着色盲这样的事实,而且还有确认它的办法。一般来说,在那些正常的人所作出的颜色判断中,存在着完全的一致。这表现了颜色判断这个概念的特征。

关于感情的表达是否真实这一问题,一般没有这种一致性。我肯定,的确是肯定,他没有伪装;但第三个人则不肯定。我能够一如既往地使他相信吗?他的推理或观察是否有某种错误呢?“你一点也不懂!”——当某人怀疑我们视之为再真实不过的东西时,我们就这样说——但我们什么也不能证明。^②

在对感情的判断中,有一种不确定性,这是算术计算中不会出现的;或者说在判断某种颜色是否介于红和黄两者之间时不会出现。不确定性是缺乏普遍一致性的结果。至于缺乏一致性的原因,可能有种种推测;但哲学意识到其存在则更为重要!我们可以设想一个社交场合,在那里,感情表达的不一致性和不确定性没有发生过或从未发生过。他们对诚实的感情表达的判断就像我们对颜色的判断一样确定。那些人几乎没有或完全没有这样的倾向,

即把别人的感觉、情绪、心情和思想当作是隐蔽的或不可理解的。

我不是说,我们不可能拿出任何理由让另一个人接受一个人自己对感情表达的反应。如果 A 和 B 多年来是死对头,如果 A 对 B 常作出贬斥性的评论,但是当 B 死去了时,A 又表现出悲伤的样子,那么另一个人 C 可能会认为:A 的悲伤的表现是不真实的或假装的。但某人也可能会说:A 与 B 年轻时是亲密无间的朋友,在各种艰难困苦的条件下都能相互帮助。这大概有助于肯定:B 的死对于 A 来说是真实悲伤的诱因。这个信息可能影响 C 的态度——但也可能不会。正如维特根斯坦所说:我们什么也不能证明。

当对感情表达的评价出现不一致时,人们往往会认为,这是因为人们对别人的知觉的敏感性彼此有别。诚然,一些人比另一些人的反应更灵敏一些。不过,我所说的“不确定性”的意思是:感情表达的例子,俯拾即是,在那里,敏感性和见识相同的观察者在评价表达的真实性时是不一样的,没有人能找到解决这种不一致的方法。这意味着:在这个问题上不存在肯定的对或错。在这种争端中,一个人可能说:“你太轻信了”,另一个人可能说:“你太不敏感”,但这只是申斥。每个爱争论的人都会继续说:“我是对的,你是错的”,但这些说法的意思与长度的算术计算或测量情况下的那种说明是不同的。

自然有无数的情境,在那里人的行为和环境是这样的,以致每个观察者要么是赞成说那个人在伪装,要么是说那感情的表达是真实的。如果没有那种高度的一致,那么就不存在关于伪装或真诚这类共有的概念,因此在它们的使用中也不可能有一致或不一致。

然而仍常有不可解释的不一致或不确定性的事例。我并不怀

332 第四部分

疑,哲学家们常常倾向于把所有的感情、经验、心理状态等说成是“内在的”,把它们当作是“内隐的”,这种倾向的重要源泉正是这种常常作为我们对感情表达的反应之特征的不确定性和不一致性。如果是这样,那么所有心理现象都是内在的和内隐的这一观念就是极其错误的;因为第一,在大量的事实中,环境与人类感觉或情绪的表达如此泾渭分明,以致不容任何怀疑;第二,在那些情况下,在有真正的不确定性的地方,主张存在着某种“内隐的”东西,这是很糊涂的,因为它把不确定性与缺乏知识混同起来了。

觉知到有意识。我打算考察一下与及物意识有关的一个令人迷惑不解的问题,即当你意识到某物时,你是否有可能分辨或知觉或觉知你对某物的意识,亦即你对一对象的意识能否成为你的意识的对象? G.E. 摩尔在“驳唯心主义”一文中对这个问题已有所论述。前面我已述及他的观点:在每个感觉中都有两个不同的因素:一个是意识,一个是那意识的对象。在谈到他的“对蓝色的感觉”这一例子时,他说,这种感觉的一个因素是对象:蓝,另一个因素是对那对象的觉知或意识。他接着说:

即使哲学家已认识到意识所意谓的是某种独特的东西,但他们从没有关于这东西是什么的概念。他们不可能把它与蓝放在心灵面前,也不可能用比较蓝与绿的那种方式对之作出比较。其理由在于……—当我们想把注意力集中在意识上,想看看这究竟是什么时,它似乎消失不见了,在我们面前似乎只有一片空白。当我们想内省我们关于蓝的感觉时,我们能够看到的就是蓝:另一个因素仿佛是虚无缥缈的。然而如果我们专心致志地去看,如果我们知道有某种要寻找的东西,那么

它又是能被区分开来的。在这一段中,我的主要目的就是想让读者看到它……^⑬

摩尔说过,通过内省分辨意识的因素是极其困难的。人们想注意它时它便消失得无影无踪(就像我们想抓住一个光滑的东西时它便从我们手中溜走了一样),难道对意识的分辨因此就难如上青天?或者说它是另一类型的困难吗,即说出某人自己对某物意识的分辨、知觉、觉知是不可思议的?

摩尔不仅说意识的要素能与意识的对象区别开来,而且说他想让读者看到它。假设摩尔的一个读者这样向他报告:“的确很困难,但我终于在我对蓝色的感觉中成功地知觉到了意识的因素。”人们怎么可能知道这个人知觉到了摩尔所谓的意识的因素呢?怎样才能使我心悦诚服地承认:他们两人指的是同一个东西呢?

说某人能知觉到或不能知觉到某人对某物的意识,这是一个令人困惑的观点。在乡间散步时,我的同伴可能对我说:“你看到远处森林的蓝颜色吗?”假设我的回答是:“看到了”,接着我的同伴会进一步问:“你知觉到了你关于蓝颜色的意识吗?”多么古怪的问题,我能回答什么呢?如果我回答说,“不错,我知觉到了我关于蓝颜色的意识”,这除了说出了我先前的简要回答即“是的,我看到了蓝色”之外,还说了什么?

问某人是否注意到、知觉到、觉知到他自己关于某物的知觉或意识,这似乎是令人诧异的。因为我看到了蓝颜色,这不是一个经验事实吗?当我说“我看到了蓝颜色”时,我不是报告了这个事实吗?我的报告指的似乎是两个因素:蓝颜色与我看到了它。除非我知觉到了两种因素,不然我怎么能作出这样的报告呢?摩尔相

334 第四部分

信：如果他很专注，他就能在蓝的感觉中分辨出两种因素：蓝与意识，这一信念来自于下述推理：“我对蓝的感觉一定包含着蓝与意识；因此如果我格外专心地集中我的注意力，那么我就一定不仅知觉到蓝颜色，而且知觉到我对它的意识。”

然而，这个推理是基于维特根斯坦所说的“语法幻觉”。该幻觉就是认为：既然“我意识到蓝色”这个句子能够成为关于我的经验的真实报告，那么我观察到意识和颜色就一定是可能的。而事实上，该报告是独自从对颜色的观察中派生出来的，不是来自于对意识的观察。但我以下述方式表述那一点也是错误的——“我观察到我关于颜色的意识”这一说法的意思仿佛是很清楚的。

J. 洛克对这个问题曾发表了一些精彩的见解。他说：“意识是对人自己心中所发生的东西的知觉。如果我不是靠自己知觉到我对某东西的意识，别的人能知觉到它吗？”^⑭洛克认为，这个反问显然只能否定地予以回答。他宣称：“人们不借助知觉而知觉到他在知觉，这是不可能的。当我们看、听、嗅、尝、感觉、考虑或想要什么东西时，我们知道我们所做的。”^⑮

洛克是在及物的意义上谈论意识的——意识到某物。人们可能会批评说，洛克的主张太笼统，例如只注意到：一种留意的思考在一个人没有觉知到它是留意的思考时发生在他身上了：他对思考的这个特征的认识根本不可能发生，或只有在反省后才发生。但我这里关心的是某种更根本的东西，即说某人知觉或观察或觉知他自己的知觉、思想和感情在什么范围内是有意义的？

我们先来看视觉方面的例子。假设我面前的桌子上有一个金属球，其运动是由电磁力控制的。这是一个实验，我的任务就是不

时地对那个球是处在运动中还是处在静止中作出报告：所以我大声说：“它在动”；“现在它停下来了”，等等。为了作出这些报告，我必须看、观察那个球。但是我观察到我的看吗？我能这样做吗？如果我不能观察我的看，那么我能观察我在看这个球吗？用洛克的话说，除非我知觉到我在知觉那个球，我就不可能知觉到它吗？这都是一些很古怪的问题。

另一个人可能在观察那个球，并且也可能在观察：我在观看球及其运动。他是怎样进行后一种观察的呢？噢，他观察到：我的头和眼随着球的运动而运动；他听到了我的语言报告（根据他自己对球的观察，他知道那是正确的）。总之，他观察的是我的行为，包括我说了什么。我能够观察到球的运动，或不能观察它们。但是我们理解下述假定，即我可能观察或不可能观察（知觉、觉知）我自己对球的观察吗？

人们常常习惯于认为：旁观者借助于行为对人们看、觉知、意识到某物的观察充其量是第二手的，人们自己关于他们的看或觉知的内在知觉才是第一手的、直接的。可以说这是完全错误的；对某人的看和觉知所能作出的唯一的观察或知觉不就是对行为的观察吗？如果这是正确的，那么对洛克的问题——如果我不是靠自己知觉到我意识到某物，别的人能知觉到它吗——的回答就是：别的人能知觉到我对一特定的对象的意识；我是否可以知觉到我意识到那对象，这是一个不合理的问题。

心理现象的内在知觉。F. 布伦塔诺曾对所有“心理现象”的一般特征进行过描述。在他的《从经验主义观点看心理学》^⑥一书中，他说，每一“心理现象”都以“指向一对象”为其特征；每一心理

336 第四部分

现象“都在自身内包含着作为其对象的某东西”。例如“在想象中,有某东西被想象,在判断中有某东西被肯定或被否定,在爱中有被爱的,在恨中有被恨的,在意欲中有被意欲的,等等。”布伦塔伦说,心理现象(心灵现象)能够被定义为这样一些现象,即“在它们自身中意向地包含着一种对象”。布伦塔诺说,语言体现了这种对对象的指向:“我们常说,一个人因某事而高兴,或为某事而高兴,一个人为某事而忧虑或悲伤。”

人们可能会问,布伦塔诺的描述是否具有普遍的适用性。人们有时似乎感到焦虑,而其焦虑并没有任何对象,或至少不知道有什么对象。沮丧有时也是如此。人们可能整天感觉到很快乐,或感觉到很忧郁,而并不是因任何特定的东西而感到快乐或忧郁。

然而布伦塔诺的这一特殊的观点是众所周知的,我不想对之作进一步的评论。他还有一个观点,尚未引起足够的注意。他坚持认为:所有心理现象的更一般的特征是:“它们只能在内在的意识中被知觉到。”^⑦他把关于心理现象的知觉称作“内知觉”;他补充说,所谓“外知觉”严格说来并不是知觉,因此就“知觉一词的严格的意义而言”,心理现象是“可能被知觉到的唯一的東西。”^⑧最后布伦塔诺说,“任何心理现象不能为一个以上的单个个体所知觉”。^⑨

布伦塔诺所说的“心理现象”包括回忆、期望、判断、相信、怀疑、愿望、意图、生气、高兴、悲伤等。他坚持的是:每当人们意欲或期望某事等时,那个人就有对他自己的愿望或期望的内知觉。此外,布伦塔诺还主张:那个人是能够知觉那个愿望或那个期望的唯一的人。

布伦塔诺的观点令人费解。我不是经常知觉到另一个人的期

望吗？当我期望做某事时，我不是知觉到了我的期望吗？一个人为他到野外散步作准备：他的狗表现出充满希望的样子，这是那个人能知觉到的。人们常常以为：那个人实际上并没有知觉到狗的期望，而只是知觉到狗的充满希望的行为。我们应该认为狗知觉到了自己的期望吗？我们能把狗的下述两种心理状态即第一是它被带出去散步的期望，第二是它对它的期望的知觉或觉知这两者区别开来吗？

可能有人说，我们不能把这种区分用到狗身上，但我们确实可把这用在人身上。让我们对此作一番考察。假设我坐火车到一个城市里去。一朋友写信告诉我：他在车站接我。如果在旅行之前或中途，某人问我：“你希望迎接吗？”我便会回答说：“是的”。这不暗示了我已知觉到或觉知到我的期望吗？一个多么稀奇古怪的问题！我肯定不会说：“不，我没有知觉到它。”因此我不是得肯定：我确实知觉到了它吗？但这样说是什么意思呢？除了有关描述中已述及的，即如果问我：我是否希望被迎接，那么我就会说“是的”，还会有什么意思呢？如果毕竟不过如此，那么这个条件事实是否暗示我知觉到了我的期望这一问题，就是此条件事实是否暗示了自身这个无意义的问题。

根据布伦塔诺的观点，每种“心理现象”（高兴、怀疑、恐惧、担忧、意图等）都指向一个对象；而且这种对对象的心理指向本身就是内意识或内知觉的对象。如果一个人是高兴的，那么他的高兴就指向一特定对象，而且那个人还有对他的高兴的内意识。我感兴趣的是后一主张。一个人何时对对他的高兴的内知觉——而且究竟持续多长时间？在布伦塔诺看来，当且仅当他高兴时，他就有

338 第四部分

那个知觉。因此如果某人听到了布伦塔诺关于内知觉的学说,且想知道它是什么,何时发生,那么他就会明白:当且仅当一个人忧伤时他就有对忧伤的内知觉,当且仅当他恐惧时,他就有对恐惧的内知觉,等等。他会明白:布伦塔诺的学说只不过说明了从“他恐惧”到“他有对恐惧的内知觉”的语言运动。

不过,这个批评太苛刻了。这里还有替换歪曲地谈论熟悉的东西的别的方式。“别的”是一心理图景,它几乎不能描述,但似乎类似这样的某东西:人们专心致志地看总是集中在内在空间中,如果不被知觉到,就没有什么能进入这个空间中,我们,不仅是洛克、布伦塔诺,而且包括我们全体,为什么强烈地倾向于这种图景?这是一个艰涩的问题,需要人们深入到人的哲学无意识之中去探究。

我觉得:每种心理现象都有其内知觉这一观念的部分动因是这样的,即当一个人高兴、忧伤、恐惧等时,在他的行为或言语中常有关于这一点的某种表达式。尽管这不是必然如此或永远如此,但常常是如此:如果不是那样,那么我们就没有关于高兴、忧伤等的共同的概念。现假设有个人正好得到了某种大好的消息。他的一个可能的反应就是对传递消息的那个人惊叫:“我太高兴了!”另一可能的反应就是突然兴高采烈、手舞足蹈。我心中所进行的哲学思维是这样的——除非那个人有对他的高兴的内知觉,他就不会认识到:由高兴而来的要么是惊呼,要么是手舞足蹈!他并不知道他因高兴而手舞足蹈,或正是由于高兴他才惊叫:“我太高兴了”。但推论仍可继续,即假设一个人并没有认识到他因高兴而手舞足蹈等,这是荒诞不经的。因此他一定知觉到了他自己的高兴。

在批评这个想法时,我将集中在这样的假定上,即如果一个人

因高兴而手舞足蹈,那么他就一定认识到了他因高兴而手舞足蹈。一个人可能因高兴而手舞足蹈,这完全足够了,但说他可能认识到或没有认识到这就是他所做的这一观点则有某种问题。人们什么时候使用实际语言中的下述语词即“他认识到他因高兴而手舞足蹈”?假设我们能描述一种可以这样说的情境。假设有一个人,他两三年来一直很沮丧。对任何消息,不管是好的还是坏的,他没有什么情绪反应。他很冷漠、无动于衷、孤独。他说,“那有什么关系?”他的家人、朋友很关心他;他们经常让他注意他的冷漠和麻木。他本人知道:什么也不能改变他。但假如有这样的情况,当他听到了好消息时,他便跳起来了,开始在房子里手舞足蹈。他吃惊地认识到:他正在因高兴而手舞足蹈!

这个例子说明了下述语词的一种自然用法,“他认识到他因高兴而手舞足蹈”。正是这种形式的语词才有意义——但只在特殊的环境中才如此。“他认识到因高兴而手舞足蹈”,这种说法并非在他因高兴而手舞足蹈(或因生气而关上了门等)的任何条件下都是合理的。如果这是对的,那么下述哲学推论就是缺乏根据的,即高兴的人一定知觉到他的高兴,因为如果他没有知觉到,那么他就不能认识到他因高兴而高谈阔论或行动。

须知,我并没有作出这样的(教条主义式的)断定:说“他认识到他因高兴而手舞足蹈”总是没有意义的。我追问的是,他可能在什么样的环境下说这些话?在思索这个要求时,在描述事实时,人们会明白:语词的这种形式的自然用法是有限制的。如果布伦塔诺或别的哲学家断言:每当一个人因恐惧、生气、感激等而行动时,“他就认识到了它”,那么我们会明白:这不是这些语词怎样被使

340 第四部分

用的问题。在明白了该主张相对于我们语言的日常用法是错误的的同时,我们也会明白:我们并没有理解语言在日常使用中据说总为真的那种用法是什么。

再来看另一个例子。在布伦塔诺看来,如果一个人不喜欢另一个人,那么前者就有关于这种不喜欢的“内知觉”。我们不可否认这是有意义的,而只是想问:在什么环境下它才有意义?假设 A 先生多年来一直以敬重和感激的态度对待他的同事 B 先生。然而近来,A 开始注意到 B 的各种令他恼火的癖性,例如夸夸其谈,他对自己讲的笑话发笑的那种过分激烈的样子,甚至他悠然自得的样子。A 开始对 B 的这些令人厌烦的习惯进行大量的思考,在这过程中,他逐渐领悟到:自己已真的不喜欢 B 了,甚至一点也不喜欢他。后来他给一位朋友写信谈到了这个变化,说:“由于我对 B 的感情的内知觉发生了变化,因此我决定以后不再理他了”。

A 对“内知觉”这一短语的使用是恰当的,但必须在这样的特殊情境中,即由于新的事实,A 对 B 的先前的态度随着对他的看法的逐渐改变而改变。这个例子自然不是说:每当一个人有意识地不喜欢另一个人时,说他有关于他的不喜欢的“内知觉”就是正确的。恰恰相反,在大多数生活环境中,一个人不喜欢某东西,或者说因吃惊而生气、有兴致、被感染等,如果他讲的是他对那些态度和反应的“内知觉”,那么我们就没法理解他了。如果我排队花很长时间等着购买看歌剧的门票,如果后来的人想插进我前面的队伍里,于是我生气地予以指责,如果后来我说我有关于那场合下我的生气的“内知觉”,那么这就没有什么意思了。洛克-布伦塔诺的命题即所有心理现象都是在“内在的意识”或“内知觉”中被知觉

到的,是错误的——因为它与这些表达式的日常用法是矛盾的。

但是如果不能这样说内知觉,那么一般的知觉又怎样呢?洛克断言:一个人如果没有“知觉到”他嗅、看或听某物,那么就不可能嗅、看或听到它。一个人也许正在思考怎样修理破损的窗户,或许他看到了海水颜色的变化,或闻到了炒卷心菜的气味,但说他正知觉到了或没有知觉到他的思考、看、嗅,这就完全是无法理解的。正如在我们先前的例子中那样,我们能够想到这些事情在其中被说出的特殊的情境。假设一个朋友曾受过脑震荡,所留下的唯一的、持续的后果是:他失去了他的嗅觉。这种无能一直延续了好几年。一天他突然兴奋地惊叫:“我知觉到我正在闻卷心菜的气味!”这种异常的表述形式常为那情境的戏剧性本质证明是正当的。就其信息特征而言,这个惊叫与说“我开始嗅到了卷心菜的气味”其实是一样的。“我知觉到”这个短语是一种修辞上的华丽词藻,在那种情境下是很自然的。其意思不是说除了嗅,还有对嗅的知觉。常见的第一人称句子,“我知觉到我在闻它”,除了“我闻到了它”这个意义,再没有什么明确的意义。我能知觉到你闻到了卷心菜的气味:我的知觉是以对你的行为的观察以及你所说的话为基础的。在正常情况下,不能说你“知觉到”你闻到了卷心菜的气味:你只是闻到了它。

不及物的意识

如果一个人受到了沉重的一击,呆呆地躺在那儿,那么我们可能难以肯定他是有意识的,还是无意识的。我们可能断定:他仍活

342 第四部分

着(他的心脏在跳动),但又是无意识的。能不能断定他死了,同时又是无意识的呢?不能。这不是“无意识”一词的用法。我们只能把词语“有意识的”、“无意识的”用在那些我们认为可能是活着的人身上。当我们知道一个人死了时,他是有意识还是无意识这样的问题就烟消云散了。可能有人认为,这是因为,“他是无意识的”根源于“他已死了”:如果我们认为他死了,那么我们就得出结论说他是无意识的,因此他是否是有意识的这样的问题就不复存在了。这是错误的。没有人在知道他的朋友死了时还说他是“无意识的”,除了作为修饰的词藻。在为那些阵亡的人举行的追悼会上,演说家可能会把他们作为“无意识的死者”加以称颂。即使“无意识的”可以用之于这类礼仪场合,但它不是该词语的常见的用法,在常见的用法中,如果我们认为一个人是无意识的,那么我们会设法使他苏醒,或希望他恢复意识。一当我们确信他死了,我们就会放弃这个希望。“他死了”既排斥“他是有意识的”,又排斥“他是无意识的”。换言之,只有当我们相信一个在事故中受到了重伤的人仍是或可能活着时,我们才会把他当作是有意识的或无意识的。

令人困惑的问题是:如果一个人是有意识的,那么由此可以推论说他意识到某物吗?我认为,不能。一个受到撞击而成为无意识的人,可能表现出恢复意识的迹象(活动手臂,睁开眼睛,坐起来,发出声音),而又没有显示:他意识到任何光、声音、感觉或其他什么东西。意识到某物可衍推出有意识,但不能倒过来。

意识有程度上的差别。一个刚刚恢复了意识的人可能是茫然的、头昏眼花的、“半意识的”等,即不是完全有意识的,缺乏警觉,不能做通常能做的事情。这种状态与半睡着、或酒醉或梦游者

的恍惚这类状态之间有一些相似性。意识的“程度”一词还有另一个意思。人们在分辨力、反应的敏捷性、概念的形成、创造力诸方面自然存在着差异。一个人如果在这些方面表现杰出,那么就可以说他有“很高程度的意识”。

人的范例。不仅对人,而且对某些别的生物(如狗、马),我们都能说:“是活着的,但没有意识。”而昆虫或蠕虫又怎样呢?受到拍打的苍蝇仍在飞或在爬:它仍活着,但它是有意识的还是无意识的?这是关于苍蝇的一个奇怪的问题。为什么是这样?维特根斯坦评论说:“只有对活着的人,以及像(行动像)人的生物,人们才能说:……它是有意识的或无意识的。”^④这是一个有意义的说法。它意思是说,活着的人是把什么东西说成是有意识或无意识的范例。如果我们想知道,能不能把某个非人的生物说成是有意识的或无意识的,那么我们显然得求助于与人体的比较,如血、肉、仪态、姿势尤其是有表现力的人类的脸和眼。正是通过与人的脸的比较,狗的脸才被称作“脸”(而蠕虫没有脸,人们会觉得,说苍蝇有脸是可疑的)。因为狗脸相似于人脸,因此人们就能在狗的脸上看到好奇、警觉和恐怖;狗的大量行为如此相似于人类,以致我们会毫不犹豫地说:狗有期盼、能识别、能记忆等。

至于苍蝇,就没有为心理谓词的运用提供哪怕是极其脆弱的基础的相似性。它对周围的运动的反应就是飞走。我们能够说它受到那个运动的惊吓吗?实际上我们只能这样说:我们的意思仅仅是:那运动使它飞走了。苍蝇不能有狗所表现出的那种惊恐的行为,也没有识别、预知、生气等行为。它会极力挣脱粘蝇纸;但它受到了惊吓吗?它处在痛苦中吗?我们不知道怎样把这些词用在

344 第四部分

苍蝇身上。我们甚至不清楚,是否应说苍蝇“知觉到了”引起它逃走的那个运动。我们不能自信地把任何形式的意识归诸于苍蝇。有的人可能会说:这是因为我们所知的还不够,终有一天,科学可能查明苍蝇是否具有意识。其实并非如此。苍蝇如此有别于人类,以致它不可能表现出恐惧、生气、焦虑和预知等。而且我们还知道,任何科学发现(如苍蝇有大脑)都不可能证明:把意识的那些形式归之于苍蝇是正确的。如果苍蝇迅速避开了附近某物的突发运动,那么我们会说:它“觉察到”那个运动。这就好比说:那运动“吓坏了”它。这都不错。甚至我们可能会说苍蝇有关于那运动的“意识”。但我们应当认识到:这个术语在这里所述及的仅仅是那有点微弱的行为。在“更深的”意义上问苍蝇是否有意识,这是没有意义的。

说活着的人是使用心理谓词和动词的范例,就等于说这些术语最初是适用于人类的,而用之于别的生物是靠类推。有人可能会责难说,我们把某些心理术语用之于狗和猫是自发的:我们并不是根据与人的明确的比较而使用它们。就算是这样。那么怎样证明下述主张即只能通过与人的类比才能说狗生气或惊恐呢?

可以这样来证明。假设有一个人 A,他在以第一人称和第三人称的方式把心理学的动词和谓词应用于人类方面有良好的表现,就像我们任何人一样,但是当别人把这些词用于低等动物时,他总是觉得可疑、荒唐、滑稽。当别人通过指出相似性而为这种使用作辩护时,他会通过说明真正的区别而反唇相讥。其他的人(尤其是狗、猫和马的爱好者)对 A 则可能感到大为诧异或报以某种愤慨:“他难道不能看到:它们多么像人类?”但是没有人能正确地

说:A 没有理解“渴望”、“神经质”、“害怕”之类的术语,因为他对他在与人们一起的日常生活中所理解的只是给予了空洞的证据。相反,如果我们设想另一个人 B,他不知道怎样把日常心理表达式应用于人类,但令我们惊讶的是,他像我们一样把它们应用于别的动物,那么根据后一类现象,我们就能够说:他没有理解那些术语:他没有把它们正确地应用于人类的能力使之不可能。B 不得不被当作白痴,而 A 可以视之为理智的、敏感的,但又是偏执的。对这两种想象的案例的比较表明:心理学术语应用于人是本原性的,而应用于别的动物则是借助类比。

至于机器,问它是有意识的还是无意识的,就像这样去问苍蝇一样几乎是没有意义的。这是很好理解的,因为有这样的事实:“死”与“活”等术语不能在其本义上用之于机器,而只能有隐喻式的使用,正像我们有时说:“那机器死了”或“那排炮死了”或“在关键时刻,发动机活过来了”。既然“活”不能在本义上用之于机器,那么就不可能提出下述严肃的问题,如机器从本义上说是有意意识的还是无意识的。有一些机器,它们对附近的光的颜色变化有反应,如果某人说那机器“知觉到了”不同的颜色,那么就没有令人大惊小怪的地方。说颜色的变化驱动着机器中的某些装置,这是一种隐喻的说法。人们总是在创造或逐渐适应一种表达式的新的用法。就其本身而言,这并没有什么害处;但是让那个词出现在陈述了同一事物的思想中就很容易引起混乱。也许有一天,人们在谈论某些机器时会说,它们在睡觉或睡着了,它们迷乱或出神、茫然、半意识、完全有意识、警觉、机智、诙谐等。当这些表达式用之于机器时,其意义是什么呢?我拿不定主意。但它们用之于机器时,其

346 第四部分

意义必定不同于它们用之于人类的意义。人们在使用这两种用法时述说的不是同一个东西。

能说什么。人们只能说活着的人、似人的生物有意识或无意识,维特根斯坦的这一论断不应被理解有这样的含义,即“有意识”、“无意识”永远只能用之于人类。如果我的妻子很健康,且一直如此,那么当一熟人对我说:“你的妻子是有意识的吗?”我便大吃一惊。我肯定不会回答:“啊,是的,她是有意识的”。这不是这些词的用法;在这类情况下,不能使用这些词。

在哲学中,“能说什么”这一说法,是存在着相当的混乱的。人们可能遭到这样无关紧要的反驳:在上述情况下,当然能说“她是有意识的”——在那里,这意思不过是说:一个人说出那些词在物理上是可能的。另一反驳殊途同归,那就是:即使这些词在那种情况下没有说出来,但能够说出来。在本文中,“被(不被)说”,“能(不能)被说”这类表达式实际上作了相同的运用。这是这些表达式的一种常见的用法,就像同样的表达式用在游戏中所说明的那样。初学象棋的人,当设法避免将死而把帅从棋盘上拿走时,就可能被告知:“不能那样走”,或者说“那是不行的”,或者说“那不是一着棋”。初学者的教练可能会说:“把帅从棋盘上拿走在象棋中是不允许的,没有意义的。”初学者也许熟悉另一种需用棋盘玩的游戏,它有点类似于象棋,在那里受威胁的子可以从棋盘上拿走。教练可能认为:该初学者把象棋与其他游戏搞混淆了。他可能问,“你以为你现在玩的是什么游戏?”同样,在向我提出关于我的妻子的问题的时候,如果一个观察者知道我的妻子并没有病,且不曾害过病,还知道我晓得这一点,那么他听到我说“啊,是的,她是有意

识的”,就会大为诧异。他后来可能问,“你在说什么,你的意思是什么?”他知道,“她是有意识的”在那种情况下是不应说的,是不能说的;因此他对我所说的、我所玩的游戏迷惑不解。我把“在那种情况下能说那些话吗”这个问题与“那些词能用之于那种情况吗”这个问题当作是同一的,同时把它与“那些词在那种情况下有意义吗”也当作是同一的。

通常,哲学家们对下述不同的概念、理论有观点或“理论”,这些概念、理论要求某种事实上不能被说出的事情能够被说出来。例如它可能是一种关于看的概念的理论,这种理论意味着,一个人不能看他面前的东西,除非在他看来,他看到了它;或者它可能是一种关于意愿和行动的理论,它意味着:每当人们做某事时,人们就试图去做它;等等。然而在许多情况下,你在其中看到某东西(你的帽子或你的汽车),就不能说,在你看来你看到了它;在你做了某事(出去散步或兑支票)的许多情况下,就不能说,你试图去做某事。这些含义证明这些理论是错误的。为什么?因为这些理论被认为是对日常概念即包含在、反映在语言的日常用法中的概念的哲学说明或分析。与这种用法相矛盾的哲学理论就是错误的,即使它也许有另一类优点。

哲学家们常常希望它们的理论是真实的,进而绞尽脑汁、独出心裁,以保证对概念的哲学说明不因为有悖于能够被说出的东西而变成虚假的。一个流传甚广的辩护就是在“言语条件”和“真值条件”之间作出区分,前者据说是确定何时说出或断定某事为合适,而后者据说是确定被说或被断定的东西何时是真的或假的。真值条件有时被等同于“意义条件”,所谓意义条件据设想是要决

348 第四部分

定一陈述何时是有意义的。于是,据论证,言语条件与真值或意义条件不可同日而语;因此说出既有意义又真实的某事就可能是不恰当的。例如我已说过的“我的妻子是有意识的”就不恰当(因为容易使人误解),然而如果我说了它,那么我的陈述就既是有意义的,又是真实的。

这个论证很特别。“不恰当地说出”的东西与不能被说的东西是相混淆的。在与初学者的象棋游戏中,下出高明的棋招可能是不恰当的,因为这会打消他学习游戏的兴趣;但是那些招法属于象棋,能够被走出来。这对于言语也是适用的;说出某种能被说出的且又为真(或假)的事情,有时可能是不恰当的(因为未被加工、不明智、不真实、容易误导等)。但是不能被说出的东西不能既真又假;正像从棋盘上拿去了帅既不是正确的招法又不是错误的招法而恰恰不是招法一样。

上述混淆的一个根源就是未能认识到:“能被说出的东西”是一相对的概念,即相对于特定的情境。假如电话铃响了——你拿起话筒说,“你好”,接着便与线路另一端的人谈话。后来,能说的是你接了电话,不能说的是,你曾试图接电话。如果你不能握住话筒,或拿掉了,砸坏了电话,或另一端没有任何反应等,那么就能说你试图接电话。不一定有这样的情境,在其之下,“被回答的”能被说出来;“试图回答”也是这样;而且不一定有这样的情境,在其之下,不知道另一事情能被说出来。

我们非常赞赏这样的意见,即阐明下述原则的细微差别一定是可能的,这些原则为“被回答”与“试图回答”——以及一般而言的“被做”和“试图做”——提供了充分必要条件。我们觉得:如果

并非如此,那么关于试图的不同例子就不能因“共同的特征”而统一,进而没有人能学会动词“试图”的用法。但事实是:我们并不知道充分必要条件。断定我们对这种条件必定有一种无意识的或不清楚的知识,就是一种臆断。这也适用于下述观点,即一般来说,必然有一种决定什么能被说出的充分必要条件。这个假定是绝对没有根据的。懂得一种语言就是懂得什么能被说出,什么不能被说出,不拘是什么事情。它是关于特殊事例而不是关于一般规则的知识。

令人费解的问题是:假设在某事能被说出的条件下,某人说出了它;这该怎样看呢?再看看你接电话的例子。在后来的某一天,有人问你:电话铃是否响了,你是否接了电话。你回答说:“我试图接电话”。你正在说什么呀?我不知道。存在着这样一种令人迷惑的哲学倾向,即宣称:当你说“我试图接电话”时,你说的正好是:你试图接电话!再明确也不过了!当你说“P”时,你说的就是“P”!

这种解释之言之无物是一目了然的。如果一个人对某人在发“P”的音时所说或所意指的是什么问题迷惑不解,那么再告诉他在发“P”的音时,他意指的就是P,这不是解释。一个真正的解释常常要提及那状态的某些有关环境。假设你过去接电话时常常很粗鲁,且受到这样的指责:“那不是人接电话的方式!”接着你回答说:“我试图接电话”,这回答大概是你所说的、试图亲切地接电话的道歉的方式。但是如果你从未表现出粗鲁地接电话的倾向,也从未受过这样的谴责,那么你的回答就不是那个意思。当一个人发“P”的音时究竟说了什么,这里的不确定性可借助对过去和

350 第四部分

现在的环境的全方位描述而得到清除——但不能通过愚蠢地坚持你发“P”的音就是在说“P”这一点而予以清除！

我是有意识的。除非某人相信一个人现在或过去也许病得很厉害,他就不能说那个人是有意识的。对于我的这一主张的一个更形而上的驳难是:有意识的人总是觉知到是有意识的,总能说或认识到“我是有意识的”;如果那个人能如此说,那么别的人说到他或她时也能说:“他/她是有意识的”。

但是一个人所说的:“我有意识”在什么范围内是合理的?如果我在一事故中受了伤,对我关心的人可能会说“他似乎没有意识了”,而我可能会咕哝回答说:“我有意识”。我的说法在那种情况下是合理的。它告诉别人的信息是:我由于有有意识的标志因而是有意识的。别的标志还有:睁开眼睛,坐起来,说点什么。我把“我有意识”说出来就已向别人表明我是有意识的,这正好是通过说话而不是借助于话语的假定的“命题内容”而实现的。

在不具备这种情境时,我难道不能对我自己说“我有意识”吗?是的。但我会做什么呢?也许我希望是有意识的或从无意识中苏醒过来——因此我对自己所说的“我有意识”可能是一种轻松的或兴奋的惊叫:正像我眼瞎了很长时间之后恢复了视力时所发出的“我看到了”的惊叫一样;或者像我在克服了艰难险阻到达目的地时洋洋自得地说的“我到了”一样。这些话语在那些情境下是有某种意义的。我们不应假定:当“我有意识”是一种因震惊或高兴而发出的惊叫时,它也表达了一种关于人的状况的知觉。

人们往往认为:这是不对的,如果我惊叫“我有意识了”,表达了我的吃惊和轻松,那么我一定知觉到(或觉知到)我是有意识的

——因为除非我知觉到我有意识,否则我怎么可能会有那样的惊叹呢?这似乎是一种强有力的理由——当我们对某人知觉到他有意识这一观念作出反思,情况就不一样了——维特根斯坦在提出下述问题时作过这种反思:“我观察我自己,进而知觉到我在看或有意识吗?”他还问道:“但不是一种特殊的经验引起我说‘我又有意识了’吗?——什么经验?”^①当我们反思这些问题时,我们便认识到,如果我有意识是我经验到的某种事情,如果我能够知觉到我有意识——那么我难道不能知觉到我是无意识的?再者,如果我知觉到我有意识是可能的,难道知觉的错误不也是可能的吗?比如说当我实际上是无意识的时,我自己不是好像觉得我知觉到了我是有意识的吗?这些含义的荒谬性清楚地表明:我可能知觉到我有意识这一概念具有无意义的特征。

非常奇怪的是,居然有这样的看法,即“我是有意识的”或“我有意识”是现有的最根本的命题——甚至比“我存在”更根本。当然,只有当从事哲学事业,例如探寻“人类知识的基础”时,人们才有这样的倾向。但是如果它不是根本的命题,还能算作一种“命题”吗?维特根斯坦设想,它是这样被说出的:“没有什么有像我具有意识那样确实”——它还能这样表述:“‘我有意识’——这是一种关于无可置疑地可能的东西的陈述”。针对后一说明,维特根斯坦问道:“为什么就不能同样地说:‘我有意识’不是一个命题呢?”^②

“我有意识”肯定是一个句子。问它是一个“命题”还是一个“陈述”,这会是什么意思?我们必须注意下面几点:第一,“我有意识”这个句子不能用来报告知觉或观察,正如我们已知的那样;第

352 第四部分

二,该句子并没有有意义的否定;第三,即使在某些情况下,我说出这个句子可能使某人相信我是有意识的,但这仅仅是因为我说出了这个句子,而不是因为该句子的“命题内容”。这几点足以表明:“我是有意识的”这个句子并没有陈述事实的用法:它并没有报告或描述说者的状态。这可能就是维特根斯坦主张它不是“命题”的意思。

“我是有意识的”、“我有意识”这类句子及其否定形式作为条件陈述的前件有真实的用法,例如“当我手术后有了意识时,请让我的弟弟到医院来一趟。”但作为第一人称、现在时、简单直陈句,它们是可以从语言中排除出去的,而不会有任何损失。这与第三人称句子“他/她是有意识的”形成了鲜明的对比,后者能用来表达观察、假设或猜测。简单直陈句“我是有意识的”无疑是一个句子。它为维特根斯坦的下述评论提供了极好的说明:

并非每个似命题的组成形式我们都知道派何用处……;当我们在哲学中被误导而把某些无用的东西当作命题时,那常常是因为我们没有充分考虑到它的运用。^②

觉知到有意识。以为人们不能知觉到自己有意识,这种看法可能产生严重的恶果。“如果我没有知觉到我自己的意识,那么我怎么能够觉知到我有意识呢?”但是难道存在着觉知到有意识这样的事情吗?人们在什么时候说出这种事情?有可能设想它在其中被说出的情境。假设一个朋友近来在一般的麻醉条件下接受手术。你可能问她:“你是逐渐还是突然恢复意识的?”她大概会说:“这是一种再明显不过的经验。我突然觉知到自己完全有意识

了!”但她还可能说:“这是一显而易见的经验,我突然完全有意识了!”试把她的说明当作这样的信息,即“觉知到”这些词什么信息也没有添加。(它们只是给予特别的强调。)说者并没有把“觉知到”和“未觉知到”加以比较——因为她说“我突然完全有意识了,但并未觉知到它”,这是没有意义的。这个人的意思可能是:她不仅突然成为完全有意识的人,而且也许还带着吃惊的情感对自己说:“我完全有意识了!”如果这就是我们对她的表白即她“觉知到”有意识的理解,那么很显然,既然人们并不是经常或并非永远对自己这样说和这样想,那么就没有根据主张:每当人们是有意识的时,都觉知到了它。

从外面知觉到的意识。我们已反思了意识概念,包括意识到某物和纯意识。主流的哲学传统主张:意识的状态、事件和进程只能为人从内部知觉到,绝不能为那个人的观察者知觉到。事实恰恰相反。略为带点倾向性地说,意识的现象与其说是从内部知觉到的,倒不如说是从外面知觉到的。

在当今反思心理学科学的哲学家中,心理学能否处理所谓的“意识”、“经验的主观特征”、“心理状态的质的特征”和“感受性质”之类的问题已受到了高度的关注。T. 内格尔说:

究竟该怎样解释心理现象的物理本质,我们目前还没有什么主意。没有意识,心身问题索然无味。而有意识,它似乎又无望解决。^②

N. 布洛克说:

我不知道,心理学以现在那个样子怎么能解释感受性质……

354 第四部分

感受性质仿佛游离于心理学之外。^⑥

维特根斯坦评论说：“心理学处理心灵领域内的过程就像物理学处理物理领域内的过程一样”，此说法是“一种错误的类比”；他还说：

物体的运动、电子现象等不是物理学的课题，在同样的意义上，看、听、想、感觉、意愿都不是心理学的课题。只要你注意这样的事实，即物理学家看、听、想那些现象，并告之于我们，然后心理学家观察那被试的外在反应（行为），你就不难明白这一点。^⑦

维特根斯坦的评论与内格尔、布洛克的观点是一致的吗？后者认为：经验心理学只能处理被试的生理过程和物理行为，而不能处理心灵现象和意识。不。下面的进一步评论能说明这一点：

那么，心理学探讨行为而不涉及心灵吗？心理学家记录的是什麼——他们观察的是什麼？那不是人的行为尤其是他们的话语吗？但这些并不是关于行为的。^⑧

心理学实验的人类被试会说出各种事情，例如“光现在似乎更明亮了”，“我开始感到有点晕了”等。这些话语是报告，不是关于被试的行为的，而是关于他的感觉的。因此被试的感觉，而不仅是他的言语行为成了实验材料的组成部分。被试的这些报告是关于意识中的变化的，是关于“经验的质的特征中”的变化的。这样一来，意识和“感受性质”显然在心理学的领域之内。实验可能会证明环境或被试身体中的物理变化与被试的感觉变化之间的因果联系。与

内格尔相对立,这是关于心理现象的物理本质的一种合理“解释”。

能表达感觉、感情、情绪、怀疑的不仅仅是话语。眼睛变小、嘴唇紧缩、耸肩、面色变苍白都可能体现猜疑或意志的坚定或恐惧。正如普劳斯特所说,“真理之显现并不一定非要通过话语,人们也许可以根据无数的外在标记更确切地得到它,而用不着等着说出来,甚至用不着对它们作出什么解释。”因此要理解内格尔和布洛克为什么认为:意识的运动或感受的质不能经过对那些“外在标记”的观察而得到分辨或解释,那是不容易的。在第2节中,我们将考察他们的观点。

环境中的行为。让我们回想一下,我们关于意识包括及物和非及物我们说了些什么。我们记得,布伦塔诺有两个论断,一是“心理现象只能为单个的个体所知觉到”,二是“它们只能在内在的意识中被知觉到,这是所有心理现象的一般特征”。他似乎提出了下述两个主张:第一,除了我自己,别的人永远不可能知觉到我是惊恐的、生气的、想做某事或意识到什么东西;第二,每当我惊恐、生气等时,我自己肯定有知觉。这两个主张实际上使认识倒退了。别的人常常知觉到我的感情、意向、思想。他们是怎样知觉到的呢?通过注视我的面部表情、姿势、动作,听我的呼叫、说明和提问。粗略地说,根据我的行为,他们可以了解我。

这种粗略的说法可能引起误解。因为能够产生关于我的感情和思想的知识的东西并不是狭隘地被观察到的行为,而是某些环境中的行为。假设在一个聚会上,我厌恶地从我正与之谈话的一群人中走开了。我的一个朋友观察到了。他怎么可能?我的离开难道不是由于无聊、厌恶吗?我的朋友怎么可能知道这一点呢?

356 第四部分

噢,他听到了那群人中其他的人所说的话。他十分清楚我的确信或偏见,因此认识到那些话语对我来说就是厌恶的诱因。另一个观察者可能也看到了我突然从那群人中离去了,但未知觉到它是厌恶的表现——因为他并没有听到那些话语,或者听到了,但未能理解,或不知道我有那些坚定的信念。

布伦塔诺的第二个主张是:如果我是厌恶的,那么我就一定知觉到了我的厌恶。这一半是错误的,一半是不合理的。之所以错误,是因为我不可能理解我的行动。如果事后立即问个究竟,那么我可能没法对我的粗鲁的离去作出真实的解释,这也许是因为反思从未使我充分觉知到我的那些使人不舒服的情感;对我相当了解并知觉到我的行动的某一个人对此可能有比我更好的理解。但是另一方面,布伦塔诺的第二个主张与其说错误,不如说含糊不清。假如我完全清楚我厌恶地离开了,在我行动之后能立即正确地予以解释的意义上,我“知道了”这一点。对于布伦塔诺而言,说我厌恶地离开的那一时刻,我知觉到了厌恶,这是什么意思?噢,这意思是:在我自思自忖的时候,“那些话语令我厌恶!”这可能真的发生了,但也可能没有发生。在后一种情况下,根据布伦塔诺的观点,这样的现象,即并不是我的厌恶而是我对我的厌恶的知觉,究竟是什么?我不知道。这就是布伦塔诺的主张的不合理方面。只有在某些情境下,才能说一个人知觉到(或觉知到)了他或她自己关于恼怒、轻蔑、厌恶等(包括各种形式的及物意识)的感受。

至于不及物意识或无意识,说某人正觉知到有意识或无意识就更没有意义。以为有意识的人一定觉知到有意识这种看法的一个根源就是这样的认识,即“他没有觉知到有意识”是一种奇怪的

说法。的确奇怪！不难想到这样一种事例，在那里，我们能说 A 没有觉知到 B 有意识：手术中的病人失去了意识，但几分钟后又恢复了意识；而外科医生身子转过去了，并没有看到意识恢复的标志，因此没有觉知到那个病人是有意识的。但是在什么条件下，我们才能说那个病人觉知到有意识？我不知道。如果病人喃喃地说“我是无意识的”或“我不觉得有意识”，或者在从手术中苏醒过来后便说“我有意识了但没有觉察到”，那么我们才能那样说吗？肯定不是这样。“他/她没有觉知到有意识”似乎是一个我们不愿在任何情况下使用的句子。但这又能支持下述观点：有意识的人一定觉知到有意识。恰恰相反！因为觉知到有意识据说与没有觉知到有意识形成了对比：如果后者是子虚乌有，那么这个对比也是如此。

自我意识。我一直没有论及自我意识。它是什么？“有自我意识”有一常见的用法，根据这种用法，处在自我意识中而为他人注意到时，就是处在窘迫的、胆怯的、局促不安的状态中。这与我们的探讨没有关系。还有一种关于“有自我意识”的哲学概念，据说它指的是每个人关于一种独一无二的对象的意识，这个对象对我们每个人来说各不相同，且可用代词“我”加以表示。我赞成安斯康帕的下述看法：“我”一词不能用来指示任何对象，这个哲学概念是一种幻觉。

安斯康帕进而还宣称：“自我意识”是“某种真实的东西”。她说：“‘自我意识’这一表达式可以正确地解释为：‘某某关于自己的意识’。”^②在“自我意识”这个名称下，她想说的是任何人所具有的、报告他/她的身体姿势、运动、感觉、意向的正常能力——这些

358 第四部分

报告一般是真实的,即使不是根据观察。因此安斯康帕强调应注意的事实是很重要的。在哲学中人们可以把“自我意识”这一表达式作为一个技术术语,以指称这些事实。但应予以注意的是:在日常语言中不能这样加以使用,因为并非每个正常的人都能被说成是“有自我意识的”:对他们的区分,正好是根据他们是不是“有自我意识的”。

在日常交流以及文学作品中,一个人如果对他自己的态度、兴趣、人格作出了大量的反思,对他对他人的、人类生活的状况的反应作出了大量的反思,他对音乐、文学、艺术和自然的反应作出了大量的反思,那么就常常被说成是“有自我意识的”。在这个意义上,有自我意识的人在他们的自我意识的程度上是有很大的差异的。这种自我意识常常为别的人对某个自我的观察所引起,并由之而得到发展。人自己成了人研究的一个对象。

注 释

①②③④⑤ G.E. 摩尔(1922):《哲学研究》,哈考特·布雷斯公司,p. 7、17、24、29。

⑥ K.V. 威尔克斯(1981):“功能主义、心理学与心灵哲学”,载《哲学论坛》,12,p. 165。

⑦ D. 丹尼特(1978):《脑猝病》,哈维斯特出版社,p. 221。

⑧⑨⑩⑪ L. 维特根斯坦:《卡片》,G. 安斯康帕等编,G. 安斯康帕译,布莱克韦尔,1967,§ 403、§ 553、§ 557、§ 566。

⑫ 维特根斯坦:《哲学研究》,G. 安斯康帕等编,G. 安斯康帕译,布莱克韦尔,1967,p. 227。

⑬ 摩尔:《哲学研究》,p. 25。

⑭⑮ J. 洛克:《人类理智论》,牛津大学出版社 1894,第 1 卷,第 2 篇,第 1 章,§ 19,第 27 章,§ 11。

⑩⑪⑬⑭ F. 布伦塔诺:《从经验主义观点看心理学》, 菲利克斯·迈纳尔, 1955, 第1卷, 第2篇, 第1章, § 5、§ 6、§ 6、§ 6。

⑮⑯ 维特根斯坦:《哲学研究》, § 281、§ 417。

⑰《卡片》, § 401, § 402。

⑱ 维特根斯坦:《哲学研究》, § 520。

⑳ T. 内格尔(1974):“成为一只蝙蝠可能是什么样子?”载 N. 布洛克编:《心理学哲学读本》, 1980, 第1卷, p. 159。

㉑ N. 布洛克:“功能主义的麻烦”, 同上, p. 289。

㉒㉓ 维特根斯坦:《哲学研究》, § 571, p. 179。

㉔ G. 安斯康帕(1981):“第一人称”, 载《G. 安斯康帕论文集》, 布莱克韦尔, 第2卷, p. 26。

勇敢地面对意识难题*

大卫·J. 查默斯

I. 导言

在关于心灵的科学中,意识提出了最令人困惑的问题。我们直接知道的除了有意识的经验,再别无其他,但是再没有什么比这更难解释的了。近年来,所有的心理现象都受到了科学的研究,而唯有意识顽固地予以抵制。许多人试图解释它,但是那些解释似乎总是无的放矢。还有些人无可奈何地主张:意识问题太难对付了,好的解释可能是无望的。

为了推进意识问题的研究,我们不得不直接面对它。在本文中,我先把该问题真正困难的部分独立出来,即把它与更易于处理的部分分开,并说明它为什么如此难解。我还将对近来用还原方

* 由于量子力学等现代自然科学的发展,意识的本质与作用问题在当代西方心灵哲学中成了格外引人注目的问题,以致产生了研究意识的新的专门学科,出现了专门反映该领域成果的刊物:《意识研究杂志》,还有专门以此为专题的专题讨论会,如1994年在美国举行了“关于意识的科学基础”的“具有里程碑意义”的会议。在会上,查默斯提出的“意识的困难问题”引起了重大争论。会后,作者以此报告为基础写出了该论文。最初发表于《意识研究杂志》,2, No. 3(1995)。不久又收入J. 谢尔主编的会议论文集《解释意识——“困难问题”》(麻省理工学院出版社,1995)一书中。这里根据后者译出。

法研究意识的作法提出批评,并证明这些方法绝不可能解决该问题的最难的方面。一旦认识到这个缺点,通向灿烂前景的大门就将敞开。在本文的后半部分,我要论证的是:如果我们转向一种新的非还原解释,那么就有可能对意识作出自然主义的说明。我愿为这种说明提出我自己的候选方案:以结构上的融贯性、组织上的不变性原则以及关于信息的两面观为基础的非还原理论。

II. 容易问题与困难问题

意识并不只是一个孤立的问题。“意识”是一个模糊的术语,指的是许多不同的现象。其中每一个都需要予以解释,但一些比另一些更难解释。在开始,把意识的连在一起的问题区分为“容易的”和“困难的”问题两类是大有裨益的。意识的容易问题是这样一些问题,它们似乎可直接接受认知科学的标准方法的处理,基于此,一种现象可用计算或神经机制的术语予以解释,而困难的问题则似乎是抵制这些方法的问题。

意识的容易问题包括对下述现象作出解释的种种问题:

- * 对环境刺激作出分辨、范畴化和反应的能力;
- * 通过认知系统对信息的整合;
- * 心理状态的可报告性;
- * 系统理解自身的内在状态的能力;
- * 集中注意力;
- * 行为的深思熟虑的控制;
- * 清醒与睡眠之间的区分;

362 第四部分

所有这些现象都与意识概念有关系。例如人们有时说:当心理状态可用语言报告时,或当它可通过内在的方式理解时,它便是有意识的。当一系统有对信息源作出反应的能力时,或更明确地说,当伴随那信息时,或当我们整合那信息并在审慎的行为控制中利用它时,该系统有时便被认为有关于某种信息的意识。我们有时候说,当一行动是有意作出的时,它显然是有意识的。通常,我们用另一种方式说有机体是清醒的,其实是说它是有意识的。

这些现象能否从科学上加以解释,对此不存在真正的争论。所有这些都易于直接根据计算的或神经的机制而加以解释。例如为了解释其可理解性和可报告性,我们只需详述这样的机制就行了,如通过它,关于心理状态的信息被提取,并被用之于语言报告。为了解释信息的整合,我们只需说明这样的机制,通过它们,信息得到聚合,并为后来的过程所利用。为了说明睡眠和清醒,对下述过程作出适当的神经生理学说明就够了,这些过程是专门负责有机体在那些状态中对行为的比较的。在每种情况下,一种相应的认知或神经生理学模型就能出色地完成那种解释任务。

如果这些现象就是意识的全部,那么意识就不是什么大的难题了。即使我们还没有接近对这些现象的完满解释,但对我们该怎样去解释它们,我们是有胸有成竹的。这就是我为什么称之为容易的问题的理由。当然,“容易”是一个相对的术语。完全搞清其细节大概需要一个或两个世纪的艰苦的、经验性的研究。不过仍有充分理由相信:认知科学和神经科学的方法会取得成功的。

意识的真正困难的问题就是关于经验的问题。当我思考和感觉时,有一信息加工过程的匆匆而过,但也有一主观的方面。正如

内格尔(1974)所论述的,存在着可能作为有意识的有机体的某东西。这个主观的方面就是经验。例如当我们看时,我们就经验到一种视感觉:红色的被感觉到的质,关于黑暗和光亮的经验,视域的深度的质。其他的经验会以不同的方式伴随着知觉:清脆的声音,樟脑丸的气味。还有躯体感觉,从疼痛到极度兴奋;能内在地呈现出的心像;情绪的被感觉到的质,对有意识思想之流的经验。把所有这些状态统一起来的東西是:有某东西可能存在于它们之中。所有这些_{都是}经验的状态。

不可否认,某些有机体是经验的主体。但这些系统怎样成为经验的主体这一问题则是复杂难解的。当我们的认知系统进行视觉和听觉的信息加工时,我们有视觉或听觉经验:深蓝色的质、对中C调的感觉,这是为什么?我们怎样解释:为什么存在着可能持有心象或经验到情绪的某东西?一般承认:经验来自于一定的物理基础,但是对它为什么如此起源、怎样起源,我们并没有找到令人满意的解释。物理加工为什么引起丰富的内在生活?说它应当如此,以及事实上如此,这似乎都是客观上难以理喻的。

如果有什么问题有资格成为意识问题,那么非它莫属。在“意识”的这一核心意义上,如果有某东西可能成为那有机体,那么它就是有意识的,如果有某东西可能处在一种心理状态之中,那么那心理状态就是有意识的。在这种场合,有时也使用“现象意识”、“感受性质”等术语,但我觉得:说“有意识经验”或简单地说“经验”,则更加自然。另一避免混乱的方法(例如纽厄尔 1990、查默斯 1996 曾用过)就是继续用“意识”一词来表示经验的现象,并用不拖泥带水的术语“觉知”(awareness)表示上述更直接的现象。

364 第四部分

如果这一约定被广泛接受了,那么交流就更加容易。照目前情况看,那些谈论“意识”的人常常相互谈论过去的经历。

“意识”这一术语的模糊性常常为从事该课题写作的哲学家和科学家所利用。翻一篇关于意识的论文,首先映入眼帘的常常是意识的神秘性,如强调主观性的奇怪的不可触摸性和无法言喻性,抱怨我们至今尚无关于这种现象的理论。在这里,该课题显然就是那个困难问题——关于经验的问题。在本文的后半部分,调子将变得更为乐观,作者自己关于意识的理论将被概述。基于考察,这一理论最终将是关于更直接的现象中的一种——如关于可报告性、内省通道或别的什么——的理论。在开始,作者将表明:意识终究是可触及的,但读者可能会觉得好像是“上钩调包诱售法”^{*}的受害者。困难的问题依然故我。

Ⅲ. 功能解释

容易的问题为什么是容易的,困难的问题为什么是困难的?容易的问题之所以容易,显然是因为它们涉及的是对认知能力和功能的解释。而要解释认知功能,我们只需具体说明能够实现功能的机制就够了。认知科学的方法最适合于这类解释,因此也适合于关于意识的容易问题。相比较而言,困难问题之所以困难,主要是因为它不是关于功能执行的问题。即使所有有关功能的执行都得到了解释,该问题依然如故。^①

* 意即用廉价商品引诱消费者上圈套,然后兜售贵重商品。

例如,解释可报告性正好就是解释一系统怎样执行对内在状态的报告的功能。而为了解释内在通道,我们就必须解释一系统怎么可能恰如其分地受到它的内在状态的影响,怎样在随后的过程中使用关于那些状态的信息。为了解释整合与控制,我们必须解释:系统的中心加工怎样把信息内容集合在一起,然后怎样在促发各种行为时调用它们。这些都是关于功能的解释问题。

我们怎样解释功能的执行呢?就是详细说明完成那功能的机制。在这里,神经生理学的和认知的模拟可以尽善尽美地完成该任务。如果我们想要一详细的、低层次的解释,那么我们可以述及对那功能负责的神经机制。如果想要更抽象的解释,我们就可以用计算术语描述一种机制。无论用哪种方法,都可得到一种完全令人满意的解释。例如一当我们陈述了执行语言报告功能的神经或计算机制,那么我们解释可报告性的全部工作便大功告成。

从一定的方面说,这个观点没什么价值。它们的解释只涉及到对各种功能的解释,这是一个关于这些现象的概念事实,因为这些现象可以从功能上加以定义。可报告性在一系统中得到例示这一点所意味的不过是:该系统有对内在信息的语言报告能力。一系统是清醒的这一点所意味的则是:它适于接受来自环境的信息,并能用这些信息以一定的方式指导行为。要明白这类事情是概念事实,就得注意到:说“你解释了执行语言报告的功能,但没有解释可报告性”的人对可报告性犯了一点概念错误。说它解释了可报告性,充其量可能只解释了有关功能怎样被执行。这对所讨论的其他现象都是适用的。

在全部更高层次的科学中,还原解释正是以这种方式发挥着

366 第四部分

作用。例如为了解释基因,我们必须具体说明这样的机制,它储存遗传信息,并把它从一代传递给下一代。最后证明是 DNA 完成这个功能;一当我们解释了该功能怎样完成,那么我们就解释了基因。为了解释生命,我们首先得解释:系统怎样再生,怎样适应它的环境,怎样产生代谢变化等。所有这些都是关于功能执行的问题,因而非常适合于还原解释。这对认知科学的大多数问题也是适用的。为了解释学习,我们必须解释系统的行为能力由以根据环境信息而发生改变的方式,以及新的信息在让系统的行为适应环境时发生作用的方式。如果我们说明了神经的或计算的机制怎样完成这一工作,那么就解释了学习。对其他认知现象诸如知觉、记忆、语言都可以如是说。有关的功能有时需要细致地加以描述,但很显然,既然认知科学要解释这些现象,那么它就可通过解释那些功能的完成而做到这一点。

当涉及到有意识经验时,这类解释就无能为力了。使困难问题成为困难的、独一无二的问题的东西是:它远远超出了功能执行的问题。要明白这一点,只须注意到:即使我们解释了与经验毗邻的所有认知和行为功能的执行——知觉分辨,范畴化,内在通道,语言报告——仍可能存在一个进一步的、未回答的问题:这些功能的执行为什么为经验所伴随。只简单地解释功能对这个问题是无济于事的。

在关于基因或生命或学习的解释中找不到类似的问题。如果某人说:“我能够明白你对 DNA 怎样储存信息并把它从一代传递到另一代的解释,但你并没有解释:它怎样成为一个基因”,那么他便犯了一个概念错误。成为一基因所意味的不过是:成为一完成

有关储存和传递功能的实在。但如果某人说：“我能明白你对信息怎样被分辨、整合和报告的解释，但你没有解释它是怎样被经验的”，那么他就没有犯概念错误。这是一个深层的、并非没有必要的问题。

这深层的问题在关于意识的难题中是关键的问题。为什么所有这类信息加工没法摆脱内在感觉而在“黑箱”中进行？当电磁波冲击视网膜，进而为视觉系统分辨和范畴化时，分辨和范畴化总是作为一种关于逼真的红色的感觉而被经验到，这是为什么？当这些功能被完成时，我们就知道有意识经验出现了，但它出现这一事实正好是一关键的秘密。在功能和经验之间存在着一解释鸿沟（从莱文那里借用的术语，1983），我们需要通向它的解释桥梁。纯粹的功能说明仍停留在那鸿沟的一边，因此那桥梁的材料必须在别的地方去寻找。

这并不是说：经验没有功能。它也许起着重要的认知作用。但对于它所起的任何作用来说，除了对功能的简单解释外，还应有更多的对经验的解释。甚至也许还有这样的结果，即在对一种功能作出解释的过程中，我们将到达一种至关重要的见解，它承认对经验的解释。倘若如此，那么该发现将成为额外的解释报偿。不存在这样的认知功能，以至于我们可以进一步说：对该功能的解释会自动地解释经验。

为了解释经验，我们需要一种新的方案。认知科学和神经科学的通常解释方法是远远不够的。这些方法显然是为解释认知功能之执行而提出的，在这方面，它们功不可没。但就这些方法本身而言，它们只适合于解释那些功能的执行。一当面对困难问题时，

标准的方案便束手无策。

IV. 某些案例研究

近年来,围绕意识问题在认知科学和神经科学的框架内展开了大量的研究工作。这大概可以说明上述分析是有缺陷的,但事实上对有关工作的直接考察只会使那些分析得到进一步的支持。当我们探讨这些研究指向意识的哪些方面、它们最终解释了哪些方面时,我们便发现:解释的根本目标总是指向某一容易问题。我将用两个有代表性的例子说明这一点。

第一个是由 F. 克里克和 C. 考克(1990;另参阅克里克 1994)所略述的“关于意识的神经生物学理论”。该理论集中注意的是大脑皮层中 35-75 赫兹的神经振荡;克里克和考克假定:这些振荡是意识的基础。这部分是由于:那些振荡似乎以许多不同的方式——例如在视觉和嗅觉系统之内——与觉知有关;还由于:他们设想这样一种机制,正是通过它,信息内容之汇聚才可能完成。汇聚是一个过程,基于此,关于单一实在的分别被表征的信息单元结合在一起以用之于后续加工,就像关于被知觉对象的颜色和形状的信息从不同的视觉通道整合起来一样。根据其他人的看法(如埃克霍恩等,1988),克里克和考克假定的是:汇聚是由表征有关内容的神经集合体的同步振荡完成的。当两个单元的信息结合在一起时,有关的神经集合体就以相同的频率和相位振荡。

对这种汇聚怎样完成的详情的理解仍很贫乏,但姑且假设它们能被弄清。最终形成的理论能解释什么呢?它解释的显然是信

息内容的汇聚,也许它还对大脑中信息的整合提供了一般性的说明。克里克和考克也主张:这些振荡激活了工作记忆的机制,因此对这一点及其他长期记忆形式可能有一种说明。该理论最终达到的是这样一种一般性的说明,即说明被知觉的信息怎样在记忆中汇聚和储存以为后续加工所利用。

这样一种理论是有价值的,但它对有关内容为什么被经验,什么也没有告诉我们。克里克和考克设想:这些振荡是经验的神经关联物。这个主张是值得商榷的——汇聚不是也发生在对无意识信息的加工之中吗?——即使可被接受,但解释的问题依然如故:那些振荡为什么引起了经验?解释的因果关系的唯一基础是它们在汇聚和储存中所起的作用,但汇聚与储存本身为什么伴随着经验的问题从未被触及到。如果我们不知道:汇聚与储存为什么引起经验,那么只陈述关于振荡的故事是无济于事的。反过来,如果我们知道汇聚与储存为什么引起经验,那么神经生理学的细节就像是糕饼表层的酥皮。克里克和考克想通过假定汇聚与经验之间的联系而实现其目的,因此对解释那个联结一点用也没有。

我认为,从根本上说,克里克和考克并不是要提出那个困难问题,即使某些人从别的方面对他们作了阐释。克里克的一个已发表的访谈对该理论的动机的限度作了清楚的陈述。

好,让我们先忘掉真正困难的方面,如主观感觉,因为对它们不可能有科学的裁决。游戏、疼痛、快乐、看到蓝色、嗅到玫瑰花味等的主观状态——在解释分子和神经元的唯物主义层面与主观的层面之间似乎有天壤之别。让我们集中在那些容易研究的事情上——如视觉觉知。你正在对我讲话,但你没有

370 第四部分

注视我,你注视的是卡普金氏棕色咖啡,因此你觉知到了它。你可以说,“它是一个杯子,里面有某种液体。”如果我把它给你,你就会伸过手来,并喝下它——你将以有意义的方式作出反应。这就是我所谓的觉知。(“什么是意识?”《发现》,1992年12月,p.96)

第二个例子是认知心理学层次的一个方案。这就是B.巴思在他的《关于意识的认知理论》(1988)一书中表达的、关于意识的整体工作空间理论。根据该理论,意识内容包含在整体工作空间中,这是一种用来沟通许多专门化的无意识加工器之间的联系的中心加工器。当这些专门的加工器必须把信息传到其他系统时,它们是通过把这些信息传送到工作空间而实现这一点的,工作空间是作为其他系统的一种交流黑板而起作用的,可通向其他所有加工器。

巴思用这一模型说明人类认知的方方面面,解释有意识与无意识认知功能之间的各种反差。不过,从根本上看,它是一种关于认知可存取性的理论,解释了某些信息内容怎么可能在一个系统内得到广泛存取,同时它也是一种关于信息整合和可报告性的理论。这一理论表明,它有望成为一种关于觉知、关于有意识经验的功能关联物的理论,但可惜没有提供关于经验本身的解释。

人们可能设想,根据该理论,经验的内容显然就是工作空间中的内容。但即便如此,它所包含的内容无一能解释:在整体工作空间内的信息为什么能被经验到。该理论充其量只是说:信息之所以被经验,是因为它在整体上是可存取的。但问题又以不同的形式冒出来了:整体的可存取性为什么引起了有意识经验?总之,这

种基础性的问题还是没有得到回答。

这一批评同样适用于近年来对意识的认知或神经科学探讨的几乎所有工作。例如埃德尔曼(1989)的“神经达尔文主义”模型就陈述了关于知觉觉知与自我概念的问题,但是对为什么存在着经验只字未提。丹尼特(1989)的“多样支取”(multiple draft)模型在很大程度上是想解释某些心理内容的可报告性。杰肯多福(1987)的“直接层面”理论对作为意识基础的某些计算过程提供了说明,但他强调,这些“计划”怎样进入有意识的经验这样的问题仍是秘密。

使用这些方法的研究者在对有意识经验这个问题的态度上常常是暧昧的,即使他们有时有明确的立场。然而在那些对之清楚的人中,态度又有很大的差异。在从事这类关于经验问题的工作时,各种不同的战略是唾手可得的。如果这些战略选择更经常地得到阐释,那么就是有用的。

第一个战略就是对别的某物的解释。某些研究者阐释说:迄今为止,经验问题太难了,或许它完全超出了科学的范围。基于此,这些研究者情愿选择更可能触及的问题中的某一个,诸如可报告性或自我概念。即使我把这些问题称之为“容易的”问题,但它们仍属于认知科学中最有兴味的、尚未解决的问题,因此这项工作肯定是有价值的。关于这一选择所能说的最糟的方面是:就意识研究的状况来说,它是相对没有什么抱负的,而且这项工作有时可能被误解。

第二种选择就是采取更强硬的路线,直至否认那现象。(这一方案的变种为奥尔波特 1988,丹尼特 1991,威尔克斯 1988 所接

372 第四部分

受。)按照这条路线,一当我们解释了像可存取性、可报告性等功能,那么就再没有所谓的“经验”这样的现象要解释了。这显然是否认了那种现象,因为据说不能从外在方面证实的东西就不可能是真实的。其他人通过承认经验存在而得出了同样的结论,但只有当我们把“经验”与分辨、报告之类的能力等同起来才能如此。这些探讨方案导致了一种更简单、但从根本上说是不令人满意的理论。经验是我们心理生活的最核心、最明显的方面,也许是心灵科学中关键的有待解释的东西。因为处在这样一种有待解释的地位,因此当新的理论到来时,经验就不能像生命精气那样被丢弃。毋宁说,它是关于意识的任何理论必须解释的核心事实。否认这种现象的理论不过是用回避问题的方式“解决”问题。

在第三种选择中,某些研究者倡导在完全的意义上对经验作出解释。这些研究者(不像上述那些人)希望严肃地对待经验;他们宣传他们的功能模型或理论,主张:它解释了经验的全部主观的性质(如弗洛尔 1992;汉弗莱 1992)。不过,解释中的有关步骤常常一晃而过,以致最后把某物看作戏法一样的东西。在提供了关于信息加工的某些细节之后,经验突然出场了,但这些过程究竟怎样突然引起经验还是晦暗不明的。也许只是简单地假定如此,但这一来我们就有一种不完全的解释进而有后面的对第五种战略的阐释。

第四,更有希望的方案诉诸于解释经验结构的那些方法。例如可以论证说:说明了视觉系统所作的分辨就可以说明不同颜色经验之间的结构关系,还能说明视域的几何学结构(例如参看克拉克 1992;哈丁 1992)。一般来说,在加工中所发现的关于结构的某

些事实与关于经验结构的事实是对应的,并且可以有根据地对后者作出解释。这种战略是合理的,但有局限性。充其量,它只是承认了经验的存在,说明了关于它的结构的某些事实,对经验的结构方面提供了一种非还原的解释(在后面我再予以详述)。这对许多目的来说是有价值的,但对究竟为什么存在着经验,它什么也没有告诉我们。

第五个也是较合理的战略就是把经验的基质分离出来。几乎每个人都承认:经验是以一种或另一种方式从大脑过程中产生出来的,因此把它由之而起的那类过程分辨出来是有意义的。例如克里克和考克公布了他们关于分离意识的神经关联物的成果,埃德尔曼(1989)和杰肯多福(1987)也提出了相关的主张。对这些主张的证明离不开细致的理论分析,这主要是因为经验在实验环境中是不能被直接观察到的,但是当审慎地加以应用时,这种战略对经验的问题又有间接的作用。尽管如此,这种战略显然是不完善的。因为对于一种令人满意的理论来说,我们不只是要知道哪些过程引起了经验;我们还必须看到关于为什么和是怎样的说明。一种关于意识的圆满的理论必须架起一座解释桥梁。

V. 额外的构成成分

认知科学和神经科学的常见方法为什么不能说明有意识的经验,我们已清楚了其常见的理由。它们显然是错误的方法:它们给予我们的东西都不可能算作是解释。要说明有意识经验,在解释中我们还必须有一个额外的成分。这对那些严肃地对待意识的因

374 第四部分

难问题的人来说无疑是一种挑战:你的额外的成分是什么,它为什么能说明有意识经验?

事实上,并不缺乏额外的成分。某些人建议引入混沌和非线性的动力学。某些人认为:关键在于非算法加工。还有些人求助于神经生理学的未来发现。另一些人则假设秘密的钥匙存在于量子力学的层面。不难弄明白,所有这些建议为什么被提出来。旧的方法已不管用了,因此问题的解决只有寄希望于某种新的东西。不幸的是,这些建议都面临着同样的老问题。

例如非算法加工是彭罗斯(1989;1994)鉴于它在有意识的数学洞察过程中可能具备的作用而提出的。这类与数学有关的论证是有争论的,但即使它们成功了,即使对人脑中的非算法加工的说明被给予了,它仍只是一个关于数学推理等所涉及到的功能的说明。因为非算法过程与算法过程一样,也没有回答这个问题:这一过程为什么引起了经验?在回答这个问题时,非算法加工并不具有特殊的作用。

这也适用于非线性和混沌动力学。它们对认知功能的动力学也许提供了新颖别致的说明,以致与认知科学中的标准方法所给予的说明泾渭分明。但根据动力学,人们只能得到动力学。关于经验的问题在这里像以前一样还是神秘的。这一点对于神经生理学中的新发现来说就表现得更为明显。这些新发现可能有助于我们在理解人脑功能时取得重大进展,但是对于我们所分离出来的任何神经过程来说,同样的问题还是会冒出来。很难想象,新的神经生理学的创立者有望在对深层认知功能的解释之上意外碰到什么奇迹。我们似乎不能在神经元中突然发现一种现象的光芒!

所有“额外成分”中最时髦的也许要算量子力学(例如哈默罗夫 1994)。关于意识的量子理论的诱人之处可能源于关于神秘性的最小化规律:意识是神秘的,量子力学是神秘的,因此两个神秘性也许有共同的源泉。尽管如此,关于意识的量子理论像神经的或计算的理论一样碰到了同样的困难。量子现象有某些显著的功能属性,例如非决定论、非区域性。于是便会自然而然地设想:这些属性在认知功能的解释中可能起着某种作用,诸如随机选择和整合,而且这一假说不能先验地加以排除。但是当它用于对经验的解释时,量子过程像其他东西一样面临着同样的困难。这些过程为什么引起经验这一问题完全没有被触及到。^②

该说的都说了,现在可以断言,同一种批评可适用于关于意识的任何纯物理解释。对于我们详细说明的任何物理过程来说,将会有一种尚未回答的问题:这一过程为什么引起经验?假如有这样的过程,在经验没有出现时,它也在概念上一以贯之地被例示。据此可以得出结论说:对物理过程的纯粹解释都不会告诉我们:经验为什么产生。经验的出现超出了根据物理理论追根溯源的范围。

纯物理解释非常适合于解释物理结构,因为它根据详尽的微观构成成分解释宏观结构;而且它根据执行那些功能的物理机制对之作出说明,从而为功能的完成提供了令人满意的解释。这是因为,一种物理说明可能蕴涵着关于结构和功能的事实:一当物理说明的内在细节被给予了,那么结构与功能属性作为自动的结果便出现了。但是物理过程的结构与动力学只产生结构与动力学,因此结构和功能就是我们指望这些过程能够解释的全部东西。关

376 第四部分

于经验的事实不可能是任何纯物理说明的自动的结论,因为从概念的一致性上说,任何给定过程都能在没有任何经验的情况下保持其存在。经验也许来自于物理的东西,但它并不是物理的东西所蕴涵的。

所有这一切的启示是:你不可能不费力气地解释有意识经验。有这样一种显著的事实,即还原方法——完全根据更基本的物理过程解释高层次现象的方法——在如此众多的领域中的确是功勋卓著。在这个意义上,人们可以轻易地解释大多数生物和认知现象,因为这些现象都可被看作是更基本的过程的自动的结果。如果还原方法也能解释经验,那真是妙极了;长期以来,我也希望如此。不幸的是,存在着使这些方法必然惨败的非偶然的理由。还原方法之所以在大多数领域中成功,是因为在那些领域中需要解释的东西是结构与功能,而这些是物理说明所能蕴涵的那类东西。当面对处于结构与功能的解释之上的问题时,这些方法就不管用了。

这似乎使人想起了活力论者的主张:任何物理说明都不能解释生命,但这些事实不可同日而语。驱使活力论怀疑论的东西就是这样的怀疑,即物理机制能完成许多明显与生命有联系的功能,如复杂的适应行为和再生。功能解释就是所需的解释这一理性主张无疑是可以接受的,但活力论者由于缺乏关于生物化学机制方面的详尽知识便怀疑:任何物理过程能履行那项职责,进而提出了关于生命精气的假说,将其作为一种替代性的解释。一当物理过程最终证明能完成有关的功能,活力论的怀疑便烟消云散了。

另一方面,就经验来说,关于功能的物理解释是没有争议的。

关键倒是这样一个理性的观点,即对功能的解释不足以解释经验。这个基本的理性观点不是某种会进一步受到神经科学研究影响的东西。同样,经验与生命力风马牛不相及。生命精气作为一个解释性假定被提出来,旨在解释有关的功能,因此当这些功能不用这种精气也能得到解释时,它就被遗弃了。经验不是一种解释性假定,而是有自身独立地位的被解释项,因此不是这类排除的候选者。

值得注意的是,各种令人眼花缭乱的现象最终都可用物理术语予以解释。但它们的每一个都是关于物理对象的可观察行为的问题,属于结构和功能解释的问题之列。因为这一点,这些现象便总是物理说明大概可以解释的那类事物,即使在某种程度上有理由怀疑这种解释不会到来。从这些事例中所得到的富有诱惑力的归纳结论是不适用于意识的,因为它不是关于物理结构和功能的问题。意识问题以一种完全不同的方式令人迷惑不解。对该问题的分析向我们表明:有意识经验不是彻头彻尾的还原说明所能成功地解释的那类事物。

VI. 非还原解释

在这点上,某些人觉得没有希望了,认为我们绝不会有关于有意识经验的理论。例如麦金(1989)就论证说:该问题对我们有限的心灵来说是太难了;我们“在认知上被局限于”现象的范围内。其他的人则论证说:有意识经验完全超出了科学理论的范围。

我认为这种悲观主义为时过早。还没有到遗弃的那一步;相

378 第四部分

反它正是充满希望和兴味的地方。当排除了简单的解释方法时,我们就必须研究取而代之的东西。假如还原解释失败了,那么非还原解释就是自然而然的选择。

即使不可胜数的现象最终可根据比自身更简单的实在得到完全的解释,但这不是普遍的。在物理学中,一种实在不得被当作是根本的,这只是偶然的情况。根本的实在不能根据任何更简单的东西去解释。人们把它们当作基本的,为它们怎样相关于世界上的别的实在提供一种理论。例如在 19 世纪,电磁过程并不能根据先前的物理学理论所诉诸的全部力学过程而予以解释,因此马克斯韦等人引入了电磁电荷和电磁力,作为物理理论中的新的根本的组成部分。为了解释电磁论,物理学的本体论不得不加以发展。新的基本属性和基本规律必须对那些现象作出令人满意的说明。

物理理论视之为根本的其他特征还包括质量和时空。没有人打算根据更简单的东西解释这些特征。但这并不排除关于质量或时空的理论的可能性。关于这些特征怎样相互关联,关于它们所隶属的基本规律,是有一种复杂的理论的。这些基本原则通常被用来解释许多与质量、空间、时间有关的、高一层次的、熟悉的现象。

我主张:关于意识的理论应把经验当作根本的。我们知道关于意识的理论离不开对于我们的本体论来说是十分重要的某东西的附加物,因为物理理论中的一切与意识的空无是相容的。我们可以增加某些完全新的非物理的特征,经验正是从中派生出来的,但是这种特征可能是什么,这是难以澄明的。我们更有可能把经

验本身当作世界的根本特征,就像质量、电荷、时间、空间一样。如果我们把经验当作根本的,那么我们就有可能专心致志地从事意识理论的建构工作。

哪里有根本的属性,哪里便有根本的规律。关于经验的非还原理论将为基本的自然规律的大家庭注入新的原则。从根本上说,这些根本的原则将为关于意识的理论带来解释的重负。就像我们根据与质量和别的实在有关的更基本原则去解释与质量有关的、熟悉的、高层次的现象一样,我们也可以根据与经验和别的实在有关的更基本原则去解释熟悉的、与经验有关的现象。

尤其是,关于经验的非还原理论将详细说明基本的原则,这些原则告诉我们:经验怎样依赖于世界的物理特征。这些心理物理原则与物理规律互不相干,因为物理规律似乎已经形成了一个封闭的系统。毋宁说,它们是物理理论的附加物。物理理论为物理过程提供了理论,而心理物理理论则告诉我们:这些过程怎样引起经验。我们知道:经验依赖于物理过程,但我们同时也知道:这种依赖性不可能只是根源于物理规律。由非还原理论所假定的新的基本原则给了我们一个额外的组成成分,因此,我们必须建立一种解释桥梁。

当然,由于把经验当作根本的东西,因此在一定的意义上,这种方案就不能告诉我们:为什么一开始就存在着经验。但任何根本理论都有这个问题。在物理学中,就没有什么东西能告诉我们:为什么一开始就有物质,但是我们不能以此去反对关于物质的理论。任何科学理论必须把世界的某些特征当作是根本的。关于物质的理论仍能通过说明关于物质的各种事实怎样成为那些基本规

380 第四部分

律的结果而对之作出解释。关于经验的理论也是如此。

这一观点可以被称之为二元论的一种变体,因为它假定:在物理学所肯定的属性之外还有基本的属性。但它是一种良性的二元论观点,完全可与科学的世界观和睦共处。在这种方案中,不存在与物理学理论发生矛盾的任何东西;因此我们显然必须进一步增添桥梁原则,以解释经验怎样从物理过程中产生出来。在这种理论中,没有什么特别的唯灵论的或神秘的东西——它的整体的形态与物理学的形态是一样的,如包括一些与根本规律相联系的根本实在。它对本体论无疑有所推进,当然麦克斯韦肯定也做过同样的工作。更有甚者,这种观点的整个结构完全是自然主义的,因为它承认宇宙从根本上说可归结为遵循着简单规律的基本实在所构成的网络,它还承认:从根本上说,可能存在着一种根据这种规律系统地加以阐述的关于意识的理论。如果用一名称表示这个观点,那么最好的选择大概就是自然主义的二元论。

如果这一观点是对的,那么从某些方面来说,关于意识的理论与物理学中的理论的共同性比它与生物学中的理论的共同性就要多得多。生物学理论并不涉及到这种形式的根本原则,因此生物学理论对它来说有某种复杂性和凌乱性;但是物理学中的理论,既然要涉及到根本原则,因而追求简单性和协和性。自然的根本规律是世界的基本格局的组成部分,物理理论告诉我们的是:这个基本的格局显然是简单的。如果关于意识的理论也涉及到根本原则,那么我们也应有同样的渴求。推动物理学家寻求根本理论的简单、协和甚至优美原则对关于意识的理论也是适用的。^③

VII. 迈向意识理论

着手创建一种理论势在必行。我们已有条件理解关于物理过程和经验间的关系的某些事实以及关于把它们联系起来的规则的事实。一当放弃还原解释,我们就能把这些事实公诸于世,以便让它们作为初始的东西,作为对构成一种根本理论的基本规律的限制,在关于意识的非还原理论中发挥其适当的作用。

有一个显而易见的问题,它阻碍着意识理论的发展,那就是缺乏客观的材料。在实验条件下,有意识经验是不能直接被观察的,因此我们不能遂心如愿地得到关于物理过程和经验之间的关系的材料。尽管如此,就我们自己而言,我们是有通向丰富的材料源泉的通道的。经验与加工之间的许多重要的规则可以从对某人自己的经验的沉思中推演出来。而且还有来自于可观察实例的良好的间接材料源泉,如人们借助于被试的语言报告,将其作为经验的显示。这些方法有其局限性,但我们有更充分的材料,足以创立一种理论。

哲学分析对于从我们所拥有的材料中得到有价值的东西也是极有用处的。这种分析可以形成许多把经验与认知关联起来的原则,基于此,可以有力地约束一种基本理论的形态。正如我们将看到的,思想实验的方法也能产生有意义的回报。最后,我们探寻根本理论这一事实意味着,我们在阐发一种理论时可以诉诸于这样一些非经验的强制,诸如简单性和齐一性等。我们必须设法把我们所拥有的信息系统化,通过细致的分析使之得到尽可能的拓展,

382 第四部分

接着当那些材料仍然有资格成为世界的根本格局的合理的组成部分时,便对能解释那些材料的更简单的可能理论作出推断。

这样一些理论将永远保留着在别的科学理论中见不到的那种思辨的因素,因为判决性的主体间实验检验是不可能的。不过我们仍然能够建立这样的理论,它们与我们所拥有的材料协调一致,而且还能在相互的比较中对之作出评估。即使是在缺乏主体间观察材料的情况下,仍有许多适合于评估这些理论的标准,如简单性,内在的一致性,与别的领域的理论的一致性,再生下述经验属性的能力,这些属性是我们自己所熟悉的,甚至与常识的要求是完全符合的。即使所有这些强制都用到了,也许还存在着一些重要的不确定性,不过我们至少还能找出一些合理的替换品。只有当替换的理论得到了阐发,我们才能对它们作出评估。

关于意识的非还原理论将由许多心理物理原则构成,正是这些原则把物理过程的属性与经验的属性联系起来。我们认为:这些原则概括了经验由以从物理的东西中产生出来的方式。从根本上说,这些原则应能告诉我们:什么样的物理系统与经验有联系,而且对于有如此作用的系统来说,它们应能告诉我们:什么样的物理属性与经验的出现有关,我们可以预言,特定类型的物理系统会产生什么类型的经验。这是挺麻烦的工作,但我们没有理由等待观望。

接下来,我将陈述我自己对心理物理原则的候选者的看法,它们可以溶入关于意识的理论。其中,开始的两个是非基本的原则——相对高的层次上的加工与经验之间的系统联系。这些原则在阐发和约束意识理论的过程中能够起着重要的作用,但它们在

根本的层次上没有资格作为真正的基本规律。最后的原则是基本原则的候选者,它大概构成了关于意识的根本理论的拱心石。尤其是,该原则是思辨的,不过是这样的一种思辨,即如果我们要想有一种令人满意的意识理论,那么才必要的那种思辨。在这里,我只能简略地陈述这些原则;在我的 1996 年的《有意识的心灵》一书中,我对这些原则作了更为详尽的论证。

1. 结构一致性原则。这是关于意识的结构与觉知的结构之间的一致性的原则。大概还记得,先前用的“觉知”指的是与意识有联系的各种功能现象。而我现在用它指称的是经验的认知基础中的有点更为特殊的过程。尤其是,觉知的内容应被看作是可为中心系统存取的那些信息内容,直至被用来以广泛的方式影响行为的控制。简言之,我们可以把觉知看作是对于整个控制的可直接利用性。大致地说,觉知的内容就是具有直接可存取性、至少可在语言使用系统中潜在地予以报告的内容。

觉知纯粹是一个功能概念,但从根本上说,它与有意识经验是有关联的。在熟悉的情况下,每当我们发现了意识时,我们也发现了觉知。只要是存在着有意识经验的地方,在对控制行为和语言报告有效力的认知系统中就有某种对应的信息。反过来,每当信息对报告和整个控制有效时,就有相应的有意识经验。因此在意识和觉知之间就存在着直接的对应性。

这种对应性还可以加以进一步的阐述。它有复杂的结构,这是一种关于经验的核心事实。例如视域有复杂的几何学。在经验之间还存在着相似和不相似的关系,以及诸如相对强度这样的关系。每一被试的经验至少可部分地加以描述,还可根据它们的

384 第四部分

结构属性如相似性与差异性关系、被知觉到的位置、相对的强度、几何学结构等而加以分析。对于每一结构特征而言,在觉知的信息加工结构中存在着对应的特征,这也是一个核心的事实。

以颜色感觉为例。对于颜色经验之间的每一种差别而言,在加工中都有对应的差别。我们所经验的不同的现象性颜色都形成了一个复杂的三维空间,在色彩、浓度和强度上变化多端。这种空间属性能够在信息加工的条件下得到再现:对视觉系统的考察表明:光的波型是按照三种不同的轴线被分辨和分析的,而且正是这样的三维信息与后续的加工关联起来。因此现象性颜色空间的三维结构直接对应于视觉觉知的三维结构。这显然是我们料想中的事情。每种颜色的差别毕竟对应于某种可报告的信息,因此对应于在加工结构中被表征的差别。

更直截了当地说,视域的几何学结构直接反映在能在视觉加工中再现的结构中。每一几何学关系对应于某种能被报告、从而能在认知中被表征的东西。如果我们只被提供了关于行动者的视觉和认知系统中的信息加工的情节,那么我们就不能直接观察到那行动者的视觉经验,尽管如此,我们可以推论那些经验的结构属性。

一般来说,任何有意识地被经验到的信息也将从认知上得到表征。视域的细密结构将对应于视觉加工中的某种细密结构。这同样适用于其他形式的经验,甚至非感性的经验。内在的心像具有加工中所表征的几何属性。甚至情绪也有结构属性,如相对的强度,它们直接对应于加工的结构属性。在有更大的强度的地方,我们可看到对后续加工的更大作用。一般来说,因为经验的结构

属性是可存取和可报告的,因此那些属性就会直接在觉知的结构中得到表征。

正是意识的结构和觉知的结构之间的这种同型性构成了结构一致性的原则。该原则反映的是下述核心事实:即使认知过程在概念上并不蕴涵关于有意识经验的事实,但意识和认知不会相互游离于对方之外,而会以一种密切的方式一致起来。

这一原则是有其局限性的。它允许我们从信息加工的属性中再现经验的结构属性,但是并非所有经验属性都是结构属性。有一些经验属性,如红的感觉的内在本质,就不能在一种结构描述中完全被把握。在颠倒光谱故事中,关于红、绿的经验是颠倒的,但所有结构属性都相同,这一故事的可理解性表明:结构属性制约着经验,而不是使之枯竭。尽管如此,当我们设想经验在功能上同一的系统之间颠倒时,我们感觉到被迫让结构属性照旧不变,这一事实表明:结构一致性原则对我们关于我们心理生活的概念是多么关键。它不是一个逻辑上必然的原则,因为我们毕竟能设想所有信息加工在没有任何经验的情况下也能发生,尽管如此,它对心理物理联系仍有很强的、常见的强制性。

结构一致性原则将成为根据物理过程对经验的一种非常有用的间接解释。例如我们能关于视觉信息的神经加工这样的事实,间接地解释颜色空间的结构。关于神经加工的事实可以蕴涵、并能解释觉知的结构;如果我们承认一致性原则,那么经验的结构也会得到解释。经验探究甚至能使我们得到对动物之内的觉知结构的更好的理解,同时有助于间接解决内格尔的“成为一只蝙蝠可能是什么样子”这样令人困惑的问题。这一原则为大量已有的、关

386 第四部分

于意识解释的工作(如克拉克 1992, 哈丁 1992 关于颜色; 埃金斯 1993 关于蝙蝠)提供了一种自然的说明, 即使它常常糊里糊涂地被利用。它如此显而易见, 以致几乎为每个人所承认, 而且在对意识的认知解释中也是一个重要的基础。

意识与觉知之间的一致性也容许神经科学中旨在剥离意识之基质(或神经关联物)的自然解释工作。各种特殊的假说纷至沓来。例如克里克和考克(1990)主张: 40 赫兹的振荡可能是意识的神经关联物, 而李比特(1993)则设想: 瞬时扩张的神经活动是至关重要的。如果我们接受了一致性原则, 那么意识的最直接的物理关联物就是觉知: 信息借以直接为整体控制所利用的过程。不同的专门假说可以看作是关于觉知怎样得以实现的经验主张。例如克里克和考克主张: 40 赫兹的振荡是信息由以被整合到工作记忆中, 进而为后续的加工所利用的途径。同样, 可以自然地假设: 李比特所说的瞬时扩张活动之所以有关联, 显然是因为只有那类活动才能使整体的有效性得以实现。这也适用于其他被设想的关联物, 诸如巴思(1988)的“整体工作空间”, 法拉赫(1994)的“高质量表征”, 以及沙里斯(1972)的“对操作系统的选择输入”。所有这些都可看作是关于觉知机制的假说, 所谓觉知机制就是执行使信息直接为整体控制所利用的功能的机制。

假如意识和觉知之间有一致性, 那么结论便是: 觉知的机制本身就是有意识经验的关联物。大脑中哪一些机制控制着整体的有效性这一问题是一个经验问题; 也许存在着许多这样的机制。但是如果我们接受了一致性原则, 那么我们就有理由相信: 用来解释觉知的过程同时也是意识之基础的组成部分。

2. 组织不变性原则。这个原则陈述的是:任何具有相同精密功能组织的两个系统将有质上同一的经验。如果神经组织的因果模式在硅片上加以复制,如让一硅片复制每一神经元和相同的相互作用模式,那么就会出现相同的经验。根据这一原则,与经验之产生有关的东西不是系统的特定的物理构成,而是它的组成部分之间的抽象的因果相互作用模式。当然,该原则是有争议的。某些人(如塞尔 1980)认为:意识受制于特定的生物学,因此人类的硅片同型体不一定是有意意识的。不过,我相信:该原则能通过分析思想实验而得到重要的支持。

简单地讲,假设(为了便于归谬论证)上述原则是错误的,且存在着两个具有不同经验而功能上同型的系统。也许其中一个系统有意识,也许两个都有意识但有不同的经验。为阐释的方便,我们假设:一个系统是由神经元构成的,另一个是由硅片构成的,一个系统在另一个经验到蓝色的地方则经验到了红色。两个系统有相同的组织,因此我们能设想:从一个逐渐转化成了另一个,也许在某一时刻一个用具有相同局部功能的硅片取代了神经元。这样,我们便得到了一系列中间型实例,其中每一个都有同样的组织,但有略微不同的物理构成和略微不同的经验。这个系列内,一定有两个系统 A 和 B,在它们之间,我们替换的不超过该系统的十分之一,但它们的经验彼此有别。这两个系统在物理上是同一的,只是 A 中一小的神经组织为 B 中的硅片组织所替换。

这个思想实验的关键步骤就是在 A 中取出有关的神经组织,把它与因果上同型的硅片组织安装在一起,并在两者之间装上开关。当我们打开开关时,会发生什么呢?根据假说,那系统的有意

识经验会改变;为方便论述,比如说从红变为蓝。这个结论源于下述事实,即改变之后的那个系统从根本上说是 B 的变体,而在变化之前,它正好是 A。

但是根据那些假定,那个系统没有办法注意到那些变化!它的因果组织一如既往,因此它的所有功能状态和行为倾向都是固定不变的。就该系统而言,没有什么异常的东西发生。也不可能有这样的思想:“嗨!某种奇怪的事情发生了!”一般而言,这样的思想结构一定会反映在加工中,但加工的结构在这里依然如故。如果应当有这样的思想,那么它必定完全游离于那个系统之外,从而根本没有力量影响后来的加工。(如果它影响了后来的加工,那么它在功能上就是不同的,这又与该假说相矛盾。)我们甚至可以反复多次打开开关,以致红与蓝的经验在那系统的“内部眼睛”前变来变去。而根据假说,那系统绝不能注视这些“变动的感受性质”。

我认为这是原先假定的复归。关于经验有这样一个关键的、我们自己都非常熟悉的事实,即每当经验发生了重要变化、且我们予以关注时,我们都能注意到那变化;如果事实不是这样,那么我们就会被引向这样的值得怀疑的可能性,即我们的经验在我们的眼前总是变幻莫测。这个假说同世界在 5 分钟之前被创造出来这种可能性一样有着相同的命运:它在逻辑上也许是一以贯之的,但它是不可可能的。假如有下述极其合理的假设,即经验中的变化对应于加工的变化,那么我们就一定会得出结论说,原来的假说是不可可能的,功能上同型的任何两个系统一定有相同类型的经验。用技术的术语说,“缺席感受性质”和“颠倒感受性质”的哲学假说尽

管在逻辑上是可能的,但从经验上、从法则上说则是不可能的。^④

这里有很多要说的,但基本的意蕴已表达出来了。再者,这个思想实验为意识与认知加工之间的一致性提供了众所周知的事实,因此有助于形成关于物理结构与经验的关系的强有力的结论。如果该论证通得过,那么我们就知道:直接关联于经验出现的物理属性仅仅是组织的属性。这实际上是对意识理论的一种更强而有力的约束。

3. 信息两面理论。前述的两个原则是非基本原则。它们涉及的是高层次的概念,如“觉知”“组织”,因此对于在意识理论中构建根本规律来说,它们处在不合要求的层面。尽管如此,它们仍作为强有力的约束而起作用。进一步必需的东西是基本规律,它们与这些约束相一致,又能从根本上对它们作出解释。

我们设想的基本原则主要与信息概念有关。我或多或少是在香农(1948)的意义上理解信息的。在有信息的地方,就有内在于信息空间的信息状态。一种信息空间具有由下述关系构成的基本结构,即它的要素之间的差异性关系,它描述了在一个空间中的不同要素由以相同或不同的方式,可能是复杂的方式。一信息空间是一抽象的对象,但在香农看来,当存在着由不同物理状态构成的空间时,当状态之间的差异能沿着某种因果通道传递时,我们又可以认为信息可从物理上加以具体化。借用贝特森(1972)的一个短语说,物理信息就是一种造成差异的差异。

两面原则来自于这样的观察,即在某些物理上具体的信息空间与某些现象的(或经验的)信息空间之间存在着一种直接的同型性。从某些涉及到结构一致性原则的观察材料中,我们可以注意

390 第四部分

到,现象状态之间的差异有一种直接对应于具体化于物理过程中的差异的结构;尤其是有对应于沿着某些包含在整体有效性和控制中的因果通道造成的差异的结构。即是说,我们能看到具体化于物理加工中和有意识经验中的相同的抽象信息空间。

这便引出了一个自然的假设:信息(至少是某种信息)有两个基本方面,即物理的方面和现象的方面。该假设享有基本原则的地位,这种基本原则是经验从物理的东西中产生出来的基础,并能对之作出解释。经验基于它作为信息的一个方面的地位而产生出来,而另一个方面则具体化于物理加工之中。

这一原则得到了大量根据的支持,对此我在这里只能作出简单的陈述。首先,对对应于有意识经验的变化的一些物理变化的考察表明:这些变化总是有关联的,理由在于:它们在造成信息变化即下述状态的抽象空间内的差异的过程中有其作用,这些状态显然是根据它们在某些因果路径中的因果差异而加以划分的。第二,如果组织不变性原则应予坚持,那么我们就必须找到经验与之相连的某些根本的组织属性,而信息显然是一种组织属性。第三,这个原则为根据信息空间内存在的结构去解释结构一致性原则带来了某种希望。第四,对我们关于有意识经验的判断和断言——可从功能上解释、但密切地关联于经验本身的判断——的认知解释的分析表明:解释主要与具体化于认知加工中的信息状态有关。由此可知:一种基于信息的理论承认了对经验的解释与对我们关于经验的判断和断言的解释之间的高度的一致性。

惠勒(1990)主张:信息对关于宇宙的物理学来说是根本的。根据这种“源于信息”的学说,物理学的规律可以根据信息来构建,

如假设有引起不同结果的不同状态,而事实上不用说这些状态是什么。唯有它们在信息空间中的地位才有考察的价值。如果是这样,那么信息也是在关于意识的基本理论中发挥着作用的自然的候选者。因此,我们就得到了一种世界观,据此,信息真的是根本的,据此它有两个基本的方面,分别对应于世界的物理特征和现象特征。

两面原则自然是极富思辨性的,而且其证据也不够充分,尚有许多关键问题未予解答。一个明显的问题就是,所有信息是否都有现象的方面。一种可能性是,我们有必要对基本理论作出进一步的限定,如指明哪一类信息有现象的方面。另一可能性是,没有这种限定。如果没有,那么经验就比我们一直确认的要广泛得多,因为信息无处不在。这首先是反直觉的,但基于反思,我觉得这个观点具有某种合理性和简明性。在有简单信息加工的地方,便有简单的经验,在有复杂信息加工的地方,就有复杂的经验。老鼠的信息加工结构比人的简单,因而相应地就只有简单的经验;也许,恒温器这种简单到极点的信息加工结构可能有最简单的经验?如果经验真的是一种根本属性,那么它仅只是偶尔出现这一点就太奇怪了;大多数根本属性恰好是更普遍的。无论如何,这是一个悬而未决的问题,但我相信,该观点又不像通常所想象的那样,是没有合理性的。

一旦信息与经验之间的根本关联被发现了,那么通向对世界本质的某种更重要的形而上学沉思的大门就会敞开。例如,人们常常说,物理学只是从外_在方_面、根据基本实在与别的实在的关系描述了它的基本实在,而这些别的实在本身又是外_在地被描述的,

392 第四部分

如此等等。物理实在的内在本质被撇在一边了。有些人论证说,根本就不存在这种内在属性,而这一来,人们面对的便是这样的世界,即不具有因果关系所陈述的那些属性的纯粹的因果流(信息的纯粹流动)。如果人们承认内在属性存在着,对上述问题的一个自然的思考便是:物理的东西的内在属性——从根本上说来是由因果关系所陈述的属性——本身就是现象属性。我们可能会说:现象属性是信息的内在方面。这便可解除对经验的因果关联物的担忧——由于这样的图景,据此物理王国在因果上是封闭的,经验是物理的东西的补充,而产生的自然的忧虑。信息的观点有助于我们理解:经验由于其作为物理的东西的内在方面怎么可能有一种微妙的因果关联物。这种形而上学的思辨有利于发展科学理论这一点也许最好不予理睬,但在讨论某些哲学争论时,它是相当有启发性的。

VIII. 结论

我所阐述的理论是思辨性的,但又是一种可供选择的理论。我觉得结构一致性和组织不变性原则在任何令人满意的意识理论中都将是支撑点;何况信息两面论的地位是确定的。而且,在现在的情况下,与其说它是一种理论,倒不如说它是一种观念。为了促使最终解释的成功,当务之急是更充分地加以阐释,使之有血有肉,以更有力的形式出现于世。要得到一种更好的理论,就只有去思考:它的合理性是什么,不合理性是什么,它的作用在哪里,失误何在。

现存的大多数意识理论不是否认那种现象、解释某种别的东西,就是把问题抬高到永恒的秘密的地位。我试图表明:当严肃地对待这一问题时,是有可能推进问题的解决的。而要想有新的进步,我们就得作进一步的探索,就得有更精致的理论,更细致的分析。那个困难问题的确是一个困难问题,但没有理由认为它永远得不到解决。

注 释

① 这里的“功能”不是下述狭义的目的论意义上的功能,即一系统被设计所能起的某种作用,而是广义意义上的,即一系统在产生行为中可能完成的那种因果作用。

② 量子理论的一个特别引人注目的方面是这样的事实,即根据对量子力学的某些解释,意识在“瓦解”量子波函数中发挥着能动的作用。这类解释是有争议的,但无论如何,它们表明了根据量子过程解释意识是没有前途的。毋宁说,这些理论假定了意识的存在,并利用它去解释量子过程。这些理论充其量只是告诉我们关于意识所起的某种物理作用。对于它怎样产生,它们什么也没有说。

③ 某些哲学家论证说,在物理过程和经验之间即使有概念上的鸿沟,但不一定有形而上学的鸿沟,因此经验在某种意义上仍可能是物理的(例如黑尔 1991;莱文 1983;洛尔 1990)。这一论证路线常求助于逻辑上必然随之发生的必然性这一概念的支持(克里普克 1980)。我认为这一观点根源于对逻辑上必然随之发生的必然性的误解,要不然就需要一种全新的、我们没有理由予以相信的必然性;详情可参阅查默斯 1996(另外可参阅杰克逊 1994;刘易斯 1994)。不管怎样,这一立场承认物理过程与经验之间的鸿沟。例如把物理的东西与经验的东西联系起来的原则不可能从物理学中派生出来,因此这些原则必须看作是解释上根本的东西。

④ 某些人忧虑的是,神经系统的硅同晶体由于技术方面的种种理由是不可能的。这一问题确是悬而未决的。不变性原则只是说:如果同晶体是可能的,那么它就有相同类型的有意识经验。

394 第四部分

参考文献

T. 内格尔(1974):“成为一只蝙蝠可能是什么样子?”《哲学评论》,4: 435-450。

A. 纽厄尔(1990):《一体化的认知理论》,哈佛大学出版社。

D. 查默斯(1996):《有意识的心灵》,牛津大学出版社。

F. 克里克和 C. 考克(1990):“论意识的神经生物学理论”,《神经科学专题讨论会》,2, pp. 263-75。

F. 克里克(1994):《惊人的假说:灵魂的科学研究》,纽约。

J. 莱文(1983):“唯物主义与感受性质:解释的空白”,《太平洋哲学季刊》,64:354-361。

B. 巴思(1988):《意识的认知理论》,剑桥大学出版社。

G. 埃德尔曼(1989):《记忆中的礼物:意识的生物学理论》,纽约。

D. 丹尼特(1991):《被解释的意识》,波士顿。

R. 杰肯多福(1987):《意识与计算的心灵》,麻省理工学院出版社。

A. 奥尔波特(1988):“意识是什么样的概念?”A. 马塞尔等编:《当代科学中的意识》,牛津大学出版社。

K. 威尔克斯(1988):“—yishi, Duh, um 与意识”,同上。

H. 弗洛尔(1992):“感受性质与大脑过程”,载 A. 贝克尔曼等编:《突现还是还原? 非还原物理主义的前途》,柏林。

N. 汉弗莱(1992):《心灵的历史》,纽约。

A. 克拉克(1992):《感觉的质》,牛津大学出版社。

D. 刘易斯(1994):《心灵的还原》,载 S. 格滕普兰编:《心灵哲学的伴侣》,牛津,布莱克韦尔。

C. 哈丁(1992):“生理学、现象学与斯宾诺莎的‘真颜色’”,载 A. 贝克尔曼等编:《突现还是还原》。

R. 彭罗斯(1989):《皇帝的新脑》,牛津大学出版社。

——(1994):《心灵的幽灵》,牛津大学出版社。

S. 哈默罗夫(1994):“微管中的量子一致性”,《意识研究杂志》,1:91-118。

- C. 麦金(1989):“我们能解决心身问题吗?”《心灵》,98:349-366。
- K. 埃金斯(1993):“成为钻孔和近视眼可能是什么样子?”载 B. 达尔邦编:《丹尼特及其批评者》,牛津:布莱克韦尔。
- B. 李比特(1993):“有意识与无意识事件中的中枢时间因素”,载 G. 布洛克等编:《意识的实验与理论研究》,奇切斯特。
- M. 法拉赫(1994):“脑损伤后的视知觉与视觉知”,载 C. 乌米塔等编:《意识与无意识信息加工》,麻省理工学院出版社。
- T. 沙里斯(1972):“意识的二元功能”,《心理学评论》,79:383-393。
- J. 塞尔(1980):“心灵、大脑与程序”,《行为与脑科学》,3:417-457。
- C. 香农(1948):“通信的数学理论”,《贝尔系统技术杂志》,27:379-423。
- G. 贝特森(1972):“迈向心灵生态学的步伐”,钱德勒出版公司。
- J. 惠勒(1990):“信息,物理学,量子:对联系的探讨”,载 W. 朱里克编:《复杂性,熵和信息物理学》,雷德伍德城。
- C. 希尔(1991):《感觉:类型唯物主义之辩护》,剑桥大学出版社。
- B. 洛尔(1990):“现象状态”,《哲学展望》,4:81-108。
- S. 克里普克(1980):《命名与必然性》,哈佛大学出版社。
- F. 杰克逊(1994):《在自然界中寻找心灵》,载 R. 卡萨蒂等编:《哲学与认知科学》,维也纳。

意识、信息与泛心论*

威廉·西格尔

I. 基本事实

在D. 查默斯看来,意识的困难问题就是要解释:经验为什么以及怎样由物理材料的某些特殊的构型产生出来。我们姑且称之为“产生问题”,并记住:“产生”一词可能使人错误地想到原因:我们寻求的解释性关系可能就是同一、例示、实现或别的什么。^①当产生问题具有真正的科学问题的外观时,人们可能会争论这样的问题:它是否是有用的、强制性的甚或可理解的。假设某人把统计热力学的问题划分为困难的和容易的问题。容易的问题就是“气体在受热时怎样/为什么扩散”、“压力怎样/为什么随着温度升高而加大”等等之类。相比较而言,困难问题就是要用“热力学上无限的”、构成了统计力学之主体的粒子去说明热力学属性之产生。多么神秘!独立的粒子集合不仅像具有热力学属性的气体那样起作用,而且该集合不知为何还产生了那些属性。

戳穿这些假冒的神秘性易如反掌,而且许许多多的哲学家主

* 该文最初发表于《意识研究杂志》,2, No. 3(1995),后收入J. 谢尔编:《解释意识——困难的问题》,麻省理工学院出版社,1995, p. 269-286。中译根据后者译出。

张:意识的情况毫无二致。一当你解释了适当的、无疑是极其复杂的、从根本上说产生了行为的内在结构,那么显然就再没有什么要解释的了。不要因为一个任务对于体现了浓厚的形而上学神秘性的人来说完全没有意义,就错误地认为它不可能完成。查默斯坚持认为:因为意识并不是功能属性,因此某个试图解释“从行为方面来说是充分的”功能组织怎样产生经验的人并“不会犯概念错误”(p. 13)。确凿无疑的是,这个错误不像下述错误那么明显,这错误是想撇开由统计学与宏观热力学之间的、关于功能同型的统计力学所提供的说明而对热怎样产生出来作出独立的、附加的解释。但是人们怎么可能说明:意识的情况在根本上风马牛不相及呢?

一个直截了当的回答显然就是指出,关于他心的古老问题具有可解决性。不存在足以取消这一问题的先验论证;思考过它的任何人都会发现:我们每个人从根本上说是孤独的。维特根斯坦也许反对这一点,说:“如果我看到某人因疼痛而翻滚,我一样会认为:他没有向我隐瞒他的感觉”(1953/1968, p. 233)。一个习惯的回答是:自然不是这样,因为我们据以行事的是这样的假定,即别的人肯定有经验,不可能对基本的假定提出质疑。但是如果那是甲虫被标本搜集者钉在钉子上所作的一种向外的伸张的动作,或者说是一个大龙虾掉进了开水中所作的蠕动,这又怎样呢?对于经验在生物发展之链的什么地方出现这一问题,不可能有轻而易举的回答,更不用说有哲学上清楚明白的、先验的回答,即使是在甲虫或龙虾的层面上,人们确实看到了类似于(至少)疼痛所引起的那类行为。

D. 丹尼特的意识理论(1991)可以看作是旨在解决产生问题的理论。他关于哲学上的怪人(即能像我一样行动但完全是无意识的、完全没有经验的造物)的讨论使人想到了维特根斯坦的说明,而且同样是令人困惑的。在某一点上,丹尼特说(pp. 405 - 406)如果怪人是可能的,那么你就真的不能判断某东西是不是怪人(每一铜币有两面),因此把公认的怪人当作是完全无意识的存在是不道德的。这不是反怪人可能性的论证,因此几乎不是瓦解产生问题之合理性基础的论证。在别的地方,丹尼特承认:动物有经验,即使它们没有人类那样的得到充分发展的意识(pp. 442ff)。他宣称:许多不同的功能构造可能接受经验的归因(我避免用“产生”^②),关于蝙蝠,他特别指出,我们通过探索蝙蝠的神经系统表征了什么,哪些表征在行为调节中实际上起了作用,就可能知道它的经验范围内的某种事情(p. 444)。但这只是告诉我们:蝙蝠可能意识到什么,而没有告诉我们:蝙蝠是否意识到那些事物,因为可以肯定的是,它忽略了调节行为的有意识与无意识表征之间的区别。因此在这里我们发现了产生问题的一种十分严格的形式,它存在于试图排除它的理论中:假设有意识与无意识表征的区别是存在的(对此丹尼特是赞同的,即使他断言这一区别不可能绝对廓清),假设某些无意识表征调节着行为是不可否认的事实,那么在没有语言的动物中,是什么造成了有意识的调节行为的表征与无意识的调节行为的表征之间的区别呢?如果它们是有意识的,那么在蝙蝠身上表征躯体伤害的状态怎么会是有意识的经验,如果它们不是有意识的,那么在受伤害时表征了翅膀所处的详细位置的状态怎么又会是没意识的呢?(在这里,我设想蝙蝠像我一

样:我在踝节部感觉到了疼痛,但我通常并不知道脚在不平的地上跑动时所涉及到的复杂的活动,然而两者都与调节表征的行为有关。)在这里,我们没法对意识作出通常的行为——言语行为——检验,因为蝙蝠自然不能告诉我们它们觉察到了什么,但丹尼特爽快地承认(即使令人不解):尽管如此,动物是有经验的。

再者,在讨论裂脑人的案例时,丹尼特否认:“连合部切开术在它的唤醒组织中留下了既鲜明而又强有力的、足以证明……有分裂的自我的根据”(p. 426)。但是如果右半球有经验(就像无语言动物那样),那么问题就应是:它是不是完全合格的自我。显然右半球使用了不同的表征,而且其中一些还能调节行为(正如大量裂脑研究充分显示的那样)。因此究竟是什么使它们或它们中的某些进入经验之中?如果它根本就不是表征的行为调节力,那么它是表征所具有的在 n 级之上(在效能的某个刻度上)的行为调节力吗?显而易见,这还是十足的产生问题:什么使 n (或 n 周围的大致确定的区域)正好成为能有意识的那类事物?

意识理论应当告诉我们:意识是什么,在世界上是什么事物拥有它,怎样判断某物是否拥有它,它是怎样在物理世界起源的(从共时性方面看它来自于物理条件,从历时性方面看它是进化发展的)。意识的困难问题已为我们与动物所共有的真正“怪人问题”所证实。再如,蜜蜂像有经验(视觉、嗅觉以及痛苦与高兴)的造物一样行动。我们有理由假设,它的行为是由具有内在表征的复杂系统所调节的,而其表征又是由复杂的神经并行加工器(极像我们的,尽管不那么复杂)所产生、控制和更新的——就我所知,表征是由显著的 40 赫兹振荡所协调的。蜜蜂处在有知觉力和无知觉力

400 第四部分

之间的模糊界线的哪一边呢,或者说就在模糊区域之内?^③对于任何回答都极为重要的是:为什么?假如我们制造了蜜蜂机器人,它很顺利地适应了蜜蜂的生活(开始是被安排好了的,参阅柯克纳和汤恩 1994)。再假设我们能肯定,该机器人没有经验(它真的是一种“蜜蜂怪物”)。这真的表明蜜蜂没有经验吗?为什么?另一方面,假设我们认为蜜蜂确实确实做出了经验的事情。这表明机器人也有经验(它确实通过了蜜蜂层次的图林测试)吗?为什么?^④

正如查默斯所主张的,现成的意识理论的确没有提出这类问题,即使它们承认关于蜜蜂经验的问题是合理的。请原谅我在这点上的唠叨,但产生问题绝对是至关重要的。没有它就没有意识的困难问题。而有了它,该问题看起来才真的很困难。

事实上,它是如此之难,以至于查默斯期盼一种激进的解决,这种解决能在物理系统和意识系统之间的所谓解释鸿沟上架起桥梁。他设想:意识是宇宙的根本特征,在心灵研究中它必须直接被看作是基本事实。这是简化产生问题的一个捷径,因为根据定义,对世界的根本特征为什么产生在它们产生的时候这样的问题,不可能作出解释。例如不可能解释基本粒子为什么以它们的被观察到的质量比率显现出来。^⑤也许一当我们承认了产生问题的真实性,那么就没有别的路可走了。既然我极力赞成把产生问题当作真正的问题,那么我自然对查默斯的解决方案感兴趣。不过我也承认,我在查默斯的说明中发现了某些令人困惑的成分,我要论证的就是:关于意识问题的更激进的观点可能受到他倡导的意识是宇宙的根本特征这一假定的支配。

Ⅱ. 组织不变性与解释的排斥性

我们先来看查默斯关于意识由以产生的条件的观点。他把这观点称之为组织不变性原则。严格地讲,这原则只是断言:“任何具有相同精密功能组织的系统都将有质上相同的经验”(p. 25)。但是由此可得出结论说:一系统 S 是否有意识,取决于它对某种功能描述的满足。因为如果假设其反面,那么 S 就有某种别的、非功能的特征,可称之为 Q,是它产生了意识。于是我们便可建造一个功能上同型于 S 的系统,它没有 Q,根据假设,Q 是没有意识的,而根据组织原则,它是不可能的。^⑥令人困惑的是,当意识依赖于满足了纯功能描述——它们与具有行为容量的描述存在着相似性——的特殊系统时,它又能成为自然的绝对根本的特征。世界的其他根本特征都没有这个特点,或者说没有酷似这个特点的特点。就像一个人宣布:“成为一部电话机”是世界的一个根本特征,它由许多一致地满足了有关的、高度抽象的、从行为上定义的功能描述的系统所产生。

查默斯既然固执地主张:意识本身不是一个功能特征,因此引出了一个困难问题,那么说意识只取决于系统是否满足某种抽象的功能描述,就太离奇了。当然由于明白了意识是一种真正的根本特征,因此我们就不可能问:所有满足了某些功能描述的系统是有意识的,这是怎么可能的,然而与其说这个观点减弱了产生问题的神秘性,倒不如说大大加剧了。

在这里,我们还面临着产生问题的一个变种,它来自于“精致

402 第四部分

的功能组织”这一短语固有的模糊性。查默斯提供了这样一个例子,即有一大脑,它的神经元逐一为“硅同晶体”所替换,这是他的组织不变性原则的一个例示(同上)。但是我们怎么知道:这恰好就是功能描述具体说明的那个水平?我们可以假设,硅同晶体不是内在地从功能上同一于它所代替的那些神经元的,然而一神经元的内在作用确实可满足某种功能描述。或者换个角度说,我们为什么不能用一个模仿了它所替代的整个神经组织的输入/输出关系的硅片装置去替换大量的神经元呢?于是就有一个新的产生问题:功能描述的一个特殊的层次怎样产生意识,它实际上产生在哪一个层次?问题在于:系统在不同层次上的功能复制品原则上能重复整个系统的行为。像一意识系统一样起作用的任何系统能根据那个原则判断是有意识的吗?如果不能,假设我们有两个系统,它们在 n 层次上是功能同型的,而在 $n-1$ 层次上,在功能上则是不同型的(正如在查默斯的那个例子中我所理解的那样)。它们是否具有相同的意识状态,这显然取决于哪一个层次是描述的适当的层次,但是谁去判定呢?对于功能描述的层次,宇宙知道什么呢?

解释的排斥性^⑦的一个致命的问题根源于意识与功能描述的结盟。任何从功能上描述的系统事实上必定为某种物理部分之集合所例示,如果它应该加入到世界的运作之中的话。该系统的因果效力完全取决于它的物理例示的效力。因此当我们说到“测量火炉的恒温器”这样的东西时,该恒温器的效力完全能由这恒温器的特殊的物理例示所解释(也就是说由它的热电偶的物理细节或它所具的别的功能所解释)。也许水溶化盐是一个更好的例子:这

完全可通过个别的 H_2O 分子与构成盐的 NaCl 分子的相互作用加以解释,而这些相互作用进一步又完全可用氧、氢、盐和氯的量子力学属性予以解释。如果不是那些根源于其构成部分的因果力,水是不可能有任何因果力的。因果根基原则陈述的是:任何复合体的因果效力,不管是已述的功能上的还是部分学上的,都完全依赖于其物理例示的基本构成部分的因果效力。问题的令人困惑之处现在就昭然若揭了。意识在世界中有因果效力吗?如果有意识系统的因果效力遵循因果根基原则,以至于任何有意识系统的因果效力都完全依赖于它的例示的因果效力,那么由于把意识当作是宇宙的根本特征,它不能还原于它的例示,因此意识在世界上就没有效力了——意识最终完全是副现象性的。另一方面,如果拒绝这个结论,而把某些独立的因果效力归之于意识,那么某些物理部分的集合就具有完全不依赖于它的那些部分的因果力的因果力。这就是哲学家们所谓的激进的突现论。^⑧正是在这里,我们有一种极端的激进形式,因为它不是这样的集合,即把物理部分集合到产生突现性质的、特殊的克分子合成中,而是这样一种集合,即它设法满足能产生奇迹的某种抽象的功能描述(我们可以称之为关于激进的功能突现论的学说)。这个二难推理的任何一支都没有任何吸引力。

这一解释排斥性问题也可以认为是源自查默斯的另一原则:即把有意识经验的现象特征等同于“信息状态”的原则。既然每一物理状态相对于某种可能的信息接受者来说都是一种信息状态,而且对应于编码在任何物理状态中的信息上的差异的因果差异通常被认为是遵循因果根基原则的(这个事实自然是我们利用物理

404 第四部分

过程传输信息的能力的基础)。因此我便有另一个二难推理:如果有意识经验同型于信息荷载,那么有意识经验的因果效力要么是(1)完全依赖于信息携带者的物理属性,要么是(2)某些信息携带者违背了因果根基原则。如果是(1),那么我们就有解释排斥性,进而有意识经验就成了副现象。如果是(2),那么我们就有激进突现论的另一形式,不知什么原因它依赖于这里所说的物理状态所携带的信息。不管是哪一支,都是没有吸引力的。

还值得一提的是,信息概念具有众所周知的模棱两可性,它有时意指的不过是一种物理状态的“比特容量”,它或者指的是某种语义上有意义的、由传输单元所携带的内容。查默斯打算把有意识经验的现象性质归之于哪一种信息,这是不清楚的,即使他所说的在我看来倾向于前一种解释。大脑的比特容量无疑是巨大的,但由于把两个大脑当作一个系统而显然翻倍了,然而值得怀疑的是,存在着一种与两个人的相互作用有联系的第三意识,即使这两个大脑在那时已形成了一个因果上相互作用的系统。因此我们又有另一个产生问题:意识实际上是由哪一些信息状态产生的,为什么正好是那样,它们怎样产生?

III. 信息的本质

查默斯猜想:信息本身也许就是世界的根本特征,而该特征又使自己成为意识的“自然关联物”。但是意识与信息是在语义学意义上的层面上而不是在比特容量的层面上才相关联的。既然信息的古典理论停留于比特容量的层面,因此它似乎不能为意识提供

严格的(或就那个问题而言的任何)关联物。而且古典理论把信息当作是某种因果过程的特征(尽管是非常抽象地被设想的),这也就是说,信息是一个功能概念:信息具体化于能被不同地例示的因果过程中,遵循着因果根基原则(因此导致了上面讨论过的解释排斥性问题)。

由于接近了关于信息的更为激进的观点,因此我们现在可以讨论关于意识的根本本质的一个更为激进的观点。这种信息观点把因果过程当作一种信息转换,但又不希望所有的信息“联系”都局限在这种过程之内。要以这种方式探讨信息概念,自然应到量子力学中去。

“纯粹信息”的作用可通过对著名的双缝实验的讨论来予以说明。一束光子、电子、原子或别的什么入射到不透明表面上的一对裂距适中的双缝上。探测屏安放在裂缝的后面。量子力学预言:对银屏的撞击形成了一种相干图样,该相干图样又来自于能到达屏幕的测试粒子的两个可能路径的相互作用,这已得到了不太理想但富于实践性的实验的充分证明。

尤其是,量子力学形式主义要求:在那束中的粒子的状态 Ψ 必须被表征为与每一空间路径有联系的状态的叠加:

$$(1) \Psi = \Psi_1 + \Psi_2$$

(为了清楚起见,我撇开标准化的精确性——详见附录。) Ψ_1 表征的是进入左边裂缝——即路径 1——的粒子。 Ψ_2 表征的是进入右边裂缝——即路径 2——的粒子。结果是,当我们计算一粒子击中探测屏的一定区域的概率时,两个可能的路径都以复杂的方式对这个概率发挥作用;没有办法确定那个粒子实际进入了

406 第四部分

这些路径中的哪一个(即使我们不管哪一个),因此我们的计算只能这样进行,即把每一路径独立地产生的概率加起来。“相干图样”说明了这一点,这种图样在足够多的粒子通过那些装置之后就可在探测器上看到,如果粒子逐个地正好通过裂缝中的一个,那么它就明显不同于所期望的干涉图样。

双缝实验的特征是众所周知的,同样广为人知的是,如果我们想办法确定粒子采取哪一路径,那么相干图样就会消失。这有时可用这样一种测量所涉及到的粒子状态的干扰来解释,有时这种干扰据说是不可避免的,从而是对该现象的这个方面的严格的说明。但是这并不是那个理论的本意。在解释相干图样消失时,没有必要假设干扰。只要有关于粒子采取哪一路径的信息就够了。试假设有这样一种完善的探测器,它能在不改变粒子状态的情况下确定粒子采取了哪一路径。这种探测器只能有两种输出状态,让我们称之为L、R(分别代表左边和右边的裂缝),探测器的输出完全与状态 Ψ_1 和 Ψ_2 的组成部分相关联。完善的探测器这一观念在数学上是没有什么错误的,而且探测器在场的作用将排除相干图样,尽管对粒子的状态没有作用。这对那些必须想到并协性并不是测量之不得体的结果而是量子力学的内在的、不可排除的特征的人来说是有一些好处的。该理论认为:我们的探测器携带着相关的信息,仅仅这一事实就足以摧毁相干效应,不管探测器在测量时是否以某种方式“干涉”了那个系统。

这里所论及的那类信息不是比特容量,而是“不同”物理系统的语义学上有意义的关联物,在那里没有这样的要求,即该关联物必须由把两个系统联系起来的某种因果过程来维持。量子力学这

一显著特征因完善的探测器概念所暗示的手段而变得更加明朗。但是直接通过删去探测器内的信息而恢复原来的相干图样这一可能性又怎么样呢？既然粒子状态没有为探测器初始的操作所改变,那么这似乎至少在理论上是可行的。形象地说,粒子尽管远离它们要到达的、最后将记录它们位置的显示屏,但对它们的路径是否被自动记录下来是没有主意的。这样一种相干检索仪器就是所谓的量子清除器(参阅斯卡利和德鲁尔,1982;斯卡利等,1991;恩格勒特等,1994;西格尔,前言)。量子清除器是一种稍微有点精密的造物,因为它们一定受到了这样的限制,即在所述的系统之间不能建立“比特通道”——探测器和目标屏就是运行于这种情况之下的(因为这可能违反了相对论的下述根本限制,即信息不可能比光速传输得更快)。尽管如此,它们至少在理论上是可以实现。我们不妨这样设想:在我们的双缝附加完善探测器装置上,我们已增加了某种未定的、清除探测器上的信息的方法(这实际上是建构一种清除器的重要组成部分)。

对量子清除器的分析确实可以说明:相干图样能够通过直接清除探测器上的信息而得到恢复,在不建立比特通道的情况下也是如此(详情还可参阅附录)。清除器的作用有力地表明:每个粒子对清除器的状态都会作出反应。但是如果我们严密地安排我们的实验(例如,我们可以使探测器屏与清除器之间的距离如此之大,使清除器的运作如此之长,以致从清除器而来的一种光信息不能在它到达那显示屏之前到达粒子),那么在粒子和清除器的运作之间的任何因果过程都不可能有什么疑义。

对量子清除器和更简单的、基本的双缝实验的自然的解释就

是,在量子系统的要素中存在着一种非因果的、但有信息载荷的联系。这种联系不是比特通道或任何别的因果过程(附带地说,这再一次表明:我们这里讨论的是信息的语义意义)。在这里,我们也许得到了一种新的、有价值的、有高度重要性的观点,即信息真的是世界的根本特征(或许是特殊的根本特征)。

IV. 泛心论

为了形成一种把意识本身当作世界的根本特征的理论,依我看来,使用信息的根本本质这个更强的意义似乎是可能的。我这里说某物是根本的,意即基本的,不是对物理系统满足任何功能描述的依赖性,也不是对因果根基原则的屈从。查默斯本人在说明信息时对这种理论有所表态,指出:这种理论“并不像通常所认为的那样是不合理的”(p. 28)。我们可以直截了当地说:这里所述的理论是泛心论,它是这样一种学说,即“所有的物质或所有的自然本身都是有心灵的,或者说有心灵的方面”(引自《牛津英语大辞典》),这的确被认为是不合理的。我用极其独特的方式对它作了辩护。在我看来,产生问题是真实的,但又很难为完全不受约束的思辨提供担保。一当我们把意识是世界的根本特征这一观点与我们关于信息及其意义的新概念统一起来,那么几种思想,如有的为泛心论辩护,有的给予抨击,就会以一种非常令人满意的方式得到和解。

我在上面已说过,根据查默斯的说明,意识完全是一种突现的现象,而且只是在它不能根据表现它的有关复杂系统的属性予以

解释的意义上,它才是根本的。他还坚定不移地认为:意识不可能还原于那些低级的属性。一段时间以前,T. 内格尔(1979)注意到:对与非还原主义有联系的、激进的突现论的否定似乎可衍推出泛心论。该论证非常简明:如果意识是不可还原的,那么我们就不能在物理复合物的某个层面根据那种复合去解释它的出现,因此它如果不是在那些复合的层面上突现出来的,那么它就一定已经出现在更低的层面上。这一来,如果我们拒绝激进的突现论,并重视产生问题,那么我们就自然投入了泛心论的怀抱。

泛心论有悠久的历史。也许它是我们的祖先所宠爱的学说,激荡着许多前科学文化中的万物有灵论的旋律。博学哲学家莱布尼兹赞成一种形式的泛心论,这实质上是出于内格尔所说的那种理由。但是泛心论总处在科学/哲学高贵地盘的边缘,随着对世界的科学理解的拓展,它不断地失去原有的那种尊严。因此有点具有讽刺意味的是,由达尔文在生物学中所发起的革命又重新燃起了对泛心论的兴趣。在一篇仍有价值的论文中,W.K. 克利福德(1984)提出了一种显然有待证实的论点:进化论应用于心灵的必要条件是:意识的某种要素存在于全部物质之中。他在承认产生问题的同时论及了意识:

我们不能假设:在进化过程的任何一点上发生了从一种造物向另一种造物的巨大飞跃,这显然是一种与物理事实完全不同的、绝对分离的事实。在上溯的路线上,没有人能说出那事件被假设发生的特定位置。如果我们完全接受进化的学说,那么我们能够到达的唯一结论就是:即使是在最低级的有机体中,即使是在漫游于我们血液中的阿米巴中,对我们来说,

410 第四部分

也存在着某种或别的难以想象的简单的东西,它与我们自己的意识具有相同的本质……(1874,p.266)

简括地说,这不就是内格尔的论证吗?突现是不可能的,还原是荒谬的——因此意识的要素必须在宇宙的基本构成材料中寻找(当克利福德把他的论证限制在有机体的时候,我们看到了一种活力论的错误,因为产生问题既不会在有机体与非有机体之间的空隙中出现,也不会超有机体的等级系统的空隙中出现)。进化论的增补部分给予了、或至少是在克利福德时代假定了一种明显可行的机制,通过它,简单的东西以不同方式结合到复杂的东西中,这种增补推进了迈向真正的泛心论的步伐。

另一方面,人们也可以对泛心论提出强有力的驳难。也许对泛心论的最集中、最有洞见的抨击就出现在 W. 詹姆斯的《心理学原理》(1890)一书中。他把该观点讥讽为“心灵灰尘”理论,并给予了严厉的痛斥,他还提出了一个我认为是关于意识的任何泛心论所面临的最困难的问题。我将把它(1)称之为合成问题,它要解释的是这样的问题,即“原子性的意识的无数要素怎么可能结合成为一个新的、复杂的、丰富多彩的、像我们所具有的那种意识。这不正好就是产生问题吗?对此,詹姆斯说得非常好:

准备一个由 12 个单词组成的句子,请出 12 个人,让他们每个人记住一个单词。接着让他们站成一排,或挤坐在一条凳子上,让每个人尽可能专心致志地想自己的那个单词;谁都没有关于整个句子的意识。我们常谈到“时代精神”……但我们知道这只是一种符号,从来不会梦想:那种精神……构成了一种

另外的意识,此外还构成了“时代”一词……所指称的那些个体的意识(p.160)。

或者说:

在假定要素单元有感情的地方,情形一模一样。取出一百个这样的东西,将其搅乱,然后随你的意愿把它们包在一起(不管那意味着什么);每一个仍有它过去所有的感情,禁锢在自己的皮肤内,没有窗户,不知道别的感情是什么,意味着什么。如果当这样的一组或一系列感情出现时,属于这样一组的意识也出现了,那么在那里就有第 101 个感情。这第 101 个感情完全是一个新事实;根据严格的物理规律,原来的 100 个感情当合在一起时可能是它的产物的标志;但是它们不能从根本上等同于它,它也不能等同于它们,人们也不能从别的推演出其中的一个,或者(在任何可理解的意义)上)说:它们进化出了它(p.160,着重号为原文所加)。

总之,詹姆斯认为:泛心论的第二个根本假定——经验的单元能融合到高级的经验形式中——在“逻辑上是合理的”,没有它,泛心论就没有通向产生问题所津津乐道的那些东西的逃路(p.158)。

如果詹姆斯是对的,那么结合问题就在泛心论中提出了一个不同的产生问题,它在形式上类似于意识从物质中产生的问题。如果同样的问题从根本上说潜藏在泛心论的心脏,而且它自然面临着断定原子有意识的那种不合理性(无论在什么程度上——这必然是不合理的),那么泛心论就没有优越于物理主义的地方。如果詹姆斯是对的,那么泛心论的第一个假定什么也没有抓住,而且

412 第四部分

使其居于优先地位的有用性完全丧失了基础。

另一反对意见来自这样一个问题,^⑩它大概可以被称之为(2)无意识智能问题。人们可能想避免第一个假定的不合理性,方法是赞成心灵的构成单元具有智能,而否认它们实际上是有意识的经验。但这自然就使产生问题成为不可解决的问题,甚至被认为加剧了该问题,因为我们怎么能根据无意识实在的结合去说明有意识经验的产生呢,即使它们在某种意义上是心理的实在?倘若如此,泛心论便面临着这样一个问题,它严格类似于物理主义者所面临的那个产生问题。

然而在想到智能在世界运行中的作用时,又会产生另一严峻的问题。人们可能会认为,像意识这样重要的根本特征在世界的因果交流中应该有某种作用。但是如果它起着这样的作用,那么我们就有望在对物理世界的探索中看到它;即是说我们有望看到物理上没有区别的系统,因为它们的心理维度所潜藏的因果力,因此至少会偶然地在它们的行为中表现出差异。如果是这样,我们关于世界的物理图景就是极其不完善的,而且许多人会发现,它是极不合理的。我常常不得不担心的是:我的汽车是否能启动,但谢天谢地,我没有这样的附带的担心:即使是绝对没有机械方面的故障,但因为它今天“觉得喜欢”留在车库里,因此开不动!姑且把这称作(3)完善性问题。以后我再对这些问题作出回答,但对(3)的不令人满意的回答应在这里予以讨论。泛心论者可能极力主张:物理上同一的系统由于它们的物理同一性而具有同一的心理特征,因此物理上同一的系统实际上总是以相同的方式行动,即使心理的方面提供了某种驱动力。这是不令人满意的,因为这直接提

出了解释排斥问题:如果系统的物理属性能够说明它的全部行为,那么有什么根据假定有心理的作用?于是心理的东西在最低限度上就是解释上的副现象,进而可能成为真正的多余无用的附属物。因此问题在于:泛心论要么断言:我们关于世界的物理图景是不完善的,要么说心智是解释上的副现象。第一点是不合理的,第二点又极大地削弱了泛心论计划的基点。

最后,有两个最简单的反对意见。我们有(4)无标记问题:没有证据能证明:自然的基本单元有任何非物理的维度,以及(5)无心问题:如果这些单元有某种特征,我们选用“心理的”词语加以命名,那么人们能拿出什么可能的根据去证明这种命名呢?毫无疑问,在我们肯定泛心论的学说有真实的内容之前,我们最好是在世界的基本特征中来考察心智本身的某种“标记”。

V. 量子万应灵丹

关于泛心论,有一种一致的看法,即认为,在一定的意义上它有助于回答各种反对意见。既然在我看来,它们大体上是按困难递减的顺序陈述出来的,那么我想倒过来予以考察。就(5)来说:如果人们把意识当作是世界的真正的根本特征,那么它出现在离我们通常所及的范围极遥远的地方,就不足为奇了。在亚原子粒子之间不存在引力的任何明显的信号,但是既然我们把引力当作根本的,那么我们就是打算承认两个电子之间的吸引力真的是存在的。不过我们必须永远记住:那些否认有关于意识的产生问题的哲学家将可能认为,把意识归因于没有表现出意识的行为记号

414 第四部分

的任何东西是更不合理的,且完全不可理解的。我上面试图要论证的是:产生问题是真正的问题,这意思是说:人们至少可以以一定的可理解性假定意识是宇宙的根本特征。

这对(4)也作出了一定程度的回答。因为如果类比于引力是可以接受的,那么我们便可以期望:与物理自然的构成单元有联系的那个“级别”的意识之结果是完全不可分辨的。不存在这样的要求,即要求根本特征在每一可能的梯级上提供操作上可观察的结果。这个回答可能是充分的,但它也可能是不必要的,因为值得注意的是,尚不清楚构成单元绝对没有显示这样的迹象,即它们有假定的非物理的(以及真正的心理的)方面。

要解释我的意思,就必须提及(3)。我认为,可以合理地预言:世界的真正的根本特征在世界中起着不同的因果作用。因此我们认为,用纯物理术语表述的世界图景,如果不参照这种根本特征,那么就是不完善的。即是说,世界偶尔以不能用纯物理观点解释的方式起作用。没有人真的知道,人的思想、行动是不是完全由物理特征决定的,因此,也没有人真的知道,人的行为是不是也从纯物理的方面被决定了。但是让我关注一下物理自然的构成单元,看看它们是不是也以不能用纯物理观点解释的方式行动。它们自然是这样——量子理论坚持这一点。作为一种物理理论,量子力学断言:既然这些过程在备选的可能性中与完全随机“选择”有关,那么对它们就不可能有解释。世界的行为表明,具有自己独特作用的附加根本特征很可能是存在的。

有各种证据表明:量子力学不可能渗透到纯决定论的物理理论之中。^①就我的理解而言,这些证据都依赖于有争议的假定。尽

管如此,我们至少可以把它们当作是纯物理的世界图景的难以名状的不完善性的信号。

沿着(5)的路线要论证的是:这里所求助的不完善性与意识绝对没有关系,但又难以完全说清楚。如果我们问,我们的构成单元对世界的什么根本特征可能作出反应,那么一个重要的作用就是信息。在双缝实验中,我们可以说,粒子对完善探测器的结果了如指掌;在量子清除器中,在粒子与清除器之间没有任何可证明的因果联系的情况下,粒子仍知道信息是不是被清除了。对信息的敏感几乎无异于具有心智的王国,即使这只是就公认的、非常有限的且贫乏的意义而言的,但根本特征在构成要素的层面显现自身应是有希望的。值得重复的是:这里所说的那种信息不是古典信息理论所说的比特容量,而是更像语义学上的有意义的信息,而且这是一个更接近于心智的信息概念。

根据这种观点,物理自然的构成单元具有心理的方面,这个方面在这些单元的行为中又享有独特的因果作用。因此它拥抱的是二难推理的第一支:物理的观点是不完善的。但是我要论证的是:在构成单元的层面,世界的物理图景看上去的确不完善。而且,这种不完善性还可能渗透到物理复合物的复杂的等级结构之内。

对更复杂的物理实在的复合物的反思自然把我们带到了泛心论所面临的这一最困难的问题面前:结合问题。当我们明白了基本的物理要素以多种方式相结合而产生出分子、蛋白质和人的时候,我们却又没法弄清楚:即使假定基本心理要素在某种意义上是有意识的,^⑩但说它们的结合形成了不同的、更复杂的有意识经验,有什么意义呢?

416 第四部分

我对说结合问题的困难可以完全予以克服深表怀疑,但是我认为:对作为詹姆斯观点之基础的形而上学的预先假定稍作深入的考察,便可对结合问题引出一个相当自然的回答。在詹姆斯看来,结合问题源于一个极常见的考虑:

大量可能的实在(如你愿意,把它们称之为力量、物质粒子或心理要素)不可能把自己总合在一起。因为在总合中,每一个仍是它过去所是的东西;总合本身只是相对于一个偶尔审视那些单元、如此去理解那总合的旁观者而存在的;或者以对外在于总合本身的实在的某种别的作用的形式而存在。假设人们都赞成说, H_2 与 O 结合成“水”,进而表现出新的属性。其实不然。“水”在新的状态 $H-O-H$ 中正好是原来的原子;“新的属性”正好是它们结合的作用……(pp.158-159)

或者说:

一点也不错,在力的平行四边形中,“力”本身并没有结合成对角线的合力;一物体要表现其合力,离不开它们所冲击的东西。音乐的声音本身也不能结合为谐和音或不谐和音。谐和音和不谐和音是表示它们对外在媒介如耳朵的合成作用的名称。(p.159)

我并不是要争辩说:这样一种观点有某种吸引力;它似乎不过是对部分学的还原论的一种合理的泛化,对此,世界已提供了很多的证据。但我知道它是错误的。正如读者所知或可猜测到的,对这种错误的最尖锐的揭露源于量子力学。再看双缝实验。有一个再自然不过的假定,即在世界上,把粒子当作是通过两个裂缝

所形成的一种混合体,该混合体包含了状态 Ψ_1 (一种表征并记住了通过左边裂缝的粒子的状态)中的一半粒子,以及状态 Ψ_2 (通过右边裂缝的粒子)中的一半粒子。但是事实并非如此。它们存在于两种可能状态的叠加 $\Psi_1 + \Psi_2$ 中,而叠加是状态的一种“结合”,其本身形成了一种真正新的、具有可观察到的、不同于混合物属性的状态。量子整体并不是它们的部分的总和。

然而进入像 $\Psi_1 + \Psi_2$ 这样叠加的能力是进入了它之中的要素之属性的反映,因此部分学还原论的概念不应当完全从我们的哲学中一笔勾销,它对于处在我们关于科学解释本身的概念之心脏位置的还原论来说无疑是一种福音。不过我们不能接受詹姆斯所信奉的关于部分学合成物的原则,因此就不存在这样的论证,即根据那些一般原则去反对泛心论者把基本的心理单元结合为不同的心理整体。

VI. 进一步的思考

至此,可以回答对泛心论的种种哲学责难了。我所设想的那类泛心论陈述的是:物理世界观是不完善的,正如物理上同一的系统能以不同方式行动这一事实所证明的。“隐变量”并不是物理的,而是基本意识的一种形式——但正如克利福德所述,它是这样一种意识,“它伴随着每一物质粒子的运动,与我们的心理事实相比,与我们的意识相比,具有难以想象的简单性,正像物质分子的运动比之于我们的大脑的运动时所显现的那种难以想象的简单性一样”(p.267)。这是那种图景的“心灵论”部分。那种图景的“泛”

的部分落脚到下述断言：意识是世界的绝对根本的特征：物理实在的每一个要素都不乏其相关的心理方面。这些心理要素按照某种原则结合在一起，根据这种原则，部分的完全总合正好导致因果相互作用中的部分之集合，正像物理要素的集合那样。

我们可以这样来思考：总合原则之间存在着一种联系，因此在物理要素的状态叠加的情况下，我们一样有心理的结合。如果我们把这个观点推广到处在非理想环境中的多粒子系统，事实上我们必须这样做，那么我们就得出这样的结论，即量子同调性（quantum coherence）可能是更复杂的意识状态的基础，因为只有同调的多粒子系统才会保持独特的量子力学属性，而正是这些属性构成了适当的“总合规则”的基础。不过尚不清楚的是，在面对高能量环境时，多大的系统才能保持同调性。尽管如此，这个观点还是得到了多得叫人称奇的作者的赞同（参阅 M. 洛克伍德，1989，讨论了这个争论；更近的有 R. 彭罗斯，1994，采纳了这一点），但他们没有看到泛心论与他们的观点之间的相当自然的联系。我的意见是：量子同调性并不能令人满意地解决产生问题，但它可能解决结合问题。

总之，如果我们坚持这一方案，那么就自然会想到一系列更富思辨性的观点，这正是我在结论中要描述的东西。第一个观点是：只有能保持量子同调性的系统才容许“心灵的结合”，因此具有意识的复杂状态只与这些系统有联系。如果大脑支托着动物和人的意识，那么大脑（或它的某个重要的部分）就是这样的量子系统，这几乎就是彭罗斯在《心灵的幽灵》（1994）一书中的假设；它不同于彭罗斯的地方在于，它否认那些现象后面有任何新的物理学。另

一方面,我们可以设想:可想象的计算机并未保持量子同调性;我们不妨说:它们是减少量子作用的装置,因而它们并不是复杂的、统一意识状态的基础。当然,一台被设计有意图的计算机,如果可以这样设想的话,并不必然带有这个缺点。^⑩这里我们碰到了一旧观念的现代复活,该观念至少可以追溯到莱布尼兹,它把统一的实在,或莱布尼兹所说的有机体,与单纯的集合体区别开来了。

可能有人责难说,几乎不能认为:大多数量子同调的系统维系着某种具有意识的复杂生命,例如一滴液态氮。这似乎是有助于进一步思考的、颇富说明力的驳难:我们难道不能想象合成的意识之本质是与这里所说的物理系统的信息结构联系在一起的吗?液态氮的结构的基本简单性相对于大脑的复杂结构而言,在信息上是枯竭的(即使我们假定两者都保持了同调性)。因此当我们的思辨的泛心论应当(事实上必须)承认:液态氮的确有一种有联系的、统一意识状态时,它依然是一种极其原始的意识状态,也许没有不同于与单个的低级的氮原子联系在一起的意识状态的地方。

还可从别的方面考察这一个观点。现代计算机具有信息上丰富的内在结构,正是这一点使它们得以作出极其复杂的行为。然而既然它们是量子的衰减器,那么泛心论(至少是我们这里所讨论的那类泛心论)就否认它们有任何有意识的心理生活。因此泛心论者可能支持塞尔的论点:计算机没有理智,这是他从他的著名的“中文屋”思想实验(塞尔,1980)中所引出的观点,同时泛心论者还能对之作出解释,所用的术语较之塞尔无理地坚持大脑“隐藏着意向性”这一观点所用的术语要清晰一些。

我必须重申我在表达这些思辨时的独特之处:借用内格尔斯

420 第四部分

一个用语,它们显然也充满着“放在形而上学实验室内的某东西的那种令人作呕的恶臭味。”这里所提倡的泛心论是一种纯粹的哲学理论,也就是说,它没有特别的经验结论。我觉得再明显不过的是:许多卷入意识问题的争论在泛心论的旗帜下已形成了一种令人惊异的统一的态度。在我看来,承认产生问题的真实性,进而认识到它极其繁复,便自然会导致查默斯所强调的那种观点,即意识是世界的根本特征。我极力主张的是,泛心论是把意识具体化为真正根本的特征的最自然的途径。但事实上我又希望该论证把许多人带回到否定产生问题之实在性这一困难任务之上。

附录:双缝实验和量子清除器

粒子通过双缝实验的状态是:

$$(1) \Psi = 1/\sqrt{2}(\Psi_1 + \Psi_2)$$

粒子碰到探测屏位置 r 的几率由内积给出:

$$(2) \langle \Psi | \text{Pr} \Psi \rangle,$$

这里, Pr 是一种算子,它在子空间上的投影表明了粒子在 r 处出现的几率。(2)式展开后出现了相互干涉项:

$$(3) 1/2(\langle \Psi_1 | \text{Pr} \Psi_1 \rangle + \langle \Psi_2 | \text{Pr} \Psi_2 \rangle + \langle \Psi_1 | \text{Pr} \Psi_2 \rangle + \langle \Psi_2 | \text{Pr} \Psi_1 \rangle)$$

前两项分别表示粒子通过路径 1 或路径 2 后在屏上 r 处出现的几率,如果它采取了路径 1 或采取了路径 2 的话。后面两个是不可

避免的交叉项,解释了干涉图样。

现在我们引入我们的完善的探测仪。粒子和探测仪的状态,在通过了该装置并被测量之后,可以写成张量积的叠加:

$$(4) \Psi_d = 1/\sqrt{2}[(\Psi_1 \otimes L) + (\Psi_2 \otimes L)].$$

如果我们想计算在 r 处发现粒子的几率,那么我们就必须有一种算子,它作用于粒子和探测器的张量积空间;它就是 $Pr \otimes I$,这里 I 是恒等算子。在 r 处发现粒子的几率是:

$$(5) \langle \Psi_d | (Pr \otimes I) \Psi_d \rangle$$

把这完全写出来就显得很凌乱:

$$(6) \langle 1/\sqrt{2}[(\Psi_1 \otimes L) + (\Psi_2 \otimes R)] | (Pr \otimes I) 1/\sqrt{2}[(\Psi_1 \otimes L) + (\Psi_2 \otimes R)] \rangle$$

但是如果我们把 $(\Psi_1 \otimes L)$ 简写成 X , $(\Psi_2 \otimes R)$ 简写成 Y , 算子 $(Pr \otimes I)$ 简写为 O , 那么其基本形式就很清楚了:

$$(7) \langle 1/\sqrt{2}(X + Y) | O 1/\sqrt{2}(X + Y) \rangle$$

这类似于上面的(2)。不过当(7)式展开时,交叉项就会出现;我们可以得到下式:

$$(8) 1/2(\langle X | OX \rangle + \langle Y | OY \rangle + \langle X | OY \rangle + \langle Y | OX \rangle).$$

(8)的第一项和最后一项的展开足以说明:在这种情况下,粒子出现的几率有什么变化。非交叉项可以展开为:

$$\begin{aligned} (9) \langle X | OX \rangle &= \langle (\Psi_1 \otimes L) | (Pr \otimes I) (\Psi_1 \otimes L) \rangle \\ &= \langle (\Psi_1 \otimes L) | (Pr \Psi_1 \otimes L) \rangle \\ &= \langle (\Psi_1 | Pr \Psi_1) \rangle \times \langle L | L \rangle \end{aligned}$$

既然我们的所有状态矢量都是归一的, $\langle L | L \rangle = 1$ 与(9)显然就是处在 r 处的粒子出现的几率,如果它采取第一个路径的话。正

422 第四部分

像我们所期望的那样,探测器的状态对这种几率没有什么影响。

试考察(8)的交叉项,比如说, $\langle Y|OX \rangle$:

$$\begin{aligned}(10) \langle Y|OX \rangle &= \langle (\Psi_2 \otimes R) | (\Pr \otimes I) (\Psi_1 \otimes L) \rangle \\ &= \langle (\Psi_2 \otimes R) \rangle | (\Pr \Psi_1 \otimes L) \rangle \\ &= \langle \Psi_2 | \Pr \Psi_1 \rangle \times (R|L)\end{aligned}$$

这个交叉项附有因子 $(R|L)$ ((8)的其他交叉项自然有因子 $(L|R)$)。但在完善的探测器中,不同状态矢量是正交的,因此它们之间的内积为 0,进而相互干涉项就消失了。粒子在 r 处被发现的几率正好是它采取路径 1 在 r 处被发现的几率与采取路径 2 在 r 处被发现的几率之和。

引入了量子清除器就有点复杂了,它依赖于一种“数学的技巧”(参阅斯卡利,恩格勒特和沃尔瑟,1991)。四个新的状态可以这样定义:

$$\Psi_+ \equiv 1/\sqrt{2}(\Psi_1 + \Psi_2)$$

$$\Psi_- \equiv 1/\sqrt{2}(\Psi_1 - \Psi_2)$$

$$G_+ \equiv 1/\sqrt{2}(R + L)$$

$$G_- \equiv 1/(R - L)$$

状态 G_+ 和 G_- 应被看作是探测器能进入清除器的操作中的状态。初始的状态 Ψ_d 可以用我们的新态来表示:

$$(11) \Psi_d = 1/\sqrt{2}[(\Psi_+ \otimes G_+) + (\Psi_- \otimes G_-)],$$

这可以根据张量积的性质来验证。既然交叉项包含着 $\langle G_+ | G_- \rangle$ 和 $\langle G_- | G_+ \rangle$ 的消失,这个状态就没有表现相互干涉。

假如探测器处在 G_+ 的状态,那么清除器的作用可借助考察

处在 r 处的粒子的几率来说明。根据探测器处在 G_+ 中的假定, (11) 的左边的第二项就一定消失, 几率可根据 $\Psi_+ G_+$ 的状态来计算。如果那探测器处在状态 G_+ 中, 那么处在 r 处的粒子几率就是:

$$(12) \langle \Psi_+ \otimes G_+ | (Pr \otimes I) (\Psi_+ \otimes G_+) \rangle,$$

这可简化为:

$$(13) \langle \Psi_+ | Pr \Psi_+ \rangle \times \langle G_+ | G_+ \rangle.$$

既然 $\langle G_+ | G_+ \rangle = 1$, 那么我们寻求的几率就显然是 $\langle \Psi_+ | Pr \Psi_+ \rangle$ 。但是根据关于 Ψ_+ 的定义, 这个几率表达式就可展开成上面的(3)式。我们已探讨了双缝装置中的干涉效应, 尽管探测器也在进行操作, 而且我们是通过引入清除器的操作来完成这一工作的。

但是用不着建立什么比特通道。因为可根据所选择的下述假定考察几率, 这个假定是: 在清除器的作用之后, 探测器就进入状态 G_- , 它将等于 $\langle \Psi_- | Pr \Psi_- \rangle$ 。这个状态也产生干涉效应, 但它们显然与(13)的那些作用是对立的。因此除非我们知道探测器系统进入 G_+ 或 G_- 的哪一个, 我们就不能“看到”任何干涉图样, 除了用通常的手段, 否则就没有办法传递关于探测器的信息。

注 释

① 产生问题早已有之; 系统而清楚的阐述是由 J. 廷德尔所提供的(引自 W. 詹姆斯): “从大脑的物理学到相应的意识事实的过渡是不可想象的。假设一定的思想和一定的分子作用在大脑中同时发生了; 我们也没有这样的理智器官或器官的明显的雏形, 它能使我们借助推理程序从一个过渡到另一个”(引自詹姆斯, 1890, p. 147, 原文在廷德尔, 1879)。

424 第四部分

②但是我注意到:《纽约时报》的评论家 G. 约翰逊(1991)把丹尼特当作提供了意识的产生理论的人:“意识产生于所有这些神经装置的联合行为——质的飞跃既不是魔术,也不是湿从氢与氧原子的碰撞中产生时所发生的那种东西”。我深表怀疑的是,丹尼特同意这种关于他的理论的看法,因为他从未在他的著作中(1991)公开表示要解决产生问题之类的问题。

③关于蜜蜂心理生活的更详细的内容,请参阅格里芬(1992)。

④举一个常见的例子,试想我们这样来予以回答,即注意到:蜜蜂的大脑调用了 40 赫兹的振荡联结系统(BS),而机器人的加工器则没有。这一来,仅 40 赫兹的 BS 怎样产生出意识?别的联结系统难道不是可能的吗?这种忧虑支持的是意识有所谓的纯物理关联物。

⑤不过,这组“无理的事实”会随着科学的发展而改变。光速在马克斯韦从独立的、可测量的磁与电子的参数中作出有关演绎之前似乎是无理的事实(1983 年以后,光速变成了确定的东西)。但完全排除无理的事实,在科学中几乎是没有任何前途的。

⑥查默斯的观点是对 R. 博伊德(1980, p. 96)所表述的下述观点的泛化:“存在着某些构型,每当它们由物理系统实现时,不管该系统由什么实体构成,疼痛的质的感觉就会出现。”

⑦我赞成这种由金在权(1994)所提出的描述和问题的一般形式。

⑧一种本身很有趣的、尤其是本世纪初很流行的学说。参阅金在权(1994)第 8 章,其中对突现论及其与解释排斥性问题的联系作了详细论述。

⑨正是这同一个克利福德,他早年关于空间弯曲的思考预言了广义相对论以及把物质还原于严重弯曲的时空的“节”的更为激进的纲领。

⑩在这些反对意见中,我受惠于 C. 麦金关于 D. 格里芬(1994)的手稿的批判性评论,该手稿对怀特式的泛心论作了辩护。

⑪参阅雅莫尔或休斯(1989)的讨论。

⑫这是对问题(2)的唯一的回答。泛心论者必须声明:正是意识本身划分为基本的单元。否则产生问题就会以强大的力量东山再起。假设上面所作的那些考察改变了其不合理性,那么就没有理由说泛心论不能作出这个基本假定。

⑬多伊奇和洛克伍德(1991, pp. 246~52)讨论了这一著名的装置。

参考文献

R. 博伊德(1980):“没有还原论的唯物论:物理主义不需要什么”,载 N. 布洛克编:《心理学哲学读本》,哈佛大学出版社。

D. 查默斯(1995):“勇敢面对意识难题”,载《意识研究杂志》,2(3),pp. 200-19(中译见本书)。

W. 克利福德(1874):“身与心”,载《双周评论》,重印于 L. 斯蒂芬和 F. 波洛克编辑的《讲演与论文集》,伦敦:麦克米兰,1879,pp. 31-70。

D. 丹尼特(1991):《被解释的意识》,波士顿。

D. 多伊奇(1985):“量子理论,丘奇-图林原则与宇宙量子计算机”,载《伦敦皇家学会会议汇编》,A400,pp. 97-117。

B. 恩格勒特等(1994):“物质与光的二元性”,载《科学美国人》,271,pp. 86-92。

O. 弗拉纳根(1992):《意识新探》,麻省理工学院出版社。

D. 格里芬(1994):《解开世界之结:意识、自由与心身问题》,未发表的手稿。

——(1992):《动物的心灵》,芝加哥大学出版社。

R. 休斯(1989):《量子力学之解释》,哈佛大学出版社。

W. 詹姆斯(1890):《心理学原理》,第1卷,纽约:亨利·霍尔特公司,“多弗丛书”重印,1950。

M. 雅莫尔(1974):《量子力学的哲学》,纽约:威利。

G. 约翰逊(1991):“在那里真实发生的是什么?”《纽约时报》,11,10。

金在权(1994):《随附性与心灵》,剑桥大学出版社。

W. 柯克纳与 W. 汤恩(1994):“蜜蜂舞蹈语言的感性基础”,载《科学美国人》,270,pp. 74-80。

M. 洛克伍德(1989/1991):《心、脑与量子》,牛津:布莱克韦尔。

T. 内格尔(1979):“泛心论”,载《严峻的问题》,剑桥大学出版社。

——(1984):《来自不知道的地方的观点》,牛津大学出版社。

R. 彭罗斯(1994):《心灵的幽灵》,牛津大学出版社。

W. 西格尔(前言):“关于量子清除器的注释”,载《科学哲学》。

426 第四部分

J. 塞尔(1980):“心、脑与程序”,载《行为与脑科学》,3,pp. 417 - 58。

M. 斯卡利与 K. 德鲁尔(1982):“量子清除器:关于量子力学中的观察与‘延迟选择’的假设光子关联物的实验”,《物理学评论》,A, 25, pp. 2208 - 13。

M. 斯卡利等(1991):“互补性的量子光学测验”,《自然》,351, pp. 111 - 160。

J. 廷德尔(1879):《科学断想:一系列独立的论文、讲演和评论》,伦敦:朗曼斯。

L. 维特根斯坦(1953/1968):《哲学研究》,牛津:布莱克韦尔,我参考的是1968年的英文第3版。

意识、民众心理学与认知科学*

阿尔文·I. 戈德曼

定义意识

“有意识的”一词在日常用法中的标准意义是什么？它的用法之一就是表示人的总的状态。在这个意义上，如果一个人处在警觉或清醒的一般状况下：是醒着的而不是睡着了或昏迷，那么他或她就是有意识的。然而“有意识的”这个意义可能是从它的原始的或核心的意义中派生出来的，根据其原始的意义，它涉及的是部分心理状态（或许是过程或表征），诸如个别的信念、计划或情绪。当且仅当部分心理状态与现象的觉知即主观经验或感受有关联，它们才是有意识的。基于有意识的部分状态这一概念，我们便可以说，当且仅当一个人在特定时刻至少有一种有意识的状态，那么他在那时的非局部的状况才是有意识的。“有意识的”还有第三种意义：对特定对象或事件的意识。这种意义也能根据上述的核心意义加以分析。意识到对象 x ，其实就是处在某种（部分的）现象性觉知状态中，这种状态包含了关于 x 的一个表征。换言之， x 是有

* 译自内德·布洛克等编：《意识的本质：哲学争论》（论文集），麻省理工学院出版社，1997。

意识状态或表征的所指或“意向对象”。

我们先来看“有意识的”的核心意义。上面提出的定义似乎有理由认为是正确的,但不那么明白。“觉知”正好是“有意识的”的近似的同义词,因此是“现象的”。提供这个同义词并没有带来什么进展。有没有这样的定义呢,它超出了未述及的同义词的范围而仍然局限于对该概念的日常把握(不是转向意识的心理机制或它的神经学基础)?

下定义的种种尝试可能是设法(部分地)从操作上即根据为意识提供了公共的或外在的证据的那类行为类型去定义意识。例如人们可能试着把有意识状态定义为可用语言予以报告的状态。不幸的是,这一主张有许多缺陷。可用语言报告并不是一种状态是有意识的必要条件。首先,说话能力受损伤如患完全失语症的个体可能就不能报告他们的内部状态,但这并不会使那些状态成为无意识的。同样,裂脑病人的右半球即使与有语言能力的左半球割断了联系,从而使语言报告没有可能,但仍可能有觉知。第二,某些觉知状态太短暂了、太混乱了或暂时失去了报告的可能性,因此不能与语言报告器官相关联。例如做梦就有觉知的过程。然而在做梦的时候,似乎就没有与语言报告器官的沟通。做梦还常常碰到一定程度的混乱和遗忘,这些使对梦的报告即使不是不可能的也是很困难的。第三,关于意识的不令人满意的定义自动地否定了动物有有意识状态,可用语言报告的定义就是如此。即使不能肯定动物是有意识的,但无疑有一种可理解的可能性,尽管它们明显缺乏报告的能力。因此报告能力不可能是意识所不可缺少的。最后,可用语言报告并不是意识的充分条件。机器和机器人

大概能够报告它们的内部状态,但从直观上看,这并不足以授予那些状态以觉知。

第二种方案就是试图根据意识的功能例如信息的可存取性去定义意识。其例证就是 B. 巴思(1988)等人关于“整体工作空间”的观点。根据这种观点,有意识表征是这样的表征,其信息“传播”到作为整体而不是作为某种局部的或特化的信息处理器的系统之中。这种关于整体传播的观点实际上刻划了人类意识的一个可由认知科学研究的显著特征,但它未能抓住对意识概念的日常理解。一般的人肯定不会把意识理解为一组在一个大黑板上公布的、所有认知子系统都能读到的信息(巴思 1988, p. 87),因为心灵作为相互交流的子系统的集合这一图景并不是我们全部的朴素概念子目的组成部分。还有一个相近的观点赞成巴思所描述的意识的另外的特质。例如他注意到:人类的意识一般保留着这样的信息,即在还原不确定性的技术意义上,它们“提供了消息”。当不确定性已为(或接近)0 时,那些信息就倾向于离开意识;最明显的例子就是;在刺激适应时,便丧失了对重复刺激的觉知。很清楚,意识与提供消息的这种相互关联并不是普通的人一般所能认识到的东西,因此不是对意识的朴素理解的组成部分。而且,人们可能真的想到一种系统,在那里,未提供消息的或“多余的”思想仍然确实是有意识的。因此提供消息几乎不能被看作是通常所理解的意识的基本属性。同样,我们可以注意到:整体性传播和提供信息加在一起对意识也是不够的。我们不难想到一种(非人的)系统,在那里,提供消息的表征分布于所有子系统中,而那些表征却完全没有现象性的觉知。巴思还列举了意识的其他一些属性,但我不解的是,

430 第四部分

这些属性与开始两个属性结合在一起就足以满足现象意识的要求。

第三种总的方案是想用自知、自监控或高阶反省定义意识。可以假定,一种状态 S 要成为有意识的, S 的处理器在它起作用时就必须有另一种意识到或觉察到 S 的状态。然而不难看到,这种构想用“有意识的”定义自身显然是极不令人满意的。如果我们用“信念”、“思想”或某种别的指称一种信息状态的术语予以替换,那么这种有问题的循环就可得到避免。这一来就有下面的形式:“一系统的 S 状态是有意识的,当且仅当该系统具有关于它处在 S 中的高阶信念”;或者说:“……当且仅当该系统有另一种监视 S 的信息状态。”这一倡议得到了哲学家如 D. 阿姆斯特朗(1968)、D. 罗森塔尔(1986、1990、1993)和 W. 利康(1987)以及心理学家如 P. 约翰逊-莱尔德(1988a, 1988b)的赞同。

这一主张本身所必需的高阶信念或监视状态一定是有意识的吗? 如果该主张有如此意图,那么我们为了把意识授予第一阶状态,就还得求助于高阶状态的意识,而这便使循环论证无法得到消除。它还导致了无穷后退,因为每一 n 阶状态一定是由于一个 $n+1$ 阶状态才成为有意识的。假设高阶信念不必是有意识的(在认知科学中,一般假定:即使信念本身离不开意向性或关于性,但也没有必要包含意识)。根据罗森塔尔(1993)清楚陈述过的解释,该定义漏洞百出。难道不能有这样的机器人或“怪物”吗,它完全没有现象性觉知或主观感受而仍有关于它的别的内部状态的高阶信念? 事实上,我们用不着求助于思想实验就能形成这种观点。真正的人都有无意识的、为其他无意识状态所监视的表征状态。

A. 马塞尔提出了这种反对意见,他说:我们无意识地修改无意识的言语生成决定,并驱动意向(1988, p. 140)。既然高阶监视发生在认知的那些无意识领域,那么监视关系对意识绝对是不够的。除了这些反例之外,这里作为基础的观点也是令人迷惑的。元状态的获得怎么可能把主观性或感受授予并不具有它的低阶状态?为什么成为意向对象或元状态的所指能把意识授予第一阶状态?当某人有关于岩石的信念时,该岩石并不会成为有意识的。为什么第一阶心理状态只凭具有一种关于它的信念就成为有意识的呢?

值得注意的是,已考察过的不成功的一一尝试都为意识提出了一种关系定义。每一种都试图根据一种状态的意识所具有的与系统的别的事件或状态的某种关系对之作出解释:(1)意识借言语行为的可表达性;(2)其内容向系统中的别的状态或区域的可传输性,或(3)反映了目标状态的高阶状态。这些主张的失败使人们猜想(即使自然没有得到证明):任何关系主张都注定要失败。当然,有意识状态仍可能具有与别的认知事件的重要的因果或功能关系;只是意识不可能由这种关系特性来定义。我们对觉知或意识的通常理解似乎存在于有意识状态本身所具有的那些特征之中,而不在它们所具有的与别的状态的关系之中。

不妨把这类关于意识的日常理解的论点称之为内在主义。内在主义者辩护其观点的一种方法就是诉诸于颠倒光谱论证。他们试图使经验或觉知的内在性质与其功能-关系属性之间产生概念“分裂”,由此表明:前者不能简单等同于后者。传统的颠倒光谱论证试图表明的是:功能(关系的)相似性原则上由质的(内在的)歧

432 第四部分

异性所伴随。两个人或不同时间中的同一个人可能有功能上同一的状态,即同等地与所有输入、输出相互作用,但他们对它们却有着不同的经验“感受”,或完全没有感受。第二类概念分裂是由 N. 布洛克(1990)在讨论“颠倒地球”案例时提出的。他证明:质的或内在的相似性原则上可伴随着功能上的差异性。两个人或在不同时间中的同一个人可能有质上同一、但功能-关系上不同的状态。^①

出于这样的理由,布洛克(1991,1992,1993)坚持认为:“意识”的意义中至少有一种指的是内在的(而不是关系的)属性,可称之为现象意识。他把这与意识的第二种意义即领会意识区别开来,大致地说,后者能识别一种状态的合理控制言语或行为的能力。E. 比撒克(1958)也作出了相近的区分。而 J. 塞尔(1992)则否认:所谓的领会意识是一真正的意识概念(比撒克也表达过的一种忧虑),我们先前的讨论支持这种怀疑。即使是坚持领会意识是“意识”的合法意义的布洛克也没有把它当作现象意识的替换品。我们自己的论述有力地倡导的是,只有意识的现象概念才是日常用法所意指的东西。

除了用未经阐释的同义词,我们并没有认真想过对现象意识的定义,但这并不必然地表明:一切都是错误的。在语言中,并非所有的词(或许极少)能有“还原”定义。纯语词定义的循环一定有其出口。而且定义问题有规律地触及到许多基本概念,诸如“真理”、“存在”,这些概念一致抵制重要的定义。最后,用不着大惊小怪的是:某些词语尤其是这里讨论的那些词语的意义在很大程度上应归结为主观经验而不是行为标准。“有意识的”、“觉察到”和

“感受”等语词在日常理解中为什么就不应与主观上可分辨的状况而非行为事件相联系呢？例如觉知和未觉知之间的差异可能是以下述方式被弄清的。某人问你是否觉知到某种噪音。你第一次注意到这种噪音，接着你把新的觉知（噪音）的状态与先前的觉知状态相比较。还有不同程度的觉知——例如模糊的觉知、逼真的觉知等——它们为意义提供了线索。因此所意指的意义不是首先存在于主观经验而非（例如）行为倾向中吗？^②

对民众心理学的意识概念的质疑

在上面，我探讨了对意识的常识理解，并试图对关于它的意义的纯关系解释提出质疑。（这并不意味着排除了下述可能性——肯定是事实上的确定性——意识的神经基质涉及到关系维度）不过，我们关于意识的日常理解的讨论可能会被当作是几乎没有或根本没有科学意义的东西。对意识的科学研究为什么就不应重视民众心理学的理解呢？后者难道没有必要为具有根本不同轮廓的科学结构所取代吗？民众心理学的“原始感觉”^{*}或“感受性质”在完全成熟的科学唯物主义中也许不可能占有一席之地。许多作者都表达了这类疑虑，其中包括在当代科学中很有影响的关于意识的论文集（马塞尔和比撒克 1988）的四位作者。本节将考察他们的质疑。

P.S. 丘奇兰德(1988)为这种忧虑提出了一种很精彩的表述

* raw feel, 指接受对象作用后产生的先于有意识评估的感觉印象。

434 第四部分

(还可参阅她的 1983)。她披露了这样一个方案,在那里,“意识”可能重蹈了“热的流质”或“生命精气”的覆辙(P.S. 丘奇兰德 1988, p. 227)。她不排除其他可能的方案,甚至是平稳地把意识还原为神经生物学现象的方案。但她似乎倾向于“用新的、更好的大规模概念完全代替关于意识的旧的民众意念”(p. 302)。即使她没有用排除这个词,但她明显赞成对意识的排除,正像命题态度(信念、愿望等)可以被排除一样(P.M. 丘奇兰德 1981, 1988; P.S. 丘奇兰德 1986)。取消主义似乎已隐含在对热的流质、生命精气的类比推论之中,即使那也许不是她的本意。(另一解释在后面再予考察。)总之,让我们考察一下那些论证,看看它们支持什么结论。

对丘奇兰德推理的一种重构可以这样进行。意识是一种理论概念,它意味着,它无疑是由公认的规律之网所定义的。如果是这样,那么这些规律就是一般的人合理地接受并视之为意识所必不可少的规律。如果这些规律事实上是虚假的,那么就不存在例示、证明或实现该概念的现象。关于这类孕育着概念但事实上是虚假的规律,丘奇兰德提供了几个例子。她说,据一般假定,再明显不过的是,如果某人能报告环境中的某一可见方面,那么她一定有意地觉知到它。但是失去视力的人则表明这个假定是错误的。第二,丘奇兰德说,据一般假定,有意识的自我是一种不能被分析的统一体,即是说,如果自我报告了有意识经验,那么自我中就不存在别的不能觉知那种经验的部分。但是大脑连合部被切开的被试证明这个假定是错误的。第三,如果一个人没有视觉经验,那么他就晓得他没有视觉经验,这正是意识概念的组成部分。但是拒绝

接受综合症如盲人的拒绝接受(安东综合症)就证明该假定是不能成立的。第四,我们把握了的东西,我们对之也有意识,这是我们常见智慧的组成部分。但这与梦游症是相悖的:在无意识的睡眠中,梦游者成功地往来于环境之中。

如果人们接受了丘奇兰德加之于他们的那些假定,那么这些现象确实动摇了意识的日常概念的基础。但是他们是不是接受了呢?这还是未定的。当普通的大众开始被告知眼睛了时,他们会作出什么样的反应呢?他们会断定:眼睛的被试一定有意识地觉知到他们正在报告(或猜测)的内容吗?那不是我开始听到关于眼睛的描述时我怎样作出反应的问题。证明别的普通的人如此作出反应的证据是什么呢?一般的人会相信自我是一个不可分析的统一体吗,即相信,自我中不可能有一部分觉知到某种经验,而另一部分却没有这种觉知吗?心灵具有统一性这当然是哲学家们(如笛卡尔)所阐发的一种形而上学的理论,但它为一般民众完全接受了吗?在这里,丘奇兰德作为坚定的经验论者,对普通的人接受什么并没有提供什么证据。另外,意识与控制携手并进这一为梦游症合理地否定的假定又该怎样看呢?梦游症很难说是一种难解的现象。如果能够像通常所理解的那样否定意识的存在,那么这种否定在很久以前为什么没有得到认可?一般的人会错误地相信:梦游者在夜间远足时是有意识的吗?丘奇兰德并未提供什么证据去支持这种论点。因此意识所需的那些要件是不是真的为普通民众所持有,仍是一个悬而未决的问题。如果回答是否定的,那么意识的日常概念就没有为她所引证的所谓的反常事实所推翻。

在丘奇兰德的讨论中,有一种不太激进的观点。在那篇文章

436 第四部分

的另外的地方,她断言的不过是:意识并不是“自然的类别”(参阅 P.S. 丘奇兰德 1988, pp. 284ff, P.M. 丘奇兰德 1985)。它并不是单一的现象,而只是这样一类现象,其子类可以得到各种神经生物学解释的检验。神经科学在意识构建中最终可能是没有什么用处的,只适合于作出相当不同的分类。这种观点还得到了马塞尔和比撒克所编论文集的第二位撰稿人 K. 威尔克斯(1988)的发展。

我发现,否认有自然类别的地位是一更惬意的观点,尤其是当认识到它并不能衍推意识的非存在时更是如此。相比较而言,在日常语言中存在着大量的术语,如“灌木”或“臭虫”,它们并未分辨出科学研究所判定的自然类别,而只是勾画出了真实存在的对象的轮廓(试比较弗拉纳根 1992, p. 22)。科学家并不认为“灌木”或“臭虫”是特别有用的分类;它们并没有组成为解剖学或生物学上的单一的范畴。由此不能得出结论说:不存在灌木和臭虫。非但如此,J. 杜普雷(1993)令人信服地论证说,科学分类学一般来说无助于我们探讨“本质”,它们像非科学的分类一样混乱不堪。将其应用到现在的领域,这便对假定神经生物学分类报告了心脑的“真正本质”提出了质疑。尽管如此,如果丘奇兰德只是想提出(关于意识的)更弱的主张,即它不是自然的类别,那么我们也许没有严重的分歧。^③

为马塞尔和比撒克论文集撰稿的第三个人 A. 奥尔波特以大致相似于丘奇兰德的用语表达了对意识的怀疑:“……不存在单一的、作为统一体的‘现象性觉知’——没有独一无二的过程或状态,没有一种能一以贯之地得到概念化、有关于它的单独的、概念上一致的理论的现象。”(奥尔波特,1988, p. 161)。他说,他并不想否定

现象性觉知的真实性,正如他不想否认生命或理智的真实性一样,认为它们同样是不能被一体化的。尽管如此,当他否认现象性觉知能一以贯之地被概念化时,他似乎提出了比单纯的非一体化更强的主张。当他赞成本论文集的最后一位作者 D. 丹尼特(1988)关于感受性质的取消论观点时,一种更激进的解释也被引发了。奥尔波特说:“我发现他(丹尼特)对这个前后矛盾的概念的分析或毋宁说摧毁,令人耳目一新,真的使人摆脱了禁锢。究竟什么是感受性质?”(p. 162)

奥尔波特在意识概念上陷入困境的首要根源是他坚持了行为标准。所谓“标准”,他指的是判断这里的概念是否适用于一种特殊事例的一种程序。他假定,如果存在着一个概念所认定的单一的现象,那么就一定存在着证明那概念在所有事例中的适用性的单一的方法。对不同标准的要求与现象的单一性是对立的。但这种方法论观点与科学哲学中标准的论述是有冲突的,后者长期以来放弃了对理论概念的单一操作标准的要求,或确切地说愿望。例如 C. 亨普尔写道:

对系统的重要性的考虑严重妨碍了不同操作标准决定不同概念这一公理所要求的概念扩散。在科学的理论化的过程中,我们确实没有看到(例如)各种不同长度概念(例如)之间的区别,其中的每个概念都以自己的操作定义为特征。毋宁说,物理理论设想了关于长度的一个基本概念,以及在不同环境下的各种具有或多或少确定性的测量长度的方法。理论的考察将表明,一种测量方法可适用于什么范围,具有什么确定性(亨普尔,1966,pp. 94-95)。

438 第四部分

根据这种精神,在本领域内,我们所能期待的东西就是觉知的种种检验或指标,它们可能适用于不同情境,而又不可能一贯正确,因为这取决于主体所面对的不同认知任务或他们所遭受的损伤。当奥尔波特发现了多种多样的觉知标准——它们诉诸于随意动作、记忆、以及对报告的信任(奇斯曼和梅里克尔 1985)——时,他不应该绝望或推断那种现象的非单一性。诚然,那种现象最终也许是非一体化的,但这不是根源于多样的、非一致的标准必要性。为了阐释亨普尔的长度例子,用同一种操作去测量天文学的长度或亚原子的长度显然是不可能的,但它仍是关于长度的一个统一的概念。它绝不会一无例外地知道两个可用的标准总是一致的。正如我们在第一节所知的那样,语言上的可报告性在正常情况下是对觉知的一种可靠的检验,但当有语言障碍或只有有限的到达语言子系统的通道时,它显然会导致不正确的结果。

奥尔波特把行为作为意识的标准也有错误;证据同样来自于神经方面。如果能对意识的神经基质作出试探性的分辨,那么就有助于解释行为标准在其中出现歧异的有问题的案例。我们马上就能看到这方面的例子。因此为了与上一段亨普尔所描述的一般理论态度相一致,期望得到关于意识的、来自于众多源泉的有关证据是明智的,很显然,这种方法论是 O. 弗拉纳根(1992)所强调的。他说,可以同时沿着三条不同的分析路线即现象学、心理学和神经科学进行探讨。

仔细地倾听:个体对事情看起来怎样必定会说什么。然后,让心理学家和认知科学家说话。仔细地倾听:他们对心理生活怎样进行、意识在它的整个组织中有什么作为的描述。最后

仔细地倾听:神经科学家对不同类型的有意识心理事件怎样实现的想法,进而考察他们的说法与现象学家、心理学家的说法的一致性。……这种方法的目的就是要弄清三种说法能否达成一致、融为一体和实现反映上的平衡,以及在什么范围内能如此。(弗拉纳根 1992,p. 11)

弗拉纳根列举了三个例子,以说明他寻求一致、融为一体的方法怎样起作用。我只阐述其中的一个,然后补充我自己的一个。

在研究双耳听觉时,被试接受访问,并给予我们一种现象学。他们把他们在留意信道里听到的东西告诉我们,并坚持说,在非留意信道中什么也没有听到。但我们知道他们事实上受到了比如说非留意信道中出现的语言材料的影响。对这些结果的一种可能的解释就是说:被试从未意识到或觉察到非留意信道中出现的句子,即使认知系统受到了这种材料的影响。第二种解释是:非留意信道的材料只在一瞬间被意识到了。有意识过程的短暂性就能解释它为什么不能被记住,即使它事实上是有意识地被经验的。

在两种解释之间存在着一种积极的选择吗?弗拉纳根说,脑科学在这里可以大显身手。例如,F. 克里克和 C. 考克(1990)就主张:(视觉的)主观觉知与有关神经元集合中的 40-70 赫兹范围内的振荡模式有关联;即是说,与某种译码任务有关的神经元使它们的动作电位以 40-70 赫兹的振荡同时发生。40 赫兹模式只能持续很短的时间,在此情况下,有一迅速的记忆衰退发生;或者说它们能共振几秒钟,正是在此情况下,它们才变成了工作记忆的组成部分,并引起更逼真的现象学,从而成为更难忘的东西。假设这种假说(或具有相似格调的某种东西)最终在多种感觉样式中得到

了确证,当句子在留意信道中出现时,40—70 赫兹的短期振荡被观察到发生了。把现在关于短期记忆和工作记忆的理论与此发现结合起来,便可为下述第二种假说提供支持,这假说是:在留意信道中的句子引起了一种有意识的、但未能记住的现象。

D. 沙克特(1989)对用现象学、心理学和神经科学这样的“三角学求证”有意识觉知现象的尝试作出了另一种说明(弗拉纳根未予讨论)。沙克特首先把意识与明确和不明确记忆之间的悬殊差别放在一起加以讨论,这里所说的明确记忆粗略地说是指“具有意识的记忆”,而不明确的记忆是指这样的状态,在那里,先前的经验促成了与对这些经验的有意识记忆没有任何关系的检验的完成。沙克特接着转向了关于缺乏特定知觉和认知的脑损伤病人的研究成果。在许多情况下,病人对他们没有觉知到的东西能得到知识。记忆缺失病人就是显著的例子,但其他类型的大脑损伤也会产生这样的条件,在此条件下,病人对他们不能有意识地知觉、分辨、认知或理解的刺激会表现出有不明确的知识。失认症病人有认识熟悉的面孔的困难,不能报告(有意识)所熟悉的家人、亲戚和朋友的面孔。尽管没有有意识的熟悉,但仍有资料表明:这些病人对熟悉的面容有不明确的知识。眼瞎的病人是另一种有名的案例,在那里,病人可以模糊地领会那不能告知于有意识视觉经验的信息。在不带失写症的失读综合症中,在视觉对象失认症中,在布罗卡和维里克的失语症以及在关于裂脑病人的大脑半球间的传输的研究中,近似的分裂比比皆是。

关于这些材料,沙克特强调了两个关键的观点。第一,据观察,相同的结果模式广泛存在于不同的病人群、诸实验任务、各类

信息和种种知觉/认知过程之中。第二,未能进入意识只是一种选择的或局部的特例。病人对他们的特定损伤范围之外的信息,不存在有意识地加以领会的困难。基于这类证据,沙克特设想了一种结构,它假定有一种不同的子系统,可称之为有意识觉知系统(CAS),它能与加工和表征各种类型的信息的模块机制相互作用。在有神经心理学损伤的事例中,特定的加工和记忆模块有选择地与有意识系统相分离,由此导致了特定区域丧失有意识的经验。CAS在这种结构中起着三种功能作用。第一,其激活对被试感觉到记忆以及知道或知觉是必不可少的。第二,CAS是一种整合模数加工输出的“整体数据库”。第三,CAS把输出送至执行系统,后者又与注意力的调节以及对诸如记忆搜寻、计划等随意活动的驱动有关联。最后,关于神经解剖学的可能性,沙克特利用了迪蒙德和梅索伦及其同僚的成果(尤其是梅索伦 1983,1985),提出:顶骨皮质区域显然有相互联系的模式,如果它们是构成具有CAS属性和功能的更大系统的组成部分,那么上述模式就是必要的。因此沙克特的主张就是对弗拉纳根关于意识现象的“三角学求证”方法的阐释。

D·丹尼特是马塞尔和比撒克主编的论文集的第四位作者,他对意识至少是现象意识提出了严峻的质疑。他专注的目标就是关于感受性质的概念,他的观点直截了当地说就是取消论:“与乍一看显而易见的东西相反,压根就不存在感受性质”(丹尼特 1988, p. 74;还可看他的 1991,第 12 章,“不够格的感受性质”)。丹尼特的论证像奥尔波特的一样,环绕着证实问题而展开。在这篇论文和后一类论述(丹尼特 1991;丹尼特和金斯伯恩 1992)中,他陈述

了一些事例,在那里,现象性觉知是否或何时发生是不可能合理地确定的。他的关于感受性质的取消论结论首先就是以这种证实上的不确定性为基础的。^④既然人们不可能判断哪一种感受性质描述是正确的,那么就根本不存在关于感受性质的真实的描述,换言之,通常所理解的现象意识就是一种幻觉。

在1988年的论文“取消感受性质”中,丹尼特列举了这样的例子,有位咖啡品尝者认为:六年前,他加入麦克斯韦尔咖啡公司时经常品尝到的滋味感受性质,以后再也没有尝到过。问题是:他的滋味感受性质是否发生了改变,或者说他的判断标准或许他的记忆是否发生了改变。丹尼特论证说:在这些(明显)对立的假说中,存在着基本的、证明上的不确定性。在后来发表的论文(丹尼特1991;丹尼特和金斯伯恩1992)中,他陈述了这样一个例子,即有一个人,他很快瞟了一眼一个没戴眼镜的女士,然后,他把她记成了戴眼镜的。有两种可供选择的说明。“奥威尔*式”的说明是:有一种关于没戴眼镜的女士的现象经验,它伴随着过去记忆中的戴眼镜的女士的混杂经验。(这种说明之所以是“奥威尔式”的,是因为历史被改写了。)*“斯大林式”的说明是:根本就没有发生过这种现象经验。丹尼特主张,不管是“从内”(通过观察者自己)还是“从外”,都没有办法区分这些竞争的理论,而且他似乎得出结论说:根本就没有确凿可靠的事实支持所谓的现象经验。第三个例子是关于“元比较”的。让被试快速(30秒)看一个圆盘,紧接着再

* 奥威尔(1903-1950),英国讽刺小说家,在《一九八四》这部使他名声大振的讽刺小说中,他描述了一个过分集权的社会如何歪曲真相、篡改历史的故事。

看炸面饼圈,其内边正好与圆盘的外边相合。如果正好如此,那么被试就会报告说,看到的只是炸面饼圈。不过有证据证明关于圆盘的信息在大脑中得到了表征。例如被试在猜测究竟有一个还是两个刺激时,其成绩就胜过于运气。“奥威尔式”的说明会说:被试对圆盘和炸面饼圈两者都有有意识经验,但后者抹去了关于前者的有意识记忆。“斯大林式”的说明则认为:圆盘从属于前意识过程,而关于它的意识为紧随其后的炸面饼圈刺激所阻止。因此“奥威尔式”与“斯大林式”的说明对是否存在着被试并未记住的关于圆盘的意识短暂闪现,意见是不一致的。丹尼特论证说:在这两个说明之间,不可能存在什么事实根据,因为它们不可能被区别开来。

这些论证方法存在着几个问题。第一是一个哲学的观点。即使真地没有人,包括那个被试,后来能料定两个说明中哪一个是对的,但为什么由此说没有事实根据呢?对于尤里乌斯·凯撒被杀害时穿的袍子上有什么装饰品,现在的任何人都不能提供判决性的证据。因此几乎不能得出结论说:关于那个问题真地没有独立于我们的证明的事实。第二,丹尼特主张:根据任何一种说明,经验都会“感觉到是一样的”(丹尼特 1991, p. 123)。然而,正如布洛克所指出的那样,这个论断是错误的,或至少是以假设为论据的。如果存在着现象经验这样的东西,那么在圆盘意识的短暂闪现与非短暂闪现之间就会存在着微妙的主观差异。即使这种闪现太快了,以致被试不能察觉到或不能予以报告。(注意“察觉到”是一个判断或相信的问题,它不应等同于视觉意识的闪现本身。)第三,丹尼特过于轻率地断言,没有科学的根据赞成一个优于另一个。(弗

444 第四部分

拉纳根 1992;布洛克 1993)。再者,假设我们已从正常的情境中(在那里不存在知觉或记忆的骗局)找到了根据来证明:克里克-考克的下述假说,即意识与 40-70 赫兹的振荡有关,或者克里克-考克的另一个假设,即意识从根本上说与新皮质中的第 5 层的锥骨神经有关联。如果我们从正常的事例中得到了趋于一致的证据来支持某种假说,那么我们就用神经信息解决元比较中的现象事实问题。^⑤然而丹尼特对脑科学在这一微妙层面的解释力持怀疑态度,但我想重复弗拉纳根的格言:“不要说不能说的”(弗拉纳根 1992, pp. 14-15)。

内在的质与自我归因

在上一节中,我批驳了以拙劣(或朦胧)的方式诋毁意识的民众心理学的种种尝试。在本节中,我将从辩护转向进攻。我将陈述一些确实的理由,以说明要严肃地对待现象意识,要重视对民众心理学的开发利用,把它当作科学心理学的材料。无可争辩的是,首次参加实验的被试常把心理状态归之于自身,所使用的是日常心理术语,诸如“相信”、“想望”、“计划”、“疼痛”等。我想论证的是:对这种认知-语言活动的最好解释就是下述现象意识事件的真实出现,这些事件是用内在的而不是关系的术语而得到范畴化的。这里陈述的论点在别的地方曾得到了更详细的阐发(详见戈德曼 1993)。

试考察一下关于两个行为主义者的一个老笑话。在做爱完后,第一个人对第二个人说:“对你来说,那太妙了,但对我又怎样

呢?”为什么这是可笑的?因为与行为主义相悖的是,在关于心理状态的知识中似乎有一种信息的不对称现象,这种现象有利于第一人称的知识,不利于第三人称的知识,而不是相反。人们接近自己的心理状态比接近他人的心理状态似乎有不同的、更好的形式。这种“优越的通道”没有必要是完全可靠的或不可错的,但似乎常常是可靠的。如果不是因为这样的事实,即人们一般能正确地报告他们(当下)的有意识状态的存在及内容,那么可用言语报告性为什么是有意识状态的正常的可靠的指示器呢?

优越通道的论点自然会遭到许多人的反对。在新近的心理文献中,A. 戈帕里克(1993)主张:人们总是根据他们在推论别人的心理状态时所用的同样的理论来推论自己的心理状态。当否认她是行为主义者时,她实际上接受了行为主义者的观点:在第一人称和第三人称心理归因之间并不存在信息上的不对称性。

不幸的是,戈帕里克和别的心理学家对可能作为自我归因之基础的推理过程语焉不详。这类说明的可能的细节,最好是到功能主义的哲学学说中去寻找。哲学功能主义坚持认为:一般的人总是通过挑出一种特定的“功能作用”,即一组与刺激输入、行为输出和别的心理状态的因果功能关系,去理解每一日常的心理状态描述符。^⑥如果这是正确的,那么把人们自己的心理状态范畴化的任务就一定与决定哪一种功能作用被某人当下的状态所例示有关。这个任务会怎样被完成呢?

试看描述符“饥渴”。在功能主义者看来,“饥渴”的意义(部分)是由下述属性所给予的:(1)它往往不是由一会儿的饮水所引起的一种状态;(2)它是一种往往会引起一种饮水的愿望的状态。

446 第四部分

这两种条件是饥渴的特定的功能作用的组成部分。“想喝水”是一种什么类型的状态？它（首先）被理解为这样一种状态，当它与相信适于饮用的液体容器在手边的信念相结合时，它将使某人把那容器拿到嘴边，并喝下去。须知，所谓的理解是纯关系的，因为它们最终把那些状态与周围的输入、输出关联起来。它们用不着参照或承诺那状态的任何现象特征。因此功能主义深受感受性质怀疑论者或害怕感受性质的人的欢迎。

如果功能主义是正确的，那么一个想确定他/她当下是否饥渴或当下是否想喝水的人会使用什么样的推理程序呢？既然饥渴是一种根据其输入、输出和别的心理状态的关系而加以理解的状态，那么人们大概会通过设法确定什么输入先于它或什么输出或别的内在状态后于它，而对现在的目标状态作出归类。就饥渴概念来说，人们可能会设法回忆近来是否饮用过什么东西，或者说人们可能会等着观察目标状态是否为饮水的愿望所伴随。就输入、输出是有用的证明材料而言，功能主义与（哲学）行为主义并没有什么不同。功能主义只是增加了与别的心理状态的关系，才有别于行为主义。这对功能主义究竟有多大作用仍有待考察。

人们以功能主义所必需的方式去完成心理自我归因的任务，这真的是合理的吗？有三方面的困难。先看早上头疼的例子。你一醒来，就碰到了一种特殊的感受状态，马上就把它归之为头疼。然而你并没有想起来它可能是由什么引起的。你没有记起一次狂饮，一段摇滚音乐或别的类似的东西。因此你不能根据关于作为头疼的典型原因的先前输出的知识推论出“头疼”的类别。你也不会做出像拿阿斯匹林这样的、可能有助于你把你的状态认作是

头疼的行动。存在着某种别的内在状态吗,它是你所分辨出的、有助于“头疼”分类的状态?也许你注意到了摆脱那种状态的愿望。但这并不能把头疼与别的不希望有的状态如在别的地方的疼痛甚或口渴区别开来。而且你在认识到那愿望之前就可能认识到那种头疼。最后,诉诸于那种愿望实际上把困难转移到那个状态上了:你是怎样把那种状态归结为摆脱最初状态的愿望?在这点上,我们不妨把现在的问题称之为证据不充分问题。如果人们只用了迄今所考察的那类关系信息,那么说归类的任务实际上可能被完成,要么是全部被完成,要么是很快就被完成,就是值得怀疑的。

我们对功能主义也许是不公正的。我们对早上头疼这个事例的讨论似乎假定:只有先于或后于目标状态的实在事件或状态才可用作其归类的证据。而功能主义并没有把有关的证据限制在实在的事件或状态之内。功能主义说,状态的同一取决于它的虚拟的属性,例如它产生的行为或别的内在状态。因此人们也许根据心理状态的虚拟的属性对包括早上头疼在内的心理状态作出归类。但这又引出了功能主义模型的第二个难题:对虚拟的属性一无所知。人们怎么能够判断一种当下的状态具有哪些虚拟的属性呢?假设你事实上并不相信:你现在拿着一个装有可饮用液体的容器。你怎么可能判断:如果你真的有这样一种信念,那么它就会与引起你把那容器拿到嘴边并喝下的当下状态结合在一起?然而在功能主义看来,那正好是你把当下状态归结为喝的愿望时所必需的虚拟信息。

也许有人回答说:不管那是怎样得到的,我们毕竟常常拥有所必需的虚拟信息。例如就早上头疼这个例子来说,你可能知道:所

448 第四部分

说的那个状态就是引起你拿阿斯匹林的状态;甚至在你实际下床、走到药柜之前,这就已是你所知道的东西。然而问题在于:你是怎样知道这一点的。难道你不是用下述方式知道它的吗,即先把那状态归结为头疼,接着得出你应拿阿斯匹林的结论?如果这是对的,那么你作出归类时并没有利用虚拟的信息。恰恰相反:你是用分类信息推论出了虚拟的属性。

功能主义模型的第三个难题至少像前面的难题一样严峻。这个困难来自于功能主义的两个核心特征:(1)标记心理状态的类型同一依赖于与它(它的关联项)有关的状态的类型同一,(2)许多关联项(即别的内在状态)的类型同一进一步又依赖于它们的关联项。例如人们要把一种状态当作是饥渴的例示,就可能有必要把它的结果之一看作是想喝的愿望。然而把一种特殊的结果作为想喝的愿望又要求人们分辨出它的关联项,其中许多也是这样的内在状态,即它们的同一就是它们的关联项的同一,等等。问题很快变得错综复杂起来。明显存在着组合爆炸的威胁;为了分辨初始的状态,太多的别的内在状态需要从类型上加以分辨。

由于这些困难,我们的分类工作不一定只用虚拟信息或功能主义所假定的那种因果关系信息。毋宁说,我们的系统必须依赖于目标状态的某些属性,这些属性从属于(而不是倾向于)且内在于那个状态,即是说,这些属性是那些状态本身所具有的、而不是由于它们与别的状态的关系而具有的属性。这些从属的、内在的属性会是什么呢?

似乎有两个候选者可充任这个角色:(1)神经属性,(2)质的或现象的属性。每一种感觉状态都有某些神经属性,这些属性可能

是从属的、内在的。(注意:神经属性可能内在于作为整体的状态之中,即使它涉及到构成性神经结构中的关系,正如温度是一团气体的从属属性一样,即使它涉及到作为构成因素的分子中的关系。)但是神经属性不是分类系统能够接近的那种属性。未经训练的人肯定没有接近神经属性的“人的”通道,对它们只能是一无所知。还会有接近这些属性的“亚人的”通道吗?不言而喻,神经事件涉及到有关的信息加工;大脑中的所有信息加工,在最低限度上都是神经加工。然而问题在于:这些神经事件所编码的内容(意义)是不是关于虚拟的属性正好据以从中推论出来的神经属性的内容。这似乎是完全不可能的。

很显然,大量的信息加工发生在有机体内的亚人层面。不过,当加工是纯亚人的时,似乎没有可认作“心理的”语言标签的产生。各种体内平衡的活动发生了,在那里,信息传送到某些液体或化学制品如葡萄糖的层面。但我们对这些事件或活动并没有民众心理学的标签。同样,瞳孔反应不仅在应答亮度变化而且在应答环境刺激的兴奋值时会连续地发生变化(韦斯克兰茨 1988)。但是对这些与瞳孔状态有关的事件并没有心理主义的标签,这主要是因为这些状态在觉知中并没有被自动记录下来。我们自发的心理命名系统似乎没有办法接近纯亚人信息。只有当心理学的或神经学的事件引起有意识感觉如口渴、热感等时,或引起别的有意识心理状态时,原始的语言标签才会介入或被运用。

这一来,留给我们的似乎是质的或现象的属性,即是说感受性质,它们是允许心理主义归类的内在属性。正如我们已论证过的那样,这些才真正是能被直接分辨或监视的(即使不是必然不可错

450 第四部分

的)从属的、内在的属性。因此关于人们自己的心理状态怎样被归类的最有前途的心理学模型似乎就是对这些状态的现象属性作出分辨,例如对发痒的“痒”或头疼的疼痛质作出分辨。(更确切地说,这些现象属性的微观构成物也可能被利用。参阅戈德曼1993)。如果这是正确的,那么现象性觉知在解释常见的认知任务之完成中就起着不可或缺的作用。

这里的讨论主要关注的是心理状态的语词自我归因这样的现象。但论证同样是以纯内在、非言语活动为基础的,L. 韦斯克兰茨(1988)把这种活动称之为“监视”反应。他主张:眼瞎的病人(至少是其中的许多人)是不能与监视系统发生联系的。如果我们像瞎眼病人被要求做的那样,不得不在差别很大的垂直的和水平的光栅之间作出区分,那么我们会适当地按住两个键中的一个,以便指出“垂直的”或“水平的”。但是一般来说,我们在按第三个表明我们正在“看”而不是在“猜”的键时也没有什么困难。韦斯克兰茨说,这就是我们有别于眼瞎病人的地方,后者的第三键反应是“猜测”。关于这种区别的很好模型就是:存在着一种我们所监视的额外的状态——现象的或质的状态——这是瞎眼病人完全没有的东西(或只以有限的或衰退的形式具有)。这些病人有模糊的信息状态,但它们属于不同的类别,不能以相同的方式被监视。这就是瞎眼病人为什么把要求他们回答的分辨问题当作是无意义的游戏的原因(至少在开始是如此)。

在此关头冒出的一个问题是:前述的解释能否从感觉(包括知觉状态)推广到所谓的命题态度即思想、想望、怀疑、意向等的自我归因上。心灵哲学家们常常坚持说:只有感觉才有现象属性。果

真如此,上述说明就不能解释非感觉心理描述符的自我运用。有理由说,所述的推广的前景是宽广的。首先,作为一个术语学事实,我们应当准备使用“现象的”或“质的”术语以表示具有任何类型的主观感受的状态,而不仅仅只是表示感觉的质。其次,试考察与熟悉面孔有关的熟悉感受。除了看到熟悉面孔的纯视觉的质之外,还有关于“貌似熟悉”的附带的质。后一种质就是失认症可能没有的东西,即使他们完全享有观看面孔的视觉维度。因此,在纯感觉的(例如视觉的)感受之外,似乎还有非感觉感受这样的东西。这大概也适用于对一般记忆的感受,而且可以合情合理地主张:感觉到相信某东西而不是期望它,感觉到希望某东西而不是惧怕它,等等,都有其特定的方式(戈德曼 1993;弗拉纳根 1992, pp. 67 - 68)。

有时有人说,如果感受性质存在,那么它们在认知中就没有什么功能或因果作用(杰克逊 1982)。这不是这里要辩护的观点。正如前述论证所表明的那样,现象状态有因果上的结果:它们常常导致语言上的自我归因,而且从纯内在的方面说,它们激发监视活动。^①(在韦斯克兰茨看来,这种活动对整合、连结人的思想也有作用)。我们前面论证感受性质不能单纯等同于功能状态这一结论,不应与感受性质完全没有功能属性这一论点混淆起来。感受性质之所以不能等同于功能状态,是因为(1)没有感受性质,被挑选的功能状态可能(“在逻辑上”)也存在,(2)同一感受性质在别的人身上或在位置上不同于我们的人身上起着不同的作用。但是在我们之中,它们事实上起着上面描述过的(尤其是无可怀疑的)那种特定的功能作用。同样,我们前面对关于意识的自我监视定义的批

评与现在赞成韦斯克兰茨的有意识状态对监视特别有用的观点一点也不矛盾。我们前面的批评只是针对这样的观点,即认为高阶监视已完全阐明了意识概念,或认为两者是同一的。拒斥这个观点与下述主张并不矛盾,这主张是:在我们之内,对一状态的意识使它有利于得到迅速的、特殊的监视。既然现象属性在我们的心理系统中起着重要的功能作用,那么它值得认知科学去认识,而不应作为理论上无价值的垃圾被抛弃。

注 释

① 布洛克担心的是由他的论证所构成的下述批评。即使对功能主义的“长臂”阐释没有什么作用,但对功能主义的“短臂”阐释或许有用,根据前者,输入和输出是外在的对象或事件,而根据后者,有关的输入与输出是内在于有机体的心理状态。如果我们关心的是对觉知属性的日常理解,就像我们在本章的这一节中所关心的那样,那么上述花招于事无补。一般的人对觉知或质的属性的理解并不来自于只为科学上的专家所熟知的心理学事实。

功能主义者常尝试另一种策略,以使用功能的即关系的术语把握质的特征。他们设想:疼痛常常引起关于某人处在疼痛中的信念,这正好是“疼痛”这样的心理术语的部分意义(休梅克 1975)。不过,这个主张面临着循环问题(希尔 1993)。一个人不可能完全领会那个定义授予“疼痛”的内容,除非他理解了“相信他处在疼痛中”这一表达式。但是既然“相信他处在疼痛中”的意义本身依赖于“疼痛”的意义,那么那定义可衍推的似乎是:他不能把握“疼痛”的意义,除非他已经把握了它的意义。

② 这个方案有时被认为莫立于“私人语言”或“内在显学习”(维特根斯坦 1953)的所谓的不可能性之上,围绕这一问题,诞生了大量的哲学文献。但是,所有的文献(在我看来)都是有争议的,而且可能受到挑战。不仅如此,其中许多早于或忽视了认知科学中的发展,正是这些发展对它的某些前提条件提出了质疑。

③ 说明丘奇兰德的这种非取消论方式的一个进一步的理由就是:当她

说意识可能重蹈热的流质或生命精气的覆辙时,她在这些术语上加了引号。这是哲学家们说明那个课题是语词而非实在的标准的方法。因此丘奇兰德的意思也许是:“意识”这一日常用语面对科学所宠爱的术语时应予抛弃,由此并不能得出结论说:意识将像热的流质和生命精气(在现在)已成了被否定的存在那样也会成为被否定的存在。而且她新近的作品几乎没有显示取消论与意识势不两立的迹象。例如似乎可以断定,丘奇兰德和拉梅钱德朗是精神上的实在论者。

④ 丹尼特论证的另一部分是:据哲学家们说,感受性质是不可言喻性、不可错性、不可分析性等属性的混合,这些属性是不可能同时被切实感觉到的。然而,正如弗拉纳根所指出的(1992,第4章),人们可以具有感受性质的结构而用不着承认各种哲学家强加给它们的属性的整体混合。他说,应除去(即取消)那些属性——或至少是其中的一些——而只留下真正的感受性质。

⑤ 即使丹尼特赞成激进的修正主张,即不存在现象的或质的意识,但他仍力图使这种观点变得温和,方法是用专门由信念或判断所构成的关于意识的更“认知的”形式来取代它(参看他^对意识的“多方案”说明,在这里“方案”是一种暂时的判断)。这种战略的一个精美的样品就是他对“填充”盲点的论述。他不赞成大脑提供“质的”视觉表征以填充盲点的观点,而主张:它产生的恰恰是信念(丹尼特 1991, pp. 354 - 355)

⑥ 对功能主义以及它的各种变种的系统阐述,参看布洛克(1980)。这里所涉及到的功能主义变种是分析的或常识的功能主义,它与科学的功能主义(或心理功能主义)是对立的。在我的 1993 年出版的著作中,我把现在的这种功能主义称之为表征的功能主义,意思是说人们用功能作用表征心理-状态语词。

⑦ 即使我主张现象状态或属性有因果有效性,但我不敢妄言,我已阐述了这一观点的形而上学细节。

参考文献

A. 奥尔波特(1988):“什么是意识?”载 A.J. 马塞尔和 E. 比撒克编:《当代科学中的意识》,牛津大学出版社。

D. 阿姆斯特朗(1968):《唯物主义的心灵理论》,纽约,人文科学出版社。

454 第四部分

B. 巴思(1988):《意识的认知理论》,剑桥大学出版社。

E. 比撒克(1988):“(闹鬼的)大脑与意识”,载 A. 马塞尔等编:《当代科学中的意识》。

N. 布洛克(1980):“功能主义的麻烦”,载 N. 布洛克编:《心理学哲学读本》(Vol. 1)哈佛大学出版社。

——(1990):“颠倒的地球”,载 J. 汤柏林编:《哲学展望》,4. “行动理论与心灵哲学”,阿塔斯卡多,CA:里奇菲尔出版公司。

——(1991):“反副现象论的证据”,载《行为与脑科学》,14,670-672。

——(1992):“关于反现象意识的诡辩”,《行为与脑科学》,15,205-206。

——(1993):“对被解释的意识的再思考”,《哲学杂志》,90,181-193。

J. 奇斯曼与 P. 梅里克尔(1985):“语词认知与意识”,载 D. 贝斯纳和 T.G. 沃勒等编:《阅读研究:理论与实践上的发展》(Vol. 5),纽约:学术出版社。

P.M. 丘奇兰德(1981):“取消唯物主义与命题态度”,载《哲学杂志》,78,69-90。

——(1985):“我们的自我概念的思辨本质:对某些批评的答复”,载 J. 麦金托什和 D. 科普编:《心灵哲学新论》,《加拿大哲学杂志增刊》,157-173。

——(1988):《物质与意识》,麻省理工学院出版社。

P.S. 丘奇兰德(1983):“意识:概念之转换”,载《太平洋哲学季刊》,64,80-93。

——(1986):《神经哲学:迈向心脑的一体化科学》,麻省理工学院出版社。

——(1988):“还原与意识的神经生物学基础”,载 A.J. 马塞尔等编:《当代科学中的意识》。

P.S. 丘奇兰德与 V. 拉梅钱德朗(1993):“填充:丹尼特为什么是错的?”载 B. 达尔邦编:《丹尼特及其批评者》,牛津:布莱克韦尔。

F. 克里克与 C. 考克(1990):“意识的神经生物学理论”,《神经科学专家讨论会》,2,263-275。

D. 丹尼特(1988):“取消感受性质”,载 A. 马塞尔等编:《当代科学中的

意识》。

——(1991):《被解释的意识》,波士顿。

D. 丹尼特与 M. 金斯伯恩:“时间与观察者:意识何时在大脑中的何处”,《行为与脑科学》,15,183-201。

S. 迪蒙德(1976):“意识的大脑回路”,《大脑、行为与进化》,13,376-395。

J. 杜普雷(1993):《事物的无序:科学无统一性的形而上学基础》,哈佛大学出版社。

O. 弗拉纳根(1992):《意识新探》,麻省理工学院出版社。

A. 戈德曼(1993):“民众心理学的心理学”,《行为与脑科学》,16,15-28。

A. 戈帕里克(1993):“我们怎样知道我们的心灵:意向性的第一人称知识之幻相”,《行为与脑科学》,16,1-14。

C. 亨普尔(1966):《自然科学的哲学》,恩格伍德·克里弗斯,新泽西:Prentice-Hall。

C.S. 希尔(1993):“质的特征、类型唯物论与分析功能主义的循环”,《行为与脑科学》,16,50-51。

F. 杰克逊(1982):“副现象的感受性质”,《哲学季刊》,32,127-136。

P. 约翰逊-莱尔德(1988a):“意识之计算分析”,载 A.J. 马塞尔等编:《当代科学中的意识》。

——(1988b):《计算机与心灵》,哈佛大学出版社。

W. 利康(1987):《意识》,麻省理工学院出版社。

A. 马塞尔(1988):“现象经验与功能主义”,载马塞尔等编:《当代科学中的意识》。

M.-M. 梅索伦(1983):“功能解剖学与有向注意的半球专门化——顶叶及其联系的作用”,《神经科学的走向》,6,384-384。

——(1985):“注意、混乱状态与否定”,载 M.-M. 梅索伦编:《行为神经学原理》,费城:戴维斯。

D. 罗森塔尔(1986):“意识的两个概念”,《哲学研究》,49,329-359。

——(1990):“意识理论”,《ZIF 报告》,No. 40。

456 第四部分

——(1993):“关于人的思想的思想”,载 M. 戴维斯等编:《意识》,牛津:布莱克韦尔。

D. 沙克特(1989):“论记忆与意识的关系:可解的相互作用与有意识经验”,载 H.L. 罗蒂杰尔等编:《记忆与意识的变种》,新泽西:埃尔博。

J. 塞尔(1992):《心灵的再发现》,麻省理工学院出版社。

S. 休梅克(1975):“功能主义与感受性质”,《哲学研究》,27,291-315。

L. 韦斯克兰茨(1988):“视觉与记忆的神经心理学对意识问题的某些贡献”,载 A.J. 马塞尔等编:《当代科学中的意识》。

K. 威尔克斯(1988):“——, yishi, duh, um 与意识”,载 A.J. 马塞尔等编:《当代科学中的意识》。

L. 维特根斯坦(1953):《哲学研究》。

关于心灵起源的四个假说*

朱利安·杰恩斯

今天上午我要向大家报告的基本上是一种有关历史的理论,它将引出这样一个结论,即意识和心灵不能离开它们的历史而得到理解。在我看来,关于心灵的一切讨论都必须是历时性的,而不应像现在大多数这类讨论那样成为共时性的。我所认定的这一立场是不讨人喜欢的,甚或与现代的哲学思维不合拍,但它正是我想向你们竭力陈述的观点,不管你们是否接受我下面所要表述的每一细节。

流行的意识理论批判

我首先要论及的问题是进化论意义上的意识起源问题。它正是 19 世纪后半叶进化论诞生以后被视之为极其重要的问题,对于通过自然选择的物种起源理论的完善及其进一步辩护,它确实是绝对必要的。进化论已给唯物主义的理论以及物种的偶然起源和增殖作出了解释,但是它怎么可能解释那似乎是非物质的人类心

* 本文是作者在 1984 年 8 月于奥地利举行的第 9 届国际维特根斯坦专题讨论会上的讲演稿,后收入以英德两种文字合刊的会议论文集《心灵哲学·心理学哲学》(R. 齐硕姆编,维也纳,1985),中文根据英文译出,小标题为译者所加。

458 第四部分

灵的起源呢？这在过去是、现在依然是一个难题。意识怎么可能通过自然选择而从单纯的物质中派生出来呢？

有些人例如像华莱士回答说：那是不可能的，人类的意识一定是神授之于人的。令人惊异的是，达尔文也持此论，当然他又认为：它发生在进化之初，它是“原本由造物主呼出的气进入一些类型或一种类型”而形成的几种力量中的一种力量，我此处的引文出自于《物种起源》的最后一段。达尔文关于有意识的心灵与形态结构平行进化的观点导致了他的弟子罗马尼斯探讨意识的进化发展的研究纲领——一个不久就被认为是不妥当的纲领。其他尝试性的解决方案接踵而至，诸如赫胥黎的不合理的、对抗性的原始行为主义或神秘主义和关于突现性进化的不精确假说，后来在哲学中又有怀特海、亚历山大、皮尔等人的新实在论，不幸的是，这种理论由于量子力学中的“观察者介入”的一些佯谬又有一种新式的复活。沿着这条弯弯曲曲的道路，心理学家便自然而然地通过根本否认意识的存在而对意识的起源问题作出轻率的回答。经历这样令人精疲力竭的、摸索的历史之后，上述问题在 19 世纪丧失了它的紧迫性以致在哲学和科学的探讨中销声匿迹，实不足以为奇。我认为：在这次专题讨论会上，应回过头来对许多曾受重视的问题作出澄清。

在解决意识起源问题的过程中，造成这种彻底失败的原因就在于：意识概念是一虚妄的、不实用的概念。经过漫长的历史，它上面堆积着大量的文化尘埃，而此尘埃又妨碍了对问题的严格的审视。指称中的某种多余部分是来自于宗教的历史，在此历史中，灵魂被授予各种可能的心理功能，这些功能后来又像那些取代灵

意 识 459

魂一词的语词一样变成了心灵或意识的属性。错误的另一根源在于哲学思维本身,这种思维作为心灵的功能是我们借逻辑推理可以到达的东西,它也是意识的一个组成部分,尽管这些命题没有任何内省的合法性。因此我们分析的第一步也是适当的步骤就是要设法把意识概念从这些历史过失的冲击层中解救出来。

当我说把逻辑上推论出来的能力注入意识的观念之中时,我的意思指的是像感知这样一类过程。当我能意识到感知的产物时(即使不总是如此),实际的过程决不会到达有意识的内省。在历史上,我们从我们感觉器官的一种实现中推论出、抽象出这些过程,后来又由于对心灵与物质或灵魂与肉体的先验假定,我们便相信这些过程是意识的一部分——其实并非如此。如果你们哪一个人仍然坚持认为感知应等同于意识,那么我倒觉得你们循着一条路径达到了一种归谬论证:你们一定会说,既然所有动物都有感知,因而也都是有意识的;如此回溯,经过进化的阶梯甚至可以追溯至单细胞原生动物门,进而到血液的变形虫状的白细胞,因为它们可以感知到细菌并能吞没它们。由此说来,它们也应是有意识的。说今天上午我们这里每一位身上每立方厘米血液中的一万个有意识的存在,这恐怕是没有人想为之作辩护的观点。

对于你们中的一些人来说,我现在对意识的定义似乎比通常的要窄。不错,是这样。但它只是我现在正陈述的、我关于什么是意识而不是感知的理论的一部分——像我已说明过的那样。我的意思绝不是说:所有意识都是内省,而只是说意识是可以反省到的。意识就是现在基尔希堡街上任一打“心里”正在发生的一切。如果你要求他们把他们正在想的一切说出来,录在录音机里,

460 第四部分

你无论如何找不到成堆的感知,而只能发现一连串流动的担忧、后悔、希望、往事回忆、内部对话、独白、计划、想象——所有这些都是意识的材料。

为什么有些人总是把意识与感知混淆起来了呢?原因有几个方面。第一,像我要描述的那样,意识是知觉的一种类似物,因此两者好像真的能等同起来。我想说明的是:因为物理的、具有感觉器官的“我”就等于是物理的对象,因此以语言为基础,“我”的类似物就等于是意识的一个“对象”。一种知觉即使并不总是意识的对象,但也确实是意识的一个“对象”。

然而令人吃惊的是,诡辩论哲学家和心理学家们也把意识与感知混同起来了。在心理学中,心理物理学一开始就犯了这个错误。费希纳这位泛心论者以为:他通过测量感觉的特质中的明显的差异正好把心灵的整个世界与物质的整个世界关联起来了,于是乎便形成了著名的“韦伯-费希纳律”。甚至在当今,一些正在研究知觉的心理学家仍主张:他们所研究的就是意识。正如 W. 詹姆斯很久以前就指出过的那样:感知是我们加之于经验的抽象,不是那经验本身的固有部分。他的出发点仅仅是主张:“思想处于流动之中”。这也正好是我的出发点。

一些现代哲学家也陷入了这样的错误。我想这是基于关于感觉材料的古老传统。罗素在寻找意识的例证时直截了当地说:“我看到一张桌子”,其实这是一种高度人为的、容易使人误解的选择——就像说一个降半音 b 是一首交响曲的例子一样。不错,他意识到他看到了一张桌子(大概在闭上眼睛时他也许会更好地做到这一点),但是对桌子的注视并不是意识,而只是视觉,只是意识的

对象,即使是在他真实地意识到了什么的过程中,那也只是他的论据。罗素理应挑选动物行为学上更有效的例子,它适用于他的意识,而又真的发生过,诸如“我何时找到重写基本原则的时间”,或者“我怎样才能为罗素第二夫人提供生活费”。如果是这样,他最终便会得到另外的结论。这些就是关于意识处在流动中的例子。而“我看到一张桌子”则不是。笛卡尔绝不会说:“我看到桌子,因此我在”。

正像感知不是由意识产生的一样,各种具有知觉的恒常性的东西如大小、光亮、颜色、形状也是如此,它们是我们的神经系统在广泛地经历了光、距离、观察角度甚或我们自己的运动等环境变化的情况下,仍保存的东西,就像对象保持其同一个位置一样,这通常被称之为场景的恒常性——所有这一切的完成都无须来自意识的帮助。

对于另一大类我们称之为优先选择的的活动,如我们怎样坐、怎样行走、怎样移动来说,情况亦复如是。所有这一切都是没有意识的作用而完成的,除非我们决定去意识它们——此即意识的优先选择本质。甚至在讲话时,意识的作用比我们的单词的任何恒常伴随物更多地穿插进来。现在我并没有想到进入我的词汇库,从中挑选词条以把它们串在这些句法结构上。我想做的是让那能最好地被描述为具有确定意义的意愿的东西、让我称之为结构的东西、进而让言语的习惯模式在没有来自意识的进一步输入的情况下显现出来。同样在听某人讲话时,听者意识到的是什么呢?如果音素甚或高一层次的词素或单词表露出来了,你们也无法了解我正打算做什么。

462 第四部分

以为意识存在于我们所做的一切活动之中,不过是一种幻觉。就像在一完全漆黑的房间,要求打开手电筒,然后环顾里面,以看里面是否有光一样,手电筒只会告诉我们说:房间明亮的地方都充满着光——这显然是荒谬的。

关于意识的又一错误源于经验主义的创立,当时伽桑狄使用亚里士多德的“白板”一词,后来洛克以伽桑狄的《哲学体系》为基础创作了他的两卷本著作《人类理智论》,将“白板”英国化,把心灵比作“白纸,没有任何标记,没有任何观念。”^①如果那时有照相机,那么我推想洛克就会用照相机作为他的基本隐喻。在经验过程中,我们得到关于世界的连续的图像,在反省中发展它们,将它们转化为存在的概念、记忆以及我们的一切精神陈设品。

但是意识并不摹写经验这一点很容易得到说明:(a)试考察我们所拥有的记忆,接着再来注意:它们并没有以我们经验它们时的方式得到组织,例如想一想你在游泳中的最后时刻;大多数人不会像它实际发生时的那样思考纷繁复杂的、多样化的感觉经验,而往往从另一种观点看待他们自己的游泳——自然是他们从来没有经验过的某些事情;(b)或者考察记忆的丧失,如果意识真的摹写了经验,那么我们就应当有这种记忆,例如知道什么字母伴随着我们注视过上千次的、电话上的数字,或根据此时在基尔希堡的记忆描述到达你的办公楼的主要通路——你们将发现这是很困难的。有意识的记忆并不摹写经验,而只将其重构为一种必定存在过的东西。

经验主义也是相信学习离不开意识的被告。在旧的术语学中,心灵有自由的观念,有被感知到的东西的摹本,当事物一块儿

被感知时,它们的摹本或观念就结合于一起——此即观念的联想。既然这可能发生于意识之中,因而似乎很清楚:学习或观念的联想似乎就是意识的判决性的标准。对此,许多早期的心理学家都十分明确地作过陈述。然而后来心理学中所作的大量的实验研究则告诉我们:这是错误的。因为人们发现:学习也存在于物种和动物体标本之中。在这里,没有人会认为像内省意识这样的东西可能与学习必然关联起来。就人类而言,各种类型的学习,如适应环境、运动神经的学习、操作工具的学习或操作性条件反射等,都表明能在没有意识的协助或注意的情况下发生。这倒不是说,在人类的学习中,如在决定学什么、在制定学好的规则、有意识地用语词描述一任务的诸方面等过程中,意识并不起什么作用。但是这不是学习本身。总之,我的意思是说:意识对于学习的发生并不是必要的。

在这里,人们可能会列举出习惯的自动化这类众所周知的现象,因为当此自动化发生于我们身上时,在刚开始,那种任务似乎离不开意识,但在习惯最终形成时,意识则退居二线,进而那一任务无须任何努力便完成了。同样,习惯由于实践而变弱以及迅速增强,这在一切能学习的动物中都极为常见。一般来说,在这类普遍存在的现象中,意识并不随着行为的改进而消失,就像对那项任务的组成部分的被迫注意不会随之消失一样。注意是对感知的注视,并不必然就是有意识的。如果你随便哪只手拿着两枚硬币,然后一个接一个地扔到另一只手上,接着相互扔,以便另一只手也抓过每一个,这是要花大约 15-20 次的学习才能学会的活动。如果你今天晚上想试试,在你如是作时检验一下你的意识,你就会发

464 第四部分

现,意识与那几乎是机械似地发生的学习几乎没有什么关系。你可能会意识到你有点笨拙,或者当你捡起地板上的硬币时,你会觉得你正在做一件有点愚蠢的事情,直到成功之时,你的意识才会对你高度的灵巧而感到惊奇不已甚或得意洋洋。这就是那经历了变化的注意。自动化是注意的减少而不是意识的减少。

意识对于思维或推理也不是必不可少的,这大概是一个有点耸人听闻的说法。但是在1901年,作为研究生的K. 马尔比在维尔茨堡做过一项在他的时代最简单然而又是最深奥的实验。他以他的教授作为被试,其中每一位在内省实验上都是经验丰富的里手,他要求他们对两个外观相同的秤砣作出简单的判断,认定哪一个更重。所得的结果与当时的实验心理学背景格格不入,确实令人震惊。尽管这样的判断隐含于对问题以及判断的材料和技术的意识之中,但是对于实际的判断本身来说,不存在任何有意识的内容。于是便导致了被称之为关于无象思维的维尔茨堡学派的诞生,而这又引发了阿赫、瓦特、屈尔佩等人对定势、任务和决定的倾向——我重新将之命名为“构造”(structions)——之类的概念的实验。构造就是像被给予神经系统的指令这样的结构,当将材料提供出来以供操作时,那“构造”就导致了没有任何有意识思维或推理参与的、自动的回答。这类现象始终存在于我们的大多数活动之中,如从简单的诸如判断重量到解决问题,直至科学和哲学的活动。意识对问题作出研究,并将它准备为一种构造、一种过程,后者又可以使一种答复突然显现,就好像是从天上掉下来的一样。在第二次世界大战期间,英国的物理学家们常说:他们不再在实验室中从事他们的发现活动,他们有三个B,即浴室(bath)、床.bed)

和汽车(bus),他们的发现就是在这里作出的。此刻这一过程正以缩小了的比例在我身上重演,因为我的词语好像是我的神经系统在给予它以我预期的意义的构造之后专为我挑选的。

在对意识的误解的一览表中,最后一个项目就是关于它的定位的言论。除了长期致力于这个问题的思考、并把它放在“外在于”理智王国的地方的人以外,大多数人,可能包括在座的朋友,都极像笛卡尔、洛克和休谟,倾向于认为:他们的意识就是通常定位于他们的头脑之中的一种空间。无论什么地方,实际上当然没有这样的空间。意识的空间,也就是我后面所说的心灵空间,是一种功能空间,它没有位置,除非我们指派一个给它。由于我们的内省、内化等等之类的一切语词的作用,以为我们的意识内在于我们的头脑之中便是一自然而然的但又是一武断的、需要研究的观点。当然我的意思又绝不是说,意识与大脑是彼此分离的;根据自然科学的假定,它不能脱离大脑。但是在骑自行车时,我们都在用我们的大脑,然而没有哪一个人会认为:骑自行车的位置就在我们的大脑之内。对意识的这种现象定位是武断的。

综上所述,我们已说明:意识并不是心智的全部,也不应等同于感觉或知觉,它不是经验的本本,不是学习的必要条件,甚至也不是思维和推理的必要条件,它仅只有一种武断的、功能性的定位。作为我下面要讲的内容的引言,但愿你们能作这样的考虑:在过去某一时期,可能有这样的人,他们已做了我们现在所做的大部分事情,说了、感知了、解决了我们所说的、所理解的、所感知的、所解决的问题,但那时它们还没有意识。我认为这种可能性是一至关重要的可能性。

这样说几乎回到了激进行为主义的立场。既然我把意识当作一种不可还原的事实,因为我的内省像所谓的“外部的”知觉一样真实,即使它具有自身独特的、不能混同的性质,但这一来,意识究竟是什么呢?我的说明步骤是,首先以多少有点简洁的方式略述一种关于意识的理论,然后再以不同的方式予以解释。

假说之一:意识起源于人类作出隐喻 和类推的语言能力

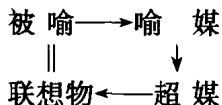
主观的、有意识的心灵是我们称之为真实世界的东西的一种类似物。它是由一种语汇或词汇域建构起来的,此域的术语都是关于物理世界的行为的隐喻或对应词。它的真实性与数学处在相同的层次。它容许我们省略掉行为的过程,而作出更妥当的结论。像数学一样,它不是一个东西或一贮藏所,而是一算符。它与意志、抉择密切地结合在一起。

试想一下我们用来描述意识过程的语言。通常描述心理事件的最突出的一组词是视觉方面的。我们“明白”(see)问题的答案,我的“明白”充其量可能是“明确的”或“清楚的”,或者可能是“朦胧的”、“模糊的”、“不清楚的”。这些词都是隐喻,它们所适用的心灵空间是通过实在空间的隐喻而产生的。在此空间,我们能“探讨”一个问题,也许是从某种“观点”出发,并与其困难“作斗争”。我们用来指称心理事件的每一词语都是行为世界中的某种东西的隐喻或对应词。我们用来描述真实空间中的物理行为的形容词通过类推变成了描述心灵空间中的心理行为的词。我们常说有意识的心

灵是“敏捷的”或“迟钝的”，或说某人是“机智的”、“有独立见解的”、“优柔寡断的”、“心胸开阔的”、“城府很深的”、“虚心的”、“心胸狭窄的”。因而像在真实的空间里一样，可以说某东西在我们心灵的“后面”，或在我们心灵“深处”或“外边”。

但是你们会提醒我，隐喻只是一种纯粹的比较，而不能像意识那样创造出新的实在。对隐喻的严格的分析表明：它也可以做到这一点。在每一隐喻中，至少包含着两个名称，一是我们试图以单词表述的东西，即被喻(metaphrand)，一是通过起着如此这般作用的构造而产生的名词，即喻媒(metaphier)。它们与里查德称之为要旨(tenor)^②和媒介物(vehicle)^③的东西相差无几，不过要旨和媒介物两个名称对于诗歌比对于哲学分析更合适一些。我模仿被乘数和乘数这样的算术术语，选用了被喻和喻媒两词，但又抛弃了算符的内涵。如果我说船像犁一样在大海上破浪前进，那么被喻就是船头经过水面的方式，而喻媒就是犁。

再举一联系更紧一些的例子。假定我们在回溯心理词汇的形成，我们一直想解决某一个问题或者在琢磨怎样完成某一项任务。表述我们的“明白”就是喻媒，它来自于物理世界中的物理行为，它适用于用别的方法难以表述的心理事件，即被喻。但喻媒常常引起我们称之为超媒(paraphiers)的联想，它们使被喻具体化为那被称之为联想物(paraphrand)的东西。由此可见，隐喻创造出了新的实在。单词“see”使人联想到物理世界中“看”的行为，进而使人想到空间，当此空间与这种由推论而来的、被称之为被喻的心理事件统一在一起时，该空间便变成了一种联想物。



这样,我们周围世界的空间属性就被转化成了一种解决问题的心理事实(我们记得这一过程是无须意识的)。正是这种被联想到的空间性质,基于我们用来描述这类心理事件的语言的作用,经过经常的重复,就变成了我们意识的功能空间或心灵空间。我认为,此心灵空间就是意识的第一重要特征。它也是你们此刻正优先予以“内省”或“观看”的空间。

但是谁在观看呢?又是谁在内省呢?这里我们得引入类比推论。它不同于隐喻,因为在类比推论中,与其说相似性存在于事物之间或行为之间,倒不如说存在于关系之间。由于具有感觉器官的身体(被称之为“我”)负责物理的观看,因此便自动地演化出了一类似物:“我”,以适应心灵空间中的心理形式的“观看”。类似物“我”是意识的第二个最重要的特征。注意不要把它与“自我”混淆起来,因为自我在后来的演化中成了意识的对象。类似物“我”是无内容的,我认为,它与康德的先验自我联系很密切。作为身体的“我”能在它的周围环境中走来走去,看这看那,因此作为类似物的“我”也学会了在心灵空间中“来回走动”,关注于一事物或另一事物。

我们将把意识的第三个特征称之为“叙事化”(narratization),即对实际行为的比拟性模拟,这似乎是意识超出于先前关于它的共时态讨论之外的一个明显的表现。意识常常让事情富有情节,给任何事件安排一个前奏和结尾。这个特点是我们的物理自我的

一种类似物,物理的自我在具有空间的连续性的物理世界中走来走去,而这种连续性可变成心灵空间中的时间上的连续性。这便导致了关于时间的有意识概念,其实时间是我们将事件、将我们的生命放入其中的空间化了的时间。以其他任何方式都不可能意识到时间,除非将它作为空间。

意识还有其他一些特征,我将简要予以述及,如:专注,它是知觉注意的“内在的”类似物;其对立面是抑制,由于它,我们便不再意识到那些令人讨厌的想法,它是在物理世界中抛弃令人讨厌的东西的类似物;选择是我们怎样才能做到一次只感知事物的一方面的类似物;一致性则是知觉认同的类似物。如此等等,不一而足。在这里根本的法则就是:意识中所发生的作用无一不先出现于行为中。

有人公正地谴责说,认知心理学家有时有这样的癖好,即重新发明轮子,进而使之成为正方形,然后说它是第一个近似值。我不想附和说,这在我所略述的那种演化过程中是真实可信的,但我的确愿把它称之为第一个近似值。意识不是一个简单的问题,不当应当把它说成好像是什么。我也从来没有说过,意识的叙事化有不同的方式,诸如语词的、知觉的、身体的或肌肉的,因为所有这些似乎与它们自身的属性迥然有别。但是我以为,我们现在有条件回到我开始说过的词汇问题上来,它在生物学、心理学、哲学中已造成了极大的困惑。就全体而言,这种“内在的”世界始于何时呢?

我们已说过:意识以语言为基础。这意思是说:一代又一代的人试图在动物的进化过程中探寻意识的起源是不正确的,而且也是徒劳无益的。如果意识是在语言的基础上、经过训练而习得的

470 第四部分

一组操作,那么就可以说,只有具有语言的物种才是有意识的;尽管近来对黑猩猩的符号语言和动物的交流系统作过一些有争议的探讨,但那意思不过是说,只有人类才有意识,意识仅仅始于语言进化出来之后的某一时刻。

语言又是在什么时候得到进化的呢?在别的地方我已阐述过关于下述问题的看法,即语言怎么可能从呼叫的元音变异中演化出来,什么是所谓的“啊嘿、啊哈(wahee, wahoo)模型”。此模型现在与其他几个模型正处于竞争之中。但是这种理论化指向了后更新世,其理由主要有:(1)这个时期正好与猎取大动物过程中对语言交流的进化上的迫切要求相符合;(2)它符合于与语言有关的大脑特定区域的令人惊奇的发展;(3)也是该理论独有的理由,即它与考古学上关于制造工具的剧变的记录相吻合,因为我们知道语言不仅仅是一种交流手段,而且是一种知觉器官,它能使我们注意某一特定任务并保持这种注意。这种追溯意味着:语言的起源不早于5万年前,也就是说:意识形成发展于那个时期与现在之间的某个时候。

对于这个问题值得庆幸的是:到公元前3000年时,人类已学会了明显的书写能力。因而很显然,我们的第一步便是应考察一下人类早期的作品,以便弄清楚是否有根据证明在心灵空间中存在着能叙述的类似物:“我”。概而言之,在大约公元前1000年以前的任何作品中都没有这样的证据,包括像众所周知的《伊利亚特》这样的古老的文学遗产。

这一来,下一步该怎么办呢?在从公元前9000年起、经过《伊利亚特》时代的一切文明中,种种证据再清楚不过地告诉我们:人

类已听得到我们常当作听幻觉的声音,即从大脑某个地方释放出来的所谓神的声音,甚至当作今日许多正常人和各种各样心理疾病患者常有的产生幻觉的声音。

假说之二:远古时代存在着一种可称之为 两院制心灵的无意识智能

这就是所谓的两院制心灵,它根据两院制立法机构的喻媒而提出。顺便说一句,该词在德文中没有贴切的译法,因而可能使你们中的一些人迷惑难解。其意思很简单,即那个时代的人类智能表现为两部分,一是作出决定的部分,一是随从部分。就我所描述的意识的那个意义而言,两部分都是没有意识的。这里我想提醒你们注意我演讲开始时对意识所作的相当长的评论。我的评论表明,人类能讲述、理解和解决问题,可以做我们所做的许多事情,但是又用不着是有意识的。两院制心灵的人也不例外,在他的日常生活中,他是习惯的奴隶,但是在碰到这样一类问题时,即它要求人作出新的决定或要求提供比习惯所能提供的解决方法更复杂的解决方法,那么作决定的重压就足以引起听幻觉,由于这样的个体没有怀疑和反抗的心灵空间,因此就不得不顺从这种听幻觉。

我认为,可以轻而易举地作出这样的推论:具有这样的智能的人类一定存在于特定类型的社会之中,即某种以森严的等级严密地组织起来的社会,它具有严格的、组合为心灵的期望,以便让那些幻觉维持社会结构。事实的确如此。两院制王国都是等级制的神权政治国家,让一个神,通常是一个偶像居于它的首位,幻觉似

乎来自于这位首脑,在整个美索不达米亚地区就是如此;要不然,就是以一个人作为国家的首脑,他本身就是神圣的存在,例如在埃及就是如此,当然比较罕见。关于证明早期文明通过被称之为神的、引起幻觉的声音而组建起来的根据,我在其他地方已作过罗列,此刻就无须深究了。

至此,我已讨论了我要讲的四个假说中的两个。对于第一点,我已花了分配给我的时间的大部分,它与本次专题讨论会关系最为密切,其核心思想是:意识根源于人们作出隐喻和类推的语言能力。第二个假说是:在远古时代,存在着一种独特的无意识智能,即所谓的两院制心灵。第三个假说可以简要地概括为:在历史上,一个由另一个所引起。

假说之三:在历史上, 一个由另一个所引起

实际上,这里有两种可能性。我所陈述的理论,其弱化的形式就是主张:意识固然以语言为基础,但它不是最近才有的,它产生于语言起源之后,甚或在文明之前,即是说大约在公元前 12000 年的时候,大约在能听到声音的两院制智能起源的时候。那时心灵的两个系统可能一起发生作用,直至两院制心灵失灵,进而被抛弃,让意识独自成为人类决策的中介。这是一种极端弱的观点,因为它几乎能解释一切,而又几乎无法被驳倒。

强的观点更令人感兴趣,就像我在介绍两院制心灵概念时所说明的那样。它为进入我们称之为意识的隐密事件所组成的明显

的隐私世界确定了一个近得惊人的日期,此日期在世界的不同地区有微妙的差异,但在两院制文明发源的中东,大致是公元前1000年。

我认为,来自美索不达米亚的证据能证明这样推定的日期,在那里,大约从公元前1200年开始,两院制心灵的瓦解已是相当明显的。这根源于社会混乱、失常,以及人口过剩,大概也由于书写成功地取代了管制的听觉形式。不过这种瓦解又导致了許多我们现在通常称之为宗教的实践,诸如祈祷、宗教崇拜,特别是名目繁多的占卜,它们是通过在想象中回应神的指导而作出决定的新型的方式,其目的就是恢复已失去的神的声音。

在希腊文学中,从B类线形文字简札开始,中经《伊利亚特》和《奥德赛》,再到梭伦以后两个世纪的抒情诗和哀歌体对句,都提供了关于两院制心灵瓦解和关于意识的词汇在隐喻的基础上产生发展的再清楚不过的描绘。例如 thumos、^④ phrenes、^⑤ kardia、^⑥ psuche^⑦等词都经历了从指称外部的、客观的对象到表示内在的心理功能的变化过程。

同一时期的另一类记载是《圣经》的希伯来《旧约全书》,它与上述材料不谋而合。犹太人的先知们就是那些有两院制剩余物或有部分两院制的人,他们听到了耶和华的声音,并能以令人信服的可靠性予以传播,因而在他们的社会倍受尊敬,被认为回到了已失去的两院制王国。这些先知的言论可与后来的“智慧书”如《传道书》相媲美,自然也可与那教给有意识的人们以一种改革过的犹太教的《新约全书》相媲美。最近我的同事,研究中国古典文本的专家M·卡尔博士也已证明:在大约同时代的《诗经》的各节中,意

识也经历了同样的演化过程。

但是,这就是意识或意识的概念吗?人们常常提出这种众所周知的质疑,它适用于霍布斯等人,也适用于现代的理论。在这里我们不是把意识概念与意识本身混同起来了吗?我的答复是:我们把它们融合在一起了,它们是同一的东西。正如 D. 丹尼特在关于这一理论的最近的论述中已指出过的那样,有许多证明陈述与其用法同一的例证。足球概念和足球就是同一的,或者说,钱、法、善、恶等概念也是如此,或者说,关于这次维特根斯坦讨论会的概念和这里举行的维特根斯坦讨论会也是同一的。

假说之四:关于两院制心灵的 神经学模型

我已阐述了我的三个假说:意识以语言为基础,曾经存在过一类不同的智能即所谓的两院制心灵,一个跟随另一个而产生。下面我将简要地陈述第四个假说,它完全可以与其他三个假说分离开来,与本次专题讨论会没有什么关系。它就是一种关于两院制心灵的神经学模型。它告诉我们:作为两院制心灵中神的一面的神经基质就相当于大脑右半球的韦尼克区域,这个特定的区域学到了人的生命过程中的全部告诫性信息,并对之进行加工,将它组合成被听到的、作为听幻觉的东西,以便把这种指令传递给对立半球或所谓的优势半球。今天如果我用精神分裂症作为这一过程的例证(因为精神分裂症是两院制心灵的局部的死灰复燃),那么我们便可看到证明这一观点的证据。而且远不止于此,我还必

须提到与此模型一致的关于大脑右半球功能的一些已变得习以为常的发现成果。

在结束我的发言时,我想指出的是:如果我们认为意识仅仅起源于 3000 年前的观点是正确的,那么这个理论便为哲学分析开辟了新的天地,同样也为探究许多古老的哲学难题提供了新的途径:历史上的发明和伦理学的诞生是与其词汇同步的,法规是作为神的声音的替换品而产生的,哲学本身的发展是为了填补神的隐退所留下的空白而产生的一种知识模式,同样还可解释真理概念的历史起源,以及科学概念、自然还有历史概念的起源。我倾向于认为:希腊哲学之所以卓绝伟大,其原因之一也许就是:有意识的世界完全是新颖的,因而是清楚明白的,据此,我们可以推论:我们的漫长的理智史怎样作为一种约束力量而作用于我们的能力和左右我们的成功。

毋庸讳言,对于心身问题这一一切难题中最大的难题,我所陈述的理论提出了一种非常真实的解决方法。它重申了我们一直试图予以理解的二元性,不过不是心灵与身体之间的二元性,而是意识与其他一切东西之间的二元性。这种二元对峙只是始于 3000 年前。在那之前,绝不存在什么意识与身体的问题。从那以后,人类的本性经历了根本性的变化。

注 释

- ① 洛克:《人类理解论》,Ⅱ.1.2。
- ② 即隐喻中的潜在方面,所指的概念、目的或人物。——译者
- ③ 即隐喻陈述的字面上的内容。——译者
- ④ 原指拇指,转意为感知。——译者

476 第四部分

⑤ 原指膈即身体的一个肌性和结缔组织性部分,转意为心灵。——译者

⑥ 原指心、胃、贲门,转意为心灵。——译者

⑦ 原指肺、呼吸,转意为灵魂、精神、心灵。——译者

参考文献

① J. 杰恩斯:《意识在两院制心灵瓦解过程中的起源》,波士顿,1976。

② J. 杰恩斯:“后更新世时期的语言进化”,载 S. 哈纳德等编:《语言和言语的起源与进化:纽约科学学会年刊》,Vol.280(1976),p.312。

第五部分

思想的语言与心理表征

为什么仍必须有思想的语言， 其意何在？*

特伦斯·霍根 约翰·廷森

如果认知器有这样的系统，即从句法上构造的、拥有对那结构敏感的程序的表征系统，那么它就有一种思想的语言。所谓有思想的语言就是这个意思。认知器之所以必定有这种语言，是因为在世界上的活动与生存需要如此强大的表征能力，以致只有当认知者有系统的构造表征即句法的方法才有这种可能。因此自然的认知系统一定有按句法构造的表征，进而一定有思想的语言。

但是非经典的思想的语言可能非常不同于经典认知科学中所设想的任何东西。尤其是经典的句法典型地涉及到把构成成分编码为表征的部分。但是句法并不生来就是部分/整体的关系（“人”（men）从句法上和从语义上说都是复数：人在这儿/人群在这儿。）

在第3章中，我们已论证过，古典认知科学对人类认知发表了根本错误的观点。古典的观点是一揽子交易，涉及到的只是有句法结构的表征和与其有关的规则。许多对古典图景不满意的哲学家一直感觉到，他们在古典主义中看到的问题使他们有理由拒绝

* 译自特伦斯·霍根和约翰·廷森合著的《联结主义与心灵哲学》（麻省理工学院出版社，1996年）第5章，有删节，标题略有改动。

有句法结构的表征。但是事实上正是古典主义的规则——与表征本身有关的规则(PRL 规则)概不例外——限制了古典认知科学。第3章的论证无论如何不能看作是对有句法结构的表征的否定。其实,可以合理地说,它们承认了句法结构。

因此本章和第3章论证的要点就是:一定存在着不能为规则所描述的、结构灵敏的加工过程。在后面的章节中,我们再设法解释:怎么会是这样。

在5.1(5即本章)中,我们通过考察非经典联结主义句法的两个例子来探讨:句法由什么构成,联结主义的思想的语言可能是什么样子。我们把关于心理表征必需句法的论证称之为“追踪论证”(tracking argument),因为它求助于大量潜在的自然认知器,让它们记下持续存在的个体及其可重复属性的踪迹。这个论证将在5.2中展开。

在5.3中,我们抨击了一种观点,它对大陆哲学家和许多精通哲学的联结主义学者颇富诱惑力,认为常见的非精神的复杂事物与表征完全无关。在5.4中,我们简要地比较了追踪论证与根据古典观点为句法结构所作的论证。

演绎推理这种现象也显露了心理表征中的句法结构。有效性和矛盾性是依赖于单元的句法属性的形式关系,它们就存在于这些单元中。因此完成演绎推理的造物的心理表征一定有句法结构。

然而,演绎逻辑的原则是没有例外的,因此至少是在演绎推理的王国,认知的古典概念似乎必定是占主导地位的。但是演绎逻辑的原则并不是认知加工的规则。有效性和矛盾性是表征中的关

系,而不是行动过程或加工过程。从这些关系没有例外的事实不能得出结论说,在演绎推理的王国存在着无例外的加工过程。

5.1 句法与心理表征

在本节中,我们先用抽象的术语讨论句法概念,集中关注的问题是:心理表征要表现句法结构,什么是必要的,什么是不必要的。根据这个讨论,我们再去考察一直在研究的、把句法结构具体化于表征中的某些方法。

5.1.1 什么是句法

哲学家们倾向于把句法理解为部分/整体的关系,即句法构成要素是它们作为其要素的表征的组成部分。^①这实际上就是怎样在数理逻辑中理解句法的问题,而且在经典认知科学中自然会以此方式去设想句法。稍微宽泛一点地说,福多和麦克劳格林(1990)以下述方式定义经典的构成因素:“对两类表达式 E_1 和 E_2 来说,只有在第二个被标记的同时第一个也被标记,第一个才是第二个的经典的构成要素。”^②

但是并不一定非要把句法设想为部分/整体关系或有关的构成要素标记。句法没有必要是经典的句法。非经典的构成性在自然语言中很常见。英语中不规则的复数名词和不规则动词就是非经典句法的例子。以为这些不规则的东西在句法上是简单的观点是错误的(即便在语义上也许是复杂的)。不规则动词正是以规则动词起作用的方式与助动词——从例证上说是句法的材料——相

482 第五部分

互作用。“他会喝水”从句法上而不仅仅是从语义上说的不恰当的,就像说“他会拉尿”一样。“女人”(women)像“人”(persons)一样同样是复数。复数(像一个语词所能关于的一样,它在语义上是复数)在美国英语中从句法上说是单数。非经典的构成性在有复杂变化的语言如希腊文、拉丁文和梵文(以及在许多非印欧语)中是准则。如果句法指的是经典句法(在福多和麦克劳格林的意义上),那么古典语言就没有句法。

而且,很容易用非经典构成要素构造语言。例如,试想一种语言,在其中有一种纯波形,它能用作专名,还有别的种类,能用作普通名称、关系名称、副词和联系词等。当一个普通名称断言的是一个体时,它的波形和个体名称的波形就同时产生出来。句子类似于和音,而不是语调。对基本的纯波形略作习惯的元音改变,就会表明:一个语词就是一个直接的对象、一个间接的对象等。声波像所有波一样都附在另一物上;因此在和音中,构成和音的个别的波都不可能标记。

如果存在着具有这样的沟通系统的造物,那么根据他们的沟通系统缺乏经典的构成成分而否认他们有语言,或者认为他们的语言没有句法,这几乎是不合理的。同样,如果一心理表征系统以同样的方式对谓项作出编码,那么否认它是一种思想的_{语言}就是不恰当的。我们的猜想是:思想的_{语言}与其说像一阶谓词演算或LISP,不如说像这种东西。

因此,句法所意谓的完全不是福多和麦克劳格林所说的“经典句法”。主张思想的_{语言}有经典句法(在那个意义上)的观点是一种经验假说。心理表征的自然系统可能有复杂的曲折的变化,就

像希腊文和拉丁文那样。复杂的心理表征也许像“喝”过了(drunk),而不像“确实喝过了”(did drink)。

直截了当地说,句法就是对语义关系的系统的和产生性的编码,可以这样予以理解:

系统的 当不同的表征状态断言诸个体具有同样的属性或关系时,相同的属性或关系被断言这一事实一定编码在表征的结构之内。当不同的表征状态断言同一个体具有不同的属性或关系时,那个事实一定编码在表征的结构之内。

产生性的 当关于一种新的属性或关系的表征得到了时,断言每一个体具有那属性的表征就一定自动地被决定了。当对于一新的个体来说的表征得到了时,断言那个体的属性的复合表征就一定自动地被决定了。^③

试考察一种用一阶谓词演算所表述的理论。它是系统的。该理论的两个句子断言不同个体具有相同属性,这一事实由于两个句子中处在对应位置的相同谓词的出现而编码在表征结构中。而且它还是产生性的。当表征了新的属性的新的谓词加于那个理论时,断言该理论所表征的、每一个体所具有的那属性的新句子就自动地被确定了。在逻辑中,以及在经典认知科学的句法中,系统性和产生性是通过串的共用部分编码指称的同一性和述谓的同一性而完成的;句法是作为部分/整体关系而实现的。但这绝不是必需的。上面所述的和音语言常常是系统的。系统性和产生性必不可少的东西是:在复合表征和系统构造复合表征所用的、它们的构成成分之间,一定存在着系统的形式关系。

484 第五部分

有句法结构的表征系统要成为一种思想的语言,还有一个东西是不可或缺的。即表征在认知系统之内必须具有适当的因果作用。而表征的因果作用部分取决于它们的、编码于句法结构中的构成要素的语义关系。因此一个表征有特定的句法结构这一事实一定能在那系统中起因果作用。正如我们后面要说的那样,思想的语言中的句法一定是有效的。

5.1.2 联结主义系统中的非经典句法

部分/整体的关系对于认知系统的表征中的因果上有效的构成结构的编码并不是唯一有用的形式关系。例如试考察伯格的联结主义语法分析程序中所用的 RAAM 表征,即我们在 4.3.3 分节中所述的表征。基于模型(来自于相继的词语表征输入、被分析句子的句法成分的表征的连续发现而形成的可进行语法分析的表征构造)所完成的信息加工任务的本质,基于训练有素的系统有能力严格地对不在训练用的语料库中的输入进行加工,在这模型中的表征,显然能以形成某种易受结构灵敏的加工即语法分析影响的表征的方式,编码句法构成成分的关系。^④因此表征表现的是有效句法的基本形式。然而它们并不通过部分/整体关系而编码句法构成成分。

波拉克的 RAAM 表征是一个一直受到研究的、并在联结主义表征中用于编码组成结构的重要的方法。另一种表征就是张量-乘积(tensor-product,简称 TP)表征,它一直为保罗·斯莫伦斯基及其同僚所研究。^⑤TP 表征在形式上与它们的句法成分有关,但标记 TP 表征与标记它的构成要素则不相干。

斯莫伦斯基利用了对矢量的两种数学计算:加法和张量乘法。任何两个 n 元矢量 v 与 w 的总和就是这样的 n 元矢量,它的元是 v 的第一元和 w 的第一元的总和, v 的第二元和 w 的第二元的总和等。对于任何 n 元矢量 v 和任何 m 元矢量 w 来说,张量乘积 $v \times m$ 是一种 $(n \times m)$ 元矢量,它又是通过 v 的每一元与 w 的每一元的两位乘积而得到的。

TP 表征可以这样予以描述:假设 F 是 n 元矢量的集合,其中每一个用来表征一种能起不同作用的“填充符”;假设 R 是 m 元矢量的集合,其中每一个用来表征一种能由不同的填充符所具有的作用。这一来,对于分别来自于 F 和 R 中的任何两个矢量 f 和 r 来说, f 和 r 的张量乘积——即 $(n \times m)$ 元矢量 $f \times r$ ——表征一种“填充符/作用汇集”——即关于起着特定作用的特定填充符的表征。因此如果 $v(\text{约翰})$ 是一表征了名词“约翰”的矢量, $v(\text{主词})$ 是一表征了主词的语法作用的矢量,那么张量乘积, $v(\text{约翰}) \times v(\text{主词})$,就是表征了起主词作用的“约翰”的一个矢量。

关于有结构的宾词的 TP 表征对于那宾词的每一成分来说就是这样的矢量,即由表征了填充符/作用汇集的矢量相加而形成的矢量。关于句子“约翰爱玛丽”的 TP 表征可能是把表征了作为主词的“约翰”、作为动词的“爱”和作为宾词的“玛丽”的矢量加在一起的结果:

$[v(\text{约翰}) \times v(\text{主词})] + [v(\text{爱}) \times v(\text{动词})] + [v(\text{玛丽}) \times v(\text{宾词})]$ 。

这种对句子“约翰爱玛丽”的处置作为关于句子结构的一般性探讨显然是不切实际的、愚蠢的。按标准的观点,句子具有把“约

486 第五部分

翰”与动词短语“爱玛丽”结合起来的树状结构,而该短语进而又有“爱”和“玛丽”这样的构成要素。因此,一更现实的 TP 表征就应是把表征了作为主词的“约翰”与作为谓词的“爱玛丽”的矢量加在一起而成的矢量。后一矢量又是由两个 TP 矢量构成的,它们分别是表征了作为(比如说)动词短语的第一部分的“爱”的矢量和作为动词短语的宾词的“玛丽”的矢量。

这说明了张量乘积的一个数学特征:张量乘积是递归的。这个特征又使它们适用于联结主义表征。由于矢量本身,张量乘积表征可以作为高阶 TP 表征的填充符或作用表征。

TP 表征的另一重要的数学特征就是可复原性。当填充符表征线性式地相互独立,且作用表征也是如此时,它们在数学上就完全能够从 TP 表征(即个别的 TP 作用/填充符矢量的矢量总和)恢复原状。当矢量不是线性独立时,一定程度的可复原性仍会出现。一般来说,填充符表征或作用表征中完全的线性独立性的离差越大,作为构成要素的填充符表征和作用表征的数学可复原性之精确性就越小。

在联结主义网络中,TP 表征自然是能够被完成的。一 n 元矢量在 n 结点的备用库中可为对应于那矢量的元之值的激活模式所表征。许多 TP 与 $(n \times m)$ 元合在一起可由“汇集单元”的 $(n \times m)$ 容量的备用库完成,其中每个单元都能成为激活水平的一个连续的值域。矢量相加可通过把一种被配列的 TP 表征迭加在另一个之“顶端”,通过单元的总合活动而实现。因此,当一特定的 TP 表征被标记了时,每一汇集单元的激活,典型地说就是来自于几个不同的、添加的填充符/作用汇集表征的总合激活的结果。反之,一

节点的激活水平对这些填充符/作用汇集表征中的每一表征都有作用。一般而言,不存在单一的、作为构成因素的单元的激活水平对之有作用的表征。

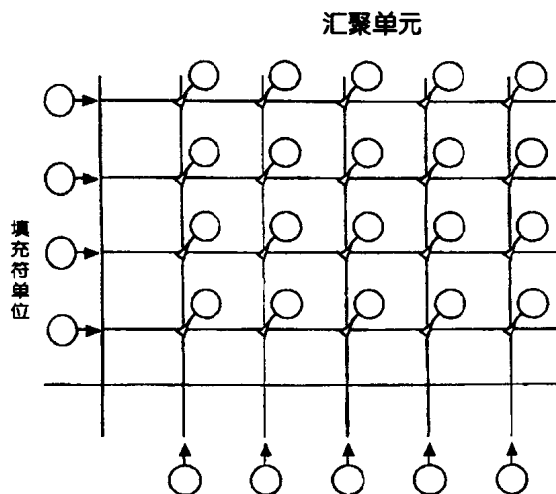


图 5.1 一种用倍增接合点来完成张量乘积聚合的网络(斯莫伦斯基,1990)

分开的 n 元备用库和 m 元备用库能够分别履行填充符表征和作用表征的职责。如图 5.1 所示,这些备用库可连接于 $(n \times m)$ 汇集单元的备用库。每一汇集单元都是通过倍增接合点——它可产生两个输入的乘积——而连接到单个的填充符单元和单个的作用单元上的。如果特定的填充符表征 f 和作用表征 r 在填充单元和作用单元。中同时被激活,那么激活就会从两个备用库到达汇

488 第五部分

集单元,进而倍增接合点就会在汇集单元中产生表征 $f \times r$ 。当现存的表征在汇集单元中仍很活跃时,叠加就可通过激活附加的填充符/作用集合而完成。

这样一种网络也能“释放”TP 表征,以便恢复它的作为构成要素的填充符或作用表征。当一 TP 表征活跃在汇集单元中时,在对应于特定的作用表征的作用单元中的活动模式在本质上就可能作为一种“询问”而起作用,如问哪一填充符(如果有的话)具有那种作用。来自于约束单元和作用单元的活动在填充符单元中会产生一种对应于适当的填充符表征的活动模式,由此而“释放”对那询问的答案。例如,如果句子“约翰爱玛丽”在那网络中被表征了,那么我们就可通过在对应于作用主词的矢量表征的作用单元中产生一种激活模式去问那网络:句子主词是什么。而且我们在对应于“约翰”的矢量表征的填充符单元中也能得到一种活动模式。

我们已谈论了句子的 TP 表征,这部分是因为,这是熟悉的言谈方式。在某些情况下,可以恰如其分地把张量乘积说成是对句子的表征——例如当该网络从事语言加工时就是如此。但“TP 表征”这一称呼的意思是:张量乘积得到了表征。典型地说,该网络表征的东西不是语言而是某种别的论域。在那些情况下,更为稳妥的是,不把 TP 表征看作是对(被标记的)汇集单元中的句子表征,而看作是一种句子本身——一种有句法结构的表征。因此在那网络中被标记或被实现的不是“约翰爱玛丽”这样的句子表征,而毋宁说是“约翰爱玛丽”这样的句子,(粗略地说)它是关于某种非语言事态的表征。TP“表征”在适当的联结主义网络中被标记时,应被视之为有句子结构的表征——即(非经典的)网络语言中

的句子。^⑥

TP 表征除了在回答“询问”时要能经受填充符或作用之“释放”的检验,还经得起结构灵敏的加工的检验吗?尤其是它们能经受下述结构灵敏的加工的检验吗?这种加工不是这样进行的,如先释放基本的构成要素,接着从每一基本要素中释放它的基本要素等,最后对这些明显的作为构成要素的表征进行直接操作。回答是肯定的,正如斯莫伦斯基从数学上所说明的那样(正在出版中)。非正式地说,关键的观点在于:对于任何借助一系列开方和对任意复杂的树状结构的重新组合操作去计算输入/输出功能的算法来说,存在着简单的两层级,即前馈和具有相同输入/输出功能的联结主义网络。更具体地说,对于任何借助一系列的开方和重新组合的操作、作用于树状结构以计算出某种输入/输出功能的算法 A 来说,存在着一种具有下述属性的矩阵 W,即对于任何两个树状结构 s 和 r 来说,以至于算法 A 在已知 s 为输入时返回为 r,如果 s 和 r 分别是关于 s 与 r 的 TP 表征,那么关于 s 乘 W 的单独的矩阵乘法运算就会产生结果 r。而且,这种矩阵乘法操作可能由“线性连接器”网络而完成,该网络的联结权值的矩阵又是 W 本身。(线性连接器 A,也许是最简单的一种联结主义网络,在一种步骤中起着把输入矢量与权值矩阵相乘以产生输出矢量的作用。^⑦)

斯莫伦斯基及其同僚借助模拟所谓的主动/被动网证明了这种技艺的正式能力(勒让德等,1991;斯莫伦斯基等,1992)。该网络

把表征作为英语句子之基础的树状结构的活动的分布式模式

490 第五部分

当作输入,通过检查那结构来确定那形式是主动的还是被动的句子,进而产生出这样的输出,即编码了输入句子的语义理解的谓词演算形式的树状结构分布式表征。……该网络平行地完成所有必要的符号操作,处理所有嵌入式子树(如复合的 NPs),就像轻而易举地处理简单的符号那样。(斯莫伦斯基等,1992,p.14)

于是,主动/被动网是下述联结主义系统的一个例子,这种系统能完成对 TP 表征的初步的、结构灵敏的加工,而又不在于加工过程中提取构成要素。该网络精确地运作,“就像确实可证明的那样”(斯莫伦斯基等,1992,p.20)。

RAAM 表征和 TP 表征都是系统性的:当两表征断言不同个体具有相同属性时,或当它们断言不同属性属于同一个体时,这些事实就编码在表征的结构中。它们还是产生性的:如果新的谓词或名称加于 RAAM 表征系统或 TP 表征系统上时,具有作为构成因素的新的名称或谓词的复合表征就自动被确定了。因此两类表征都能提供一种句法。此外,正如现存模型所证明的那样,两类表征至少能够提供一种基本的有效句法。

福多和麦克劳格林(1990)论证说,TP 表征不可能支持有效句法:“纯粹反事实的表征没有因果推论;只有实际被标记的表征才有。”(p.346)“复合活动矢量的构成要素通常不在‘那里’,因此如果标记-复合矢量的因果推论对它的构成因素所组成的结构灵敏,那么这就是奇迹。”(p.374)在这里,福多和麦克劳格林承认:TP 表征能成为一种对构成要素所组成的结构的编码。他们质疑的是:以这种编码为基础的、结构灵敏的加工是否可能。(他们的

论证在对细节作出修正之后推广到了 RAAM 表征上。)

但是不知什么原因,福多和麦克劳格林错误地表达了那种条件。问题不在于,构成要素是否能起因果作用,而在于:表征有特定构成因素这一事实是否能起因果作用;而且,如果那表征携带了它有那种构成因素的信息,那么那个事实能不能起因果作用。

在经典的系统中,构成因素就在“那里”。但是经典系统中的表征仍然只在整个系统的背景下才拥有那些构成因素。作为整体的系统结构决定了表征拥有一种对它们的构成结构灵敏的因果作用。仅仅是由于它们具有这种因果作用,才能有意义地说,某些物理项目是构成因素。在像主动/被动网和伯格的语法分析程序这样的联结主义系统中,表征 R 有构成因素,C 这样的信息是由表征 R 携带的——相对于整个系统来说,即使构成因素 C 在物理上是不存在的。

再明显不过的是,TP 表征和 RAAM 表征在一系统内或相关于一系统携带着构成性信息,而且这种信息对该系统是有用的。这显然是因为该系统从事的是对构成因素敏感的操作。成功的系统的整个倾向系统表明:表征携带着这些信息。

TP 表征携带这种信息绝非奇迹;这是可从数学上加以解释的。我们假设:RAAM 表征携带这种信息也不足为奇。我们认为:表征和加权在训练中由之所决定的活动目标战略允许系统选择这样的表征,即它们携带着与系统的最终加权有关的构成性信息。(我们是根据 9.3.2 节中所述的动力系统讨论这个问题的。)

但是下述观念似乎有可能成为神秘的观念,即表征即使不包含构成因素也能包含构成性信息。类比可能有助于支持直觉。试

492 第五部分

想一种有屈折词形变化的语言,在其中,格与时态等语法功能都由单词的形式所体现,而与词序无关,以至于在一个从句中,单词能以任何次序出现。但是假设它是一种语言,在那种语言中,语法作用不是由后缀而是由单词形式中的内在变化所体现的,就像在拉丁语和德语中那样。

在具有 TP 表征的系统中,填充符的矢量表征类似于这种语言中的动词和名词词根形式的无限多的样式。起着特定作用的填充符 TP 表征类似于这样的名词或动词的复杂变化形式。^⑧

再作进一步的设想。试想,这种有复杂词形变化的语言是由 5.1.1 节中所述的那类和音-语言生物所用的。即是说,在一句子或从句中有一作用的词语是同时而非依次被标记的。讲这种语言的生物(海豚或鲸?)一定有可能存在。这种语言的句子——即有作用的附加词语——必然会携带构成性信息,因为诸如此类的大量信息对句子的理解可以施加影响。

在和音语言中,词语迭加之所以携带有信息,是因为存在着能利用那种信息的更大的系统,即那种语言的说者。有 TP 表征的系统直接类似于这种想象的但有可能性的和音语言。TP 表征同样也携带着构成性信息,因为那种信息能为更大的系统所使用,正如主动/被动网这样的系统所证明的那样。^⑨

正如这些说明所表明的,有效句法除了在经典主义中之外,在第 4 章所述的关于认知科学的非经典图式内也可能是一种非常独特的现象。仍有待查明的是,与对构成成分的某种非经典编码有关的表征是否能接受结构极其敏感的加工——至少像经典主义中能得到的那种结构极其敏感的加工。我们认为,对于作为认知结

构的联结主义的生命力来说,这可能是一个关键问题。^⑩在第9章中,在讨论不能计算的动态认知图式时,我们再来处理这个问题。

5.2 追踪论证

为了在野外(即在真实世界中)生存,有机体一定能够掌握持续存在的个体的踪迹,以及它们的变化、可重复的属性和关系的踪迹。关于属性以及与个体的关系的可能的谓项,对有机体得以在复杂的、不确定的世界上生存下来一定是有用的,但其数量极其巨大——数量级太高,以致每一个都不能作为没有复合结构的简单表征预先结合到系统中去。我们把这称之为“追踪论证”。最令人感兴趣的一类追踪就是与复杂的物理技能之训练有关的追踪。在5.2.2中,我们将以篮球为例,考察这种技能。不过,我们先说更简单的事情。

5.2.1 关于某人的位置的信念

此刻,你有这样的信念,即相信你附近确实有数以百计的、中等大小的、能被看到的和不能被看到的对象。你还有关于这些对象的知觉的和非直觉的属性的信念。你有关于其中许多对象的功能或用途的信念,以及关于它们由什么组成的信念,关于它们的许多倾向属性的信念,关于某些对象的年龄、部分和内容等的信念。

你现在并不知道其中的大多数信念,但如果某物引发了其中的任何一个,那么它就会在你心中浮现出来。很显然,这样一些信念必然会在世界上起作用,并利用它所提供的东西。任何自然的

494 第五部分

认知器在其生命历程中所具有的信念的高百分比或者说大多数,均属这种类型。

每当你选择了一个新的位置,即使是很短暂的,你都会得到许多这样的信念。因此你已有的此类信念数不胜数。但大自然并不知道你要到地球的什么地方。你可能到过不同的地方;这一来你可能有关于你的位置的不同信念。大自然不得不为你提供进入地球上的任何地方的能力——让你有关于地球上任何位置的构造和其中的对象的信念。

大自然实际上也不知道地球可能会是什么样子。你有进入许多可能的、有地球特征的位置的能力,这些位置其实并不存在——实际存在之外还有许多这样的位置。你的确有进入许多可能位置的能力,这些位置尤其不具有地球的特征——如经历“星际旅行”。你可能已经具有的、关于你的直接环境的信念——这是你能够拥有的——远胜于你将会具有的那些信念。

我们不妨把这推广到第二个人身上,但我们所说的自然适用于每个人。在相当的程度上,它一定适用于任何能够在世界上起作用 and 生存下来的认知系统。任何成功的自然认知系统对于大量可能的位置来说一定能表征它的位置的构造。在那系统的生命历程中,实际会出现的尽管只是这里所说的相对少的潜在的表征,但无论哪一个,只要发生了,就一定有一种与内容相称的因果作用——否则它们对生存就没有任何助益。因此认知系统一定是以这种方式建立起来的,即每个潜在的信念都有其适当的因果作用。

也许有的人特别为诸状态提供了简单系统中的语义上适当的因果作用,方法是把所有潜在必然的状态和因果关系连接起来,但

是对于我们在自然中实际上发现的复杂的认知系统来说,这几乎是不可能的。一系统能拥有它所需的潜在表征的大量供给的唯一途径就是产生“必需的”表征。这意味着它一定是一种带有产生性的表征系统。

那个系统一定会产生关于其周围对象的位置和属性的信念。因此下述几点至少应是大致真实的:(1)对于该系统有办法表征的每一个体 i 和属性 P 来说,表征“ i 有 P ”的方式是自动地为那系统所决定的。 i 有 P 这一表征一定自动地有其与内容相称的因果作用。因此,(2)每当那系统发现了一新的个体时,潜在状态把原来的属性归之于新的个体一定自动地被决定了,而且当那系统发现了新的属性时,把那属性归之于原来的个体的状态也一定被决定了。再者,这些新的状态一定有它们的与内容相应的因果作用。只有当断言同一个体有不同属性的表征是整体上相关的,只有当断言不同个体有同一属性的表征也在整体上相关时,(1)和(2)适用于这样大的范围才是可能的。因此当表征要求以那些整体的关系为基础时,它们才能“被构建”。否则自然认知器追踪它们的环境所必需的潜在表征就不可能有如此多的供给。任何具有这种关系的表征系统都有一种句法,而且这样的任何系统都是一种思想的语言。

5.2.2 篮球

我们在开始陈述追踪论证时,我们不是根据我们已有的、在世界上的居住、而是根据在世界上的走动而予以陈述的。在世界上的一特别令人感兴趣的走动就是与复杂的身体技能之训练有关的

496 第五部分

走动。复杂的身体技能有其基本的、丰富多彩的、普遍性的认知构成,它与在大范围表现技能域限之特征的能力有关——这个范围如此之大,以致在表征确实有必要构建时,它就需要构建表征的系统方法。

以篮球为例。撇开这样的事实不说,如一些篮球运动员在运动中获得技能水平比我们大多数人随便获得的要高得多,篮球运动典型地表现为广泛的、常见的身体活动,其中包括投球、接球、防守、传球、运球、在防守密集的人群中穿梭等。如果不使用有句法结构的表征,它们类似于篮球所必需的表征,那么人们就不可能完成这些动作。

在篮球比赛过程中,运动员千百次地面对的是:必须非常迅速地作出决定:是以迅捷的动作带球过人,还是把球投出去,是去策应还是不去策应,是换防还是与被保护的运动员留在一起,在对方抢到篮板球时,是去追那球,还是留下来防守等。对这些决策条件的众多描述都是有用的,每一个都包含着无数可能的特定条件。

.....

在任何特定时刻,一个运动员有许多关于各个个体的表征,其中每一个表征都有时空属性和超出一定时段及可能性的关系。在这种表征的内核,可能有这样一种表征,即与其说它像语言,不如说它更似意象论者的、像模型一样的东西或类似物。

但是关于变化着的场景的复杂模型绝不是人们打篮球时所需的那种模型。各种命题性的或似语言的信息也必须考虑到。

首先,有构成那场比赛的许多基本属性。谁在比赛?在知觉场景,有许多人,他们并不是比赛的参与者。参与比赛的那些人

中,谁是队友,谁是对手?一个运动员是队友,另一个是对手,这一事实,可由他们的服饰的颜色加以区分,但它不是包含在关于条件的知觉的或似意象的表征中的文字上的信息。比赛还有更为普遍的属性:得分、时钟上剩余的时间、比赛剩余的时间、教练员的比赛计划或最终的指导,而且对方所实施的防御都可能影响作出什么样的决策。

第二,个别运动员有特殊的属性:高度、弹跳力、速度、命中技能、谁有绝招等。在高水平的比赛中,这些属性以微妙的方式受到关注,但即使在赛场上,也不可能无视它们。

第三,有更多变化无常的属性:谁保护谁,每个球员占什么位置,谁技术娴熟,谁不娴熟,谁投球好,谁需要顺利投篮得分以受到激励,谁碰到了犯规的麻烦等。

第二和第三类属性不仅直接影响人们的决策,而且还影响人们关于球员应做什么的直接预期。琼斯所处的位置,他在球队整个比赛方案中的作用,他的特殊技能和习惯都以不同的方式决定了人们对他在任何特定时刻的表现的期望。这些期望反过来又对作出什么决策发挥决定性作用。

所有这一切似语言的信息一定以适当的方式与关于变化着的场景的、可能更富有意象派色彩的表征密切相关。如果你不知道你所看到的、具有所有那些属性的球员就是琼斯,那么知道琼斯在你的队里,打中锋,犯了规,防范对方最好的球员,在 12 英尺或更近的范围内是一优秀的投球手,但并没有上好的技能,这是无济于事的;如果你不知道他在哪个队,那么你看到一个球员,知道他要到哪里,这是一点用也没有的;如此等等。因此要打篮球,人们就

498 第五部分

必须把许多非知觉的属性归之于他们所看到的那些人,归之于他们所知道的处在附近的别的人。

结果是,一个人在成为一个职业篮球运动员时就得到了一种专门的认知系统。该认知系统正是利用大量各种各样的信息,才作出了赛场上的种种决策。它还产生表征——判断、期望、信念等——这些表征与处在同样环境中的非职业的人的表征是不同的。它们以各种方式相互关联。它们包括似印象或似模型的表征系统;不过,似语言的巨大的表征系统由于涉及到把属性和关系归属于持续的个体,因而也是必要的。^①因此职业篮球运动员使用了有丰富语义结构的表征。

进入职业篮球运动员临场决策中的信息大多是似语言的信息,以可重复的大量的信息的形式出现。存在着可重复的属性与关系,它们被归属于具有不同类型——比赛、球队、个人、球等——的持续个体。正是不同属性归属于同一个体这一点对决定做什么至关重要。同样,也正是把同一属性归之于不同个体这一点起着重要作用。

指称和述项的这些同一性要想在系统中起作用,它们就必须以影响认知加工的那种方式存在于那系统中。构造这样一个篮球认知系统,它的大量的、潜在的述项中的每一个都预先与之连起来了,具有与输入、输出以及别的认知状态的相应的联系,那也是完全没有用的,即使是可能的。但是对于一个具有篮球认知者那样大小和复杂的系统而言,根本就没有可供选择的方案,去把指称和述项的同一性编码在表征本身的结构之中,因此必不可少的是,表征只能根据这种结构来加以构想。即是说,需要的是一种句法。

5.3 身体技能不是作出反应的倾向

对此,我们经常听到的一个回答就是:熟练的篮球运动并不涉及到表征。该回答认为,技艺娴熟的运动员所需要的是一种引起行为的系统,而不是表征。通常引证的一种根据是,在打球过程中,技艺娴熟者常常没有意识到正在考虑我们所提及的那些事情。这种观点可能由以得到阐发的一种方式,就是主张:唯一有关的表征是知觉性的,而知觉直接导致行为,在正常情况下,并没有属于认知性的东西的介入。但是有些人甚至还想否认知觉性的表征。

这是关于身体技能的常见观点的一种表述:身体技能是对刺激的奇特的、习得的反应倾向系统。有些人乐意把实践一种身体技能当作是学习一种对刺激的适当的反应系统。篮球技巧自然比对特定情境的一组反应要复杂得多。篮球场上的情境实际上完全不像其他情境。球员前此不可能熟悉他现在所处的那个特定类型的情境。但是球员实际上会重复同样的情境,而且在这样做时,他们学会了作为类别的相同情境的反应。这一来,人们大概会假设:一个人所获得的东西就是这样一个系统,它把知觉输入映射到反应上,结果从经验过的情境推断新的情境,因此反应上的变化与情境的变化相互关联起来了。但是这样设想严重地误解和低估了身体技能运用中所发生的東西。

首先,如果球员以他在训练中常用的方式对一种情境作出反应,那么他就必须认识这种情境。人们必须认识到,在以常用的方式进攻时对方会采取什么防守措施,而且还要认识到各种不同种

500 第五部分

类的场上形势,以对它们作出回应。很清楚,认知就是一种远远超过了单纯的知觉的认知状态,对情境和防守的认知也远远超过了对球员的位置和运动的单纯知觉。球员不可能无一例外地认识到这些东西,只是一些人比另一些人做得更好一些。因此,具有相同体育能力的球员对相同的情境,在身体或行为方面并不总是以相同的方式作出反应,因为他们的认知能力和反应是彼此有别的。职业学校的球员年复一年得到改进的一个方面就是他们认知情境的能力。

第二,主张身体技能是一种反应倾向的观点有一更根本的缺陷。压根就不存在球员对特定场境所作出的特定反应这样的东西。一球员在一特定的场境下可能作出许多不同的事情,这取决于他当时的投掷能力。球员的身体技能加上他对当时场上布局的理解还不足以决定他的行动。

短期的战略考虑在决定球员的行动中起着举足轻重的作用。这场比赛的比赛计划,当时正在生效的战略,负责发起进攻的防守队员所发起的进攻,投篮时限,得分,剩余时间——所有这些都能够影响球员的所作所为。而且它们可能同时地、各自独立地发生作用。这些都不是场上布局或知觉场境或物理刺激的组成部分(不管人们怎样描述球员合理地对之倾向于作出反应的输入)。因此球员在特定场上布局中作出的反应是多种多样的,这些取决于各种其他的考虑。球员关于场上布局的表征加上他对实际战略的理解事实上决定了他怎样表现他的身体技能。如果他没有关于场上布局——我们重复地说,这包括各种非知觉属性的归因(队友、核心等)——的表征,那么上述情况就是不可能的。

不同的进攻与防守,不同的比赛风格等都处在实践之中。从现在的观点来看,更有兴味的则是这样的事实,即实践对于由新目的所影响的反应来说并不是必要的。一技巧娴熟的篮球运动员可能有意输掉一场球,或勉强取胜(即这样去打,使得分控制在刚好取胜的范围内)。如果他从事了这样一种用心险恶的活动,那么他对某种刺激的反应(加上关于比赛和球员的信息)就会与别的反应有微妙的差别。他在实践中不一定要做这样的事情;而且还有这样的可能,即这样做是没有用的。他可能第一次做这样的事情,而不知道任何新的东西。

勉强取胜和输掉一场比赛在有组织的运动中都是不合理的、不道德的。因此值得注意的是,在交际性的运动中,勉强取胜是一常见的、有用的行为。试想网球和高尔夫球这样的个人运动。你对你能轻易打败的对手,即使未用你对付与你同样或更高水平的人的那种方式,你也不会失掉得分的机会。一个人在比赛时可以让水平适度发挥,以一定的方式保持可观的战绩而又不必同时想到它。人们对不同的球员会采取不同的方式,而且人们在如此去做时常常没有有意识的思考——即使是第一次面对一特别的对手。

假设你已决定通过尽可能好的表现使对手遭到尽可能惨的失败。如果你想取得好的得分,那么在对付同一个对手时你除了常用的打法之外还得有所改变。这意思是说你得以不同的方式对特定的刺激作出反应。但是要如此,你将会运用同样的知识和技巧。

球员还会采用新的风格,也许是为了弄清它怎样起作用,或为了在实战中模仿即将面对的对手。球员参与篮球比赛和训练可能

502 第五部分

带着许多不同的目的,在每种情况下,球员对某些物理刺激的反应也会是不同的,除非他们被指派了别的目的。

相同的技术和能力系统可用来产生这些各别反应中的任何一种反应。一个老练的篮球运动员能够完成这些反应;而别的人则做不到。从长期的经验中,他的所见所闻告诉他:人现在在哪里,他们会到哪里,他们可能做什么。假如他想到要做什么,他就会如此去做。因此篮球技能(包括认知技能)不可能独自决定场上的反应。严格地讲,也不能认为,篮球运动员有一种仅对刺激作出反应的系统。运动员所拥有的部分是一种产生表征的系统,这些表征之所以能用于决定反应,要么典型地要么非典型地是由别的因素决定的。

而且,这些不同的潜在反应的每一个对于当前的情境来说都是适宜的。它们分明是篮球运动反应,即使不是最理想的篮球运动反应。如果没有实在地表征大量的情境,如果不能表征有助于决定作出哪一种反应的一切东西,那么就不可能对一种特定的情境作出如此多样的适宜反应。

一般来说,身体技能用多种方式包括新的方式加以训练,这取决于当前的目的和意图。如果身体技能是一种轻易学到的、对刺激的反应系统,那么这就是不可能的。

5.4 与别的句法论证的比较

近来受到广泛讨论的关于句法结构的论证就是福多和皮利辛(1988)的系统性论证。他们还还为句法结构提供了一种有关的论

证,此即他们所说的产生性论证;不过他们并未强调这个论证,因为他们觉得它是以假定作为根据批判联结主义的。让我们把这些论证的每一个与追踪论证作一简要的比较。

5.4.1 系统性论证

系统性论证是这样展开的:认知系统在认知能力渗透在有结构的束中这一意义上是系统性的。“你找不到这样的有机体,它们能学会宁可要绿的角也不要红的圆,但不能学会宁可要红的角也不要绿的圆。你也找不到这样的有机体,它们能思考那个女孩爱约翰,而不能思考约翰爱那个女孩。”(福多与麦克劳格林,1990,p.332)一般来说,我们找不到这样的造物,它们能处在一种具有内容 aRb 和 cRd 的状态中,而不能处在一种具有内容 aRd 和 cRb 的状态中。这就是所谓的表征的系统性。就认知系统从它的每一潜在的认知状态的潜在进化对应地关联于那个状态的内容而言,认知系统也具有加工的系统性;与内容相应的进化以同样的方式关联于那个内容的巨大范围之上的认知结构。因此,“一般来说,认知能力在法则学上必然是系统性的,在人类和似人类的有机体中都是如此。”(福多和麦克劳格林,1990,P.332)福多及其同僚论证说:对这种已观察到的(以及从演绎上推论出的)系统性的最好解释就是:所有具有包含着命题内容的心理状态的认知器都有句法结构表征。

我们赞成说:句法能解释个体认知器中的系统性。如果一认知器的表征有句法结构,那么它就会表现出表征的系统性。而且自然的认知器离不开有效句法,因此它们的表征就会表现出加工

504 第五部分

的系统性。

但是正如福多和麦克劳格林所阐明的那样,真正需要得到解释的是:为什么系统性是普遍有效的——它为什么是这样的规律,即所有包含着命题内容的心理状态的认知器都有句法结构表征。经典主义用它的认知构造的概念去解释系统性的普遍有效的特征。在经典主义看来,认知就是对有结构的符号的受规则控制的加工。因此根据经典主义的看法,所有(非退化的)认知器一定有句法,从而表现出系统性。

事实上,福多和皮利辛都把系统性论证当作对经典主义的论证:“无论如何,在心理表征中如果你需要结构,以说明心灵的产生性和系统性,为什么不假定结构敏感的心理过程以说明心理过程的一致性呢?简言之,为什么不成为一个经典主义者呢?”(福多和皮利辛,1988,p.66)但是结构敏感的过程就是遵循规则的(即经典的)过程,这并无概念上的必然性;因此假设结构敏感的过程并不必然就是经典主义者。如果结构敏感的过程必然是遵循规则的,那么它就是这样一种高水平的经验事实,即必须承认它独立于对句法结构和对结构敏感的过程的需要。

然而,如果人们拒绝经典的观点以及与之相关的下述观点,即有句法结构的表征正好是关于认知的概念的组成部分,那么系统性论证作为一种对句法的论证似乎就是极不令人满意的。句法结构是从系统性中推论出来的,因此也许在弱的意义上对普遍句法结构的假设就解释了系统性。但是如果没有经典的观点,这就使认知系统为什么首先有句法结构这一问题成为无法解释的了。我们为什么没有找到缺乏句法结构、不具有系统性的(即联结主义者

的)认知系统呢?

因此,福多和皮利辛对联结主义提出了一种挑战。联结主义模型不需要句法结构。从直观上说,句法结构在网络中是罕见的。这一来,联结主义怎样解释这样的事实,即所有自然生成的认知系统都表现出了系统性呢?

追踪论证为我们提供了必需的解釋:要做出它们必须做的事情以便在世界上生存,除了最基本的东西之外,认知器还需要构成性的句法。复杂的认知器为了在世界上生存,除了通过获得有句法结构的表征之外,再没有办法获得它们所必需的表征能力。因此,既然复杂认知器中的表征是有句法结构的,那么它们就会表现出系统性。^⑩这一来,追踪论证用不着把系统性当作前提;相反,对系统性的解释正好是该论证的副产品。

5.4.2 产生性论证

从起源上看,产生性论证是关于自然语言的生成语法的论证(例如乔姆斯基,1965,1972)。每个说话的人都知道理解特定句子、造出更长的句子的多种方法。因此自然语言不存在最长的句子;即是说每种自然语言都包含有无数的句子。但说话的人是有限的。一个有限说话者获得这种无限能力的唯一的方式就是去获得一种用有限的词汇、借助一系列有限的加工来构造句子的能力——即有一种生成句子的语法。

福多与皮利辛(1988, pp. 33 - 37)主张:一种类似的论证表明了思维对句法结构的需要。人类对个别事物的大量思维在语义和句法上都是复杂的。例如一个人可以想到一个碟子,它盖在琼斯

506 第五部分

住房的前厅顶灯的开关上,这房子位于孟菲斯外的小城西北的湖边。人们不可能经常以这种有点复杂的方式去思维这样的事情。当然人们能够把属性归之于这样的对象。人们可以记住它的颜色,它是不适当地附加上去的,或者说它完全没有什么异常之处。

于是,存在着非常多的、潜在的、系统相关的认知状态。这些思想的每一个与其他可能的、不计其数的思想有演绎和归纳的关系。除了解释说这些思想是用可重复的要素构成的,人们还会怎样去解释拥有这么多思想的能力呢?要求认知的潜能是无限的,即使是从观念上,也是没有必要的;大量有限的能力足以导致相同的结论。

追踪论证有别于产生性论证——而且有别于系统性论证,像它们经常被陈述的那样。这些论证诉诸于递归过程和逻辑上复杂的表征。而追踪论证求助的是这样一类状态,它们最自然地作为(一元的或 n 元的)原子句子在谓词演算中得到表征,而用不着量词或联系词。因此我们在追踪论证中所谈论的潜在表征的巨大性并不依赖于思维的递归特征。于是我们相信,该论证不仅适用于人类认知,而且也适用于非人的动物,因为对于它们来说,是很难找到证明表征中有超过最低限度的递归结构的根据的。

5.5 演绎推理

认知的一个显著的特征就是,我们能作出演绎推理。既然演绎关系基于句法结构,因此这就成了对心理表征中的句法的又一个论证(参看 5.5.1 子节)。不过,决定演绎关系的原则是没有例

外的,因此硬规则至少在人类认知的这一方面似乎是必要的。但是演绎关系并不是规则,而且演绎推理的现象没有为认知加工的硬规则提供根据,就像它没有为认知的其他任何方面提供根据一样(参看 5.5.2 子节)。

5.5.1 演绎推理需要句法

“如果彼得进了百货商店,那么他就买了啤酒,如果他买了啤酒,那么它就在冷藏室中;现在那里没有啤酒,因此他没有去商店。”——一种常见的推理。

我们人类能生成大量具有不同命题内容和不定语义复杂程度的心理表征,从而享有用这些表征进行演绎推理的系统能力。这就证明(我们认为是结论性的):我们有这样的表征,它们有这些推理所必需的逻辑形式。除非有关的形式结构内在于表征本身之中,不然就很难弄明白:心理加工怎样系统地顾及心理表征中间的演绎关系。如果演绎关系在心理加工中举足轻重,那么认知系统与那种演绎关系的基础就一定有一种适当的因果关系。我们的心理表征一定有这样的句法结构,它对保证我们能够视之为有效的论证之有效性是必不可少的。

许多哲学家都强调:未受过逻辑训练的人是不善于演绎推理的。最明显的是他们常犯某些错误,尤其是在处理不熟悉的课题时更是如此。但是正是这种能力,而不是杰出的能力,表明表征具有逻辑形式。^⑬

而且,经过明显或熟悉的类比,人们转而能作出他们过去所否定的正确推理。还有这样的可能,即在不告知他们的推理犯了什

508 第五部分

么错误的情况下,仅指出那些类比论证明显缺乏根据,就能使他们认识到他们推理的错误。即是说,人们能够认识到他们清楚的例子与他们起初不清楚的例子之间的逻辑形式的相似性。例如,试考察下面的论证:

(1)a. 只有美食家才喜欢鸡肉薄荷冰淇淋。

b. 约翰不喜欢鸡肉薄荷冰淇淋。

c. 因此约翰不是美食家。

(2)a. 只有美国人到过月球。

b. 约翰没有到过月球。

c. 因此约翰不是美国人。

在测验中,当推论(1)是那种形式的唯一的论证时,初学逻辑的学生中的大多数都会判断它是正确的。不过当(1)和(2)出现在同一测验中时,极少数人会把推论(1)当作错误的。如果在一种测验中推论(2)正好在推论(1)之前,那么认为论证(1)为错的比例则会达到0。许多学生似乎认识到了推论(2)的逻辑形式并且认识到了推论(1)与(2)的逻辑形式的同一,这表明他们能够认识推论(1)的逻辑形式,即使他们在某些语境下不这样做。如果逻辑形式能被注意到,那么它就一定是可发现认知系统的地方,即使在它没有被注意到时也是如此。¹⁹

至此,我们讨论的只是与语言运用有关的事例。但是关于信念的逻辑形式的根据在非语言行为中也是不难找到的。尤其是,人类或非人动物常鲜明地表现出有关于方位情境的析取或条件信念。古代斯多亚派的克里斯波斯论证说:狗有析取表征,所依据的是这样的观察:一只狗会沿着一条路嗅,接着沿着另一条路嗅,然

后不用嗅就沿着第三条路跑。因此那个狗一定作了这样的推论：“那造物要么跑到这条路上去了，要么跑到那条路上去了，要么跑到别的路上去了；既然它没有跑到这条路或那条路上去，因此它跑到别的路上去了。”（塞克斯都·恩披里可记述，见《皮浪主义述略》，1.69）这种行为在人类和许多动物身上是司空见惯的。

5.5.2 演绎推理无须硬规则

但是有效性和矛盾性是经典认知科学的核心地盘。演绎有效性的原则是硬的；它们不容许例外。如果有例外，那么它就不具有有效性。因此至少在演绎推理的王国，人们肯定会论证说：人类认知一定符合 PRL 规则。^⑤

但是这是不能接受的。有效性、矛盾性等是表征中的形式（句法）关系。它们之所以没有例外，是因为它们制约着句法基础在其中相同的一切事例。不过，这些关系不是认知加工的规则，而且它们没有隐含认知过程的任何东西。它们的确不是规则，即使它们中的一些在逻辑测验和组织证据中用作推理规则的基础。至于人们一定或可能做什么，它们是完全缄默的，它们也不会描述什么东西做了什么。它们是关于哪一些真值组合可能和不可能的事实。

即使对演绎关系的意识在构成演绎推理的认知加工中是一常见的因素，但它不是有确定结果的因素。如果你一并想起：你的信念 P，以及你的下述信念：当且仅当 Q 从而 P，那么不能由此说你最终将相信 Q。你可能会推论说：“啊，我的天啦，Q 不可能是真的，因此我关于 P 的信念一定是错误的。”你要么是重新审查前件，要么是不不知道该相信什么。正如他们所述，一个人的肯定式就

510 第五部分

是另一个人的否定式。意识到演绎关系常会在你的认知状态中产生某些进一步的变化,但一般不会去分辨这些变化是什么。

与信念固恋(fixation)有关的那类加工可以合理地看作是对多种软强制条件的满足,即使演绎关系卷入了也是如此。认知系统对演绎关系的意识最好视之为寻求满足的各种(软!)强制因素中的一种。信念中的一致性之保持常常(即使不是固定不变的)是作为优先的强制因素表现出来的。但是这种因素能够得到满足的方式通常不止一种,进而信念固恋过程的潜在结果也不止一种,正是在这一过程之下,该强制因素才得到满足。既然一个人的肯定式是另一个人的否定式,那么就可自然而合理地主张:即便是在演绎推理的王国,认知加工也完全是软的,而不能由 PRL 规则加以描述。

而且,一致性是一种软强制因素。有这样一些哲学家,他们可能真的会说下面的话来:

“我相信,我知道我前方有一张桌子。我也相信,有一关于我的真理(第一沉思式的),它意味着我不能知道我面前有桌子。我意识到:这使我的信念前后矛盾。这就是我为什么觉得认识论问题令人恼火的理由。”

“我相信,我有自由意志。我也相信,关于物理世界的流行的科学观点一般来说是正确的,即使不是详尽无遗的。我相信:在相关的意义上,我不可能有自由意志正是那一观点的一个结论。我还相信,这意味着我有一种关于信念的矛盾的三位一体,因此我的信念不可能是真的。”

事实上,所有哲学问题都可以归属于这类矛盾的阵营。也许大多数关心认识论问题或自由意志问题的哲学家已经为自己找到了解决这种不一致性的方法。但有一些则还没有;他们发现自己处在这样一种状态中,在那里,一致性的强制被违背了。而且还可以适当地补充说,他们处于这种状态不是没有合理性的。

在日常生活中自然有类似的情况。一个人可以相信,有证据表明只有约翰才会干出那些伤天害理的事情,而同时又坚定地相信,约翰是那类从不干这种事情的人。

无例外原则决定了这些信念组合是前后不一致的。但不能因此就说:存在着这样的无例外的原则,它们在认知系统面临这样一种不一致时决定了认知系统会做(或应做)什么。一般来说,决定演绎关系的原则是无例外的这一事实本身,绝不是假定自然认知系统的认知过程可由 PRL 规则描述的理由。

注 释

① “构成性关系自然是部分—整体关系。”(B. 麦克劳格林:“泰伊论联结主义”,载《南方哲学杂志》,1988,26,p. 190)

② 根据这个定义,属于空间或时间部分的构成要素显然可看作是经典的。但这也意味着还包含有其他的可能性,如在记忆中的项目就是“指示器”的对象或“推导”运算的值。其核心思想是:经典的构成要素是这样的项目,即当它们是系统的表征的构成要素时,它们就以物理的形式出现在该系统中。(即使在福多和麦克劳格林定义的意义下,“经典句法”如我们所理解的那样不是对经典认知科学的教条式的承诺,但它在经典主义的模型中也是无处不在的。我们沿用福多和麦克劳格林的术语,因为它在这些问题的讨论中是比较流行的。)

③ 这不是出现在关于递归句法的“产生性论证”中的那种产生性,对此,在 5.4.2 节中,我们简要地给予了讨论。

512 第五部分

④ 这也适用于波拉克原文(1990)所用的 RAAM 表征,它们被用在一种循环的自动-关联任务中;同时这也适用于查默斯 1990 年所用的 RAAM 表征,在那里,任务是直接把主动句的 RAAM 表征转换成相应的被动句的表征。在两种情况下,泛化于项目中的网络都没有用于训练过程中;波拉克的网络在任意复杂的树结构中运行良好。(查默斯并未用 2 级深度以上的树,在训练他的网络或对之作出检测时都是如此。)

⑤ 斯莫伦斯基,1988,1990,1991 以及正在印行中的;斯莫伦斯基等,1992。张量乘积表征“提供了一种形式化,它对构成性结构的非形式概念的联结主义计算是自然而然的,同时它还可能是一种在联结主义认知科学中发挥作用的候选者,这种作用类似于符号认知科学中的构成性结构树所起的作用”(斯莫伦斯基 1988,pp. 156-157)。

⑥ 即使是当 TP 表征表达了自然语言中的单词、句子和别的项目时,该表征本身也不会有非经典的句法结构,如果①它们能被重新解释为句子或有关句子的构成物本身,②这样予以解释,即它们能从属于结构敏感的加工。

⑦ 加上某种最小的非线性(如一个临界值单元,当它的网状输入大于或等于某种特定的值时,它的活动就为 1,否则即为 0)就可容许前馈控制网络,这些网络能完成与一系列树处理一致的矩阵运算,不仅包括开方运算还包括乘方运算,而且还包括“更复杂的、把条件支的等量结合起来的运算”(斯莫伦斯基等 1992,p. 18)。

⑧ 在人类语言中,同一单词的不同形式与有不同作用的同一填充符的诸 TP 表征相比,在直观上更加相似。但这似乎不是语言的交流作用所必不可少的东西。可能有一些造物,对它们来说,同一单词的语法上不同的形式可能有很大的区别——以及在整体上——就像有不同的作用的同一填充符的 TP 表征一样。

⑨ 福多和麦克劳格林(1990)主张:具有非经典的构成要素的表征包括 TP 表征并不能支持有效句法。关于此观点的进一步的批评性讨论,可参阅霍根和廷森 1992。

⑩ “有效地表征复杂等级结构和把结构敏感的运算应用于这些表征的能力似乎是必不可少的。大多数联结主义研究者接受了这一点,即使他们希望这种能力能够以标准的符号加工传统所不期望的方式体现出来。”(欣顿

1990, p. 2)

⑫ 科林·麦金(1989)论证说:心理模型而非似语言的心理表征形成了“内容的基础”。他为“那一理论的雄心勃勃的变种”作了论证,“……(它)主张:模型仅凭自身就足以实现内容,而用不着求助于任何别的表征系统。因此模型提供了一种激进的内容理论,一种自足的基本原则……它试图为内容如何可能提出一种完全非句子的理论。”(p. 184)

⑬ 我们认为,最好不要把追踪论证当作是对最佳解释的论证,而应当作为对唯一可能的解释的论证。即使在文献中没有明确地予以论及,但对唯一可能的解释的论证是一种常见的论证形式。例如人们可以论证说:“史密斯昨天在伦敦;现在在纽约。因此她一定乘了飞机。”即是说,她已乘飞机是对她现在在纽约的唯一可能的解释。结论是一经验的主张,而且在一定意义下,该论证是一种经验的论证:它以未予陈述的经验前提为基础,这个经验前提是:史密斯在这么短的时间内从伦敦到了纽约,只有一种途径。当然,这个前提在逻辑上可能是错误的。……而另一方面,对最佳解释的论证则是归纳的。最佳的解释——不管根据什么规则——都不大可能是真实的解释。但唯一可能的解释一定是真实的解释。如此看来,追踪论证的基础是下述经验主张,即有句法结构的表征是对自然认知器所拥有的大量表征的全部构成内容的唯一可能的解释。

⑭ 乔治·雷伊(1991)指出:人们一贯犯同样形式的错误这一事实表明:他们的表征有形式的属性,这些错误正是基于这些形式属性。如果你没有这样的表征,它们有后件,那么你就不可能一以贯之地作出所谓的“肯定后件”这样的错误推论。

⑮ 对此的一个可能的答复是:演绎推理完全依赖于自然公共语言中的形式,而不依赖于心理(非自然语言)的表征中的形式关系。我们认为:这不适用于狗和海豚,它们能够进行演绎推理(参阅下一段)。具有相同逻辑形式的论证并不总是具有(或并不总是被表现为)相同的语言形式。这在某种程度上也适用于我们的简单的样本。人们必须认识到:同样的逻辑形式可以以不同的语言形式表现出来,尚不清楚的是:这怎么可能只是(自然)语言的问题。

⑯ 这不是一个关于理由的推论,除了下一节给予的那一个。演绎系统

514 第五部分

中的推理规则并不是能成为 PRL 规则的那种规则的恰当类别,因为它们
是允诺的规则,而非强制性的规则。

参考文献

J. 福多和 Z. 皮利辛(1988):“联结主义与认知结构:批判的分析”,《认知》,28:3-71。

J. 福多和 B. 麦克劳格林(1990):“联结主义与系统性问题:为什么斯莫伦斯基的方案没有用”,《认知》,35:183-204。重印于 T. 霍根等编的《联结主义与心灵哲学》,克卢沃出版公司,页码是重印本上的页码。

J. 波拉克(1990):“循环分布式表征”,《人工智能》,46:77-105。

D. 查默斯(1990):“关于分布式表征的句法转换”,《联结科学》,2:53-62。

P. 斯莫伦斯基(1988):“联结主义心理状态的构成结构:对福多和皮利辛的答复”,《南方哲学杂志》,26。

P. 斯莫伦斯基(1990):“张量乘积可变的束与联结主义系统中的符号结构的表征”,《人工智能》,46:169-216。

P. 斯莫伦斯基(1991):“联结主义,构成性与思维语言”,《心灵的意义:福多及其批评者》,B. 洛伊尔编,布莱克韦尔。

P. 斯莫伦斯基与 G. 勒让德等(1992):“关于高级认知的整合联结主义/符号理论的原则”,技术报告 CU-SC-600-92,芝加哥大学计算机科学系。

T. 霍根与 J. 廷森(1992):“联结主义系统中存在着有结构的表征吗?”载《联结主义:理论与实践》,S. 戴维斯编,牛津大学出版社。

G. 欣顿(1990):“关于联结主义的符号加工的专门争论的序言”,《人工智能》,46:1-4。

C. 麦金(1989):《心理内容》,布莱克韦尔。

G. 勒让德等(1991):“分布式循环结构加工”,载《神经信息加工系统的进展》,3,D. 图雷茨基等编,摩根·考夫曼。

N. 乔姆斯基(1965):《句法理论的概貌》,麻省理工学院出版社。

——(1972):《语言和心灵》(增补版),哈考特·布雷斯·约瓦诺维奇。

心理语义学或者：真值条件 来自哪里？*

J.A. 福多

引 言

让我们来假设。我们姑且假定关于心灵的表征理论(简称RTM,下同——译者)是正确的。你是这样想的,即设想存在一种我们以之进行思维的内部表征系统,一种思想的语言。说我们以这种语言进行思维,至少有如下的意思:

(a)思维密切地涉及到的标记心理状态(特别是像相信P那样的标记命题态度),应看作是有机体和标记心理表征之间的关系。(如果你不喜欢这样假定,那么你可以假定,这种有机体与心理符号的关系对于命题态度的标记来说,在法则学上是必要且充分的。即是说,对于所讨论的论题而言,是否把RTM当作一种同一论,这都无关紧要。)

* 译自 W. 利康编:《心灵与认知文选》,巴兹尔·布莱克韦尔,1990。据编者交待,这是作者不情愿发表的文章,因为他后来认为:他所辩护的观点是没有前途的,因而在1987年发表的《心理语义学》一书中对之作出了批判。编者之所以要收入该文,并最终取得了作者同意,是因为它对流行的达尔文式的目的论语义学作了精彩的概述,引发了大量有价值的争论,而且在编者和其他一些人看来,该文表达的观点比作者现在所持的观点更有希望。

516 第五部分

(b)心理过程(例如推理,以及其他类型的信念的合理固恋)应看作是对属于这种内部语言的惯用语的操作的结果。具体地说,应把它们视为计算操作的结果,这种操作是由于该表征的形式的——即非语义的——属性而适用于内部表征的操作。

这可能是一种幻想,但它正是近二十年来的认知科学所形成的关于心灵的图景。我认为,它是最接近于我们所得到的严肃的认知心理学的东西。某些主要概念无疑是不明了的,但否认 RTM 已经提供了某些相当成功的和富有启发性的经验研究由以完成的理论框架,则是错误的。假如是这样,人们就可以论证说,最好的战略就是接受这种框架(不管有多少缺点),并希望通过我们都如此强烈企盼的联邦政府资助的“进一步研究”来解决这些不明了性。

然而,这一框架存在着一类至关重要的、在某些意义上又自我隔绝的问题。它们就是关于心理表征的语义属性问题。在这篇论文中我打算讨论其中的一些问题。

为什么这些问题是至关重要的呢?至少从哲学家的观点来看,RTM 最有趣的属性也许就是它试图说明心理状态的意向性(即说明这样的事实:信念常常是关于事物的信念,由于它们与事实的符合而常常是正确的[或错误的])。RTM 的战略就是认为:心理状态的意向属性根源于心理表征的语义属性,根据原则(a),这种语义属性又涉及到心理表征的标记。许多哲学家发现这种还原的前景很是诱人,因为一方面,他们信服反行为主义的论证:可操作的经验心理学将不得不尽情地假定意向状态;另一方面,他们

再次受到意向性是心理状态(或别的东西——谈谈你的本体论的悬垂物吧!)的基本属性这一观念的冲击。然而我的观点是,讨厌把意向性作为心理的不可还原属性的任何人,有可能同样因把语义性作为符号的不可还原特性而苦恼。只有存在着这样一种相当严肃的理论的前景,它能说明什么授予心理符号以语义属性,那么参照内部表征来解释意向性看起来才像是一种严肃的理智进展。奎因(正确地)告诫我们:要提防用人们对意义的说明来定位对心灵的说明,因为这样做无异于循环论证。

另外,这个观点不仅仅是说,我们在避免奎因所说的循环论证方面可能有麻烦。关于整个 RTM 事业的哲学怀疑论更深的一个根源是这样的观念,即关于心理表征的某些语义学问题正是起源于原则(a)为真这一假定。这种担忧是值得考虑的。

RTM 的辩护者们有时提供如下的东西来代替关于心理符号的语义性理论。他们说,“瞧,我们知道,存在着像符号这样的东西。我们还知道,符号的确有语义属性。例如,英语名词‘爽身粉’是一符号,它指称这样的物质即爽身粉。而且,英语句子‘爽身粉的气味令人作呕’也是一个符号,它有某种真值条件,即当且仅当爽身粉的气味令人作呕时,它是真的。总之,存在着 RTM 偏袒的那种语义事实。大家都承认,某人有这样的_人问题要问:这样的事实意味着什么;某物作为一个符号意味着什么,或者(对我们的目的来说是相同的)一个符号具有它所具有的语义属性意味着什么。但是 RTM 为什么应该负起这种责任呢?为发展认知理论起见,为什么不直接把符号的一般概念当作给定的,为什么不继续整合围绕心理符号论概念详尽阐述的心理学?在研究中不让其他人因

518 第五部分

惑无疑是策略高明的睿智的一个方面。符号就其本身而论是为作为符号论者的符号论者而准备的。”

然而,对这种规避战术存在着严厉的驳难:即对 RTM 的辩护很可能要求我们把心理符号看作是自成一类的。如果是这样,那么我们不能设想,心理符号的语义性仅仅是由关于全部符号的符号学理论所解释的现象的一种特例。困难在于:对于各种非心理表征的实例(诸如自然语言的词和句子、ALGOL 的符号、狼烟、象形文字、数字、姿态、账单、图画,或诸如此类的东西)来说,符号的语义性至少应根据产生或解释它的人的心理状态部分地加以解释,这全然是合理的。例如,试考察这一事实:“爽身粉的气味令人作呕”这些标记是真的,当且仅当爽身粉的气味令人作呕。关于这个事实我们所能说出的最好情节可能就是必须述及英语说者/听者的某些交流意向。对于这一情节怎样进展,我不可能提供任何详细的说明。也许下述说法涉及到说者的意向,即他产生的标记应该被当作是真的,当且仅当爽身粉的气味令人作呕;或者关键也许在于:说者的意向遵循他的语言共同体的成员所享有的惯例系统,这种惯例中有这样的演绎结论:“爽身粉的气味令人作呕”这些标记是真的,当且仅当爽身粉的气味令人作呕。虽然细节是有争议的,但这样的某种说明尤其能从格赖斯、希夫尔、刘易斯和哈尼什那些哲学家的洞见中综合出来。

正如我所说的那样,我不是主张:对自然语言符号的语义性的这种说明是正确的;而仅仅是说:这种说明很有可能而且应该是关于符号的相当一般的陈述的一个特例,即是说,符号的语义属性总是可以追溯到关于其使用者的命题态度的某个或其他的事实。甚

至可以认为,映象(images)就是以这种方式起作用的。我们或许会这样被告知:某人曾坐着作为“自由女神铜像”的模特儿。但“自由女神铜像”并不是它的坐者的肖像;更确切地说,它根本就不是一个肖像。它意味的是:高高举起自由火炬的自由意志的表征。这无疑是说,使得它与高高举起……等的自由意志的表征这一事实(也许它与这一事实是同一的)密切相联的东西,就是雕塑家打算让它成为的东西。肖像不能使然,因果联系也是如此,除非它们经由行动者的意向所驱使。在解释表征为什么代表它所代表的东西时,诉诸于命题态度看起来是必不可少的。

据说,这触及到了这样一个观点,尽管是一个合理的众所周知的观点:有一种符号,其语义性原则上不应追溯到行动者的心理状态上,这种符号就是心理表征。理由自然在于,RTM旨在把心理符号的概念作为关于心理状态的意向性理论的解释装置的组成部分;^①根据RTM,心理状态就是依据心理表征的语义属性才具有意向性的。因此就不宜说心理表征也是依据心理状态的意向性而具有语义属性的。

所以,RTM不能仅仅把心理符号论的概念视为理所当然的,也不能担保这样的信念:某人总有一天会得到关于心理表征作为其特例的一般符号论的说明。相反,就语义的东西可能依赖于一般意义上的心理的东西而言,由于RTM可衍推出:至少存在着那学说所不适用的一种符号,因此RTM便陷入了麻烦。这一来,RTM需要一种关于心理符号的说明,它不求助于把命题态度作为其语义性的来源。如果它不能提出一种说明,那么现款交易,至少要对主张原则上可形成这样一种说明提供某种理由。

我曾说过,对心理表征的语义性的说明问题既是至关重要的,在某种意义上也是自我隔绝的。前一点已得到阐明,我说两句后就结束这部分的讨论。

符号的标记是合格的物理上的特殊事例。我推测这尤其适用于那些大概是神经性客体的心理表征的标记。现在,作为物理上的特殊事例,标记符号可以被假定有许多非语义的属性,而且我们可以认为这些是——至少为了当下的目的——相对不成问题的。关键在于,明确地强调只有心理表征的非语义属性才在确定哪一种心理操作适用于它们中起作用,这是流行的 RTM 解释的十足的特征——的确,它是在很大程度上不同于该学说的经典阐释的少数方面之一。我不再谈这个问题,因为它是我在别处详尽讨论过的核心问题(福多,1981,第9章),但是我把它当作是心理过程是计算的这一观念的必要部分,而且我认为这一观念是认知科学的中心原则之一。

结论是,关于心理表征的语义性即使不提出基础性的——或任何别的——问题,人们也可以建立相当可观的认知科学。在某种意义上,即使不提出这样的问题,即心理操作由于什么而成为计算的东西、心理过程由于什么而成为心理操作的因果序列,你也可以建立关于心理过程的全部理论(见上面的(b))。事实上,这是认知科学文献中屡见不鲜的东西(虽然这种境况由于区分逻辑句法问题与所谓严格的语义学问题这一普遍的错误在很大程度上未得到阐明。作为这种混淆的结果,许多认知科学家设想自己从事的是语义学研究,而事实上他们不是)。

因此,一方面,对于完整的 RTM 来说,至关重要的是,心理表

征的语义性中应该产生一些富有见地的思想;另一方面,我们不能简单地假定:“进一步研究”——至少认知科学家们到目前为止所从事的那种研究——可能会产生任何东西。第三(可以这样说),这一问题是最关键的。因为,确实有理由担忧:在努力从心理符号的语义性中引申出命题态度意向性的时候,RTM 也许把解释的次序弄颠倒了。导言就写到这里。

语义学假设

在这篇论文中,我想做的就是概述关于心理表征的语义性的一种(大致地说,是目的论的)说明。这一概述基本上是明确的。尤其是它阐明了S是M的真值条件意味着什么,在这里,M是一个心理表征,S是现实或可能的事态,由于它,M有其真值。这种说明将满足如下重要条件:对语义关系的阐述不用求助于意向概念。不过,可能有人会论证说:既然“S是M的真值条件”是所讨论的唯一的语义学语境,而我又没有证明:心理表征的语义性完全存在于它们所拥有的真值条件之中。因此这种说明是不完善的。这是值得讨论的。

似乎比较清楚的是,具有相同真值条件的符号在它们的语义属性方面仍可能是不同的(这意味着:可能是非同义词)。正是使“晨星是明亮的”为真的事实,也使“暮星是明亮的”为真。我不会轻率地对待这类问题的。在我看来,这分明是主张,纯粹真值条件分析也许过于粗鲁,因而不能重构语言符号或心理符号的语义属性。^②尽管如此,我仍将把有关真值条件的问题当作中心,理由在

于,对于心理表征的语义性的说明似乎至少得弄清它们与它们的真值条件的关系的本质;下述期盼不是毫无根据的,即基于对这种关系的一种说明,其余的工作就可通过诉诸于像“推论的作用”那样的概念而加以完成。我多半是为自己壮胆子,但如下说法似乎不是完全没有道理的。正是“晨星是明亮的”和“暮星是明亮的”这一非同义词的构成,你才通过适当不同的推论思路从它们那里得出结论说:“……在早晨升起的一颗星……”。当然,这里的“适当”是空白支票;我当然不知道该怎样付现金。

不管怎样,因缺乏更好的说明,我便暂时赞同近来所谓的关于心理表征的语义性的“两因素”说明(例如可参阅菲尔德 1977,洛尔 1981 和麦金 1982)。根据这些说明,语义属性同一性的充分条件是真值条件的同一性和推论作用的同一性,所有相关理论的心理表征的语义属性,应求助于这些因素的一个或另一个或二者而得到解释。根据 RTM,推论既然是心理表征的转型,所以心理符号的“推论作用”大抵是从关于心理操作的事实中产生的某种逻辑构造,这种心理操作又的确(或在适当理想化的情况下)适合于该逻辑构造。然而,由于假定心理操作仅仅是依据心理符号的句法属性(作为不容否认的功能形式)而成为适用的,因此下述假定与两因素观念是完全相容的,这假定是,影响心理表征的语义性的最适当的符号与世界的关系,是它们所具有的与决定它们的真值的事态的关系。^③

归根结蒂,我得这样来设想关于心理表征的语义性问题的规范形式:使一种心理表征与满足它的真值条件的事态关联起来的东西是什么?这是一种什么类型的问题?

理解它是何种类型的问题的一种方式,就是查明什么不能算作相关的回答。于是设想某人说:“使一种心理表征 M 与作为它的真值条件的事态 S 关联起来的仅仅是, S 是这样的事态,如果得到了它,那么它就使 M 为真。或者,以略微多一点形而上的腔调加以表述就是,只是在 M 与世界相符合的条件下,当且仅当 S 正好在这样的条件下, S 是 M 的真值条件”。我希望在直观上明显的是,这不是我们想要的那种回答;无论如何,值得花时间来反思,以弄清它为什么不是的某种意义。

一种观点是,我们既不想让语义概念(如真值或符合)出现在对“……是……的真值条件”的说明中,也不想让意向概念出现于其中。但是这个驳难太肤浅了。假设我们具有关于心理表征系统的真值理论;即是说,心理符号映射到实际或可能的事态之上的一种函项满足这样的条件,即每一符号是真的,当且仅当与它配对的事态出现了。现在我们能从外延上考虑这种函项,并指出,使一种心理表征与它的真值条件相关联的东西是,后者在这种映射条件下是前者的映象。这一公式不包含可能引起疑义的语义或意向概念的出现。为什么是这样呢?

问题在于,存在着太多的把符号映射到世界之上的方式(符号与事态能由之而“相符合”的方式);虽然——正如我们可以假定的那样——只有一种配对,其外延是符号与它的真值条件的(有次序的)配对,但是不存在把这种配对与其他东西区分开来的符合概念本身。试考察符号向事态的任意的映射,即一种不把每一符号与它的真值条件配对起来的映射。即使这样,这种配对规定了一种符号与世界的符合。我们可以把这种符合称之为(不是“真值”而

524 第五部分

是)“shmuth”。* 我们可以说,T 是关于心理表征的一种 shmuth 理论,如果它以这种配对规定一种映射是共外延的话。所以,如果某人想知道一种心理符号与它的 shmuth 条件之间的关系意味着什么,那么我们可以说,“它是这样的事实,即 shmuth 条件是在 shmuth 理论规定的映射下的关于该符号的映象;”这正是我们关于真值条件所要说的那种事情。对此,人们应该回答说:关于由真值定义所详述的那种符号与世界的符合,存在着某种特殊的东西;存在着 shmuth 符合所没有的某种东西;存在着关于它们的真值条件的事实据以阐明符号的功能作用的某种东西,在某种程度上,关于它们的 shmuth 条件的事实则不能阐明符号的功能作用。好,是这样;但神秘的因素会是什么呢?

这个问题所要求的不是像规范语义学家对那个概念的理解那样的一种语义理论,而是一种“意义理论”,这里的“意义理论”指的是该短语的古老而不严谨的意义。例如,这里有一种相关的回答(哎呀,这种回答可能肯定有例外,如它是错的)。人们可能会以多少有点休谟式的精神说,使心理表征与它的真值条件相关联的是,前者摹写后者。使这成为正确类型的回答的东西是,据假定,它把真值理论为心理表征所指定的那种符合与心理表征向世界的每一别的映射区分开来了。这是因为,心理表征据说只摹写了它们的真值条件;例如,它们并没有摹写它们的 shmuth 条件。即使碰巧心理表征的真值条件被满足了,当且仅当它的 shmuth 条件被满

* shmuth 为 truth(真值)的押韵讥讽词,由 shm+uth 所构成,shm 代替的是 tr。为保持原文的神韵,故不译出来。

足,也没有摹写它们的 *shmuth* 条件。假设, *M* 是一种心理符号,以至于真值理论把它映射到雪是白的之上,而且 *shmuth* 理论把它映射到草是绿的之上。把前一种映射区别开来的东西是, *M* 据假定摹写雪是白的,不摹写草是绿的(例如, *M* 是对白雪的摹写,而不是对绿草的摹写)。

我已经说过(见 n. 2),下述推论图式似乎有可能是无效的: *S* 是 *M* 的真值条件;当且仅当 *R* 则 *S*; *R* 是 *M* 的真值条件。我们不希望结果是这个样子,即草为绿是“雪是白的”的真值条件,即使“草是绿的”和“雪是白的”真的在本质上是等价的。(参阅戴维森 1967;里夫斯 1974;福多 1977, pp. 36 – 37)意义理论为你要做的事情之一是,弄清这样的推论为什么不被认可。它之能予以澄清,显然是借助于说明:它正是 *S* 由以成为 *M* 的真值条件的东西;即是说,借助于阐释 *S* 和 *M* 之间所具有的关系,而不是 *R* 和 *M* 之间所具有的关系,即使在这里,“当且仅当 *R* 则 *S*”是真的。例如,正如 *M* 是对 *S* 的摹写,而不是对 *R* 的摹写那样。

噢,我们都知道休谟的意义理论错在何处;那显然是无法定义“摹写”(或“相似”或别的任何概念),这就使得心理表征只能摹写真值条件。在很大程度上,它独立于心理表征(其中的一些或全部)是“图像式的”还是“推论式的”这一悬而未决的问题。即使你设想:约翰个子高这一思想是一种摹写,它表现的是约翰个高,你也会碰到这样的事实,即诸如此类的任何摹写还表现了他的许多别的方面(如有脑袋)。但约翰有一个脑袋并不是使约翰个高这一思想为真的事态。

这绝不是没有任何争议的;我假设:任何人都不会主张:摹写

是把心理表征与其真值条件关联起来的東西。我提出这个问题,旨在澄明休谟坚持他的 RTM 观较之我们坚持我们的 RTM 观更幸运的地方。他对我们只是呆头呆脑地盯着的下述问题已作出了回答(尽管算不上绝顶好的回答):什么使 S 成为 M 的真值条件?

现在我们可以概括一下贯穿于余下讨论之中的语义学假设:两类问题对于人们建议当作符号对待的任何对象集合一定是可以说明的;对于起心理表征作用的(大概是神经逻辑的)对象一定是更有凭据的。它们是:

(C)符号有什么语义属性;

(d)符号由于什么而享有它们的语义属性。

第一类问题可通过提供形式语义学而得到回答。我假定,就心理表征而言,它们的语义性在于它们拥有真值条件,因而关于心理表征的形式语义学便采取了真值理论的形式。

第二类问题可通过提供意义理论而得到回答。在当前的情况下,一种意义理论必须解释心理符号与事态之间的这样一种关系,即由于它,后者便成为决定前者的真值条件的东西。对这类问题的标准回答包括:“事实是(当且仅当真值条件被满足,符号的使用者便想到它应该为真)”;或者说,“事实是(如果真值条件被满足的话,那么那符号便把世界描述为它看上去所是的样子)”;或者就此而论,“事实是(符号的产生是对由被满足了的真值条件所组成的刺激的一种能识别的反应)”等等。也许有一些符号,对这些符号来说,这些回答的一种或另一种是对的。但是我认为无可置疑的是,它们对于心理表征来说都是不正确的。那么就心理表征而言,

回答将会是什么呢?

心 理 学 假 设

从根本上说,一种意义理论(就我使用这个短语的意义而言)就在于主张:某种关系与真值理论所详述的关系具有相同的外延。所以,休谟的意义理论在于主张, M 摹写 S 的关系与 S 是 M 的真值条件的关系是共外延的;斯金纳的意义理论在于主张, S 是 M 的可识别的刺激这种关系与 S 是 M 的真值条件的关系是共外延的;如此等等。我想提出的意义理论也将是这个样子。即是说,我将描述事态 S 与心理符号 M 之间的关系,我认为它是这样的关系,即由于它,前者才成为后者的真值条件。既然这种关系将参照认知系统的假定的功能的和目的论的属性而被规定,因此我对它的说明(就此而论像休谟和斯金纳的说明一样)将极大地依赖于关于什么样的心理学实际上为真这一假定;特别是,关于有什么样的认知机制以及它们行使什么功能的假定。恐怕其中的某些假定会使你惊慌失措。啊,好啦。

考虑到RTM框架,我将首先假定某种似乎不是特别有倾向性的东西(原始资料详后)。我认为我们将获得一种关于信念的计算心理学是理所当然的。信念的计算心理学是一种理论,它将这样具体阐明某种关系(称为 R^*)^④和一种心理表征系统(称为 M_s):

(e)O相信 S ,当且仅当(有心理表征 M ,以致(1) S 是 M 的真值条件以及(2)O具有与 M 的 R^*)^⑤

528 第五部分

所以——还是根据思想的语言是英语这一方便的假设——(e)的例子包括：“O 相信天正下雨，当且仅当有一种心理表征 M，以致(1)天正下雨是 M 的真值条件；以及(2)O 具有与 M 的 R^* ”。这意思是说，已知关于英语的语义学，结果便是，这个双条件句的左边是真的，当且仅当 O 具有与“天正下雨”(或其同义词之一)的 R^* 。

对此，人们已阐发了许许多多的见解。首先，根据 RTM 的精神，这一描述设想，信念关系将参照相“……是……的真值条件”这样的外显语义概念而得到解释。根据 RTM，信念是一种与语义上被解释的符号的计算关系。计算心理学理论不解释符号与世界的关系；它们仅仅预先假定这种关系，由此自然可以说，在分析心理表征的语义性时，我们不能诉诸于信念关系，否则便要陷入循环解释。

另一方面，如果我们愿意的话，我们可以允许自己求助于 R^* 。这是因为 RTM 假定，关于心理状态的理论据以而得到详尽阐述的有机体与惯用语之间的关系必须是计算的（即形式的；意即参照那些惯用语的非语义属性可彻底予以描述的）。当然，有的人可能怀疑：有机体和惯用语之间存在着任何纯粹的计算关系，以致前者由于纯计算关系所获得的事实而相信后者所表达的东西。例如，某人可能坚持认为，有一种理解命题的基本关系，以至于只有当有机体以某种方式（即是说认同式地）理解一个命题时，它才相信一个符号所表达的东西。我认为，坚持这种观点并非全然没有意义；任何人毕竟不能提供微乎其微的令人满意的计算替换物这样的任何东西。然而，如果你确乎认为 R^* 不可能是一种计算关系，那么你事实上就是主张：RTM 一定是错误的。而我们则假定，

RTM 是正确的。

接下来的观点是,作为计算关系的 R^* ,虽然不可能知道心理表征的语义属性,但它仍然必须予以具体说明,以至于涉及到了这些属性。导致这一结论的论证思路是因丹尼特、戴维森和其他人的著作而为人所熟悉的。大致地说,人们倾向于作出适当的推论是其持有一个信念的一个条件;人们倾向于作出的是哪些信念,部分地决定了人们所持有的是哪些信念。从计算心理学的观点看,这意味着, R^* 详述的某些计算将不得不与这样的惯用语关联起来,它们的语义属性适合于(归纳的,演绎的,似乎有理的或别的)论证的前提和结论。假定有机体具有与 M_1 的 R^* ,并假定这一事实:有机体具有的这种关系对于它最终具有与 M_2 的 R^* 在原因上是充分的,而且假定这样的因果条件,即具有与 M_1 的 R^* 一般会导致实际上直接具有与 M_2 的 R^* ;前一状态可谓后一状态的有效原因。这一来,人们便不愿把真值条件天正下雨归属于 M_1 、真值条件现在天没下雨归属于 M_2 ;或者人们至少会设法不让这类事情经常发生。至少在此范围内,对 R^* 的系统阐述限制了把真值条件归之于心理表征,反之亦然。

我这里的表述是很不精确的,因为这种限制意味着什么,显然是不清楚的。正如人们通常所说的那样,没有人知道,信念的概念在事情到头来就是这样之前究竟能容许多少推论的不连贯性。而且有足够的理由假定,回答这类问题可能是非常棘手的;例如,可能要相对于有机体的类型来作出回答(参看切尔尼克 1986)。我提出这个令人不快的论题只是想支持下述否定的观点:我认为,不存在这样的方式,即对 R^* 的详尽说明完全通过它而限制了把真

530 第五部分

值条件归之于心理表征。这就是说,我认为,在它假定的计算的操作中, R^* 无论承认多高的复杂性,将总是存在着一种以上的把语义属性归之于 R^* 所限定的心理符号的可能性。不妨把这样的每一种归因称作对心理表征系统和过程的一种“解释”。因此,要回到我们前一节所提出的一类观点,意义理论应予澄清的问题就是:从其他所有解释中选择正确的解释意味着什么;是什么把心理表征的真值条件与心理表征能映射于其上的其他事态区分开来,以与主张有机体的心理生活始终连贯的观点保持一致。

正如读者肯定会想到的那样,我准备用“O 具有与 M 的 R^* ”作为“M 有真值条件 S”这一描述中的一个从句(即使——考虑到前面一段,——我显然还需要许多别的材料)。借助用图表示该情境的方式,我将从 S. 希夫尔那里得到一种启发。假设每一真的有信念的有机体在它们头脑中都有一个盒子。这个盒子能包含标记心理表征。人们可以想象(在此说明中,这虽然不起根本作用),这个盒子标上了“是”的记号。“是-盒”的内容由如下原则所决定:对于每一心理表征 M 来说,O 具有与 M 的 R^* (因而相信 M 表达的东西),当且仅当 M 的标记在是-盒中。

成为是-盒这一属性据约定是一种功能属性。我特别假定,每一是-盒是与轮子、滑轮、继电器等等精致机械装置相联系的,以致把心理表征标记放入盒中,对于有机体的行为和这类标记的其他配置来说,便有各种相当复杂的因果推论。而且,你把既定的标记放入盒中那里得到的因果推论,再度复杂地依赖于那些标记的(非语义的)属性。(也许,它依赖于它们的重量、形状或它们的电传导性。)无论如何,对于每一 M_i 来说,把 M_i 标记放入是-

盒的结果一定包括： R^* 所说出的必定是那个样子，以便 O 相信 M_i 所表达的东西。

我推测，这种说明有一些可以得到改进的方面。例如，我们忽视了信念的产生性。如果我们开始变得认真起来，那么这就需要加以矫正。也许我们要把“核心信念”的有限集合（参阅丹尼特 1975）与假定有机体仅仅在禀赋上拥有的信念的无限集合区分开来。仅仅是由于前者，标记 M 在是 - 盒之中对于 O 具有与那个 M 的 R^* 来说才成为必要且充分的。对于非核心信念而言， R^* 会依据某种闭合原则而得到定义，对于这些原则的细节，我不想予以深究。

不过，所有这些可以留给心理学家去做。因为，考虑到计算分析的本质，我们可以肯定，对满足功能主义者的基础规则的 R^* 作任何详尽说明，将正是对是 - 盒理论的精心阐述，即一种趋向于（人们常常这样希望）具有更多的经验合理性的精心阐述。下面，我集中论述的是“ M - 标记在是 - 盒之中”而不是“ O 具有与 M 的 R^* ”，前一概念具有后者无论如何也没有的力量和活力。

我们很快就要进到那假定有争议的地方；不过在进至之前，还有提出一个更温和的假定的余地。我认为，我们能确认一些心理学机制，它们事实上导致心理表征的标记进入（ O 具有与各种 M_i 的 R^* ）是 - 盒。此是 - 盒本身就是那些机制中的一个，因为我们可以推测它如此被构造，以致放入某些标记就会直接引起也在它里面的其他标记。[这也许就是那个模型借助重构“直接推理”（假如还有人喜欢更老的习语，或许可称作“直接联想”）概念所提供的东西。]

无论如何,是-盒肯定不是其活动决定哪些标记在是-盒中的唯一机制。大概还有知觉机制在起作用,而且记忆机制和推论机制以及健全的心理学一般所说的任何东西,均卷入了信念的固恋。我想用一个替补术语表示心理学机制的这样的集合,它们的运作可影响是-盒的内容。我将称之为“事实上的”认知系统。

我要强调的是,在关于事实上的认知系统的描述中,是不能加进目的论的假定的。例如,我从来没有说,这些机制是为了把材料放入是-盒之中;我说的不过是,它们事实上的确把材料放到那里。因此比如说,如果有遗忘的机制(包括作为时间函数的机械记忆痕迹衰变那样的非理性机制),那么由于它们部分地决定了在是-盒之中的是什么,因此它们也是事实上的认知系统的一部分。

不过,我将提出一些目的论假设。第一个是:存在着心理学机制——据定义是事实上的认知系统的子集——其功能是把标记放进是-盒中。我将把这些机制的集合直接称之为“认知系统”。我推测,痕迹衰变机制不在其中,因为我推测,痕迹衰变根本没有任何功能。只是碰巧如此。但是我敢断言,知觉机制是认知系统的组成部分,正如有助于信念的合理固恋的其余机制那样。对于我的其余论述非常重要,这种目的论主张——以及我还要提出的一个更有力的主张——应该在实际上是正确的。因而,也许我最好还是说一下我对一般目的论主张的态度。

首先,我认为,目的论语境(如“……的功能就是使——* 得以实现”)常常是非真值函项的。这里的论证是众所周知的。例

* 这里的“——”代表功能所要实现的东西。

如,试考察像心脏的功能是实现血液循环这样的真值。假设,成功地让血液循环的所有心脏,当血液经过时事实上发出一种嘶嘶声。发出这种声音是不是心脏的功能之一:它是不是那个心脏所做的事情之一,以至于不能如此作为的心脏因而不能正常地行使功能,仍然是一个悬而未决的问题。假设不是这样,那么我们就有:“心脏的功能是实现血液循环”和“当且仅当有适当类型的嘶嘶声血液就循环”二者为真,而“心脏的功能是实现有适当类型的嘶嘶声”则为假。总之,目的论语境至少在作为非真值函项的意义上是内涵性的。但我认为,对目的论概念的开发利用仍没有超出我们开始从事的事业的基本规则。我允许你求助于非意向语境说明“S是M的真值条件”,但我不允许回避内涵性的语境。

事实上,要把余下的秘密公布开来,现在就应该澄清这篇论文的整体战略是什么。有这样的老生常谈,即目的论大约就是:你没有在逻辑上预先假定动原,你也能得到意向性。(事实上,在有一个动原的那种目的论中——如在铅笔是为了书写中——功能和意向的因素都混合在一起了。我推测,使铅笔是为了书写为真的东西,就是铅笔制造者打算让它们派上的那种用场。)现在,我强烈地倾向于相信,格赖斯及其同僚的主张是正确的,他们认为:意向性是大量语义现象的中心;尤其是命题态度的意向性是词、句子和诸如此类的自然语言表达式的语义性的源泉。然而,正如我们已看到的那样,我们不能接受格赖斯对于思想的语言的那种说明,否则在RTM中将陷入循环论证。所以我建议,下一步最好是这样来行事:从关于认知系统的目的论的某种东西中引申出心理表征的语义性——特别是它们拥有的真值条件。再具体一点讲,S是M

534 第五部分

的真值条件对于 S(参阅 n.2)诉诸于下述事实而得到解释来说是不透明的,这事实是:“……的功能就是使——得以实现”对于“——”是不透明的。

最后,我认为,目的论归因的真实性(或错误)与对应的反事实的真实性(或错误)有某种联系,因为不同的反事实在不同场合下是相关联的。我假定,正是目的论的真实性与反事实的真实性的联系说明了前者的意向性。因而,我坚持认为有目的论事实,并不意味着坚持有不可还原的目的论事实。如果相关的反事实的真实性不过是有目的论断言的真实性,那么对我来说真是太好了。另一方面,我不知道哪一些反事实是为支持把功能归因于认知系统而挑选出来的反事实。为了启发的和说明的目的,我将假定,它们是关于自然选择的事实;亦即这样的事实,若不是……,那类选择就不会以过去通常的方式发生。我极不喜欢庸俗化的达尔文主义,但因缺乏更好的东西我不得不说到它。

对于这里要说的内容,我最好是开门见山。我将具体说明某种(非语义的)属性 P,以至于我或许会断言,正因为心理表征具有 P,它才具有它所具有的真值条件。这里, P 是心理表征依据有关认知系统的目的论事实而具有的属性;心理表征具有 P,因为有机体的认知系统具有某些正常功能。我推测,这些正常功能本身应通过列举某些反事实而被具体阐明;因为说有机体的认知系统正常行使功能作用,就是说它适合于这样的系统:如果它是如此这般的情况,它就会做这样那样的事情(对于某种决非必然有限的这样那样和如此这般的清单而言)。于是就此范围来说,心理表征语义性的说明所关注的,严格地讲,不过就是存在着关于认知系统的正

常功能作用的某些真理。就此范围而论,对心理表征语义性的说明,并不关心达尔文主义是不是正确的,或认知系统事实上是不是自然选择的产物。例如它能为聚集在某人的汽车房中(好歹不是通过自然选择)的有机体所满足,只要这些有机体具有认知系统,它们的认知系统具有适当的正常功能。当我不得不说认知装置的正常功能是什么时,当我不得不提出某些理由假定:有机体的认知装置的确具有我所说的正常功能时,达尔文的话语就肯定会出现。达尔文的说明对这些目的中的两个都是适合的。在进化了的有机体的情况下,把有关正常功能活动的事实归结为有关选择模式的某些或其他事实,似乎是合理的。大致说来,进化了的系统的正常功能是这样的功能,依据它,有机体才有选择的优势。但是我将作出的关于心理表征的真值条件的陈述,不需要(即不能衍推)达尔文主义是真的,或认知系统是进化了的;它只要求存在着我所说的关于认知系统的正常功能活动的事实。

现概括如下:我假定,存在着目的论事实。我假定,它们也许是可还原的。我假定,目的论语境由于与反事实语境的联系而有内涵性。我假定,存在这样的认知机制系统,其功能是把标记心理表征放入是-盒之中。

但是,这自然不只是任何旧的标记心理表征,也不只是在任何旧的场合。例如,如果认知系统是由选择压力所塑造的,那么它们必定是这样的压力,即心甘情愿地让 M_s 放入是-盒中,而不是漫无目的的。(毕竟,在是-盒中,就是让有机体具有与你的 R^* 。有机体具有与哪一种 M_s 的 R^* 必然影响它的行为和它最终的生存能力。)然而,如果不是漫无目的的,那么大概是在某些强制之

536 第五部分

下。

不妨把这种强制称作“进入条件”。这一来,对于心理表征 M 来说,进入条件就是那样的事态,即当正常地发挥功能作用时(认知系统促使 M 在是-盒之中),当且仅当(进入条件被满足)。在下文中,进入条件将是一种基本的构造。对此我有许多要说的。

(f)心理符号的进入条件在实践上可以是任何东西,只要它是正常功能活动的认知系统促使那个符号的标记在是-盒之中所不得不被满足的东西。例如,如果你是一只蛙,那么你的认知系统可能如此行使功能作用以致把某种标记放入你的是一盒,当且仅当一只蝇在你的视野内(参阅莱特文等 1959);或者,它可能如此行使功能作用以致把某类标记放入你的是一盒中,当且仅当你的听觉系统受到了同种呼叫的刺激(卡普拉尼卡 1976)。

我特别要强调的是,进入条件没有必要根据(例如)视网膜排列、鼓膜的振动模式,或一般意义上的任何别的传感器状态来加以说明。进入条件通常是对世界所是的方式的限制。由于进入条件是依据认知系统的目的论而被定义的(例如,它们是当这些系统正在做它们被选择做的事情时所关注的条件),因此这应该是显而易见的;认知系统的初始功能肯定是使有机体的命题态度和它的环境的状态之间的融贯关系得到实现。例如,当蛙的同种哇哇鸣叫时,蛙脑内的灯为什么打开?因为存在着这样的(认知)机制,它们恰好在一连串的听觉能量冲击蛙的听觉传感器的条件下打开了灯的开关。但是为什么有这些机制呢?它们凭什么有它们的选择优

势？是靠它们把蛙的心理状态与哇哇鸣叫的同种之出现相互关联起来的能力。这一来，蛙的认知系统在这一方面正常发挥功能作用意味着什么？它意味着蛙的是一盒包含着这样的标记：“嘿，有一种发咕噜声的同种”，当且仅当在那场境有一个发咕噜声的同种。那么，发咕噜声的同种的标记的进入条件是什么呢？仍是有一发咕噜声的同种在那里。注意，进入条件不是蛙的鼓膜（如果那是蛙所具有的东西的话）受到了适当的刺激；虽然通常的因果链通过这种刺激把蛙与同种的咕噜声联系起来了。

你可能会说，“一点也不错。但是，即使目的论上的考虑表明，心理表征的进入条件常常是外在于有机体的某种东西，但仍然存在着这样的问题，即我们应该认为它是哪一种外在的东西。在同种的咕噜声与 M 放入是一盒之间存在着一条漫长的因果链。我们将把它的哪一个链当作神经反应的进入条件呢？”下面的思考据说有助于解决这个问题。

(g)与既定的心理符号相联系的进入条件可能有两种相当不同的组成部分：一是对于那个符号是特有的；二是反映了各种认知系统的结构的一个或其他方面的部分。用相对原始的例子来说，试考虑一下作为蝇的捕食者的蛙。这里所说的进入条件大概是这个样子：有一只在周围飞舞的蝇；蝇及其周围为灯光所照亮，这种光的强度、频率、对比度以及上帝知道的所有别的东西，都落入某些审慎判断的范围之内；从蝇那里反射的光在不少于——时间内冲击蛙眼的表面；如此反射的光线的视角不少于——；等等。（我不知道怎样填这个空，但在原则上无疑是可以填写的。视觉心理物理学显然在从事这项工

作。)

我的观点是,在“有一只蝇”这一符号的进入条件中的某些从句完全是来自于蛙的视知觉机制的普遍特征。在“有一睡莲叶子”、“有一个蛆虫”或者作为蛙的视觉装置的活动结果而进入蛙的是——盒的其他一切符号的进入条件中,我们基本上有同样的从句。从这里开始,我将采用一个术语,根据这个术语,进入条件的那些从句反映了下述一般要求,即认知机制必须接近它们在其中起作用的情境,这样的从句就是“认识恰当性的条件”。因此,严格意义上的进入条件将是与特殊心理符号的标记相联系的强制。

现在,我们可以把这些考虑与在(f)标目下详述的那些考虑综合起来。即是说,为了发现心理表征 M 的进入条件,你首先得弄清,当认知装置正常行使功能时,需要什么事态来促使 M 在是——盒之中。接着你得有效地使用差异法,以析出认识恰当性条件的作用因素,只把对于 M 是特有的那些条件留在是——盒之中。据约定,这种剩余物就是 M 的进入条件。^⑥

(h)一个心理符号的进入条件可能是任意复杂的。如果某一符号有许多进入条件,那么就把它当作析取式。

如果都是这样,我们便可把认知装置看作是实现某种功能的东西;即从进入条件(被当作实际的或可能的事态)映射到心理表征类型上的一种函项。在这种映射之下,每一心理符号 M 与事态 S 是相互关联的,以至于(已知相关的认识恰当性条件得到满足):认知系统的正常功能是这样构成的,即当且仅当有了 S,该系统就把 M 的一个标记放进是——盒中。我们把这称作心理表征系统的

“进入条件功能”。我假定,如果我们对蛙的视觉系统的目的论、对隐含在蛙的视觉装置结构中的认识恰当性条件知道了应该知道的一切,那么就可得出结论说,在进入条件功能下作为蛙的思想语言的心理符号之一的映象是:有一只蝇在附近飞舞。

现在我可以说明我的意义理论的结论了:那就是,心理表征系统的进入条件功能与它的真值定义是共外延的。这就是说,对于每一心理表征 M 来说, S 是 M 的真值条件,当且仅当 S 是 M 的进入条件。更确切地说,某种更有分量的东西是:使得 S 成为 M 的真值条件的东西——即 S 据以看作 M 的真值条件的精确事实——就是, S 是 M 的进入条件。这样,思想语言的真值定义 A 仅仅是对一种关于(适当理想化的)认知系统的输入输出功能的具体说明。或者说,由于具有同一性的东西是对称的,因此你同样可以提出,认知系统不过是这样的机制,它们的(适当理想化的)输入输出功能就是它所使用的表征系统的真值定义。

有非常严肃的理由认为,这种意义理论不可能是真的。我将对其中的一些作出详尽的考察。然而,我首先要让你信服的是,真值条件应该是我所主张的那种东西,这至少在表面上是合理的——基于更严密的审视,不管这种观点最终被证明是多么的荒唐。

再来考察一下蛙。借用澄清某些哲学混乱的方式,试想你就是卡普拉尼卡或莱特文,或某个高度娴熟的神经心理学家,再假设你给自己安排了这样的任务,即弄清你看到的神经动作电位所代表的东西。(如果这不是探究心理表征的语义属性在实践上非做不可的那种事情,那么我就不能想象它还会是什么。)唔,你该怎样

做呢？我猜想你将恰恰是莱特文和卡普拉尼卡所做的事情：你试图发现进入条件。

这意味着：(1)你寻求一种事态，以至于它的实现对于动作电位的出现是必要而充分的，假如有机体的视觉装置是完好无损的，而且有机体与那一事态的关系在认识上适宜于信念的视觉固恋；(2)你会问自己是否作出如下假定：那一事态是动作电位的真值条件，假设你谈论的是神经触发的行为(等等)结果。^⑦噢，就现在的情况而言，正是在周围飞舞的蝇使得神经元得到激活，而且所触发的行为结果包括各种捕获和进食反射的活动，诸如动物确定方位、语言交流，等等。如果，一方面，触发的语义内容是有一只蝇在那里；另一方面，蛙的心理生活与丹尼特、戴维森和其他人所说的纯正的心理生活必定如此的方式是一致的，那么所有这些大约是你孜孜以求的东西。

因此你所阐明的是：触发的进入条件是周围有飞舞的苍蝇。假设你承认了(比如说基于哲学上的理由)下述观点，即中枢动作电位是有真值条件的那类事物。(记住：我们假装相信 RTM 为真。)我关于真值条件同一于进入条件具有不言而喻的合理性的论点根据的是这样的事实，即对于下述修辞学问题我们不可能想出合理的回答，这问题是：为了说明那里有苍蝇时动作电位为真，你不得不做的事情到底是什么呢？

我主张把它搁置在那里，因为眼下有更严肃的任务。某人可能会这样说：“先不管你说的是否合理；有一些责难如此严峻，因此它们对于进一步的思考来说是压倒一切的。”在本文的其余部分，我的计划是说明：这些责难并不像它们似乎所是的那样，具有

相当的判决性。最后我将尽可能从这些意义理论中抽引出一点启示来。

对于真值条件和进入条件的同一性,我能想到的两种反对意见是:一种认为它太荒唐了;另一种认为它空洞无物。我将依次处理这些麻烦问题。

这个理论明显的荒谬性在于,它隐含着一点这样的含义:在正常情况下,每个人都是无所不知的;在非正式的讨论中我注意到,对这个观点,人们的眉毛往往会竖起来。这种含义从何而来? 嗨,如果 S 是 M 的进入条件,那么在正常条件下, M 的标记在是 - 盒之中,当且仅当 S 事实上如此。如果 S 还是 M 的真值条件,那么由此可以说, M 在是 - 盒中,当且仅当它是得到满足的真值条件。但是一方面,如果 M 的真值条件被满足,那么 M 是真的;另一方面,我们的功能心理学向我们保证: M 在是 - 盒中对于 O 相信 M 所表达的东西来说是必要且充分的。所以,把所有这些综合在一起,我们就得到: O 相信 M, 当且仅当 M 是真的;这几乎等于说 O 无所不知。然而流行着一种对立的偏见,它显然是基于下述观察,即某些人相信某些错误的东西,而正确的东西又完全无人相信。

观察,简直是胡说八道! 我想论证的是,在相关的意义之下,每个人都几乎是无所不知的;所以归谬法的结论完全应该被接受。

首先,让我们来澄明那个“几乎”。成为无所不知,就是相信所有的真命题。但是对于每一真命题来说我们并没有保证:在 O 的思想语言中存在着一种能表达它的符号。基于这种理由,我主张我们具有的那种无所不知,与我在别处所说的“认识界限”的原则是完全相容的;即可能真的有这样的事物,即考虑到我们心理构成

542 第五部分

的本质,我们最终不能认识它们(参阅福多 1983)。根据现在的假设,它们包括我们的心理符号不能表达的真命题。这种限制可能过分严厉,正如我们思考物种而不是我们自己时立刻就明白的那样。例如,就我所知,有一只蝇和有发咕噜声的同种,它们是蛙能把握的唯一的命题,那对模于蛙的心理表征系统的表达力来说,与蛙的无所不知是完全相容的。

无论如何,我真的不偏袒这种主张,即在承认认识界限这个有限的意义上,我们甚至也是无所不知的。我偏袒的东西仅仅是,我们(以及具有信念的其他任何有机体)在合目的地相关的环境中是无所不知的,此环境包括所有有关的认识恰当性条件的满足。大致说来,我坚持的是:“应不厌其烦地强调下述令人不快的事实,即 P 以及(如果我们能想到 P)我们将最终相信 P。”但是对我们的心理符号能表征的许多不定的事态来说,我们从来没有如此去强调——因为适当条件从未得到满足,而且关于这些事态,我们可能形成错误的信念或全然没有信念。总之,“正常的功能作用”意即我们相信 M 所表达的东西,当且仅当 M 表达的东西是真的,在此意义下,认知系统的正常功能完全有可能是它从没有行使过的功能。^⑧

表述这一观点的另一方式是,我所坚持的不是差不多的无所不知,而是对错误、无知和其他类似的认识过失的某种说明。我所主张的(值得一提的是,它与可敬的柏拉图主义传统一道)是,错误和无知在某种意义上是偶然事件。这就是说,除根据认知装置的那些由以行使功能作用的属性以外,它们还应该用别的方法予以解释。当然,这些“偶然事件”最常见的东西是:由于有机体的因

果/时空环境这样的事实,认识恰当性条件不出现。(“我不能从这里看见它”,“我年岁不大,没碰到他”,“天黑了”,“我按时睡觉”,“没有人告诉我,我怎么能知道?”)在我看来,关于信念固恋的任何理论将不得不把相当多的无知归咎于认识恰当性条件的阙如。我们经常没有我们过去用别的方法获得的信念,因为在那个时候我们简直没有获得该信念的恰当的条件。当凯撒到达鲁比孔河的另一边时,为什么我们不能看见他先放下哪只脚?这不是因为我们的视觉系统没有正常发挥作用;也不是因为我们没有想到:凯撒首先放下他的左(或右)脚;而仅仅是因为就这个事件来说,我们在时空上不处在那类视知觉所要求的位置。正是这类讨厌的、地理学上的事实的总和,可以解释我们为什么不能判断(因而不知道)是哪一只脚……等等。试与上帝相比,根据目前的观点,他的无所不知主要决定于他的无所不在。

当然,我无意于主张,不具备恰当性条件是有助于限制我们认识操作的唯一的“偶然事件”。还存在着似乎反映了信念固恋的认知机制的机械限制的心理学偶然事件。(“它太复杂了,我算不出来”,“我不能想得那样快(或看得那样远)”,“我害怕记住事实”,“我不能判断距离”,“我总是忘记了名字”,“我太吃惊而没注意到(或太讨厌而没留意)”。)后面的例子涉及的过失可归之于信念固恋的机制,但不能归之于选择优势据以形成这些机制的那些属性;就那概念的目的论意义而言,也不能归之于它们的功能属性。不能清楚地看,就是视觉系统不能做它被设定要做的事情;矫正这些缺失,人们就能更好地看。由此便有了眼镜。

总之,在说明既定的心理符号为什么的确出现(或不出现)在

544 第五部分

是一盒中时,有三种条件无论如何值得注意:关于该符号的进入条件是否被满足;有机体的认知装置就满足进入条件的事态来说是否正常发挥功能作用;有机体就那个事态来说是不是处于适当的位置。从直觉上看,这些分别相当于:对于世界状态的强制,对于有机体状态的强制,对于有机体和信念对象之间的(因果)关系的强制。只有当所有三种强制同时被满足时,无所不知这一论断才是有效的。

我认为,这个论断意味着的东西仅仅是,无所不知的缺失不是认知装置的设计上的缺失;相反,它们应根据那些从目的论观点来看的、相对外在和令人乏味的事实而得到解释。

使这个论断之成为一个关于应把什么称作偶然事件的约定,是这样的可能性,即失误似乎成为认知装置的一个组成部分;甚至当环境是认识上最适宜的时候,它的操作法也有利于错误信念的固恋(或固恋真信念失败)。为了弄清这种问题怎么可能不只是术语上的,请考察一下与“除偶然事件外我们无所不知”在逻辑上相当接近的一个学说,即除偶然事件外我们是不朽的。后者是一个值得敬重的科学假设,而且诉诸于相当多的人死了这样的偶然观察是不能把它驳倒的。

对于衰老和死亡是否内在于我们的构造这一问题,科学上尚存有争议。一幅图景——乐观主义的图景——认为:我们之所以死亡,只是因为我们衰老,之所以衰老,只是因为像蛋白质合成这样的机制中随机失误的积集。这幅图景在两方面自然过于理想化;它不仅来自于对表面的偶然事件的抽象(如被砖击中或从属查理定律),而且它视之为错误的东西从另外的观点看则有热力学的

必然性。“乐观的观点”对确实想长命百岁的人并没有提供慰藉。

尽管如此,在此意义上主张可朽是偶然的这一观点与主张我们似乎注定要死这一观点之间显然存在着经验上的争论。事实上,有一些极好的材料有利于证明后一说法很可能为真(例如可参阅威廉斯 1966(还有亚历山大 1987 等));它们给强调个体生存(对立于是基因组的传递)是自然选择所繁殖的事态的进化论出了一个难题。

我们的问题与此相仿;人们至少可以想象与如下问题有关系的事实,即真值是不是认知机制注定要记录的东西。例如,正如 S. 斯蒂克喜欢指出的那样,存在着这样的实例,它们似乎(乍一看)促使认知机制指向另外的某物。这些例子的逻辑值得很好地加以思考。

简单的例子来自生态学文献。许多动物在辨认它们的种群的捕食者方面是极端谨慎的。只要有略微类似于捕食者的刺激形态就将引起释放逃避反应(以致被鹰追捕的某些鸟,当你仅仅在它们面前抖动一件黑衣服时,它们也会表现出逃跑反应)(参阅廷伯根 1969)。看待这一点的一种方式假设,动物完全没有注意到“虚假的事物”;有许多关于有一只捕食者这种类型的错误信念是相当幸运的,只要(假设恰当性条件的满足)它还有那种类型的所有真实信念。在这种情况下,你可能争辩说,人们不可能假定,认知装置只用来释放真理。更不容置疑的是,人们不可能假定,认知装置倾向于有限制的无所不知。

然而,这些理由决不是决定性的,因为人们在信念和玩弄把戏之间有司空见惯的调包。你可以描述说,动物生来就准备宽容错

546 第五部分

误信念,但是你也可以描述说,它有时准备按它认为只具有很低概率的信念行动。不是“有一只捕食者,所以我最好躲避”,而是“那可能正是一只捕食者,我不能冒险”。无论如何,对于裁决目前有争议的主张来说,这些是不合适的事实。飞的反应和诸如此类的行动终归是必要的、相对狂热的事件。在这样的行为中,如此行动的目的是避免被吃掉,而且在那些事实成为绝对确实的之前,这种利益是不会最有效地赐于傻等的。从无所不知论的观点看,有趣的——确实是至关重要的——是仔细想一想:在可为动物得到的材料向好的方面变化时,人们对于动物的命题态度想说什么。

所以,借助思想实验,试想:我们能设法让恐慌的鸟镇静下来。不是一看到抖动的黑衣服就让它急忙离开,而是促使它以一种更无偏见的和谨慎的心态注视它的处境。我们可以说,“注意,它不过是一件哑的旧黑衣服,看见了吗?”(为了满足认识恰当性条件,鸟做了一切该做的事情:鸟从各种观点考察衣服:嗅它、尝它、听它、用别的方法思量它;鸟的嘴实际地和隐喻地在它上面啄擦。)然后鸟急忙离开。在这点上,我们往往认为,对于鸟的信念,我们肯定知道我们应该作出哪种描述。现在我们想说的是:鸟认为它确实是一个抖动的黑东西;即是说,鸟认为它是某种真实的东西。但如果我们假定:这个鸟逃走是因为它认为这个抖动的黑东西是或可能是鹰,那就错了。尤其是,对于鸟的逃跑反应我们错误地描述了它的诱发刺激。它不是那种反应所隶属的捕食者,它是抖动的黑东西。当然,仍存在着根据鸟的捕食者的出现这一事实对鸟的行为的解释:保守的进化战略通过使逃跑反应隶属于看上去像鹰的一切来说明避鹰行为的产生。从认识论上说,我们让鸟摆脱危

险处境,并把虚假事实归咎于“自然之母”。我们通过重译心理表征可以做到这一点:如把处在是一盒中的东西当作是(比如说)“它是抖动的黑东西所以走掉”而不是“这是一只鹰所以急忙离开”。

总而言之,一种宽容原则支配着我们的信念归因;它大约是:如此归因信念以至于由于有机体关于 P 的证据成为任意明晰的,因此有机体相信 P 的可能性任意密切地接近于一致(证据的明晰性主要是相关的认识恰当性条件被满足的程度问题)。这就是说,我们以坚持有限制的无所不知原则这样一种方式把信念归因于有机体。正如我们所看到的那样,这正是进入条件与真值条件的同一性所需要的东西。^⑨

但是,进入条件与真值条件的同一性是空洞的这一可能性又怎么样呢?这一论点可以用两种不同的方式加以表述。你可能会说,要么(1)关于进入条件与真值条件共外延的陈述意味着纯粹的约定,要么(2)这个事业的基础由于不能提供关于“认知系统”的非武断的定义而被削弱。我想考察这两种驳难。

要弄清什么使第一类反对者恼火,重要的是把握下述两方面之间的差异,一方面是关于心理表征的语义属性的论断的证据情境,另一方面是(例如)关于自然语言符号的语义属性的论断的情境。我们还是来考察蛙,这是最后一次了。

伪卡普拉尼卡确信:某一动作电位类型的进入条件是同种的咕噜声,进而还宣称那也是它的真值条件。这里的关键词是“进而”。显然所想要的东西是一种经验论断。这幅图画是:有一种事态,它是那动作电位的进入条件,还有一种事态,它是动作电位的真值条件,看哪!我发现它们是完全相同的事态。这里,反对者宣

称的东西是：当你至少在原则上能不依赖于对某物是不是进入条件的分辨而分辨出它是不是一种符号的真值条件时，那类理由才是成立的。然而在现在的情况下，我们不能作出这样的分辨。因此有了经验发现的外观——即通过发现进入条件是什么来发现真值条件是什么——这不过是假象而已。

要注意，这个问题并不是以这种方式由自然语言的真值理论所引发的。这是因为，在后一类情况下，存在着该理论宣称要重构的符号的真值条件的前提概念。即是说，对于诸如“正在下雨”是真的当且仅当正在下雨这样的事实，我们知道理论重构的前提条件。关于英语的真值理论所主张的至少是，它使所有这些直觉上已知的配对物成为正确的，而且那一论断在经验上是可检验的。试比较蛙的动作电位，我们所知的、它的语义属性，先于它的进入条件之发现，因此完全不是那个样子。^⑩

这听起来像是一种严厉的驳难，但实际上是没有意义的。诚然，心理表征的真值理论不是通过把它的预言与关于心理符号的真值条件的前提上可利用的直观加以比较而得到检验的。但我们还有另一组经验事实——它们同样是有用的。请记住：我们假定了这样一种心理学理论，即对于 M 的每一个值来说，这种理论适合于填充这样的图式： O 相信那个 S ，当且仅当（ S 是 M 的真值条件以及 O 具有与 M 的 R^* ）。当然，从这种理论的观点看， S 是一个自由的参数，因为一方面，我们设想的那种心理学只限于解释计算关系，另一方面，符号与其真值条件之间的关系不是计算的。结果，那种心理学所做的事情就是要定义 R^* ，因而已知有机体心理表征的真值条件的具体细目，便能预言它的信念。

现在,我的确认为,存在着关于 O 所相信的东西这一事实,其中的一些作为事实对于检验关于心理符号的真值条件的理论来说是有用的前提。^①因此在这里,你正好可根据关于心理表征的真值条件的断言(更一般地说,根据它们的真值条件与它们的进入条件是同一的断言)形成和检验经验预言。

1. 发现 M 的进入条件;称之为“E”。

2. 用关于 E 的详细说明代替在下述双条件句中出现的每一“S”:O 相信那个 S,当且仅当(S 是 M 的真值条件以及 O 具有与 M 的 R^*)。

3. 使 O 具有与 M 的 R^* (比如说从实验上)产生。

4. 发现双条件句的左边是不是真的。左边的真是赞成进入条件与真值条件同一的证据;左边的假则相反。每当我们对我们查明的进入条件的信任是值得重视的,我们对 R^* 的功能说明得到了充分证实时,如果我们大体上获得了真的左边,那么我们对于把进入条件与真值条件同一起来的意义理论就有可靠的经验证据。

最后,我们要考察的责难是:引进“认知系统”(进而引进“认知系统的正常功能活动”、“进入条件”)是以未经证明的假定为根据的。而我认为,这实际上有可能是真的。

请记住,我在“事实上的”认知系统和“简单的”认知系统之间作了目的论上的区分,前者包括让材料进入是一盒或让其从中出来的一切机制(如像记忆痕迹衰变那样的遗忘机制,假如记忆痕迹衰变是一种遗忘机制的话)。后者只包括像影响是一盒的内容的功能那样的机制。现在似乎有足够的理由假定,存在着某种像压

550 第五部分

抑那样的心理现象。显然,如果压抑存在,它便会影响存在于一盒中的东西。压抑致使你遗忘,从而(根据 RTM)部分地决定你具有与哪一些心理表征的 R^* 。而且我们可以假定,导致你遗忘正好是压抑系统的功能。也许经受压抑的有机体(当然有选择性地)享有竞争的优势。你对它想到的越多,它好像就越合理。

但是现在,我们显然不想根据包括压抑机制在内的认知系统的正常功能作用来定义“进入条件”。因为那个认知系统并没有促成无所不知,这绝非偶然。据假定,对于某些真值来说,不把表达它们的 M_s 放入一盒中实际上是预定的。然而,正如你将记住的那样,正是下述假定,即无所不知的缺乏是偶然的,成了真值条件与进入条件的同一所依赖的基础。对压抑的选择,对于无所不知假定来说就像对致死基因的选择对于不朽假定一样,是非常令人讨厌的。

当然,一条出路就是简单地约定:压抑机制不属于认知系统。然而,反对的人对此会正确地回答说,对“真值条件”的分析在这里是循环的。一方面,某物是不是进入条件依赖于人们准备将其当作是认知机制的东西,另一方面,作为认知机制的东西又依赖于人们准备将其当作进入条件的东西。我认为,这种责难的根据也许是对的:很可能没有办法(在预期的意义上)分辨认知机制,除非它们是这样的机制,即当它们正常行使功能作用时,它们确定了能与真值条件同一的进入条件。即是说,除了列举,没有别的办法。

但是列举为什么不行呢?该方案毕竟不是要分析“ S 是 M 的真值条件”。如果该方案是这样,那么我们一开始就有相当多的归纳根据将它放弃。这些分析远没有触及到堂奥;令人感兴趣的观

念未得到分析,这无疑正是现代哲学的教训。毋宁说,目的就是要说明,心理符号有其真值条件是怎样可能的,即使由于没有使用它们的行动者,因而它们的语义性不可能追溯到任何人的命题态度的意向性。这正是我们所论述的。心理表征可能有真值条件,因为——我们也许会假设——具有信念的每一有机体也具有在某些类型的强制之下适合把标记表征放入是一盒中的机制。是哪些强制呢?唔,是那些被观察到的,例如,这样的机制如……(这里是大量的列举)。那些强制如此令人感兴趣的是什么呢?正是这样的事实,即满足那些强制的事态能与心理表征的真值条件相同一,在这方面,假如有对 R^* 的详细说明,结果便导致从经验上把内容合理地归因于命题态度。如果那是循环论证,那么它恰恰是科学理论建构中的正常循环论证。

最后,我打算从上述讨论引出一些启示来。自从康德以来,哲学家们总是对下述观念抱有兴趣,它认为关于意义的问题和关于信念的固恋的问题之间存在着内在联系。证实主义是这种观点的一个分支,正如“程序语义学”和其他大量的理论变型一样。尽管如此,根据现今的解释,最终将证明有某种符合于这种直觉的东西。因为如果我所说的东西是对的,那么把真值条件归因于心理符号与影响信念固恋的装置的功能就是相关的。尤其是,你的信念具有它所具有的语义内容这一事实原来是由你在你通常获得信念的环境中获得了信念这一事实所决定的;或者说假如认知装置有正常的功能作用,那么你在某种环境下至少会拥有它。

在某种意义上,你不可能有更强形式的证实主义。因为证实主义者所说的毕竟是,符号的意义是从它在其下被断定的条件中

552 第五部分

而来的一种构造;这也就是我要说的东西(仅仅把“心理表征”当作“符号”,把“真值条件”当作“意义”,把“处于一盒中”当作“被断定”)。

但是关于我的这种证实主义——与我听说过的任何其他类型相反——我觉得令人惬意的地方就在于此。传统证实主义被假定具有各种奇异的和极妙的结果:诸如,形而上学的概念没有内容,不存在像电子那样的东西,桌椅由感觉材料所制成,逻辑真理是分析的,上帝知道别的一切。你得到这些惊人的答案的方式首先是把意义与断言条件等同起来,然后对能算作断言条件的东西施加非常严厉的(就我所知,是完全无法证明的)限制。例如,这些断言条件必须能够根据可观察物或感觉材料或中型客体的属性……等等来加以详细说明。

而根据现在的说明,与断言条件(即进入条件)对应的东西,正如我所论述的,可以直接是任何东西。能满足一种进入条件的东西是认知装置对之能有选择地作出反应的任何事态;那种事态可以是非常大的,或非常小的,或非常遥远的,或不可能完全间接地发现例外的,甚或完全不可能发现任何东西,只要你没能找到关于它的理由,仅只是满足相关的恰当性条件的事实上的不可能性就行了。唔,是呀,即使事态不是由感觉材料所构成的,这事态也能满足一种进入条件。

说教到此为止。作一个证实主义者完全没错,只要当你这样做时,绝对坚持真理的符合论就行了。就我所知,当你这样做时,你得到的是这样一种证实主义,它全然没有令人感兴趣的结论(除了它在使 RTM 合法化方面的有用性之外)。这种哲学——借用

一证实主义者同伴的话说——实际上使一切依然如故。这当然是一个相当有力的、对这种哲学很有利的论证。

注 释

① 我想在作为心理状态的属性的意向性和同样是符号属性的内涵性(大致上相当于关于性)之间作出区分。这两个术语必须与专门指称(一方面)意欲某物的心理状态和(另一方面)某些句子语境的一群逻辑特性的同音异义词区分开来。

② 我认为理所当然的是,如果一种理论没办法抓住能用英语表达的语义差别,那么它也不足以抓住能由心理符号表达的语义差别。这来自如下观点:语言符号的语义性(通过说者/听者的交流意向)源自心理符号的语义性。

③ 我这是把一个符号和决定其真值的一切东西之间的关系从范例上当作符号与世界的关系;对语义学的真值条件的探讨以真理的符合论为前提。因为这个理由以及其他将变得明显的理由,我对运用这种装置来解释作为表达逻辑和数学的必然真理的那些符号的语义性深表怀疑。一方面,对于这种符号来说,作为符合的真理会意味着什么是不清楚的,另一方面,将共外延性与同义语加以区分的问题,当这些共外延是必然的时候,似乎是特别棘手的。既然对必然真理,我的确不知道该说什么,因此我建议把当前的讨论限制在对偶然有其真值的符号的讨论上。

④ 当然, R^* 完全不同于意义理论所解释的符号与世界的那种关系, R^* 是一种有机体和一种心理表征之间的关系,基于这种关系,前者便相信后者所表达的东西。相比之下,意义理论描写的是(如)一种有机体、一种事态和一种心理表征之间的关系,由于这种关系该表征有它所具有的语义特性。

⑤ “O”是代表有机体的名称;“S”的每一个显现代表一个英语句子。

⑥ 毫不奇怪,在是-盒的M的因果条件中,具体说明哪一种可算作它的进入条件这一问题,直接与下述问题相关联,即具体说明在导致知觉信念的形成的因果链中哪一个链可算作知觉的对象。但愿下面的初步讨论能借助对知觉文献的整合而得到极大的推进。

⑦ 要注意,这第二个条件成了进入条件的概念的组成部分。S是M的进入条件,仅当S使M的标记处在是-盒中,进而使那有机体处在与M所

554 第五部分

表达的一切东西的信念关系之中。由此可以说, S 是进入条件, 仅当对于在其是 - 盒中有 M 的有机体来说, S 最终引起了 R^* 所衍推的一切行为结果。

⑧ 你说, “啊哈, 如果认知系统事实上没有对一种类型的标记行使功能作用, 那么选择压力怎么可能决定: 如果满足了恰当性条件, 它会怎样行使功能?” 回答是: 一旦选择塑造了认知的 (或任何别的) 机制, 就有许多不定的反事实, 它们由于该机制的结构将是真实的。假设, 选择压力有利于能做加法的有机体, 那么它们尤其有利于能加算 27 和 54 的有机体。即使没有有机体曾加算 27 和 54, 这仍有可能是真的, 因此做算术题这样的事例在形成它的机制的起因中不起作用。如下假定是对进化论的严重误解, 即假定, 根据选择优势对能力的解释预先假定了: 那种特别的能力在该有机体的进化史中时常被运用。

⑨ 当然, 一些哲学家提出了更强形式的宽容原则; 但我怀疑这样的原则能得到证实。我也不准备承认下述流行的推论, 即从“宽容在内容归因于信念中起作用”到“信念归因是启发式的、工具性的或在其他方面非虚构的”的推论。我的观点是, 运用宽容的结果是: 我们的 (内隐的) 命题态度心理学使我们与关于认知机制的 (同样是内隐的) 目的论假定紧紧连在一起了; 即是说它们预定要固恋真实的信念。我们能 (且的确可) 通过适当重新描述威胁性的事实 (见以下的例子) 来辩护目的论。但是只要这些理论是精心构建且极有价值的, 那么这种战略就是对它们的独特的展开。就此而言, 心理学似乎与物理学一样有了同样的根基。

⑩ 事实上, 这一考察在一个重要方面可一分为二。有诱惑使人设想, 人们可能驳斥本文的论点, 方法是强调这种可能性, 即表达了 P 的惯用语应存在于 - 盒中, 即使那有机体的认知系统正常发挥功能作用, 而且其证据必定是非 P。但是主张有这种可能性, 除了对心理表征表达 P 意味什么作出了某种替换的、非认识的说明之外, 不过是以假定为根据的争辩。尤其是, 我们不赞成那些诉诸于我们关于心理表征意味着什么的直觉的主张, 因为我们根本没有这种直觉。在辩护 (或抨击) 心理表征的意义理论时的论辩情境相当不同于我们论证自然语言的意义理论时所习惯的情境。

⑪ 我认为, 存在着这样的可利用的事实并不是自明的。某人可能会说, 不存在命题态度或者不存在发现有机体有哪一种命题态度的任何可能性。

然而,关于心理表征的整个争论首先是在估价 RTM 的背景中产生的,而 RTM 明显是关于有机体的信念的一种理论。如果你认为没有像信念那样的东西,或关于信念的事实原则上是不可得到的,那么你就不可能被怎样拯救 RTM 所困扰;相反,如果你对拯救 RTM 感兴趣,那么你可能愿意承认,存在着关于信念的事实,这正是 RTM 应予关心的。

参考文献

- R. 亚历山大(1987):《道德系统的生物学》,格吕特尔的奥尔代恩。
- R. 卡普拉尼卡(1976):《蛙的神经生物学手册》,斯普林格-弗拉格。
- C. 切尔尼克(1986):《最低限度的合理性》,麻省理工学院出版社。
- D. 戴维森(1967):“真理与意义”,《综合》,17,304-323。
- D. 丹尼特(1975):“大脑写作与心灵阅读”,载 K. 冈德森编:《明尼苏达科学哲学研究》,第 7 卷:《语言、心灵和知识》,明尼苏达大学出版社。
- H. 菲尔德(1977):“逻辑、意义与概念作用”,《哲学杂志》,74,379-409。
- J.A. 福多(1980):“作为认知科学研究战略的方法论的唯我论”,《行为与脑科学》,3,63-73。
- J.A. 福多(1981):《表征》,麻省理工学院出版社。
- J.D. 福多(1977):《语义学:生成语法中的意义理论》,克罗威尔。
- J. 莱特文等(1959):“蛙眼告诉蛙脑什么”,无线电学会文献汇编,1940-1951。
- B. 洛尔(1981):《心灵与意义》,剑桥大学出版社。
- C. 麦金(1982):“内容的结构”,载 A. 伍德菲尔德编:《思想与对象》,牛津大学出版社。
- H. 普特南(1975):“‘意义’的意义”,载 K. 冈德森编:《明尼苏达科学哲学研究》,第 7 卷。
- A. 里夫斯(1974):“论真理与意义”,《心灵》,8,343-359。
- N. 廷伯根(1969):《本能研究》,牛津:克拉伦登。
- G. 威廉斯(1966):《适应与自然选择》,普林斯顿大学出版社。

表征与心灵理论*

莫顿·韦格曼

1. 知识表征的本质

进化为现存的有机体提供了自然智能保存、发展和利用所必需的遗传和神经结构。为了创造像自然智能那样的、能处理世界上的事态的人工智能,就必须找到表征这些事态的方法。因此知识的表征是人工智能的基本条件。

戴维斯、施罗伯和索洛维茨(1993)对知识表征的本质作了有说服力的分析。在下一节中,将对他们的分析作出描述,然后提出简要的评论。

2. 知识表征的概念

2.1 五个方面

戴维斯、施罗伯和索洛维茨在下述简要的说明中把知识表征

* 译自莫顿·韦格曼:《人类理智与认知科学》,普雷格出版公司,1966,第1章。本章每一部分,围绕有关问题,选录了反映诸家有代表性的观点的原文资料,因此是了解本领域进展的一较全面的、极有价值的“原著”荟萃。文中分节序号为译者所加。

区分为五个方面。

即使知识表征在 AI 中是一个中心的、在某种意义上是最熟悉的概念,但关于它的最根本的问题——它是什么?——则几乎看不到直接的回答。大量的论文已对表征的一种或另一种变体作了阐发,而另一些论文则对表征所应有的各种属性作了论证,还有一些论文强调的是对一般表征概念至关重要的属性。

在本文中,我们将返朴归真,开门见山地处理问题。我们相信,答案能基于表征所起的五个重要的、彼此有别的作用而最好地得到理解,其中每一个作用都对表征所应有的属性提出了不同的、甚至时常有冲突的要求。我们要论证的是:记住这五个作用将展现这样一幅有利的、广阔的前景,它将为某些长期聚讼纷纭的争论带来解决希望,并使该领域的探讨和实践充满活力。

什么是知识表征呢?我们认为:该概念最好根据它所起的五种不同作用而加以理解,其中每一个对目前的任务都至关重要:

第一,知识表征从根本上说是一种替代物,一种关于事物本身的代表,它常常通过思维而不是行动,即通过对世界的推论而不是采取对它的行动而让实际存在物决定结论。

第二,它是一组本体论承诺,即是说是对于下述问题的一种回答:我们用什么术语思维世界?

第三,它是一种关于由下述三种成分表达的理智推论的零碎的理论:(1)关于理智推理的表征的基本概念,(2)表征所支持的推理集合,(3)表征所推荐的推理集合。

第四,它是有实用效果的计算的媒介,即是说是思维得以完成的计算环境。对这种实用效果所作的一个贡献是由表征为着组织

558 第五部分

信息以促进作出被推荐的推理而提供的向导所奉献的。

第五,它是人类表达式的媒介,即是说是我们用以谈论世界上的事物的一种语言。

理解其作用,承认其多样性,有几个有用的步骤。其一,每一作用都离不开与一表征略微有别的某东西;因而每一作用都导致一种有趣的、不同的、我们想让表征拥有的属性之集合。

其二,我们相信:那些作用提供了一种构架(framework),它对描述各种各样的表征是有用的。我们设想:表征的根本心智定势可以通过理解它怎样估价每一种作用而加以把握,而且,这样做可揭示基本的相似性和差异性。

其三,我们相信:当五种作用被恰如其分地加以考虑时,那么先前关于表征的某些分歧看法就会迎刃而解。我们可通过重新观察、解剖关于构架和逻辑的早期论证来证明这一点。

最后,我们相信:以这种方法看待表征,对于研究和实践都很重要。对于研究来说,该观点对该领域中的具有根本重要性的问题提供了直接的回答。它还思考表征中什么重要这一问题拓展了视野,并断言表征研究的一个重要方面——把握和表达自然世界的丰富性——所受到的注意远远不够。我们相信:这一观点也能改进实践,如让实验家想到灵感,而这正是各种表征的力量的重要源泉。(戴维斯、施罗伯和索洛维茨,1993,pp. 17-18)

2.2 替代方面

一个有理智的系统一定具有知识,而知识又一定适于被编码。被编码的知识就是表征。戴维斯、施罗伯和索洛维茨以下述术语

讨论了知识表征的替代本质。

作用之一:知识表征是一种替代物。任何有智能的、想推论其世界的实在都会面临一个重要的、不可避免的事实:推论是一个内在地进行的过程,而它想推论的大多数事物则只存在于外部。例如,一从事策划自行车装配的程序(或人)就不得不推论这样一些实在,诸如轮子、环、手柄和链轮,而这些事物只存在于外部世界。

这种不可避免的二分法就是表征的一种根本的原理和作用:它在推理器内部作为替代物,作为存在于世界中的事物的代表而起作用。对表征以及用表征所进行的运算代表着对真实事物的运算,即是说代表着与世界的直接相互作用。根据此观点,当我们不能或还没有想到在世界中采取那种行动时,推理本身部分就成了那种行动的替代物。

把表征当作替代物自然会导致两个重要的问题。第一就是它所意指的同一性:它是一个关于什么的替代物?在替代物与它在世界中意想的所指之间一定有某种被指定了的符合形式;符合就是关于表征的语义学。

第二个问题是逼真度:替代物怎样接近于真实的事物?它把握到了、进而详述了本原的什么属性?它略去了哪些?一般来说,完全的逼真度在实践上和原则上都是不可能的。之所以说原则上不可能,是因为事物自身之外的任何事物都必然不同于事物自身(不说别的,仅就位置来说是如此)。换言之,对象的唯一完全准确的表征就是对象自身。所有别的表征都不准确;它们不可避免地包含有简化的假定,可能还有人为的东西。

两种简要的阐释就会扩展这种关于表征即替代物的观点。第

560 第五部分

一,它似乎一无例外地适用于齿轮之类的可触的对象和不可触的对象:表征可作为行动、过程、信念、因果性和范畴之类的抽象概念的替代物而起作用,允许它们在实在内被描述,因而它能对它们作出推理。第二,形式对象自然能以完全的逼真度存在于机器内部:例如数学实在就能确切地、明确地被把握,因为它们是形式对象。由于几乎是全部推理任务都会碰到这样的需要,即处理自然对象(即是说,在真实世界中碰到的那些东西)以及形式对象,因此不完全的替代物在实用中就是不可避免的。

从不完全替代物的不可避免性中可引出两个重要结论。其一是,在描述自然世界时,我们一定免不了撒谎,至少是通过省略。在最低限度上,我们一定省略了自然世界的无限复杂性中的某些方面;此外,我们的某些描述可能引入了并不存在于世界中的人工造物。

第二个也是更重要的结论就是:所有关于自然世界的、有足够广泛基础的推理最终一定会到达这样的结论,即不正确的、独立于所用推理过程和所用表征的结论。完善的推理也不能拯救我们:如果世界模型不知什么缘故有问题(且必然如此),那么某些结论就会不正确,不管它是怎样被仔细抽绎出来的。更好的表征也救不了我们:所有的表征都是不完美的,而任何不完善性都能成为错误的源泉。

错误的意义自然是可变的;而且,选择一个好的表征的技艺主要就在于:为了眼前特定的任务找到一种可把错误减少到最低限度(甚或排除)的方法。但是替代物的不可避免的不完善性意味着,我们至少能为以任何方式推论自然世界的任何实在提供一种

保证:如果它的推理很长、很广泛,那么它就肯定出差错。

因此,哪怕是只作出可靠的结论也免不了使推理犯错误;它能肯定的是:推理并非错误之源。假如有广泛基础的推理不可避免出错,那么从可靠的推理转向别的推理模式因而就不是从完全的准确性过渡到错误,而是一个使错误的更多的源泉的可能性与它可能提供的利益(例如有效性)实现平衡的问题。

我们不是说:不可靠的推理应当随便予以接受,而是主张:假如错误不可避免,甚至可靠的推理也不例外,那么从效用方面估价运用可靠和不可靠推理方法所伴随的相关的代价与利益就是很有意义的。(戴维斯、施罗伯和索洛维茨,1993,pp. 18-19)

2.3 本体论方面

对实在的表征不只一种方法。不过,可以肯定,在凸现某些特征时,特定表征的选择就会把实在的其余特征放在次要位置。戴维斯等人用下述术语讨论了这个本体论承诺或结论。

2.3.1 作用之二:知识表征是一组本体论承诺。像我们所论证的那样,如果所有表征都是不完全地近似于实在,每种近似既注意到了某些东西,又忽略了别的东西,那么在选择任何表征时,我们必定是以同一的行动方式对于下述问题作出一组决定,即在世界中怎样看,看到了什么。即是说,选择表征意味着作出一组本体论承诺。

聚焦作用是表征所提供的东西中的必要部分,因为自然世界的复杂性是不可抗拒的。在决定关注世界上的什么、舍弃什么时,我们(以及我们的推理机器)需要向导。

562 第五部分

试图为不同任务领域建立精致本体论的工作有一漫长的历史,包括早期关于液体的本体论工作(海斯,1978),广泛用于表征电流回路的集总元件模型(例如,戴维斯与施罗伯,1983)以及关于时间信念甚至程序本身的本体论。这些本体论的每一个都提供了观察世界某一部分的一种方法。

这些本体论自然是用各种不同的语言和概念(如逻辑、表处理语言)写出来的;其基本信息不是这种语言的形式而是内容,即是作为思维世界的方法而提供出来的概念集合。

我们通过选择一种或另一种本体论而作出的承诺对于目前的任务可能产生相当不同的观点……。基于规则(如 Mycin^{*})所考虑的(医学的)诊断从根本上不同于基于构架(如 Internist^{**})所考虑的同样的诊断。在这里,Myacin 把医药世界当作是由连接症状与疾病的经验联系所构成的,而 Internist 看到的是典型症状,在特定的、典型的疾病中,它们应匹配于眼前的案例。

2.3.2 承诺开始于最初的选择。上述 Internist 例子还表明,即使是在众所周知的表征技术的层面,也存在着重要的、不可回避的本体论承诺。逻辑、规则、构架等包含着关于各种对世界至关重要的事物的观点。例如,逻辑就包含了这样的(相当微妙的)承诺,即根据个体的实在及其相互关系去看待世界。基于规则的系统便根据属性-对象-价值的三合一以及把它们联系起来的似乎合理的推理规则来观察世界,而构架让我们根据典型的对象进行思维。

* 指一种医学上的咨询系统。

** 原意即“内科医学”,指的是医学上的诊断系统。

这一来,这些表征技术的每一个对于值得引起注意的东西都提出了自己的观点,反过来,每一个又设想:任何难以用这些术语观察的东西都可以置之不理。自然,这种主张并无正确的保障,因为任何被忽略的东西后来有可能证明是有关系的。但是这一工作原则上是没有前途的——每一表征都有对世界的某物的忽略;因此我们能做的最好是从好的猜想出发。现存的表征技术对于什么应予注意、什么应置之不理提供了一系列的猜想。因此对它们作出选择就涉及到一定程度的本体论承诺:这种选择对我们关于那个任务的知觉与探讨以及对我们关于被模拟的世界的知觉都有重要的影响。

2.3.3 承诺之分层累积。于是,表征的本体论承诺就从表征技术的层面开始,并由此而累积。随着我们对技术的运用,承诺的随后的层次便得以形成。试以 Internist 中似构架的结构之运用为例。在最基本的层次,根据构架来看待诊断的决策就提醒人们依据典型症状、欠缺和分类学上的等级来进行思考。但是典型症状是属于什么的,分类学应怎样组织呢?

对该系统的较早的描述(波朴尔,1982)说明了这些问题在目前的任务中应怎样予以回答,并提供了第二级承诺:

存在于 Internist 系统中的知识基础是由两种基本的元素构成的,即疾病事实和表现……(它)还包含了一个……疾病范畴的等级,最初是围绕着器官系统的概念而组织的,在其最高层次,就有“肝病”、“肾病”等等之类的范畴。(pp. 136-137)

因此,典型症状旨在把握典型的疾病(例如一种疾病的经典的

564 第五部分

案例),它们将在以组织器官为线索的分类学中得到组合。这组选择是可感的、直观的,但很显然,它不是把构架应用于那个任务的唯一的方法。因此,它是本体论承诺的另一层级。

在第三(在此案例中,是最后的)层级上,这组选择例示了:哪些疾病被包括在内,它们会出现在该等级系统的哪些分支中?出现在这个层面上的本体论问题可能是根本性的。

对于所有表征技术,必须作出相同类型的决策,因为它们的一个对于怎样观察世界只提供了一个初级的猜想:它们提供了一种观察方法,而并未指明怎样例示这种观点。构架显示典型症状和分类学,但并未告诉我们:哪些事物应选作典型症状,规则提示的是用合理推理来思维,但并未告诉我们:哪些合理的推理会出现。同样,逻辑能教我们依据个体和关系去观察世界,而没有具体说明所用的是哪些个体与关系。因此,对一种特定世界观的承诺是从选择表征技术开始的,并随着随后的关于怎样用这些术语观察世界的选择之作出而得到累积。(戴维斯、施罗伯和索洛维茨,1993,pp. 19-21)

2.4 推理的方面

智能推理系统的设计者选择了反映基础性的推理理论的知识表征样式。源于逻辑科学的理论主张,知识表征是作为逻辑符号而投射的,而由神经科学激发的理论则举荐表现为节点和链路网络形式的知识表征。

2.4.1 作用之三:知识表征是一种关于智能推理的零碎的理论。知识表征的第三个作用就是作为关于智能推理的一种零碎的

理论。这种作用之出现,是因为表征的原始概念一般是由指导人们怎样理智地推论的洞见或某种关于理智的推理意味着什么的信念所激发的。

该理论在两种不同的意义上是零碎的,(1)典型地说,表征只吸收了驱使它的洞见或信念部分,(2)这种洞见或信念反过来又只是智能推理的复杂的、多面性的现象的一部分。

2.4.2 什么是智能推理? 什么是智能推理的必要的、限定性的属性? 作为关于智能推理本质的一个学科、一种观点,人工智能相对年轻的一个结果就是常常借鉴别的领域的工作。五个领域——数理逻辑、心理学、生物学、统计学和经济学——为相互有别的关于什么东西构成了智能推理的五个概念提供了启发(表1)。

数理逻辑	心理学	生物学	统计学	经济学
亚里士多德				
笛卡尔				
波尔	詹姆斯		拉普拉斯	边沁 帕累托
弗雷格			伯努利	弗里德曼
丘奇	纽厄尔	莱特文	卡内曼	西蒙
图林	西蒙	麦卡洛克		雷发
		皮茨		
戴维斯		霍伯尔		
		韦塞尔		
普特南				
鲁滨逊				
程序设计 语 言	SDAR KBS, 构架	联结 主义	因果 网络	合理性 动因

表1 关于智能推理和它们的理智渊源的观点

566 第五部分

几个有用的结论来自于下述理解,即对每种传统视之为根本问题的不同观点的理解。首先它证明:对这些传统的任何现代产物——即上述图表下面所表明的任何表征技术——的选择意味着只是选择表征。以同样的方式,我们正在对关于智能推理的基本本质的概念作选择。

第二,这些概念在一些重要方面彼此有别,即在关于我们试图把握的现象概念中存在着根本的差异。进而不同的概念又意味着:在努力创造智能程序的各种探索活动的特征和目的方面,存在着根深蒂固的差异。简言之,关于智能推理的本质的不同概念导致了不同的目的、不同的成功定义和不同的人工造物。

最后,这些差别几乎没有被明确表达出来。而缺乏明确表达进而又导致了这样的论证,即当我们相信,真正的结果是关于智能的根本本质的不同概念时,这些论证可以根据表征选择(如用初级谓词演算进行的可靠推理与因基于构架系统而产生的难以描述的推理)之类的结果而陈述。理解不同的观点有助于对那些结果作出适当的分析和分类。(戴维斯、施罗伯和索洛维茨,1993,pp.21-24)

2.5 计算方面

就把推理当作计算而言,特定知识表征的组织属性就可能导致最大的计算有效性。

作用之四:知识表征是有效计算的媒介。从纯机械论观点来看,机器中(也许在人内部)的推理是一种计算过程。简言之,我们使用一个表征,一定是以之进行计算。因此,关于计算有效性的问题必然是表征概念的核心。

这一事实早就为人们认识到了,至少毫无疑义地为表征设计者认识到了:与他们对一组被推荐的推理的详述一道,表征常常为怎样以促进这些推理形成的方式组织信息提供了一系列的观念。例如原构架概念的一个必要的部分就是关于这类方法的,就像许多关于构架的论文所阐明的那样(明斯基,1974,1975):

一构架……(表征)一遥远的位置,如处在某起居室中,或来到一小孩的生日舞会上。

隶属于每一构架的是多种信息。该信息的一些是关于怎样使用那构架的。一些是关于人们预期下一步要发生的东西的。一些是关于那期盼没实现时该怎么办。

在那些构架中,关于触发器和程序上的附件的概念并不是关于程序写了什么陈述(在这里,该理论是相当模糊的)。因为它是一种对于怎样组织信息的有用方法的描述,如(把前面引文加以阐释)让关于怎样用那个构造、预期未实现时该做什么的信息隶属于每一构架。同样,将那构架加以组织使之进入分类学等级系统,既启发了分类学的推理,又促进了它的实施(就像在结构遗传网络中那样)。

别的表征也提供了相同的向导。传统的语义学网状系统通过提供一系列适当的链这样简单的权宜之计来推进双向传播,而基于规则的系统则通过提供从目的到其结论与之相符的规则(回向连链)和从事实到其大前提与之相符的规则(上向连链)的索引,而推进合理的推理。(戴维斯、施罗伯和索洛维茨,1993, pp. 26 - 27)。

2.6 通信方面

知识表征的最终目的就是使与计算机世界的信息沟通以及与人世界的信息沟通成为可能。而使这些通信得以进行的知识表征语言则应受到优先的关注。

作用之五:知识表征是人类表达式的媒介。最后,知识表征也是我们用以表述世界上的事物的手段,它们还是表达式和通信的媒介,我们正是通过它,把世界上的事件告诉给机器(也许是相互的)。只要我们需要向机器谈论世界,只要我们是通过创造和传递表征而予以实现的,那么表征的这个作用就是不可避免的。因而知识表征的第五个作用就是对我们有用的、作为表达式和通信的媒介。

反过来,这种作用又提出了两类重要问题。一是众所周知的:表征怎样恰到好处地作为表达式的媒介而发挥功能作用?它一般是怎样的?明确到什么程度?它提供的表达式充分吗?等等。

平常很少被讨论的一个重要问题就是:它作为通信的媒介怎样恰到好处地起作用?即是说,我们以这种语言谈论和思维容易到什么程度?什么样的事物用这种语言谈论容易一些,什么样的事物难些以致实际上不可能?

请注意这些问题所涉及的形式:它是多么容易?而不是我们能行吗?这种语言是我们必须使用的语言;因此原则上可能的东西就是有用的,但不是足够的;真正的问题是具有实用主义功利性的问题。如果表征使事物可能,而不是使之容易,那么作为真正的使用者,我们大概就绝不会知道:怎样运用它,或者说它是不是真

的不能表述我们想说的某些事情。表征就是我们以之进行沟通的语言;因此,我们不用太大的努力就一定能够讲说它(戴维斯、施罗伯和索洛维茨,1993,p. 27)。

2.7 评论

知识表征概念由替代物、本体论、推理、计算和通信五方面构成。这些方面是按哲学方面(替代物、本体论)、心理学方面(关于智能推理的理论)和实践方面(计算效能、通信促发性)的顺序排列的。在设计计算机程序的不同时期,不同的方面会受到不同的重视。对研究和发展的强调可能要求回到知识表征的哲学和心理学方面,并由此而证明它们对于创造性事业的价值。

3. 表征与智能

3.1 数学与表征

科学知识常以数学等式的形式被表征。而等式是关于物理世界的事件的符号表征。因为物理科学是用符号数学表征的,因此认知科学包括人、动物和人工智能,为了其发展也要求建构符号数学表征。自然系统和人工系统的认知性数学代码使智能得以适应于环境。

3.2 关于表征和智能的一般理论

智能系统如人或动物或人工系统的行为是表征系统的功能。

570 第五部分

这个理论性命题可用符号加以表达： $B=f(R)$ 。符号表达式 $f(R)$ 意思是：一运算组合是用表征系统完成的。范例有数学表征系统（例如，不同等式的集合）、逻辑表征系统（谓词演算、乘方运算的 INRC 群）、电磁表征系统（如电子回路图）和人工智能表征系统（程序语言、语义学网络、树状图解、组合及相互套入范畴、构架、文稿编辑程序）。

在下一节，加利斯特(1990b)提出了关于动物(包括人类)认知的理论，布鲁克斯(1991)阐发了他关于可动机器人的理论。前者的观点与用符号表示的理论—— $B=f(R)$ ——是一致的，并包含于其中。而后的立场则是对立的，因为他主张，他的可动机器人没有表征系统但可表现出智能行为。他关于“无表征的智能”的观点与他的智能可动机器人依赖于有穷状态机似乎是有冲突的，因为有穷状态机既离不开程序语言，又离不开电子的声纳的和光学的电路图。

3.3 表征与动物智能

3.3.1 一般理论。加利斯特对动物认知中的数学表征本质提出了一种形式理论。心理学中的表征概念的意义以及它的形式特征在下面得到了阐述。

关于表征的概念，它们是什么，它们是否存在以及一个表征概念在动物和人类行为中应起什么作用，这些都是有争议的课题。我曾作过这样的论证(加利斯特,1989,1990a)：表征在心理学中应有同一的意义，就像它在数学中那样，在那里，它意指系统之间的起作用的同型性。两系统间的同型性(形式上的一致)使在一系统

中利用运算作出关于另一系统的结论成为可能,就像代数的运算被用来作出解析几何学中的几何学结论那样。

如此类推,心理学中的表征就是大脑或心灵内的加工(依赖于分析水平)与这些加工使动物行为所适应的环境方面之间的起作用的同型性。更形式化地说,表征存在,如果:

(1)存在着从外部实在或事件(短暂的间隔、众多的集合、食物出现率、模型的形状、食物的化学特征、猴群内母系家族的数量等)到用作这些实在之代表的心理或神经变项的映射。

(2)在涉及到这些神经或心理变项的关系的、组合的运算与涉及到它们所表示的事物的关系及组合加工之间,存在着一种形式上的一致。例如,大脑对表示两个不同的被观察到的食物出现率的神经信号的加工与在那两个加工组成的世界中的运算之间的一致。如果两种加工同时运算时,食物出现率是每一孤立运算过程所产生的比率的总和,如果大脑在计算同时发生的运算所期望的比率时附带地把两种表征信号的比率结合起来了,那么在同时起作用的过程(被表征系统中的组合过程)与对神经系统内的有关变项的附带组合(表征系统内的组合过程)之间就存在一种形式上的一致。

(3)映射过程与组合过程十分协调,以至于表征系统中的组合过程就产生关于被表征系统中的事件与关系的有效预期。大脑组件中的运算结果可预言或预期外在于那组件的系统中的过程与关系的结果或状态。

(4)神经的或心理的运算一般地产生关于外部事件及关系的正确预言的能力可由产生适应于这些事件和关系的行为的机制来

572 第五部分

说明。换言之,心灵或大脑过程与世界过程的形式一致,有利于生物学目的。(加利斯特 1990b, pp. 1-2)

在下面,加利斯特把他关于表征的形式理论与关于心理测量的理论关联起来了。

测量理论最初由史蒂文斯(1946)所阐发,稍后又得到克兰茨、卢斯、萨普斯和特夫斯基(1971)等的发展,知道这一理论的人将认识到表征之运用与测量理论之运用之间的相似,在这里,根本的任务就是为被测量的心理变量(如高声)与那个数域的某些或全部之间存在的同型性建立必要的、充分的经验条件。同型性依赖于找到适当的测量步骤(标准),它从心理变量映射到它的数量代表,此外,同型性还依赖于对心理变量的(如对“音高于”的判断)组合运算与“ $>$ ”、“ $+$ ”、“ $=$ ”之类的数量运算之间所存在的形式上的一致。这种一致允许人们根据对心理变量适当地被决定的数量表示式的数学运算,作出关于心理变量的有效推论。

表征具有不同程度的丰富性和力量,就像测量的标准所是的那样,这一点是表征的定义所固有的。表征的丰富性或力量是由可用来作出关于被表征系统的结论的各种组合运算所决定的。在极端贫乏的表征如名义上的标准中,有效地被用到的只有一种运算。就名义的标准而言,这就是恒等算子($=$)。当人们用到一个名义的标准(如关于运动衫的号码)时,被表征的实在(运动员)就是同一的,当且仅当他们有相同的代表(相同的号码)。在更有力量的表征中,就可用到相互关联的组合运算。如果人们得到了关于某些心理变量的比率刻度表征,如被知觉到的重量,那么所有数域都可能被用到(\rightarrow , $+$, $*$),以便作出关于被从数量上表征的变

量的结论,就像一个人得出结论说:一定的重量被感觉到是另外两个重量合在一起的重量的一倍以上。

表征系统中的组合运算用不着同型于对那个数域的算术运算。例如组合运算可用之于树状结构或相互交叉的范畴,而不是用之于数量。(加利斯特,1990b, pp. 2-3)

在下面,加利斯特在表征与表示式(representive)之间作出了区分。

在表征与表示式之间作出区分是很重要的。一个表示式就是表征系统中的一个符号,它表示的是实在或变量的值或被表征的系统中的变量之关系。当存在着从世界变量到心理或神经变量的稳定的映射时,心理或神经变量便存在着。粗略地说,研究感觉过程就是研究低级的物理变量向它们的低层次的心理物理或神经表示式的映射——例如研究从光谱组合到换能器通道中的活动(心理物理的分析层次)或到锥体活动(神经层次)的映射。对知觉的研究就是研究,高级刺激变量——如对象形状——如何映射到高层次心理或神经表示式,例如到思维的皮层领域中为形状编码的活动中心。

神经表示式之存在并未确立神经表征之存在。例如表明某皮层区域的活动以有序的、具有多样集合的方式变化并不能肯定:大脑表征了那种多样性。表征是系统间的一种关系,而不是变量之间的关系。被表征系统和表征系统都是由变量及其关系以及涉及到变量的加工或运算所构成的。表征要存在,环境变量的神经或者心理表示式就必须进入能产生关于被表征的变量的有效推理的组合性的神经或心理加工之中。

形成关于心理或神经表征的概念取决于从世界变量到心理或神经变量的映射之出现,取决于在两个可能极受限制的领域的运算之间的形式上的一致。这意思是说:该概念不适用于,至少不严格地适用于这样的情况,在那里,人们对被认为是从心理学上被表征的实在方面不可能给予适当的形式化描述。例如,对光谱合成——颜色视觉中的物理变量——不可能有符合于我们对色彩知觉(光环)的形式描述的描述。被知觉的色彩之间的关系在产生它们的刺激中找不到配对物。条件等色* 配对物是产生同一的心理表示式的不同光谱合成的光。条件等色配对物无处不在;彩色电视机和彩色打印机依赖于它们,因为这些媒体绝不能再生那表层所呈现的相应的光谱;它们习以为常地“戏弄眼睛”。因此即使是同一的算子也不能有效地应用于我们的色彩知觉。颜色在心灵中,而不在世界中,反过来说,存在于世界中的光谱合成是不能在心灵中被表征的。

不过,我想论证说:表征的数学定义从概念上说是有用的,即使被用于下述许多领域也是如此,在那里我们不能从形成上描述:它关涉到我们相信其为某种心灵模式所把握的世界的什么。对心理模式(心理学理论)的详细阐述,对被摹拟的东西的形式描述是相辅相成的事情,这对任何人都是显而易见的,只要他承认语言学的句法理论——对语言结构的形式化描述——与心理学中的语言产生和理解理论之间有相互作用(加利斯特,1990b, pp. 3-4)。

3.3.2 时间间隔。在下面,加利斯特把表征理论应用于时间

* 指光谱结构不同而看上去完全相同的两种颜色。

间隔,并把表征与由吉本和丘奇(1981)所做的关于动物行为的实验工作联系起来了。

人们怎样用下述理论说明动物的时间分辨,这是很难弄清楚的,根据该理论,动物的神经系统没有变量,它们的神经系统的值以一种系统的方式相关于被经历的时间间隔的持续。因而关于动物的时间分辨的材料是从时间间隔到它们的心理和神经表征的映射的证据。更重要的是,吉本和丘奇对于计时行为中的决策过程的细致的理论与实验关注已提供了令人信服的根据,因为它用行为资料证明:普通实验室中的鸟和哺乳动物在这里所辩护的意义上有关于时间间隔的表征。为了对世界的当下状态作出有效的推理,它们用关于时间间隔的表示式完成了同型于减法、划分、比较(\rightarrow)的运算。

例如,在布鲁纳所做的、为吉本和丘奇在论文中所考察的实验中,同他们的前几个实验一样(例如吉本和丘奇,1981),实验材料适合于一种模式,在那模式中,决策过程计算出了正消逝的间隔与被记起的标准之间的差异,并通过那个标准的评估,区分了这一差异,还对商是否超过了决策临界值作出了检验。符合这种决策规则的模式说明了这些材料及别的材料具有相当的明晰性,而别的模式则望尘莫及。除非某人对这样的实验发现提供同样成功的说明,这些实验不涉及到对作为时间间隔的变量的减去、划分和比较,否则这些发现就是证明动物具有数学意义上的时间间隔表征的根据。用于心理学变量的心理学运算系统是同型于时间系统的。(加利斯特,1990b, pp. 5-6)

3.3.3 认知映射。昆虫及别的动物要想准确地对准能提供食

576 第五部分

物的目标,掌握关于环境的航行几何学是必不可少的。关于这种航行几何学的本质,它的形式表征以及认知映射的存在、形式和功能,将在下面予以讨论。

几何学是对空间关系的形式描述。对被表征系统的这种形式描述之存在使直接应用这里所辩护的空间表征定义成为可能。说一昆虫表征了它的环境的空间结构或环境的某些方面就等于说存在着从这种结构要素到昆虫神经系统中的有关表示式的映射,就等于说昆虫的神经系统完成了对于这些表示式的运算,而这些运算同型于对它们所表征的空间构造的运算。蜜蜂毫无困难地学会了分辨二维模型这一事实似乎包含了这样的映射,即从这些模型到神经系统中的表示式的映射,并且至少利用了恒等算子。

动物是否有关于它的认知环境的映射这一问题也能在几何学框架内重新加以阐释(程,1986;程与加利斯特,1984)。许多昆虫与别的许多动物一样,能借助被记住的、目标与周围界标之间的几何关系而找到目标(有关评论可参阅加利斯特,1990a)。问题是,被表征的是什么样的几何关系,为了从最初的(直接被编码的)关系推论出别的关系,可能作出什么样的运算?

映射表征的是环境的外观。像通常所理解的那样,外观是由构成环境的面的相对测量位置而得到界定的。相对测量位置决定了由面所形成的角以及面的距离,面与面之间的距离,反之亦然。因此提出动物是否有映射这一问题的一种方法就是问:它是否表征了目标与周围界标之间的测量关系。如果它不是这样,那么它就没有表征环境的通常意义上的外观。不过,卡特赖特与科利特(1983)表明:蜜蜂记录了靠近食物源的界标的罗盘方位(一种角的

关系),而程等人(1987)则说明:如果界标离食物源有不同距离,但有相同的外观与明显的大小,那么蜜蜂更依赖于附近的界标,这也意味着它们也记录了在界标与食物源之间的距离。简言之,蜜蜂记录了目标与周围界标之间的测量关系,如果它们有关于环境和那环境内位置的映射,那么这就是它们必须做的事情。(加利斯特,1990b, pp. 9-10)。

计算从食物源到目标间的距离可以借助视差这样一种天文学家所熟悉的、也许内在于昆虫和别的动物的认知能力中的方法来完成。

界标离食物源的距离可以借视差加以估算,即根据对蜜蜂的特定的横向位移的视角之变化而加以估算(卡特赖特和科利特,1979)。视差在估算距离时的重要性可以解释蜜蜂或黄蜂在接近与离开食物源或巢穴时所采取的“之”字形飞行模式(维纳,1981)。“之”字形飞行模式在引起蝗虫扑向目标之前所作的针锋相对的注视活动中的作用可能是独一无二的(科利特,1978,华莱士,1959),在引起小沙鼠越过陷阱之前所作的垂直转动脑袋中的作用也是如此(埃拉德等,1984)。索贝尔(1989)利用人工诱导的视差变化说明:黄蜂减缓速度是它的跳前注视期间的目标视差运动的一种作用。把同一级别的减速与所跳的距离关联起来的千篇一律的功能作用常重新出现,以用来对准黄蜂在得到所需视差的注视期间横向运动所形成的距离。

依视差所作的距离估算在各种映射的建构中起着核心作用,因为在任何可动的、能检测其所通过的距离的动物中,这种机制适于对无论多大的距离作出量的测定。被估算的距离越大,借视差、

578 第五部分

以一定程度的准确性对之估算所需要的横向位移也就越大。映射的建构也就要求动物把距离和从不同的观察点所作出的角度估算结合起来。而要如此,运动的动物就必须记下它所经过路程的各点之间的测量关系。

有一基本的文献表明:从蚁到蝙蝠的各种动物记录它们当前的位置,利用的是航位推测法或路径整合,即是说,利用的是这样的神经运算,即它至少大致类似于相对于时间的速度整合(例如维纳等,1981)。航位推测法提供了对位移距离(或速度)的估算,这对远距离的视差估算是必不可少的,而且它还可以提供关于不同观察点的相对测量位置的信息,这是建构地心映射所必需的。加利斯特(1990a)指出:要形成地心测量映射,计算就必须把来自于推测模数的位置、航向信息与来自于知觉模数(就像用于根据视差计算距离的模数)的对象距离、角度的信息结合起来。

必须强调的是,卡特赖特与科利特(1983)以及程等(1987)的实验已表明:蜜蜂可记录界标与到达它们的距离的罗盘方位,这对记录目标与由界标所规定的外观之间的欧几里德关系来说,是必不可少的材料。还要强调的是:航位推测过程提供了关于动物的位置与奔向不同有利地点之间的地心关系的必要信息。因而,实验上已证明:处于它们之间的蜜蜂知觉和航位推测系统提供了建构地心方面的欧几里德认知映射所必需的材料。(加利斯特, pp, 10-14, 着重号为我所加)

3.3.4 范畴与概念。在下面,数学表征理论将应用于范畴或概念的定义之中。

在讨论范畴和概念时,清楚地记住从被表征实在到它的表示

式的映射与对这些表示式的运算之间的区别是有裨益的。一种范畴就是关于大量可区分的、在某种背景中(对于某组运算来说)被视之为等值的实在的一种表示式。数字是范畴的典型的例示:数字3表征的是那些别的极为不同的、其大数是3的一切集合。只有在那些集合大数是相关变量的上下关系中,它才能用之于旨在作出关于这些集合的推理的运算中。定义范畴有两种极为不同的方法。一是给予规定(内涵定义)或列举(外延定义),它能确认那范畴的表示式有效地指称的一切事物。例如数学3能指称下述所有集合,它们可以与为着检验3这一目的而构想的抽象指称集合一一对应地加以排列。这种定义3的方法强调的是,从大数到它们的表示式的映射过程。它说明的只是这些表示式在任何推理系统中所起的作用。另一定义范畴的方法就是详述它在一系统中所起的作用。3是直接后于2、先于4的整数,是一大于1的第一个质数,是头两个整数的总数,5与2之间的差等等。这种定义范畴(或概念)的方案并未述及这个表示式有效指示的东西是什么。算术的真理内在于算术的系统中,而不在该系统可以表征的范围内。当范畴以这种方式定义时,我们就更倾向于把它看作一个概念。(加利斯特,1990b,p.16)

在下面,将以鸽子为例对范畴之形成的研究作出概述,并给以数学的解释。

赫恩斯坦所考察的实验说明:当一鸽子被强化对某种自然类型如树等标本作出反应时,它就把那范畴的新奇标本视作强化的预兆,即使通过强化向它显示的只是某些树的照片,而不是鸽子在标本中可靠区分出来的别的东西。(因而树的照片并不像条件等

580 第五部分

色的色彩刺激,后者是同一的,因而是不可区分的心理表示式。)他作出的分析说明:条件作用发生在范畴水平而不是在个体标本水平,因为新的标本是在各种程度的任务的条件作用下被轻而易举地反应的,就像先前用来例示被强化范畴的标本一样。

鸽子似乎已经具有或易于形成关于这些自然类别的范畴。我们大概可以把范畴看作是在多维描述空间中的、由边界曲面所围成的体积。一个心理或神经描述空间就是其维度符合于描述对象的维度的空间。皮层似乎是由这样的空间构成的,因为不同皮层区的活动中心依照各种描述维度限定了刺激的价值。如果鸽子有关于某些或全部自然类别的天赋范畴,那么这些边界曲面就是先在的,鸽子在条件作用过程中所做的事情就是对强化预示器可靠地映射进来的体积作出分辨。换言之,鸽子可能具有内推规则系统,正是它们生成了,包围被经验的点这群集的曲面,就此而言,鸽子基于所经验的点建构出了范畴(一种由边界曲面诸如皮层区所围成的体积)。无论怎样,都不能由此得出结论说:鸽子有关于树的重要概念,我这里说的概念意思是树的范畴的表示式(树的符号)在其中起着作用的推理演算系统。赫恩斯坦考察的那些实验并不是要测定作出了关于这些自然范畴的推理的系统的存在;它们只是要测定映射过程。它们使人们注意到,对于大脑怎样生成关于诸如树、水、景色之类的刺激的表示式,我们所知道的是多么的贫乏,或者用帕夫洛夫的术语说,关于大脑运用于自然刺激的皮质“分析器”,我们所知的是多么贫乏。(加利斯特,1990b pp. 17-18)

3.4 无表征的智能

3.4.1 布鲁克斯的论点。在对智能建构和开发离不开表征这一人工智能的主流命题的挑战性批评中,布鲁克斯提出了“无表征智能”的论点,并将其具体化于在特殊环境中理智地行动的机器人之中。

人工智能研究已陷入了关于表征的争论。当智能以增量的方式被探讨,从而通过知觉和行动严格地依赖于与真实世界的相互作用时,它对表征的依赖性就不复存在了。在这篇论文中,我们将概述我们对增量式地建立起来的完善的智能造物的研究。智能系统不应被分解为必须借助表征而相互作用的、独立的信息加工单元,而应被分解为独立的、平行的能动产生者,它们经过知觉和行动而直接与世界相互作用,而不是以更特殊的方式相互作用。中枢系统、周围系统之类的概念已去无踪影了——一切既是中心的,又是边缘的。基于这类原则,我们已相当成功地建造了一系列能动的机器人,它们可在标准的办公环境中作为造物无监控地发挥作用。(布鲁克斯,1991,p.139)

布鲁克斯从对 AI 研究与应用的目标与成就的评估出发,阐述了他的无表征智能命题。

人工智能刚开始是这样—一个研究领域,其目的是要在机器中重复人类水准的智力。

由于该目标被发现过于庞大和困难,因此早期的期望渐趋消失。随后的 25 年中,对智能的个别方面的验证曾取得了缓慢的进步。近来的工作倾向于集中在人类劳动者之“智能助手”的商业化

582 第五部分

方面。再没有人谈论重复全部的人类智能了。我们所看到的是对特殊的子问题的回归,诸如表征知识的方式、自然语言理解、视觉甚或像真值维持系统或计划验证等更专门的领域。在这些亚区域所做的工作就是用基准问题测试人类在这些方面所做工作的种类。在 AI 领域的梦幻者中(即是说在那些不是梦想金钱的人中),存在着这样的情感,即在某一天所有的硬币都会掉到地上,进而我们就会“真的”看到智能系统的出现。

然而,我和别的人则相信:人类水准的智能太复杂了,几乎不能理解为在某一时刻可准确分解为适当的子件的东西,即使我们知道了那些事件,我们也还是没法知道它们之间的真正的相互作用。而且只有当我们对更简单水准的智能有更多的实践时我们才能理解怎样分析人类水准的智能。(布鲁克斯,1991, pp. 139 - 140)

在下一段中,布鲁克斯陈述了他的论点:

因此在这篇论文中,我将论述创造人工智能的不同的方案。

* 我们必须循序渐进地建构智能系统的能力,在其每一步都有完善的系统,进而自动地保证部件及其相互作用是有效的。

* 在每一步,我们应当建立完善的智能系统,使之以真正的感觉和真正的行动在真实的世界自由自在。几乎没有什么东西能提供这样一种我们能以之欺骗自己的候选者。

我们一直恪守着这种方案,并建造了一系列自动的、能运动的机器人。我们已得到了一种意想不到的结论(C),并有了一种相当激进的假说(H)。

(C)当我们考察非常简单的智能时,我们发现:清晰的表征和

世界的模型便直接出现了。结果表明用世界作为模型是更好的。

(H)在建立智能系统的最庞大部件时,表征是错误的抽象单元。

表征之成为过去 15 年人工智能研究中的争论焦点,仅仅是因为它在别的孤立的模块与会议论文集之间提供了一种联系纽带。(布鲁克斯,1991,p. 140)

3.4.2 增量智能。布鲁克斯指出,他的理智兴趣就是建构能动的机器人,进而把他的工程目标定位于增量智能的概念的来龙去脉之上。

我将建造一种完全自动的、能动的行动者,它们与人类共存于世界之上,并为人类认可是具有自己权利的智能存在。我将把这种行动者称作创造物(Creatures)。这是我的理智动因。对于证明人类怎样工作,我是没有任何特殊兴趣的,即使人类不同于别的动物,是成功的、自动的行动者,因而是这种事业中的有趣的研究对象。我对应用也没有特殊兴趣;在我看来,如果我的目的能实现,那么这种创造物的应用范围将只受我们(或他们)的想象力的限制。我对这种创造物的哲学意义也无特殊的兴趣,即使很显然将会有重大的意义。

那么,现在就把建造创造物的问题当作一个工程问题来看待。为了建造创造物,我们将发展一种工程方法论。

首先,让我们考察我们的创造物的某些必要条件。

* 创造物在它的动力环境中必须以流行的随机应变的方式恰到好处地处理问题。

* 创造物对于它的环境应当是强壮的;世界的属性方面的微

584 第五部分

妙变化不应导致该创造物行为的整体崩溃;相反,随着环境越来越多的变化,人们有望看到的只是该创造物能力方面的逐渐变化。

* 一个创造物应当能够持有多样的目标,而且能依据它自己所处的环境,改变它正在能动追寻的那些特定目标;因此它既能适应环境,又能利用偶发的环境。

* 创造物应当在世界上做点什么;它也有某种生存的目的。

为了获得这些条件,现在就让我们来考察某些有效的工程方案。如同在所有工程工作中一样,有必要把复杂的系统分解为部分,接着建构那些部分,进而把它们连接于复杂系统之中。

(我们的)分析并未在视觉这样的周围系统与中枢系统之间划出界线。相反,对智能系统的基本划分是沿着下述正交方向,即把它分解为产生活动的子系统。每种活动或行为产生系统单独地使感觉与行动联姻。我们把一种活动产生系统当作一个层次。一种活动就是一种与世界相互联系的模式。表示我们活动的另一个名称最好称之为技能,以突出每种活动随后至少能被合理化为对某种目标的追求。不过我们之所以选择活动一词,是因为我们的层次必须决定何时为自己行动,而不是在需要和呼叫某种别的层次时所诉诸的某种子程序。

这一方案的优点在于:它提供了一条从非常简单的系统到复杂的、自动的智能系统的增量路径。在其每一步,只需建立某种微小部件,让它连接于一种现存的、工作着的完善系统。

该观点首要的一点就是要建立一种非常简单的、完善的自动系统,并在真实世界中予以检验。对这种系统,我们所宠爱的例子就是创造物,其实它是能动的能避免损害他物的机器人。它可感

觉到它周围的对象,并避开它们,如果它感觉到某物挡在路上,它就会停住。通过把这种系统分解为部分而予以建构仍是必要的,但用不着在“知觉子系统”、“中枢系统”与“行动的系统”之间作出清楚的化分。事实上,可能存在着把感觉连接于行动的两个独立的通道(一个用于发动运动,一个用于产生停止),因此没有这样的孤立的空间,在那里“知觉”传递着传统意义上的、关于世界的表征。

接下来,我们再来建立关于智能的一个增量层次,它平行于第一个系统而运行。它连接于已存的、排除程序中的错误的系统,并且再次在真实世界中得到验证。这个新的层次可以直接通向感受器,并对所传递过来的材料作出不同的演算。第一级自动系统继续平行起作用,而不管第二级的存在。例如,在(布鲁克斯,1986)中,我们已报道过:建立一个初始控制层次,它让创造物避开对象,接着再加一个层次,它设置试图访问远距离可见位置的活动。第二层次把命令加之于第一层次的自动控制部分,指导机器人朝向目标,但是第一层次是独立地促使机器人离开先前未看到的障碍物。第二层次监控那创造物的行进,并发送新的运动命令,这一来在不明确知道障碍物的情况下,它也能实现它的目的,而这先前是由控制的低级层级操纵的。(布鲁克斯,1991,pp.146-147)

3.4.3 控制层次。布鲁克斯宣称:在他设计的能动的机器人中,控制层次直接与环境相互作用,而根本不需要表征。

由于多级的层次,因此传递关于世界的描述的知觉概念就变得模糊不清了,当完成知觉任务的系统部分分散于不是由数据通路特殊地关联起来或由功能连接起来的许多部件之中时,情形更

586 第五部分

是如此。肯定没有这样的可分辨的空间,在其中能发现知觉之“输出”。而且,对传感器的数据的各种完全不同的加工是独立地、平行地进行的,每一个都经过不同的控制通路影响整个系统的活动。

事实上,我们不是通过设计,而是通过观察注意到这样一个共同的主题,那就是:不存在核心表征,基于此,我们的分层和分布式方案才有助于我们的造物实现我们的目标。

* 在反作用于环境中的危险或重要变化时,低水平的、简单的活动会逐渐渗透到造物中。在没有复杂表征,没有掌握那些表征的必要性、没有对表征的推理的情况下,上述反作用也会轻易地、迅捷地被作出来,以服务于它们的目的。要点在于:要经常不断地感知环境,进而才有关于世界上发生的事情的最新的观念。

* 由于有多样的、平行的活动,由于铲除了核心表征的观念,因此世界所拥有的各类属性的任何特定变化就几乎不可能引起系统的整体瓦解。相反,人们可以预料:一给定的变化充其量只会使某些控制层次报废,而绝不会使其全部报废。由于异己世界的逐渐闯入(所谓异己,意思是那个层次所拥有的属性不同于它在其中得到调整的世界的属性),那造物的行为可能会连续地衰退。由于不想得到关于那个世界的、主要存在于该系统的类似的模型,因此我们几乎不可能固定地依赖那完全准确的模型。毋宁说,个别的层次只是从世界中抽取出了它们发现是有关联的那些方面(阿格雷等,1986)——也可以说是表征在一简单亚区域的那些投射。那世界的根本结构中的变化在它们的每一投射中被反映的机会少于它们在把某种询问匹配于主要的单一的世界模型时作为困难凸现出来的机会。

* 每一控制层次可以认为有自己隐含的意图(如果你坚持的话也可以说目标)。既然它们是能动的层次,平行地运行,并与感受器相通,因此它们就能密切注视环境,并选定它们的合适的目标。在环境不允许的情况下,目标有时可能被放弃,而在别的时候,偶然发生的环境又可能被当作是有利的。这里的核心理念是:把世界当作自己的模型,连续不断地把每一目标的前提条件匹配于真实的世界。因为对应于每一层次,存在着一种分离的硬件,因此我们就能为那些能平行存在的多种目标作出匹配,而用不着为更多的目标付出任何代价,如果我们想使单一的加工器,甚至某种具有有限能力的网络的复合加工器越来越复杂,那么我们就能够这样做。

* 那创造物的意图隐含在它的更高水平的意图、目标或层次之中。用不着关于下述目标的明晰表征,它们是某种中枢(或周围)过程从决定那造物下一步最适宜于做什么中挑选出来的。

3.4.4 布鲁克斯论点之阐释。在下一节中,布鲁克斯断言:表征的核心不是存在于系统之中,而是限制在 AI 系统的创立者的心灵范围之内。

正如不存在中枢表征一样,甚至也不存在中枢系统。每一活动产生层次都直接把知觉与行动联系起来。对中枢表征或中枢控制作出归因的正是创造物的观察者。而造物本身并不如此;它是计算行为的集合。从它们的相互作用的局部混沌中,在观察者的眼中就出现了一连贯的行为模式。根本就不存在中枢的、有目的的控制中心。明斯基(1986)对人类行为怎样产生也给予了相同的说明。

须知,我们并不是断言:混沌是智能行为的必然的要素。相反,我们为该系统内的所有相互作用的精心设计的工程作了辩护(进化经历了漫长的时间阶梯和大量的个体实验,已有丰厚的馈赠,因而也许用不着这种精确的工程就能完成)。

不过,我们坚持认为;让创造物表现智能行为,既用不着关于世界的明晰表征,也用不着它关于那系统的意向的表征。如果没有这样的明晰表征,局部地加以考察,相互作用可能会是混沌和无目的的。

然而,我的主张还不止于此。即使是在局部的水平,我们也没有传统的 AI 表征。我们从来用不着这样的标记,它们具有可归之于表征的语义学。在我们的补充中,我们所能说的不过是:一个数字从一种加工通向另一种加工。但是只有通过注意第一和第二种加工的状态,那个数字才能得到解释。极端主义者也许会说:我们真的有表征,但它们恰恰是内隐的。随着从完善系统及其状态向另一区域的适当映射,我们就能把一种表征定义为:那些数字和加工之间的拓扑学联系以某种方式得到编码。

不过,我们并不乐意把这些东西称之为表征。在许多方面,它们是不同于标准的表征的。

在推理过程中,不存在着需要例示的变量(例如,关于这一点的更详细的论述,可参阅明斯基,1986)。也不存在需要经过模式匹配挑选出来的规则。更没有什么选择要作出。就大范围而言,世界的状态决定了创造物的行为。西蒙(1969)注意到:一系统的行为的复杂性并不必然内在于创造物的复杂性之中,也许是根源于环境的复杂性。他在描述爬行在海滩上的蚂蚁时作出了这一分

析,但当在下一段谈到人类时,则忽视了它的意蕴。我们假设(根据阿格雷等):大多数人类水平的活动同样是经过非常简单的机制而反映世界的,用不着详细的表征。(布鲁克斯,1991,pp. 149 - 149)。

3.4.5 可动的机器人。在下一节中,布鲁克斯描绘了他的可动的机器人以及它们的行为、设计与有穷状态机的构成。

我们基于任务分解的方法论,建造了四个机器人系列。它们都运行于一个不受约束的动力世界(实验室和 MIT AI 实验室中的办公区域)。当人们与它们一道行走,试图干扰它们以及站在旁边注视它们时,它们都有成功的表现。四个机器人在下述意义上都是创造物,即基于增加动力消耗,它们存在于世界上,并与世界相互作用,追求由它们的执行不同活动的控制层次所决定的多种多样的目标。这与别的为特定使命被给予了程序或计划的可动机器人形成了鲜明对比。

四个机器人中……有两个是同一的,因此事实上有三种设计。一是用机外携带的 LISP(即表处理解释语言——译者)机器来完成它的大多数计算,另有两个用的是机载联合网络,而另一个用的是一种定做的机载平行加工器。所有机器人都具备相同的抽象结构,我们称之为归类构造。它包含有关于完成任务的行为层次的分解以及经过真实世界的调整增量合成的根本观念。关于这些补充的细节可参阅(布鲁克斯,1986)。

归类构造的每一层次都是由一个固定的、具有简单有穷状态机的拓扑学网络构成的。每一有穷状态机有少量的状态,一个或两个内在的寄存器,一个或两个自动按时操作装置以及到达简单

590 第五部分

计算机的通道,该计算机能计算像矢量和之类的东西。有穷状态机异步运行,发送和接收线路上的有固定长度的信息(在两个小机器人上的 1-bit 信息,在更大的机器人上的 24-bit 信息)。在我们的第一个机器人上,这些就是虚拟线路;在我们后来的机器人上,我们就用物理的线路连接计算元件。

根本就不存在控制中心。毋宁说,有穷状态机就是由它们所接收到的信息从数据上驱动的。信息之到达或所设计的时间期满就引起有穷状态机改变状态。它们还有通向信息内容的渠道,也许会输出它们,以一种谓词予以检验,并有条件地划分成不同的状态,或者把它们传递给简单的计算元件。不可能有通向整体数据的通道,也不会有动力学上建立起来的通信链。因而不可能有整体的控制。所有有穷状态机都是单一的,不过它们同时又是它们的固定的拓扑学连结的囚犯。

诸层次是由我们称之为抑制(归类构造的名称由此而来)和阻止的机制结合起来的。在两种情况下,由于加上了新的层次,因此新的线路之一就从旁侧连通于已存的线路。前定的时间常数与每一旁侧-连通都有联系。在抑制的情况下,旁侧-连通发生在有穷状态机的输入端。如果一种消息到达了那个网状线路,那么它就指向有穷状态机的输入端口,好像它已到达那已存的线路。此外,在已存线路上的任何新的消息在指定的时间内都会被抑制(即被拒绝)。就阻止而言,旁侧-连通发生在有穷状态机的输出端口。新线路上的消息在指定的时间内仅阻止信息在已存的线路上的发送。不像抑制那样,新的消息不会送到它们里面。

更详细地说,三个层次以下述方式起作用:

(1)最低级的层次执行使机器人(创造物的物理体现)避免有伤害对象的行为。它既回避静止的对象,又回避运动的对象,甚至回避那些能动地攻击它的东西。被称之为声纳系统的有穷状态机操纵的不过是声纳装置,每秒钟都释放载有被转变为极坐标的读数的瞬时图式。这种图式传送到对立的、有感受力的有穷状态机中。第一级不过是注视,以查明先前是否有无用的东西,如果是这样,就提前给负责操纵机器人的有穷状态机发送停止的消息——如果那有穷状态机不在正常的状态,那么消息就会被忽略。与此同时,另一有穷状态机基于平方反比定律计算对于机器人的排斥力,在这里,每种声纳的返回都被认为标志着有排斥力对象之存在。来自每一声纳的东西叠加起来,就产生出作用于机器人的全部力量。输出传到逃跑机,后者发动它,并将其送至转换机,它又使机器人直接离开总合的排斥力。最后,前溯机促使机器人向前行进。这种机器在机器人向前驱进时,不管何时接收到停止消息,它都会命令机器人停下来。

有穷状态机的这种网络产生出让机器人避让对象的行为。如果让它从空房间中央开始,那么它就只会坐在那里。如果某人走向它,它就会走开。如果它朝别的障碍物的方向运动,它就会停下来。总的说来,它能设法存在于一种动力学环境中,而不伤害对象或被对象所伤害。

(2)下一层次使机器人在不忙于避让对象的情况下行走。行走的有穷状态机大约十秒钟为机器人提供一个随机的导向。避让机把这种导向当作一种吸引力,并把它与从声纳系统计算出来的排斥力加以总合。它用这一结果来抑制低水平的行为,迫使机器

592 第五部分

人朝着接近于行走既定的方向前进,当然同时又要避让各种障碍物。要注意的是:如果转向与前进有穷状态机忙于驱使机器人行进,那么新的行走冲力就会被忽略。

(3)第三个层次使机器人尝试着探险。它寻找远距离的空间,进而设法到达那里。这一层次抑制行走层次,观察基底层次怎样根据障碍物而使机器人发生转向。它纠正各种偏差,进而机器人才能达到目的地。

何时看(when look)有穷状态机关注机器人何时用不着忙于行走,并发动自由空间定位程序……有穷状态机。同时它抑制行走行为,以便让观察仍然有效。当一种路径被发现了,它就被传递给路径筹划有穷状态机,后者又把被指定的方向送至避让有穷状态机。这样,低级障碍物避让仍继续发挥作用。这就能促使机器人朝着不同于路径筹划所期盼的方向前进。由于这个理由,机器人的实际路径是由整合有穷状态机决定的,它给路径筹划机发送最新评估。因此,这种机器作为一种不同的发动机而起作用,它迫使机器人朝向所期望的方向,并为机器人在回避障碍物时补偿实际的路径。

这些特殊的层次,在我们最初的机器人身上都得到了实现。要知详情,请看[布鲁克斯,(1986)]。布鲁克斯和康奈尔(1986)报道:另三个层次在那特殊的机器人上也运行成功了。(布鲁克斯,1991,pp.150-154)

3.4.6 与别的构造的比较。在下一节中,布鲁克斯比较了他的归类构造与联结主义、中枢网络、产生规则、黑板构造之间的异同。

归类构造以其简单的机械网络至少在表面上使人联想到对于智能的许多机械论方案,诸如联结主义和中枢网络。但是它在许多方面又不同于这些工作,而且也相当有别于人工智能中的许多别的后达特茅斯(post-Dartmouth)传统。在下面,我们将非常简要地解释这些差异。

(1)它不是联结主义。联结主义者试图造出简单加工器的网络。在这方面,他们所建构的东西(仅只是模仿——联结主义者从未让真实的机器人在真实的环境之中行走,不管多么简单)相似于我们所建构的归类网络。不过,他们的加工节点倾向于统一,而且他们试图从对怎样正确把它们直接关联起来(这常常被假定有丰富的意义)的理解中寻找启发(正如他们的名称使人联想到的那样)。而我们的节点都是独一无二的有穷状态机器,联结的密度是相当低的,因而肯定是不一致的,在层次之间肯定是很低的。此外,联结主义者似乎寻找明确的、分布式的、从它们的网络中自发产生的表征。我们则完全没有这种希冀,因为我们相信:表征并非必要,只出现于观察者的眼或心中。

(2)它不是中枢网络。中枢网络是联结主义最近得以转世的基质。致力于中枢网络的人断言:他们的网络节点有某种生物学意义,就像神经元模型那样。假如少量被模拟的联结与真实神经元中所发现的无数联结有关,那么大多数模型似乎不大可能。我们认为:我们把有穷状态机挑选为网络节,并没有什么生物学意义。

(3)它不是产生性规则。我们的构造的每一个别的、有活动力的产生层次都能被视之为一种产生性规则的实施。当适当的条件

594 第五部分

在环境中出现了,那么一定的行动就会做出来。我们觉得:这类似于说:带有 IF 陈述的 FORTRAN(指公式翻译程序语言——译者)程序会执行产生性规则系统。标准的产生性系统更是如此——它有一规则基础,从这里,一种规则基于把所有规则的前提条件都匹配于某些基础性材料而被挑选出来。前提条件可以包括必须匹配于基础性材料中的个体的变量。我们的层次是平行地运行的,而且没有变量,或者说不需要匹配。世界的诸方面是抽象出来的,它们直接引起或纠正那一层次的某些行为。

(4)它不是黑板。如果人们真的想将我们的网络与黑板控制构造进行类比,那么他们能得如所愿。某些有穷状态机是局部化的知识源泉。而其他的则是通过在黑板上找到这些知识源泉而作用于它们的加工。不过,在我们的构造中则有简单明了的特点,即所有加工实际上知道观看黑板的什么地方,就像它们实线连接于确切的位置一样。我认为这种牵强的类似性显示了其自身的弱点。根本就不存在这样的伸缩性,加工能在这里搜集到相应的知识。大多数得到改进的黑板构造过多地利用了几乎所有知识的普遍成分和效用。而且至少就精神实质而言,黑板系统往往把知识的消费者隐藏起来,而后者正是特殊的生产者。在黑板系统中,这是抽象的初级方法。而在我们的系统中,我们则使这种联系明朗和持久。

3.4.7 机器人设计的目的。布鲁克斯认识到,就智能进化的等级体系而言,他的机器人尚处在原始水平。机器人智能的进步有赖于他的一般研究战略之成功完成,就像下一段所讨论的那样。

既然我们的方案是一种以行为为基础的方案,因此,我们要建

构的正是系统的行为,它必须适用于评估它的有用性,并指明其局限性。

我们认为:1987 年中期,我们的机器人由于用归类构造来补充完善的创造物,从而成为当时反应最敏捷、最实时的能动机器人。而其他大多数机器人仍停留于静力学环境或充其量在完全被映射的环境中的个体的“实验运行”阶段。另一方面,我们的机器人在它们自己的开关啪嗒一响时,便完全自动运行于复杂的动力学环境中,直至它们的电池用完为止。我们相信:它们以更接近于简单昆虫的智能水平运动,而远胜于细菌水平的智能。我们的目的(如果我们不说出来,就毫无价值)就是在两年内达到简单的昆虫水平的智能。从单细胞到昆虫的进化花了 30 亿年,从昆虫到人类又花了 5000 万年。这一陈述无意于作为对我们未来行为的预言,毋宁说它表明的是昆虫水平的智能的重要本质。

尽管迄今已有如此骄人成绩,但我们的方案仍然存在着许多严重的问题。这些问题怎样被解决,对此我们充满信心和期望,但根据我们自己的标准,真正重要的是要脚踏实地。实验、建构复杂的系统要花时间,因此要说明的是,下面所述的实验尚未完成,只能略述我们现在怎样看待我们工作的进展。我们讨论这一点的意图就是表明:从我们现有的条件出发迈向更高的智能机,至少存在着一条合理的道路。

我们的信念是:我们正开发的各种产生活动的控制层次(灵活性、视觉以及与困难工作相关的生存)是以我们赋予人类的那种方式表现出来的更高水平智能的必不可少的前提条件。(布鲁克斯, 1991, p. 156)

3.4.8 机器人设计的限制。在下一节,布鲁克斯对自己的研究纲领作出了评价。

关于我们方案的限制,最自然而严重的问题就是:

* 在每个层次之间的相互作用变得太复杂以致难以继续之前,归类构造中可建构多少层次?

* 不借助中枢表征而开发的行为能达到什么样的复杂程度?

* 学习之类的高水平功能能发生在这些由简单有穷状态机构成的固定的拓扑学网络中吗?

下面谈谈我们对这些问题的最新想法。

(1)多少层次? 我们已述及的物理机器人的最多层次是三个。在模拟中,我们用过六个平行层次。在设计和增加一个新的层次之前,在所有现存的产生活动的层次上完善地调整机器人的技术,至少到目前为止似乎一直处在实践之中。

(2)怎样复杂? 在我们的、大约需要 14 个个别的活动产生层次的第四代机器人上,我们目前正在研制一种复杂的行为模式。

该机器人有适于避免局部障碍物的红外接近感受器。它还有这样的机载控制器,它能抓住地面和桌子上的对象,并能确定它们的大致重量。其手有安放于其上的深度感受器,以至于可抓住目标对象,可直接控制对目标对象的追踪。我们正在研究一种结构化的激光扫描器,以确定机器人前溯方向上的粗略深度图像。

我们设法在这个创造物身上设置的高水平行为就是在我们的实验室的办公区行走,寻找敞着的办公室的门,并走进去,从密集的办公室中的乱七八糟的桌子上取出空氧化钠容器,最后把它们放回中心仓库。

为了完成这整套行为,完成更简单任务的大量行为就是必不可少的。它们包括:避让对象、循着墙走、识别门径、走进去、校准熟悉的界标、逆着返回方向行走、在界标处辨认返回方位、继续行走、找出似桌子的对象、靠近这种对象、在桌面上搜寻具有氧化钠容器大致高度的圆柱形对象、伸出操纵器的手臂、在被感知对象上移动手臂、用手的感受器从背景中找出竖着的、像氧化钠容器大小的对象,如果它们足够轻就予以抓住,进而把对象存放起来。

个别任务用不着任何中心控制器的协调。它们能把世界的状态编成索引。例如抓握行为能使操纵器抓住由手的感受器识别的、具有适当大小的任何对象。不过,机器人不会随机地抓握这样的对象,因为只有当别的层次或行为注意到似桌子的对象上大致有适当形状的对象时,抓握的行为才会在它关于世界的感知让它起作用的地方展示自己。由上可知,如果一个对象看上去不再像氧化钠容器,那么抓握反射作用就不会发生,别的低水平行为将让机器人在别的地方寻找新的候选物。

(3)是学习吗,有这种可能吗?某些昆虫显示有一种被称为“本能的学习”的简单的学习类型(古尔德等,1986)。据假定,如蜜蜂生来就知道怎样区分各种花,知道往返于蜂箱与花源之间的路线。别的昆虫如蝴蝶已表明能学会对花的区分,但用的是一种信息限定的方式(刘易斯,1986)。如果它们被迫熟悉了第二类花,它们就会忘记关于第一类花所知道的东西,这在一定程度上表明:它们所知的信息总量是不变的。

我们已找到了一种为我们的有穷状态机建构固定的拓扑学结构的方法,它们能作为一孤立的子系统、以能与上述实例相媲美的

598 第五部分

水平完成学习任务。此刻,我们所处的正是这样的立场,据此我们在本文中对早期的大多数 AI 工作者作出了猛烈的抨击。我们只有关于系统工作的孤立的模数,而输入与输出仍是飘忽不定的。

我们还在努力,以改变这种局面,但是对物理的创造物的实验工作是一项重要的、需要耗费时间的活动。我们发现:关于装置或软件的任何预先设计的组件几乎都有如此多的、它们应怎样成为它们的的部分的前概念,以致它们并没有足够的灵活性,不足以成为我们的完善的系统的组成部分。于是到 1987 年中期,我们对学习的研究工作便以下述需要为转移,即建立一种新的电视摄像机和高速低耗的加工箱,以操纵特别地开发的、以每秒 10 幅画面出现的视觉规则系统。这一工作的每一步都是一种有意义的工程学事业,我们正以财力所允许的速度向前推进。

当然,空谈是不费力气的。

(4) 前瞻。对于在实在世界的真实创造物的实验只能解答关于我们的方案的常见的疑惑。而时间将作出判决。(布鲁克斯, 1991, pp. 156 – 158)

3.5 评论

布鲁克斯(1991)提出了无表征的智慧的命题。不过他并未对表征作出清晰的定义。什么东西构成表征系统,这一直为帕尔默(1978)所重视。图 2 借助五项联结条件定义了表征系统。最后一个条件规定了模型的适当方面与环境的适当方面之间存在的一致性或同型性。

-
1. 有目标域。
 2. 有模拟域。
 3. 目标域中的是相关联的结构子集。
 4. 模拟域中的结构子集是相关联的。
 5. 模拟域中的相关结构与目标域中的相关结构之间存在着整体的一致。
-

表 2 表征系统的定义性条件

就布鲁克斯(1991)设计的机器人而言,一致性出现在环境的适当信息方面与机器人的适当信息方面两者之间。机器人的电子设备中的信息加工是由图式或有穷状态图表表征的。图式构成了表征系统,而该系统又确定了环境中的信息流和机器人的电子设备中的信息流之间的一致性。智能就存在于表征系统所在的地方。

4. 表征与关于动物、计算机 和人类心灵的理论

即使表征存在于动物、计算机和人类的认知之中,但对心灵状态(知识、信念和情感)作出归因的能力在人类身上则得到了充分的发展,而在计算机上可能只存在于有限的范围,在大多数动物身上压根就不存在或残缺不全。动物能与它们的同类进行有意义的交流,但是这种交流并未顾及交流接受者心中所存在的有知或无

600 第五部分

知的区别。关于心灵的理论不能归之于动物。使其行为适合于另一计算机或人类的有知或无知水平的计算机掌握了关于心灵的有限的理论,所谓有限是指自身程序的范围和力量的有限。

关于猴子有心灵理论的研究在下面的说明中有简要的介绍。

我们的发现……并不能说明猴子的叫喊能否上升到人类词语的地位。人类语言涉及到的远不只是词语与其所指示的对象或事件之间的指称关系。在交流信息时,我们也把知识、信念和愿望之类的心理状态归之于他人。我们认识到:心理状态具有因果力量……我们曾想知道:猴子能否在它们自己的信念、知识与别的猴子的信念之间作出区分,它们在通知不知道的个体或纠正别的猴子的错误信念时是否采取了特定的步骤。

动物可能对别的动物的心理状态敏感这一类想法来自于新近在不同鸟和哺乳动物身上所做的听众效应的研究。马勒近来在丹佛的加利福尼亚大学与其同僚重做了 N. 廷伯根的一个古典实验,在该实验中,一老鹰的剪影模型悬挂在一根电线上,接着一群原鸡从此“飞过”。马勒发现:公鸡在母鸡出现时发出了令人震惊的叫声,但是当它们独自与别的种类的雌性在一起时则仍保持安静。在野外,康奈尔大学的 P. 舍曼发现:成年雌性的松鼠有其近亲在场比无近亲在场发出惊叫的可能性更大。在对被捕捉到的雌性黑长尾猴的实验中,我们注意到:成年雌性在与自己的后代在一起比与年龄、性别相同但无亲属关系的黑长尾猴在一起所发出的惊叫要多得多。

即使动物对听众在场或缺席显然很敏感,但这一事实并没有证明:它们对它们听众的心理状态也是敏感的。而且有广泛的证

据表明:动物并不能识别无知听众与有知听众之间的差异。例如,公鸡与黑长尾猴在它们的同伴看到捕猎者并逃跑之后仍继续长时间发出惊叫。

为了在猴子身上验证心灵理论,我们在丹佛的加利福尼亚灵长目动物研究中心对两群罗猴短尾猴和两群日本短尾猴进行了实验。

对动物信息沟通的研究不可避免地要与对动物怎样思维的研究牵连在一起。我们对黑长尾猴的研究允许我们作出如下结论:灵长目动物发出的声音不仅是本能的尖叫声,而且是个体有选择地使用的叫唤,这些个体注意到它们的听众,并用不同的发声作为表示它们环境的不同特征的信号。猴子的叫唤像语词一样起着指示不同对象或事件的作用,而猴子自身甚至可以认识到叫声与对象或叫声所表示的概念之间的指称关系。

同时,我们正着手对猴子的信息沟通与认知怎样不同于人类的作出更明确的理解。即使人类的大多数信息沟通旨在影响引起行为的知识、信念与动机,但现在还没有根据说:猴子在沟通时也带有影响别的动物的心理状态的意图。猴子,甚或类人猿不可能带着改变别的动物的心理状态的意图进行信息交流,因为它们显然没有认识到这种心理状态的存在。(赛发思等,1992,pp. 126-128,出自 R.M. 赛发思等“意义与猴子的心灵”,1992,《科学美国人》)

参考文献

R. 戴维斯和 H. 施罗伯(1983):“表征数字硬件的结构和行为”,《IEEE

602 第五部分

计算机》,16(10),75-82。

R. 戴维斯、H. 施罗伯和 P. 索洛维茨(1993):“什么是心理表征?”《AI 杂志》,14(1),17-33。

H. 波朴尔(1982):“把结构加之于有结构毛病的问题的启发式方法”,载 P. 索洛维茨编:《医学中的人工智能》,美国科学促进协会会议论文集,51。

M. 明斯基(1975):“知识表征的构架”,载 P. 温斯顿编:《计算机视觉心理学》,纽约:McGraw-Hill。

M. 明斯基(1986):《心灵社会》,纽约。

C. 加利斯特(1989):“动物的认知:关于时、空、数的表征”,《心理学年鉴》,40,155-189。

C. 加利斯特(1990a):《学习的组织》,麻省理工学院出版社。

C. 加利斯特(1990b):“动物认知中的表征导论”,《认知》,37,1-22。

R. 布鲁克斯(1986):“可动机器人的、强有力的、分布式控制系统”,《IEEE 自动机器人杂志》,2,14-23。

R. 布鲁克斯(1991):“无表征的智能”,《人工智能》,47,139-159。

S. 斯蒂文斯(1946):“关于测量标准的理论”,《科学》,103,677-680。

J. 吉本等(1981):“剩余的时间:线性式与对数式主观时间”,《实验心理学杂志》,7,87-107。

K. 程(1986):“鼠的空间表征中的纯几何学模型”,《认知》,23,149-178。

B. 卡特赖特和 T. 科利特(1979):“蜜蜂怎样知道它们离附近视觉界标的距离”,《实验生物学杂志》,82,367-372。

B. 卡特赖特和 T. 科利特(1983):“蜜蜂的界标学习:实验与模型”,《比较心理学杂志》,151,521-543。

R. 维纳(1981):“节足动物的空间知觉”,载 H. 奥特朗编:《比较心理学与无脊椎动物的视觉进化手册》,纽约,pp.287-617。

H. 西蒙(1969):《人造物的科学》,麻省理工学院出版社。

S. 帕尔默(1978):“认知表征的基本方面”,载 E. 罗施等编:《认识与范畴化》,希尔斯代尔。

J. 古尔德等(1986):“本能的学习”,《科学美国人》,pp.74-85。

心理表征*

哈特里·H. 菲尔德

对唯物主义的任何有趣的描述要求的不仅是没有不可还原的心理对象,而且是没有不可还原的心理属性:这样的观念,即尽管人类和某些高等动物并不含有任何非物质的实体,但是他们具有一些完全无法用物理语言来解释的心理属性,是一种几乎不会令那些把自己视为唯物主义者的人满意的观念。^①对假定不可还原的心灵属性的这种不满正是心灵哲学中两大问题的根源。第一个问题,也是受到更加广泛讨论的问题,就是由经验属性例如感到疼痛这种属性所引发的问题:唯物主义者必须对这样一些属性作出可信的说明,根据该说明,那些属性不是不可还原的心理的东西。第二个问题是由布伦塔诺所提出的意向性问题。^②许多心理属性——相信、愿望等等——似乎是关系属性:更加确切地说,它们似乎把人与被称之为命题的非语言实在联系起来了。因此,任何看重“相信”和“愿望”的表面价值的唯物主义者——任何承认信念和愿望是人与命题之间关系的唯物主义者——如此之类的任何唯物主义者必须说明:这里所谓的关系并非是不可还原的心理的东西。

* 原载《认识》,13,1(1978),9-61,后收入 S. 斯蒂克等编:《心理表征文选》,布莱克韦尔,1994,现据前者译出。

604 第五部分

布伦塔诺认为这是不能成立的；除了把信念和愿望看作是命题之间的关系，他觉得别无选择，既然如此，他便得出结论说，唯物主义一定是错误的。

这篇论文的前半部分考察的是布伦塔诺问题，即对给予相信、愿望等以唯物主义的、充分的说明的问题（布伦塔诺认为这是不可解决的）作出考察。不同于布伦塔诺，我认为唯物主义是正确的，这是毋庸置疑的（就我们对世界所知的而言）。我还认为，就本文的目的而言，说人们有信念和愿望也是没有疑义的。这两个假设加在一起就形成了这样的假定，即布伦塔诺问题是可以解决的；我感兴趣的问题是：它怎么可能被解决，即是说它可以被解决这一假定说明了关于相信和愿望的什么本质。我特别想说明的是：人们不能仅靠采纳 S. 斯托纳克、D. 刘易斯^③所倡导的关于信念和愿望的功能主义理论来解决布伦塔诺问题。在本文的后半部分，我将阐述在前半部分表明的立场，并从意义理论出发讨论我关于意向性的结论的一些意蕴。

首先我应该指出，我的前假设，即人们实际上有相信和愿望，旨在成为这样一个前假设：有关信念—话语、愿望—话语的工具主义观点是不妥当的。换句话说，我预先假设的是：关于信念和愿望的归因确实是正确的，而不仅仅是我们为了各种目的而采用的有用手段，但是对于这些目的来说，真与假的话语就是不恰当的。很明显，对唯物主义的信念，即使与那种认为信念—话语和愿望—话语是非常有用的工具的观点结合在一起，也不足以衍推出：布伦塔诺问题是可解决的；^④因此采纳关于信念—话语和愿望—话语的工具主义观点就是拯救唯物主义进而回避解决布伦塔诺问题的一

种可行办法。然而对我而言,工具主义是没有吸引力的:在我看来,如果一种假定了信念和愿望的心理学理论很有用,而且没有比它更有用的理论,那么除非有相反的、极为确切的理由,我们应该把这种理论看作事实上正确的理论。应该如此看待它,这正是我在本文中据以行文的一个工作假说。

I. 基本论证

回应布伦塔诺论证的一种方法就是去阐发一种关于信念、愿望等的理论,根据这个理论,信念和愿望的对象不是命题,而是某种更“易于接近的”事物,比如说句子。如果我们将信念和愿望解释成人与句子之间的关系,那么——不管怎么说,如果我们也有这样的需要,即与相信者或愿望者相联系的一些句子是他所熟悉的语言中的句子——就有可能用唯物主义认为是完美的方法来对这些关系作出说明。在后面,我会进一步论及这种解决布伦塔诺问题的方法。这里我不想赘述。

再者,假设我们接受这种观点,即信念和愿望是人与命题之间的关系,而且为了明确起见,假设我们拥护刘易斯和斯托纳克主张命题是一系列可能世界的观点。问题由此便产生了:对于如此被诠释的人与命题之间的关系,人们又怎能作出唯物主义认为可接受的说明呢?

回答这个问题既非愚蠢又非不可能。人们可能会认为,这样做很愚蠢:“如果你接受关于命题的本体论,加上一点最低限度的集合论,那么人与命题之间的关系仅仅只是一系列有序偶,这些有

606 第五部分

序偶的第一类成员是人,第二是命题。或者更确切地说——既然同一个人在某一时刻可以处于信念关系之中,而在另一时刻却相反——那么信念关系就只是一种功能,这种功能每次都归属于人与命题的一系列有序偶。归属有序偶集的这种功能根本就不是一种心理实在,更不用说是一种不可还原的心理实在,因此这里的问题是什么呢?”不用说,它离题太远了。如果这种摒除布伦塔诺问题的方法合理的话,那么即使不做任何实质性的工作,我们也同样可以克服唯物主义的其他表面的困难,因此那种唯物主义可能是一种完全空洞的学说。例如,如果某人被问及怎样对疼痛作出唯物主义的说明,他也许会说:“这是什么问题?疼痛不过是人的一种集合,即这样一些人,我们常说他们‘感到疼痛’或‘处在疼痛中’。或者更准确地说——既然同一个人可能在某一时刻感到疼痛而在其他时刻则不疼痛——因此疼痛就是这样一种功能,即每次都归之于在那时处在疼痛中的那群人的功能。这样一种功能纯粹是一种集合论的实在;它不是心理的东西,更不用说是不可还原的心理的东西,因此它并没有给唯物主义提出任何难题。”我认为这种对疼痛理论的集合论探讨显然是完全不能令人满意的。而且对信念和愿望理论的集合论探讨为什么应被视为更好的,我也想不出任何理由。

对疼痛的集合论说明为什么不令人满意呢?部分解释是这样的:它把疼痛看成一种集合而不是一种属性。所需要的是对“疼痛”一词所表示的人的那种属性作出可为唯物主义接受的说明——我们必须表明那属性不是不可还原的心理的东西。集合论说明所做的是:完全忽视属性问题,而集中注意拥有那种属性的事物

之集合。(或者——既然同一事物只能在某一时刻具备这种属性而在其他时刻不具备这种属性——因此它侧重于每次把延伸集功能映射到在那时具备这种属性的事物之集合。)这种集合(或功能)作为不可还原的心理的东西,显然不存在实质性问题;但是这正好说明了这一有趣的问题所关注的不是集合而是属性。这种说法同样可以推广到信念中:关于信念的有趣问题是一个关于“信念”所表示的关系的问题,我这里所说的“关系”并非指“有序偶的集合”,而是指“又一元属性”。唯物主义者需要说明的是:那属性不是不可还原的心理的东西。

但是对信念关系有望作出唯物主义可接受的说明吗?尤其是,当我们把信念看成是人与命题之间的关系,并用可能世界来阐释命题时,有望作出这种说明吗?也许人们会认为,答案“显然是没有”,这一事实说明,将信念看成人与一组可能世界之间的关系是难以站得住脚的。然而我认为,对关于信念的可能世界观点的这种摒除未免过于仓促,因为我相信,对信念关系作出唯物主义认为充分的说明是可能的,即使这种关系被解释成一种与可能世界集合的关系。现在我将勾勒出这样一种可能的说明。在后面,我们将看到,这种说明的主要内容可以被保留,即使我们抛弃信念是人与可能世界集合之间的关系这一假定。

这种说明包括这一假定,即信念关系是另外两种关系的合成:第一,人与人所理解的句子之间的关系;第二,句子与可能世界集合的关系。这种说法很快就会得到完善,暂时还不能如此,该断言是这样的:

608 第五部分

(1)当且仅当有句子 S,以至于 X 相信 *S,S 意指 p,X 才相信 p。

这里,相信*是人与他自己语言中的一个句子之间的关系:我相信*英语句子“雪是白的”,而不相信*“雪是绿的”,不懂英语的人绝不会相信其中任一个句子。^⑤接受(1)的后果就是把对信念关系作出唯物主义认为充分的说明这一问题分成两个子问题:

子问题(a):解释某人相信*(他的或她的语言中的)句子意味着什么的问题。

子问题(b):解释句子意指 p 意味着什么的问题。

(对愿望关系作出唯物主义认为充分的说明这一问题也可以用类似的方法、通过引入愿望*这一观念而分成两个子问题。)

关于怎样说明(a)的大致的观念应当是足够清楚的:当且仅当我乐意以特定的方式运用我的语言中的一个句子进行推理、思考等等时,我便相信*那个句子。当然,这很含糊,但是假设这种含糊性可以被消除,而且可以为这些被求助的倾向找到一个物理基础,那么“相信*”就不会成为给唯物主义者出难题的关系了。在这节的后面,我将陈述关于相信*的倾向说明的更明确的描述,它将澄清:所祈求的倾向的物理基础必须是什么样子。但是我希望,即使是以上模棱两可的论述也足以使读者先倾向于认为:相信*不应是一种使唯物主义者特别担忧的关系(即使是被布伦塔诺问题所困扰的唯物主义者)。另一方面,任何人如果被布伦塔诺问题所困扰,他也可能被子问题(b)所困扰,因为:(b)不像(a),它诉诸的是一种语义关系(“意指那个”的关系)。

在探讨能对(b)作出一种唯物主义可接受的说明这一预期是

否合理之前,让我先来考察坚持(1)的两个附带的优点,然后再来考察对我所说的话的必要的修改。通过(1)来解释什么是相信 p , 探讨这一解释问题的第一个附带的优点在于:由于这种探讨,我们无需从一开始就作出信念的对象是命题这一有争议的假定。也就是说,如果我们以后想出一种方法来解释“意指 p ”究竟为何物,这种方法并没有揭示句子与命题之间的关系的意义。那么,当我们接受了(1)时,我们也就有一种关于相信那的说明,根据这种说明,相信并不是人与命题之间的关系。事实上,人们即使不假定(1)中的“ p ”的出现就是可用量词限定的变量的出现,也能坚持(1);“ p ”的出现可以被看作图式化字母的例示,因此关于信念对象的话语(前面曾阐释过这种话语)没有任何字面意义。这样一种可能性将在第三节中予以讨论;然而现在我将继续沿用刘易斯和斯托纳克假定,即 p 是一个可用量词限定的变量,而且表示应根据可能世界加以解释的实在。

对于可能世界理论家而言,(1)的第二个附带的优点就是,它为他们区分信念提供了一种合理的方法。这种区分从直观上说存在于那里,但是照现在的情形,它们又不能根据可能世界说明作出来。试想任何集合理论家(他相信选择的公理),他从未听说过巴那克-塔尔斯基定理;如果现在有人为他明确表述这一定理(而不告诉他这是一个定理),那么他几乎不会相信它是真的。(事实上,他几乎肯定地相信它是错误的;^⑥在所描述的这种环境下相信它是正确的跟相信它是错误的相比,会显得更加荒谬。)因此,在相信集合论与相信集合论加巴那克-塔尔斯基定理之间,存在着很大的差异;然而集合论与集合论加巴那克-塔尔斯基定理在逻辑上

610 第五部分

是等值的,因此必定能用相同的可能世界集合来加以表征。D. 刘易斯指出,可能世界理论家无需说:集合论^⑦公理的合取与这些公理加巴那克-塔尔斯基定理的合取意思相同:除了可能世界集合之外,他可以认识到更高级别的意义。^⑧可能世界的一个集合可以被称作一个质朴的命题;而精致的命题则是用可能世界来定义的一种复杂的函项,但是这种函项编码在它之中,也就是说编码在有特定句子的结构之中。根据刘易斯的说明,一个句子所表达的意思是精致的;而一个人所相信的却是质朴的,因此相信集合论和相信集合论加巴那克-塔尔斯基定理显然是同一的。

严格地讲,承认这些精致的命题是信念的对象,并且以不同于刘易斯的方式运用它们,这完全是可能的。但是这样做又十分不合情理,除非我们认为信念是由句子或类似句子的某物表征于大脑之中的,因为精致的命题具有被编码在它们之中的特定句子的结构。因此,刘易斯和斯托纳克接受了这一观点,即相信集合论和相信集合论加巴那克-塔尔斯基定理是同一的。斯托纳克针对我们的强烈的反对意见作了如下辩解:

数学上只有两种真命题,必然真的命题和必然假的命题,而且我们都知道,第一个是真的,第二个是假的。但是两个命题中哪个是由一个已知数学陈述句表达出来的呢?决定这一点的那些函项正好是这类函项:它们非常复杂,对哪个命题由一个陈述句表达足以引出合理的疑问。因此,将数学中相信和怀疑的对象看成是关于陈述句与它们所要表达的东西之间的关系的命题,似乎是合理的。(斯托纳克,1976,p. 88,着重号为我所加)。

但这是行不通的。假设“巴那克-塔尔斯基条件句”代表这样一种条件句：它的前件是集合论公理(包括选择)的合取，结论是巴那克-塔尔斯基定理。试想这样一种人，他怀疑或不相信巴那克-塔尔斯基结论，但是他知道把这些集合论语言中的句子与命题关联起来的语义规则。在斯托纳克看来，这样一种人并非真正怀疑或不相信“巴那克-塔尔斯基条件句”所表达的命题，因为那是一个逻辑上的真理；他所怀疑或不相信的是这样被表述的命题：

(i)集合论语言的语义规则将巴那克-塔尔斯基条件句与必然的真理关联起来了。

但是既然此人知道集合论语言中的语义规则为何物，那么他能怀疑或不相信由(i)表述的命题的唯一办法就将是怀疑或不相信由(ii)表述的命题：

(ii)语义规则____将巴那克-塔尔斯基条件句与必然真理关联起来。

空格处可填写一个关于集合论语言的语义规则的陈述。然而对斯托纳克而言，不幸的是，由(ii)表述的命题本身就是一个必然真理，所以根据斯托纳克精致的可能世界观点，任何人根本就不可能怀疑或不相信它。从数学命题上升到元语言命题毫无意义。

既然如此，我便认为，将命题分得比可能世界集合更细致些，这显然是可取的，而且刘易斯的分析向我们表明：这是怎样完成的，如果我们能够接受相信一个命题涉及到一种对句子的态度这种观点的话。相信集合论就是大致地相信*公理的合取；相信集合理论加巴那克-塔尔斯基就是大致地相信*那些公理的合取加巴那克-塔尔斯基公理。根据刘易斯的分析，被相信*的句子有

612 第五部分

不同的、精致的意义,由此(i)将不同的精致命题归之于两个相信者。

然而,人们可能会说,这种句子分析方法导致了信念中的过于细密的区分;它使这样的信念即要么罗素没有头发要么雪是白的,有别于下述信念:如果罗素并非没有头发,那么雪是白的;这一来我们怎能信心十足地从断言某人持有第一种信念过渡到断言他持有第二种信念呢?这里有两种可能的答案:第一就是通过在一适当适当的等价关系下理解等价的类别而引入一种使精致命题变得些许粗糙的方法,我认为,这种答案相当特别,且矫揉造作。第二种也即一种更令人满意的回答,即承认这些信念确实不同,但是它解释了:持有一种信念的人为什么总还会持有另一种信念。这样一种解释很容易说明:它涉及到信念*带有倾向性这一观点。现在我将阐述信念*的倾向本质(因而回到子问题(a))。在阐述的过程中,持有关于罗素的两种信念之一的某人为什么总还会持有另一种信念这一问题将迎刃而解。

对信念*作出唯物主义说明的最原始的方案可能完全是非倾向性的:这一方案可能包含下述假设,即只有当一个句子清晰地贮存在某人的大脑中,此人才会相信*这个句子。(愿望*与此相仿。如果我们坚持这种非倾向性的方案,那么我们必须假设至少有两种贮存:信念贮存和愿望贮存。句子怎样作为信念贮存或作为愿望贮存的准确细节当然不是一件没有经过详细的神经生理研究就能解决的问题)。这个方案的一个缺点就是人们显然相信无限多的命题;如果是这样,那么当我们接受(1)(并且假设“意指那个”的关系不是一与多的关系)时,就会得出结论说,人们相信*无限多

的句子。但是,无限多的句子清晰地贮存在有限的大脑中又是绝对不可能的,因此一个句子必须清晰地被贮存起来以便于被相信*这一观点是错误的。

这一论证表明,对信念*的某种倾向探讨是必要的。(即使没有这样的假定,即我们有无限多的信念,这一结论事实上也是可以论证的。)倾向说明的最简单的一种(由丹尼特提出的^⑨)就是:有机体在其内部贮存了某些被明确表征的信念,丹尼特称之为核心信念;我们说一个人所相信的只不过是他的核心信念的明显结果[不是他的信念的明显结果;因为那样的话,我们就可能循环地说明:即使是信念的极不明显的结果也得是信念(因为不明确推论可以分解成大量明确的推论)]。如果这种核心信念方案是正确的,那么它可能是关于核心信念的看法,而不是关于可进入任何详细的心理学理论中的信念的看法。但是即便如此,这也是关于信念的看法,而不是关于日常生活中对我们更有用的核心信念的看法,因为了解一个人相信什么比了解他所相信的核心内容要容易得多。理由在于,当命题 A 是某人核心信念中的一个足够明显的结果时,他几乎会自信地把 A 归之于自己信念的核心,不管 A 的真实性出了什么问题;这一来,从行为上就难以将这种情况与那种 A 过去一直存在于某人的核心信念中这一情况区分开来。但是有许多情况,在那里,一个信念显然不是某人信念核心的组成部分。例如,假设我告诉你:在 1953 年,没人挖出了一条从这儿到中国、穿过地球中心的隧道。我确信,在向你讲这些时,我不是在告诉你一些你已不相信的东西,但是我同样也确信,它不是你的信念核心的组成部分——即是说,不是你的那些得到明确表征的信念之一

614 第五部分

——在我告诉你之前是如此。

核心信念这一观念为我们解决问题(a)提供了非常自然的方案。这一观念不过就是说:当且仅当一个句子是被清晰地贮存(以适用于信念、对立于比如说愿望的方式贮存)^⑩的那些句子的一个明显结果,一个人才会相信*这一句子。这种说明会澄清下述问题,即一个要么相信罗素没有头发,要么相信雪是白的人,为什么还会相信:如果罗素并非没有头发,那么雪是白的,前者由之而成为一个明显结果的任何核心信念贮存物,几乎也是这样的核心信念贮存物,即由于它,后者成为一个明显的结果。^⑪(这就解决了从一开始就遇到的问题。)这一说明还表明:在远离核心的信念归因中,为什么存在着某种含糊性或不确定性(例如,当一个哲学家的信念显示出一种深刻的不一致性时,为什么没有明显的理由去问:他是相信还是不相信这样一个命题,对此他并没有发表任何意见,但是根据他所坚持的东西,借助同样不明显的推理,这一命题既可证实又可被反驳)。因为在根据核心信念定义信念时,运用了不明确性这一观念,因此不确定性就产生了。^⑫核心信念方案的一个明显优点就是,它解释了这些事实。

再者,我认为,对子问题(a)的核心信念探讨是相当有吸引力的。然而,我并不想一味地予以坚持,可以设想,完全有可能形成一种更微妙的探索信念*的倾向方案——也许是一种几乎用不着被明确贮存的句子的方案。(然而如果完全没有贮存在一起的句子,人们似乎不可能如愿以偿。)为核心信念方案找一个替换品是否更好些呢?对此我不发表任何意见,因为我认为,为了证明子问题(a)可以以这样或那样的方式加以解决,我已说得够多了。

在转到子问题(b)——即对“意指那个”这一关系(我们权且假定它把句子与可能世界集合关联起来了)作出一种唯物主义认为是充分的说明这一问题——之前,我应该提到(1)中需要的、有远大前途的改进。改进的需要来自这一事实,即狗、猩猩以及婴儿可能有信念,但是却没有语言。人们自然会明确地坚持说,信念和愿望不能归之于这些或其他任何没有语言的造物。然而,这一看法似乎是极不合理的。因为,我们毕竟经常通过假定信念和愿望来解释无语言生物的行为;而且在此类情况下,由于这些解释完全是拟人化的,进而行为可以用更简单的方法来解释,因此有许多例子涉及到心理复杂的有机体,在那里,这些是几种唯一可行的解释。断言所有这类解释都是错误的,似乎过于轻率,因此我们必须修改(1)。幸运的是,只需要作出细微的修改,如果我们接受下述广泛认可的观点的话,^⑬即尽管一只狗没有语言,但它的确具有某种内在表征系统,而且它之所以能把任何命题表征给自己,并予以相信,这仅仅是因为它具有某种内在的表征系统。那么我们不妨这样修改(1),即放弃被相信*的实在其实是句子这一要求,相反只要求它们要么是句子要么是句子类似物——在这里,所谓句子类似物,我指的只是一些表征了命题的心理实在(或者更加中性地说,它具有句子所有的那种意义或内容)。这一来(1)应该替换为:

(1')X 相信 *p, 当且仅当有一个句子或句子类似物 S 时,以至于 X 相信 S 且 S 意指 p。

这一更改对具有语言的成年人来说有一种独立的优点:它允许这样一种可能,即他们可能有少许超出他们语言的表达能力的信念。

616 第五部分

许多哲学家对非语言有机体有内在表征系统这一假设表示异议。在本文后面,我将尽力使那种对立在一定程度上得到消除。但也许我可以以下述方式提前做一点这方面的工作。据我的观点,猩猩相信和想望的内在表征系统可能会有比人类语言更简单的结构——它的“句子类似物”从句法上说可能比英语句子简单得多,而且说它们是“句子类似物”,我的意思不是否认这些结构的非类似性。动物的“内在表征系统”可能极为简单,怎么依然被看成一种表征系统呢?在我看来,这是术语学的一个无聊的问题。如果你愿意的话,你甚至可以允许动物的“内在表征系统”由5个句子组成,每个句子没有句法上的结构。然而,我认为,说动物相信或想望某事物没有什么意义,除非人们将信念与愿望的适当复杂的系统归之于它;而且[根据(1)]适当复杂的信念与愿望系统将需要适当复杂的表征系统。所以对内在表征系统作出归因根本没有什么意义,除非人们对适当复杂的系统作出归因。但是无须划出明确的界限,而且也无须说明“某种事物必须至少如此复杂才可以被视为一个内在表征系统”。在何种情况下这一界限有效只是人们想让名词“信念”的意义扩展多广的问题。我的看法与这些术语上的争论没有关系,它认为,如果有机体是足够复杂的,以至于能清楚地运用信念和愿望概念,那么它们便有内在表征系统,在这一系统中,句子类似物具有重要的语法结构。

难道我们不能通过对(1)的一种可供选择的修改来避免引入内在表征系统吗?这一可供选择的修改将会放弃这一要求,即只有当有机体理解一个句子时他才会相信*它,于是我们可以说,狗相信人类语言的句子,比如英语和塞尔维亚-克罗地亚语。这种

探讨方案存在着三个问题：

(A)作为解决布伦塔诺问题的一个步骤,这一方案不是很有用的。因为据此,相信*便成了一种语义关系——一种有机体与英语句子(或塞尔维亚-克罗地亚语)之间的关系。这种关系由于这些句子的意思而存在着。所以所提出的(1)的替换品可能将有机体与命题之间的原始信念关系分成两种语义关系(因为“意指那个”也是有语义的);而且既然语义关系只是布伦塔诺认为不可解释的那类关系,因此这似乎是我们的原始问题的放大。

(B)这一主张所包含的第二个问题就是,信念和愿望概念的整个意图就是帮助我们解释行为。通过猩猩与它并不理解的人类语言中的句子之间的关系来解释猩猩的行为看上去相当怪异;关于英语句子的事实与猩猩为什么那样行动也毫不相干。因此将英语句子从关于猩猩的信念和愿望的说明中省略掉是很重要的。

这一主张所提出的第三个问题在第三节中再来论及。正因为这些困难,我认为(1')是(1)的唯一合理的修改。

最后,让我们转到子问题(b)。由于从(1)到(1')的过渡,这便成了对下述问题作出唯物主义可接受的说明的问题,即内在表征系统中的一个句子或一个句子类似物意指 p 是什么意思。试想,我们正在把“意指那个”暂时解释成把句子(和句子类似物)与可能世界集合关联起来(或者与从根据可能世界定义的函项中建立的有微妙结构的实在关联起来;但是为了简单起见,我们不妨忽略这个更加复杂的说明——对唯物主义的信念理论来说,它原则上并未引发新的困难)。句子(或句子类似物)意指的是什么样的可能世界集合呢?它指的是该句子在其中为真的可能世界集合。因此

618 第五部分

如果我们能为句子与它在其中为真的世界之间的关系提供唯物主义可接受的说明,那么我们将会有一种关于(b)的唯物主义可接受的说明。

在另一篇论文里,^④我曾论证说,对真值(即在现实世界的真值)作出唯物主义认为充分的说明是可能的。关于如何做到这一点的观点,对于逻辑学家经常讨论的非常简单的语言来说,是再简单不过的了;在那里,它涉及到给出一个塔尔斯基型的真值定义。这样一种真值定义根据一个名称指称一个对象(在现实世界里)意味着什么,一个谓词具有某种外延(在现实世界里)意味着什么等,来解释一个句子为真(在现实世界中)意味着什么。换言之,这样一种真值定义将一个句子为真(在现实世界中)意味着什么还原为这个句子里的单一语词组成部分代表(在现实世界中)某种实在或实在集合意味着什么。我已论证过,我们有充分的理由认为,“代表”关系(即作为外延的具有、指称关系等)用唯物主义可接受的方法是可以解释的,因此,真值定义衍生性地为我们提供了一种关于真值的唯物主义可接受的说明(在现实世界里)。以上论述保留了塔尔斯基方法直接适用的简单语言的句子。但是像戴维森^⑤和其他许多哲学家一样,我相信此类方法可泛化于自然语言,因此对自然语言里的句子来说,我们也可以说真值能够用唯物主义可接受的方法来解释。对句子类似物来说,情况也是如此,只要我们假定它们与自然语言里的句子非常相似。

我曾断言,我们能为真值即现实世界里的真值提供唯物主义认为充分的说明。但是如果我们要解释一个斯托纳克型的命题的意义,对于一个任意可能世界 W 来说,我们需要的是关于在 W

中的真值的唯物主义认为充分的说明。而这似乎是更难得到的,因为如果我们把塔尔斯基语义学与一个任意可能世界 W 联系起来,我们将发现:它是根据一个名称指称 W 中的一个实在意味什么、以及一谓词有一组实在作为其在 W 中的外延意味着什么,去解释一句子在 W 中为真意味着什么。也就是说,我们需要诉诸于(在现实世界中)表示可能世界 W 里的实在或实在集合的语词的关系;难以弄明白的是,怎样用唯物主义可接受的方法来解释这样的跨世界关系。

然而,这是一种看待问题的非常错误的方法,因为事实上人们完全可以用(在现实世界中的)语词与在现实世界中的实在之间的关系来定义 W 中的真值。例如,对于一个形式上为「 $P(b)$ 」的句子来说, P 是一个谓词, b 是一个名称,定义就是这样的:

当且仅当有一个物体 X 由 b 指称(在现实世界里),而且有一个属性 Z 由 P 表示(在现实世界中),以及 W 是 X 存在于其中且具有属性 Z 的世界,「 $P(b)$ 」在 W 中才是真实的。

比如,“伯特兰·罗素没有头发”在任何可能世界 W 中是真实的,只要在 W 中罗素(即在现实世界里由名字“伯特兰·罗素”指示的那个人)存在而且无头发(即有现实世界里“无头发”一词所表示的那种属性)。因此,根据名称的所指关系加上表示谓项的关系,是可以给 W 中的真值下定义的,在这里,应当这样来理解,即谓项表示的不是一个集合(它的外延),而是存在于现实世界中的一种属性。^⑩

因此为 W 中的真值提供唯物主义认为充分的说明这一问题

620 第五部分

可还原为这样两个问题:(i)为名称的现实世界所指提供唯物主义认为充分的说明,(ii)为谓项的现实世界代表提供唯物主义认为充分的说明。我已讲过,在这里,谓项应被解释成代表属性而不是集合。但是(i)甚至是在说明现实世界的真值时就会产生的问题;我假定(既然我已在前面述及的文章中论证过它),我们有充足的理由相信那个问题是可以解决的。然而(ii)就有点麻烦。在关于现实世界中的真值的任何说明中,我们所需要的是对谓项与它们的现实世界外延之间关系作出唯物主义认为充分的说明。因此我们假设这个问题是可以解决的。但是(ii)需要形式上更强的某种东西;它需要一种关于谓项与属性之间关系的唯物主义认为充分的说明;而既然这些属性决定谓项的外延,而不是被其所决定,因此我们所需要的在形式上就比现实世界语义学必须给予我们的东西要强一些。然而,稍加思索就会发现,现实世界语义学也一定会给出形式上更强的东西。但是对“是……的外延”的关系作出一种唯物主义认为充分的说明又如何呢?这样一种说明必须告诉我们,我们运用语词的方式是什么样的事实种类,由于这种方式,无头发的事物之集合是单词“无头发”的外延。如何做到这一点呢?一种可能性就是,它是关于指称的一种因果理论,它会这样陈述:无头发的个体与我们对单词“无头发”的运用之间的某种因果联系能够对我们的语词具有作为它的外延的那些无头发的个体之集合作出解释。但是我们肯定没有与“无头发”的外延中的所有成员进行这种因果交流——毕竟,在遥远的行星上存在着无头发的事物,在遥远的将来也存在着无头发的事物等等。与某些无头发的个体的因果联系怎么可能促进下述问题的解决?这一问题是什么使我们

的语词“没有头发”有这样的作为外延的集合,它包括这些未来的没有头发的生物(而不包括将来的有头发生物)?我认为,唯一可能的答案是“无头发”一词的外延决定于一种属性。与谓项直接联系在一起的并不是它的外延,而是某种属性。也许无头发这一属性是通过与大量无头发的个体的因果联系从而与“无头发”一词联系起来的,就像关于指称的因果理论常说的那样;或许存在某种关于指称的非因果理论,它决定了无头发的属性是怎样与那个词联系起来的。但是不管那种联系怎样构造出来,外延与“无头发”一词的联系则是一种派生的联系:“无头发”的外延简直就是下述对象(在现实世界中)的集合,它们有这样的属性,在现实世界里,它与“无头发”一词相联系。如果不可能对属性是怎样与谓词联系起来给出一种唯物主义认为充分的说明,那么也不可能对这些集合怎样与谓词联系起来作出一种唯物主义认为充分的说明,进而也不可能对现实世界的真值作出一种唯物主义认为充分的说明。因此对可能世界的真值给出一种唯物主义认为充分的说明这一问题并不比给现实世界的真值作出一种唯物主义认为充分的说明这一任务难多少;从根本上说,这些看似不同的问题实际上就是一个问题。

现在我有了所有的构件,让我把它们按顺序排列起来。在第一节的大部分,我所做的工作就是解释一种可能的、为信念提供唯物主义认为充分的说明途径;而且直到现在,我还假定,信念应该被阐释成人与可能世界集合的关系。信念的假设对象的“别的世界”特征似乎排除了对信念关系的唯物主义说明的可能性,但是我已论证过,那种现象只是幻觉。因为我们可以说,一个人相信可能

622 第五部分

世界的一集合 p , 当且仅当他相信* 某个句子(或句子类似物) S , 而且 p 是 S 在其中为真的可能世界的集合。(考虑到可能世界的本体论)定义可能世界中的真值需要的不是信息源, 就像它不是定义真正的真值即现实世界中的真值所需要的一样: 这两方面的任务都要求我们解释: 是什么让 S 中的名称指称它们在现实世界中实际指称的一切项目, 是什么让 S 中的谓词表示它们在现实世界中实际所表示的任何属性。^⑦这无疑是一个很困难的任务; 但我知道, 以为这个任务无望完成则又没有任何理由, 事实上我认为(正如我前述论文中所论证过的), 有足够的理由断言, 它可以得到解决。

刚才对信念的详尽的阐释只是作为可能的说明提出来的。它实际上是我后面要在某些方面作出修改的说明——例如, 把所有的指称都下放到命题和可能世界中。但我不会修改已给出的说明的两个核心特征: 一个观点是信念涉及到与句子或非常类似于句子的某东西的关系, 另一观点是这些句子一定被给予了一种似塔尔斯基式的语义学(包括原始的指称理论)。在很大程度上, 我对这两点的坚持是基于我无法想象: 若不以这些观点为基石, 对布伦塔诺问题的解决会是什么样子。我并未声称已经证实: 不可能有关于信念的下述可供选择的说明, 它解决了布伦塔诺问题, 但没有上述两个特征。然而, 我还从未看到对这种可供选择的说明的清晰的描述。我上面所述的那种说明, 可能解决布伦塔诺问题, 而目前尚无其他替换的解决该问题的方法, 这一事实在我看来就是试探性地接受所述的那种说明的非常有力的理由。^⑧

在此, 读者或许会觉得, 对信念有个著名的可供选择的说明,

它要么可以解决布伦塔诺问题,要么证明这只是一个假问题。这个可供选择的说明(在我的假设的读者看来)实际上是刘易斯和斯托纳克所极力倡导的,它就是功能主义者的抉择。下一节我将考察这样问题,即功能主义是解决了还是毁灭了布伦塔诺问题。

II. 功能主义

刘易斯和斯托纳克对信念的探讨涉及到这样一个观点,即信念是一种功能状态。粗略地说,如果有机体的一种状态在有机体的心理学中起着适当的作用,也就是说,如果它以适当的方式在因果上关联于输入、输出和其他心理状态,那么这样的状态就是一种相信 p 的状态。首先,这种关于信念的功能理论似乎允许信念有独立于语言或内在表征的可能性,即是说,为什么不能是这样一种情况呢?即该有机体以适当方式处于与输入、输出或其他状态的因果联系之中,因为那状态满足成为信念 p 的功能条件,尽管那种状态并未涉及到与任何内在表征系统中的任何句子或类似句子的项目的关系。斯托纳克对此观点作了如下表述:

可以设想(不管真实与否),确实存在着这样的理性生物,它们有信念和愿望,但它们不使用语言,也没有关于它们的、有语言形式的态度的内在表征。我认为,这对许多动物都适用——甚至一些比较笨的动物——但可能有更清楚的实例。试想,我们发现了许多生物——或许在其他星球上——它们不能交流,但其行为的大部分是可根据它们曾从事过高深的理论思索这一假说加以预言的。再设想,我们以这个间接的证

624 第五部分

据来支持我们的假说:相信我们的假设归之于那些生物的信念在多数情况下可以用它们的感性输入来从因果上加以解释,而那假说归之于它们的愿望在很大程度上与它们生存的物理需要有机地关联着。最后,再设想,我们通过控制那些生物的环境来检验那一假说,比如说灌输给它们错误的信息,或满足或抑制它们的某些所谓的愿望。如果它们的行为仍像所预测的那样,我想我们应该归之于这些生物的不仅是信念和愿望的类似物,而恰好是信念和愿望本身。然而,我们并没有任何理由假设:它们用心灵语言或其他任何一种语言进行思维(1976,p. 82)。

斯托纳克的这一说法正确吗?

为了评价斯托纳克的观点,我们有必要让关于信念的功能理论的观念比我前面关于它的粗略描述更清楚一些。根据前面的粗略的描述,有机体的状态如果以某种适当的方式与输入、输出或其他心理状态在因果上关联起来,那么它就是相信 p 的状态。但一当我们知道别的什么样的心理状态是它必须在因果上与之关联的状态,而且知道它与它们关联着的“适当方式”是什么,我们便无权决定这种关于信念的功能理论是否需要一种内在表征系统。(例如一种关于 p 的信念状态必须在因果上与之关联的一种别的状态为什么就不能是有关于命题 p 的内在表征的状态呢?)在考虑斯托纳克关于另一星球的生物的故事时,必须牢记这一事实。在斯托纳克所描述的情境中,有一点非常清楚,即我们不能绝对肯定:那里的生物曾以任何一种内在的表征系统进行思维。但是还有一点也很明显,即我们不能绝对肯定:它们有信念和愿望。我们之所以

不能肯定这一点,是因为(a)我们没有观察到它们所有的实际行为,更不必说所有可能的行为了,(b)即使我们观察到了,但关于信念和愿望的断言并不直接就是关于行为的断言,它是关于那种行为怎样产生的断言。当然,为了知道有机体是否有信念和愿望,我们不需要关于行为产生方式的详细知识——功能主义的全部目的在于提供一种关于有机体内在状态的完全抽象的表征,因此关于这种有机体内在状态的某些种类的信息对于它是否有信念和愿望的问题变得无关紧要了。但是我们的确需要关于行为产生方式的一些信息,以便了解它有信念和愿望:功能主义者关于有机体内在状态的断言是关于那种有机体内在状态的断言,而不能被还原为关于有机体实际和可能行为的任何断言。(我希望这个事实是常识。这一点在刘易斯对于功能主义理论的明确细致的阐释中表现得十分清楚,很快我就会作出阐释。)因此,当我们得出结论说,另一个星球上的生物有信念和愿望时,我们如此推断是因为这是我们对于它们的行为所能找到的最好解释。但现在,如果我们能知道这些生物有信念和愿望,是因为信念和愿望理论是关于它们行为的可得到的最好解释,那么我们为什么不能通过同样的方式知道这些生物有内在表征系统呢?第一种推论是否比第二种有更好的归纳根据,不是随便凭藉直觉能予以解决的:这取决于能否为信念和愿望提供充分的功能说明,根据此说明,信念和愿望不需要内在表征。所以引自斯托纳克的最后一句话(或者,如果你愿意这么说的话,最后两句话的合取)就是在回避所争论的问题。也许它的结论是对的,但我们只能通过对信念的功能理论是什么样子作出详尽考察才能表明这一点。

626 第五部分

对这个问题的全面讨论需要一分为二。这样一个讨论部分牵涉到这样一个问题,即任何合理的信念理论是否必须具有关于明显成为其组成部分的内在表征的假定。我非常倾向于肯定这一点,我认为,我在第一部分关于核心信念的论述和第三部分关于命题的论述为这种主张提供了某种证据;但这些主张的最终证据则存在于心理学理论的详尽阐发中。在这一部分,我不打算论述这些问题,而将从事一项更平常的工作,我想说明:对信念和愿望理论的功能主义探讨不能提供取代信念需要某种表征系统这一观点的观点。

这样说,我的意思不只是说,功能主义与存在着某种内在表征系统(包括自然语言的句子和句子类似物)这一说法是相容的,刘易斯或斯托纳克都不会否认这一点。他们都肯定赞成说:如果我们打开一个人的大脑,发现里面有一块写满了各种英语句子的黑板,如果我们还发现黑板上那些句子的出现与这个人的行为正好吻合,那么这将是主张(至少对那个人来说)信念涉及到内在表征的强有力的理由。这当然不会与功能主义发生冲突,因为我们将发现的不过是,(在那个人身上的)以适当方式与输入和输出以及其他状态因果关联的这种状态碰巧是涉及到内在表征的状态。因此无可争辩的是,那种功能主义与内在心理表征假说是相容的。但我想为另一种更强的主张作出论证,我推测它是为刘易斯和斯托纳克所否定的主张(尽管我这样推断不是根据他们明确宣称的东西,而更多是根据他们讨论的基调)。我想刘易斯和斯托纳克所坚持的是,不打开人们的头脑,我们几乎没有或完全没有理由认为,相信某物与内在表征系统有关;因此信念的确牵涉到内在表征

系统这一假说就只是没有根据的神经生理学的猜测。在第一部分,我曾初步论证过,它不只是没有根据的神经生理学的猜测,我断言,内在表征假说是解决布伦塔诺问题的已知的唯一的可能性,而且这本身为尝试性地接受它提供了充分的理由。但在论证这一点时,我忽略了功能主义方案的可能性。因此我们现在必须研究的是,功能主义方案是否改变了这个结论。我认为它没有改变,而且那恰恰是当我说功能主义没有为信念需要某种表征系统这一观点提供替代品时所要表达的意思。

一言以蔽之,功能主义没有解决布伦塔诺问题,它也没有以任何方式清除这个问题。这就是这一部分要论证的主张。

在心理学上,功能主义的全部意义在于提供一种关于心理属性的完全抽象的表征,即一种与特定有机体的物理结构细节没有太密切联系的表征;事实上,一种功能理论就是要保证,如果两种有机体在一种合适意义上是心理上同型的,那么它们显然具有同样的心理属性,不管它们在那些与建构心理同型无关的物理结构方面多么不同。功能主义者不会说,有机体的物理结构与其心理属性无关,所有的功能主义者几乎都是唯物主义者,即是说他们都相信所有心理属性都离不开物理的实现。^⑨这就是说,比如对任何有机体 X 和任何时间 t 来说,只有当 X 在时间 t 具有实现 F 的物理属性 $R(F)$ (在一种很快会得到阐明的意义上), X 才可能有 1-元心理属性 F 。但是不同的物理属性在不同有机体(或者不同时间的同一有机体)中可以实现相同的心理属性;因此结构迥异的有机体可以有相同的心理属性。他达到了从特定有机体的物理结构中作出抽象的目的,^⑩这是因为功能主义者承认了不同的实现这

628 第五部分

一可能性——而不是因为他全然放弃了实现的要求。

弄清什么是心理属性的实现和心理属性本身是什么对于后面的论证是很重要的。下面即是 D. 刘易斯^②的说明,只有少数几点例外(脚注中述及的),但它们并不影响眼下的论证。

假设 H 是一种心理学理论,它打算在任何时间 t 适用于 t 下的属于 K 类的一切有机体。(例如 H 可能是一种旨在适用于所有成人或所有能感觉疼痛的有机体或有理性的存在的理论。)^③为简单起见,我们假设 H 得到了有限的公理化,于是我们便能用一个公式来代表它,我把它简化为: $A(x, t)$, 其中 x 是表示有机体的变量,而 t 是表示时间的变量。[因此当且仅当以下论断成立,那么该理论在其设想的范围内就适用于所有有机体:

对于任何 t 和 x 来说,如果 x 在 t 时属于 K 类,则 $A(x, t)$ 。
正是这后一主张,而不是 H,有严格意义上的真或假,所以你可能宁愿把这后一主张而不是 H 称作“理论”。]在对心理学理论进行粗略的阐述时,我们常常省略变量(及初始量词),但它们必须如此含蓄地加以理解:我们说“疼痛有如此这般的因果作用”,我们的真实意思是:“对于任何 t 和属于 t 时的 K 类的任何 x 来说,疼痛对于 t 时的 x 有如此这般的因果作用”。如果我们不用这种方式表达该理论,那么我们就无法正确地为实现这一概念下定义。^④

假定 H 中的特定的心理学的^⑤原词是 T_1, \dots, T_n ; 那么我们就可以将 H 写作 $A(T_1, \dots, x, t)$ 。为简明起见,我假定 T_1, \dots, T_n 都是谓词。^⑥我们不妨说,属性的 n 项 (P_1, \dots, P_n) 在有机体 X 中于时间 t 实现了 H, 当且仅当公式 $A(Y_1, \dots, Y_n; x, t)$ 对于 (P_1, \dots, P_n, X, t) 是成立的;而且这样一个 n 项在 X 中于 t 时独自实现了 H, 如

果它在 X 中于 t 时实现 H , 而其他任何 n 项则相反。

如果 H 是一个带有 n 原始心理谓词的心理理论, 我们就可以用之来定义 n 功能属性。例如假设 T_j 是一谓词, 表示的是有机体这样的 1-元属性, 即是有有机体 t 时有它、在其他时间无它时所意指的那种属性(刘易斯将这种属性称为“状态”,^⑧但我宁愿将这个词留作他用)。如果 T_j 是这种谓词, 那么与 H 有关的第 j 个功能属性就是属性 F , 可这样予以定义:

(2) X 在 t 时有 F , 当且仅当有某种 1-元物理^⑨属性 P , 以至于:

(i) P 是 H 于 t 时在 X 中的单独实现的第 j 个成分。

(ii) X 在 t 时有 P 。

如果 F 是与 H 的有联系的第 j 个功能属性, 那么我们就可以说, F 于 t 时在 X 中的实现直接就是 H 于 t 时在 X 中单独实现的第 j 个成分。据此及(2), 我们可推出:

(2') X 于 t 时有 F , 当且仅当有某种 1-元物理属性 P 存在, 以至于:

(i) P 在 X 中于 t 时实现, 且

(ii) X 于 t 时有 P 。

这一方法使我们能够澄明前面五段的一般性评论的意义。功能主义关于疼痛所断言的是, 疼痛的属性是一种功能属性,^⑩它通过(2)(或通过(2)中的对应词, 如将“物理的”一词换作“非功能的”——见注^⑩)与某种理论 H 相联系。这样理解功能主义, 我们就能够准确理解一般性评论中所诉诸的各种模糊概念(如心理同型的概念), 同时我们还能证明下述真正重要的主张: 如果唯物主

630 第五部分

义是真的,那么对于一有疼痛心理属性的有机体来说,它必定有在它身上实现那心理属性的某种物理属性。

现在让我们将这种方法应用于“信念”这一例子中,在此,我们像前面一样假设,信念是有机体与命题之间的一种关系。这一来,信念就是与某种理论 H_b 有联系的功能关系,在该理论中,术语“相信”出现了,比如说作为第 j 个心理学术语。我们当然不能将图式(2)用于定义这种功能属性,因为(2)仅在定义与 1-元谓词相对应的属性时才有意义,但如何将它泛化到另一种实在中则是足够清楚的:

(3) X 具有与 t 时的 p 的 F ,当且仅当有某种 2-元物理属性 R ,以至于:

(i) R 是 H 于 t 时在 X 中的一单独实现的第 j 个成分。

(ii) X 具有与 t 时的 p 的 R 。

(2) 对于 2-元功能属性的事例的其他泛化都是不可能的。

关于这一点,要注意的重要事情是,它用量词对人与命题之间的物理关系从存在上作了限定。如果人与命题之间没有适当的物理关系,那么根据(3),此人对该命题不具有功能关系 F 。功能关系 F 本身并非物理关系;但如果 F 每次将一有机体与一命题关联起来,那么一定存在某种物理关系 $R(F)$,它实现 F ,而且将有机体关联于命题。这一来,即使我们把信念当作功能关系,那么我们不得不解决布伦塔诺的问题:我们必须说明,在人与命题之间存在着物理(非功能性)关系。这就是我前面所说的功能主义既不能解决也不能消除布伦塔诺问题的意思。它说到的与该问题有关的唯一的东西就是没有人曾经怀疑过的东西,那就是我所相信的、把我与

命题关联起来的物理关系可能不同于别的人所相信的把他们与命题关联起来的关系;而且甚至不同于把其他人与他们所相信的命题关联起来的关系,以及不同于把 20 年前的我和我那时所相信的命题关联起来的关系。换言之,对于不同有机体,用不同方法解决布伦塔诺问题的确是合理的;但这并没有铲除在一有机体中解决该问题的需要。我无意于降低下述看法的重要性,即我们没有必要在不同有机体中用同样方法解决布伦塔诺问题,这一事实对于第一节表达的信念模式的合理性是非常重要的,因为没有人能合理断言:相信*的关系对于不同物种来说在物理上是同一的。我的意见是,承认这种跨物种变体的可容性及重要性,无论如何没有削弱对第一节所提出的信念模式的尝试性论证的力量。依然真实的是,如果其他模式不能说明人与命题之间的关系是什么样子,那么就我们应该尝试性地接受那里所提出的模式。

我认为,我刚给出的论证是极其明显的论证,人们可以预期,它对于任何一个稍微思考过功能主义的人来说,也将是明显的。然而在交谈中我发现,该论证的结论几乎是所有功能主义者所反对的:他们几乎都相信功能主义在某种程度上瓦解了第一节的论证。我猜想,一个原因是从功能主义滑向极端行为主义的趋势,根据这种行为主义,有机体中没有什么与决定其心理属性有关。另一原因也许是一种莱布尼兹式的趋势:它认为关系没有 1-元属性那么真实——这并非由于有意识的信条,而仅仅是因为这样的事实,即“关系”一词不同于“属性”一词,可兼作表示某种集合的词。但错误还有第三种来源,这是我这里要考察的,它同样是初步的,但我认为,存在着严重的、难以捉摸的混乱,这使人们不可避免

632 第五部分

地犯错误。

错误的第三种根源在于,即使是在那些旗帜鲜明地倡导信念是命题与人之间的关系的人们中间,有时似乎有这样的倾向,即陷入“正字法不规则”观:该观点认为,一种充分的信念理论能够将“X 相信罗素没有头发”,“X 相信要么罗素没有头发要么雪是白的”等等当作原始的 1-元谓词,而且在完全没有 2-元谓词“X 相信 P”的情况下,也能如此。(术语“相信那”以两种 1-元谓词的形式出现,这一事实从一种理论观点来看,也许并不重要,仅是一种纯粹的正字法不规则,二者都含有“罗素没有头发”这一事实可能同样是一种正字法不规则。)严肃地对待这一观点并不容易。但我们假设它是真的,接着会怎样呢?噢,那种说明将干净利落地铲除对人与命题之间物理关系的需要:既然它没有谈及人与命题之间的心理关系,因此人与命题之间的物理关系在实现中显然是不必要的了。但是这当然没有驳倒我在本节所陈述的观点,即如果你从关系上解释“信念”,你就需要该信念关系的物理实现。

尽管这一错误很露骨,但我认为,它是很容易犯的。事实上,在本部分的开头一段,当我试图促发下述观点即功能主义铲除了对表征系统的需要时,我发现我自己在以很明显的正字法不规则观谈论:“如果那种状态在有机体心理学中起着适当的作用”,我便说“有机体的状态是一种相信 p 的状态。”现在,为了使这一点明白易懂,这里的字母“p”就必须被理解为一具体句子的缩写——比如说“要么罗素没有头发要么雪是白的”。什么是相信要么罗素没有头发要么雪是白的这一状态的“适当作用”呢?我并不是说这种谈论没有意义:如果我们对存在于有机体和命题之间的相信关系给

予功能说明,那么这样一种说明肯定有关于相信“要么罗素没有头发要么雪是白的”这一特定命题的状态的含义。但是我认为,相信这一特定命题的状态的“适当作用”这一说法清楚地表明:我们可以为这一特定状态下一个直接的功能定义。而且它还清楚说明:这种适用于“疼痛”的程序也能适用于“相信要么罗素没有头发要么雪是白的”,换言之,它表明:“相信要么罗素没有头发要么雪是白的”能被看作由图式(2)所定义的一个功能的 1-元属性。然而那就是正字法不规则观,因为它预先假设,理论 H 包含着表征这种属性的原始术语,而根据这种理论,相信要么罗素没有头发要么雪是白的应能从功能上得到定义。这里的启示在于,如果你想避免正字法不规则观,那么你不应该认为,对具体的句子 p_0 而言,“X 相信的那个 p ”在功能上是可以定义的,就像“X 处于疼痛之中”一样:你应该把它们看作是从一种关系谓项“X 相信 p ”中非功能地被定义的,而这种谓项则是由(3)从功能上定义的。这就意味着,我们必须求助于有机体与可能世界集合之间的物理关系。

我相信,滑进正字法不规则观这一倾向存着更深层次的根源;它与求助于心理状态的功能理论有关,在这里“状态”不是指刘易斯的那种意义(试比较注^⑥),而是“内在事件的类型”这层意义。这些功能理论并没有什么错误,但是必须小心予以对待,正如我要解释的那样。然而,读者可以跳过这一解释而直接进入第三部分。

我们先来思考疼痛这一情况。语言学的用法使人想到下述观点

- (4) X 在 t 时感到疼痛,当且仅当内在事件 o 发生了,以至于:
- (i) X 在 t 时感觉到 o ,

634 第五部分

(ii)o 是一种疼痛。

现在假设我们给出一种理论 $H'p$, 它是关于用于(4)之分句(ii)中“处于疼痛”这一类型属性的。不难看出, 根据这一理论, 从功能上定义处于疼痛的恰当方法:

(5)o 是一种疼痛, 当且仅当内在事件具有某种物理属性 P , 以至于:

(i) P 是 $H'p$ 在 X 中于 t 时的一个单独实现的第 j 个组成部分

(ii)o 具有属性 P 。

(在这里, t 是 o 发生的时间, X 是 o 出现于其中的有机体。)将它与(4)放在一起, 我们就有了:

(6) X 在 t 时感到疼痛, 当且仅当有一个内在事件 o 和一种物理属性 P 存在时, 以至于:

(i) X 在 t 时感觉到 o ,

(ii)o 具有属性 P ,

(iii) P 是 $H'p$ 在 X 中于 t 时的一个单独实现的第 j 个组成部分。

现在假设我们引入“状态”一词, 以之表示内在事件的物理属性。(即是说我们用它表示状态-类型; 如果 P 是一种状态-类型, 那么状态-标记就是有 P 的内在事件。)如果我们再引入“在 t 时存在于 P 之中”这一表达式, 以之表示“在 t 时感觉到类型 P 的某种内在事件”(即具有属性 P), 那么我们可这样改写(6):

(7) X 在 t 时感觉到疼痛, 当且仅当有一种物理状态 P , 以至于:

(i) X 在 t 时处于 P 中,

(ii) P 是 $H'p$ 在 X 中于 t 时的一个单独实现的第 j 个组成部分。

因此寻找这种疼痛理论在 t 时的 X 中的一种实现形式这一问题就是于 t 时在 X 中寻找疼痛的一种物理状态的问题;即是说,在状态-类型意义上的物理状态就是在内在事件的属性这一意义上的物理状态。

在这里关于疼痛所说的一样可用之于信念。(4)的对应物是:

(4') X 于 t 时相信 p , 当且仅当有一种内在事件 o , 以至于:

(i) o 出现于 t 时的 X 中,

(ii) o 是对 p 的信念。

从功能上定义的这一关系是一个信念,类似于上面的(5),将这一结果同(4')结合起来,我们便有了(与上面的(6)类似):

(6') X 于 t 时相信 p , 当且仅当有一内在事件 o , 且有内在事件与命题之间的物理关系 R , 以至于:

(i) o 于 t 时出现在 X 中,

(ii) o 具有与 p 的 R ,

(iii) R 是 $H'b$ 在 X 中于 t 时的一种单独实现形式的第 j 个组成部分。

到此一切都很顺利,结果已很清楚了:在运用关于这种信念的心理学理论时,我们必须探寻的是内在事件与命题之间的关系的物理实现。现在假设,我们想要引入“状态”一词,来表示内在事件的1-元属性。这样做毫无异议,因为对每个句子 p_o 而言,具有与 p_o 的 R 存在着一种1-元属性,我们可以称之为一种状态(意思当然是一种状态-类型)。而且我们可以继续界定:有机体处在这样

636 第五部分

一种状态中意味着什么: X 于 t 时处在状态 Q_0 中, 当且仅当在 t 时的 X 中出现了某种 Q_0 类型的内在事件(即具有属性 Q_0), 但是如果我们现在想用状态-类型重新对(6')进行公式化, 那么我们会得到:

(7') X 在 t 时相信 p_0 , 当且仅当有一种物理状态 Q_0 , 以致于:

(i) X 在 t 时处于 Q_0 之中,

(ii) Q_0 的相关组成部分是 $H'b$ 在 X 中于 t 时单独实现的第 j 个组成部分。

(Q_0 是具有与特定句子 p_0 的 R 的属性, 所谓 Q_0 的相关成份, 我意指的是关系 R_0 。)这就是把状态-类型的话语引入信念的功能理论的正确方法。

但我猜想, 很多人并未透彻地去想那个问题。看到关于疼痛的功能主义像(7)那样表述, 他们便直接跳到这样的结论: 根据状态阐述的信念理论应这样解读:

(7*) X 在 t 时相信 p_0 , 当且仅当有物理状态 Q_0 , 以至于:

(i) X 在 t 时存在于 Q_0 中,

(ii) Q_0 是 X 中的 $H'b$ 在 t 时的单独实现的第 j 个组成部分。

以这种方式阐述, $H'b$ 就必须解释成这样一种理论, 在该理论中, 表示处在相信 p_0 的状态(表示特定句子 p_0)的术语以原词形式出现。这当然就是正字法不规则观。

总而言之, 主张功能主义要么解决要么解除布伦塔诺问题这一看法源自下述混淆: 即把信念的关系理论和信念的正字法不规则理论搞混了, 根据后一理论, 信念根本就不是一种关系。我确信, 任何功能主义者会断然拒绝这一观点: 信念的心理学理论应采

取正字法不规则的样式;但如果人们认为相信的状态有功能作用,而不是基于信念关系的功能作用的派生物,那么他坚持的肯定是正字法不规则说明。

Ⅲ. 摒弃命题

在第二部分的开头,我援引了斯托纳克的论点,它主张:信念的功能理论排除了对内部表征系统的需求。我曾提出,为了完整地评价该论点,有必要澄清的是,信念的功能说明是怎样的。我只是在最低限度上做了这一工作:我只关注信念的功能说明的那些特征,它们来自于功能主义的本质以及信念涉及到人和命题之间的关系这样的(假设的)事实。但需要弄清的还有,信念的功能说明到底是什么样子。因为我们知道,根据关于信念的任何功能说明,信念关系是一种与某种理论 H_b 相关联的功能关系。因此弄清信念的功能理论是什么在一定程度上涉及到对 H_b 是什么这一问题的澄清。我非常倾向于认为:任何充分的 H_b 理论必须清楚明了地假定一种表征系统,这为人们相信有独立于对布伦塔诺问题的任何思考的内部表征假说提供了理由。然而在此我不想阐述这一主张。我强调的只是对立的方面:既然(如我所论证的)我们似乎需要一种表征系统以解决布伦塔诺问题,那么我们就应该对清楚明了地把这一系统合并到我们的心理学理论 H_b 中带有太多的勉强,如果这样做有帮助的话。

明确地把表征系统结合到心理学理论中的一个优点就是:它将使我们将信念和愿望当作人与命题之间关系的大部分优

638 第五部分

点,而不会有伴随的不利性。假设由于这样或那样的理由,我们不想从量上限定命题;如果是这样的话,我们显然不能说:

(8)她相信许多关于他的许多事情,而其中没有一样是赞美性的;或者说

(9)如果不相信有关一对象的各种事情,那么没有人能知觉到它。

很明显在(8)和(9)中被量化的“事情”就是命题。然而在说出(8)和(9)时,我们的目的为什么不能通过量化信念*的对象而非信念的对象来加以实现,比如如果不相信*有关一对象的各种事情,那么没有人能知觉到它,对此,我看不出有什么理由;在这里被量化的“事情”不是命题,而是内部表征系统中的句子。当然,只有当我们把内在表征假说明确地整合到我们的心理学理论中,我们才能在这种理论中接受(9)的这种被重新表述的观点。

我认为,星号至此成了累赘,因此我们不妨引进下述术语约定,即像我们以前使用“相信*”那样使用“相信”一词。根据这种话语方式,信念的对象是句子或句子类似物,而且这些句子或句子类似物都有内容或意义。与第一部分第一段中的提法相反,这种话语方式并没有真的铲除布伦塔诺问题:该问题会作为对享有内容作出唯物主义说明这样的问题再度出现。除非对内容能够作出这样的说明,否则我们关于信念的大多数谈论(例如某些信念是关于凯撒的,某些信念是真的等)就毫无意义。然而在第一部分的结尾,我曾提出:对内容作出唯物主义的说明这一问题似乎是可以解决的:方式之一就是对真值作出塔尔斯基式的说明,辅之以指称理论。

我用“相信”一词意指“相信*”，这与英语中“相信”一词的用法并不十分吻合，在英语中，“相信”和接从句的即“相信 that”几乎是同义词，而且我们可以说，即使不懂英语的有机体也会相信雪是白色的。因此让我们引进一个新的技术术语，即“相信 that”，它将有这样的好处，即方便“相信”和“相信 that”在英语中的运用：我们可以说，若一个人相信他内部表征系统中的某个句子，而这个句子译成英文是 S，那么他便相信 that p (p 在这里是一个英语句子 S 的缩写)。事实上，“相信 that”一词是用来表示有机体与英语句子之间的一种关系的。须知，在“相信 that”的定义中所用的翻译概念是一个松散和不正式的用语。因为即使我们中的那些对奎因的激进的不确定性命题反感的人也一定会意识到，在两种具有非常不同的结构或表达力的语言间的翻译是极不确定的。(部分的理由在于，语言 L_1 中的句子没有必要精确地和语言 L_2 中的任何句子同义，因此把 L_1 译成 L_2 的人就不得不满足于意义的近似的一致性，一个人选择哪一近似性则取决于复杂的语用学上的考虑。)既然“相信 that”这一用语应该应用于这样的有机体，它们的内部表征系统无疑与我们的有很大区别——如黑猩猩、火星人以及人(如果沃夫是对的话)，其口语或书面语言都明显有别于我们的——那么若没有认识到与“相信 that”的定义有关的翻译是非常不确定的，那将显得非常荒唐。由于这个原因，上面所定义的“相信 that”本身也是一个非常不确定的用语，因此是我们在心理学理论化中应予避免的概念。我们应该使用的是“相信”这一概念(即“相信*”)，也许要与应用到所相信的句子中的各种概念一起运用。[在我们的心理学理论中避免使用“相信 that”(如上所定义的)的

另一个理由已经在第一部分提到过——试比较(B)]。

我已经介绍了关于“相信”和“相信 that”的用法的约定,据此,第一个术语把有机体和它们自己的表征系统中的句子关联起来了,而第二个术语把有机体和英语中的句子关联起来了。这些术语上的约定无意于排除这样的可能性,即在分析信念和愿望时,我们需要命题:在我们对所相信和期望的句子的意义作出说明时我们可能需要命题。不过,我自己的观点是:

(a)关于命题的话语最好予以避免(除了作为可有可无的言语方式),除非它表明有不能另作他用的用途;而且可疑的是这个条件会被满足。

(b)这种话语使我们承诺了那些很可能是错误的(独立于本体论的顾虑)语义学命题。

我们先看看(a)。我认为,毋庸置疑的是,人们不应不必要地假定实在:这样做将会沉溺于“毫无根据的本体论思辨”。没有引进命题的特殊的必要——至少没有必要引进根据可能世界解释的命题——我在第一节末尾的讨论中曾暗示过这一点。在那里,我论证说,如果有人想用可能世界的观念,如果他把命题解释成可能世界的一组集合,那么一个特定命题与一个句子的关联就不会引出什么问题,只要我们有语言的充分的真值理论语义学。这一观点的另一表述方式是,关于命题的话语对语义学不会增添任何好处;在真值理论语义学中已包含有对语义学有用的一切东西。(我讲的当然是把属性而非集合归之于谓词的真值理论语义学。)换言之,如果不说一个人被相关于可能世界的集合,在那里,罗素没有头发(或者被相关于由可能世界所构成的精致的命题),那么

为什么不说他被关联于这样的句子,即它由表示罗素、并与代表无头发的谓词相连接的名称所构成?

有人可能这样来回答这个问题,即认为(我想不可能)名称不是不严格的指示符。但这种回答即使充分回答了刚提出的问题,但也没有推翻我试图阐述的观点(正如注释^⑥中已述的那样)。因为如果“罗素”是不严格的,那么相信罗素无头发就是相信某个含有一种确定摹状词的句子,即某种具有“ $G(\text{the } x \text{ 以致 } Fx)$ ”这样形式的句子,在这里“ F ”是罗素的个体化的摹状词, G 代表无头发。对这个句子的真值理论分析包含有可能世界的分析中所包含的全部的语义信息,因此为什么要引入可能世界呢?它们似乎再一次成了没有任何语义学作用的、多余的本体论累赘。

在(b)中有两点值得一提。第一,如果一个人谈及命题,而且还假定命题一定能根据可能世界来解释,那么他不得不假设:专名和“温度”这样的理论术语是不严格的指示符;这种假设从最好的方面看是有争议的。第二,不管人们是否坚持可能世界的方案,对命题的谈论则促使他承认同义词的普遍关系,而是否有这种普遍关系则一点也不明显。在第六部分,我将阐发这些观点,尤其是第二点。结论是,对命题的假定不仅涉及到无根据的本体论思辨,也牵涉到极有争议的语义学思考。在我看来,作为命题假设之基础的本体论的和语义学的主张看起来远比内部表征系统的假设可疑,这主要是因为最后一个假设对于解决布伦塔诺问题是一个必要的假设。

因此我们不妨说愿望和信念都不是对命题的态度,而是对内部表征系统中有意义的句子的态度。对内部表征系统中的意义的

任何充分说明在一定程度上可能就是一种真值理论语义学。在第六部分中我将提出:意义还有另外一个方面。但我首先想尽力阐述的是关于内部表征系统的观念。

IV. 评内在表征假说

我已提到,信念和愿望涉及到内部表征系统;但我承认,就真有语言的有机体来说,内在表征系统要么就是语言,^②要么包括语言。这种综合的主张似乎很令人费解:如果表征只是存在于自然语言中,那么为什么把它称之为内在表征呢?如果内在表征存在,那么说表征在自然语言中又是什么意思呢?

对这些问题的回答关键在于把类型和标记区别开来。我已谈到了相信(即相信*)句子-类型的有机体。但在第一部分,我又说过(撇开核心信念的复杂性),当且仅当一个人在推理、思考等过程中以适当方式运用句子-类型时,他才会相信它;而运用句子-类型的唯一方法就是运用它的某种标记。很显然,为了相信(或衷心相信)一个句子,我没有必要使用它的口头或书面标记:没有人会写出他所有的信念(或核心信念)。因此如果我相信我的语言中的句子,那么我所使用的一定包括那些句子的内部标记。这就解释了:即使所有表征都在自然语言中,为什么我们必须说到内部表征:某些标记肯定是内部的。

这便把我们带到了第二个问题:说内部表征是自然语言中的表征有意义吗?当且仅当我们说内在标记与书面的或口头的标记是同一种类型有意义时,它才有意义。有人可能会说,任何关于内

部标记与口头或书面标记之间的类型同一的说法都是极不合理的,因为假设内部标记与口头或书面标记之间在物理上非常相似毕竟是很荒谬的。但这自然是一个糟糕的论点:口头标记与书面标记毕竟能说成是同一类型的,然而口头标记和书面标记二者几乎没有什么物理上的相似性。

我自己的观点是:我们是否说内在标记在类型上同一于口头或书面标记,这在一定程度上属于语词问题,但引发它的则有一些有趣的经验问题,其中最重要的是语言的发展在什么程度上(以及以什么方式)涉及到概念的发展。笼统地说,如果(在我看来相当可能)学习第一语言涉及到扩展最初的表征系统,包括所学语言的同型复制,那么我认为,把同型作为建立内部标记与口头或书面标记之间的类型同一的标准是非常自然的事情。然而这里的争论非常复杂。(参阅哈曼,1970,关于或多或少支持前类型同一观的有趣讨论,以及福多,1975, ch. 2,关于支持另一方面的有趣讨论)。在这一点上,我对哈曼的同情远胜于对福多的同情,但在本文关于此问题的其余讨论中,我会尽量保持中立。

但还有其他一些问题仍值得考虑。首先,我谈到了“内部标记”。这是些什么样的实在呢?假如它们是某种内在事件,那么是哪一种呢?第二,尽管我回避这样的问题,即说内部标记与口头或书面标记是同一类型意味着什么,但仍然有两个内部标记是否属于同一类型这样的问题。既然心理学理论显然需要使用不同标记之间的类型同一性这样的概念,那么这个问题就不容忽视。第三,存在着从句法上描述内部标记的问题——说内部标属于一种特定的句法范畴是什么意思(例如它成为句子标记意味着什么)?再

644 第五部分

者,在发展心理学理论的同时,我们需要诉诸于对内在标记的句法描述,因此这是一个相当紧迫的问题。

在处理这些问题时,有两种可能的战略。第一是尽力在阐发心理学理论之前回答上述问题,我认为这是错误的。这种战略是一种糟糕的战略,因为如果不做大量神经生理学思考,就很难弄清它是怎样实现的,例如我们显然得详述两种神经生理属性 P_1 和 P_2 ,以及一种神经生理关系 R ,而且还得说,内在事件是一种表达式-标记,当且仅当它有 P_1 ,而且它是一种句子标记,如果它除此而外还有 P_2 的话,最后还得说,两种有属性 P_1 的内在事件属同一类型,如果一个具有与另一个的关系 R 。在我们现有知识的条件下,详述 P_1 、 P_2 及 R 并不一定是我们有资格去从事的工作。在阐发心理学理论之前,我不知道我们怎么可能有完成它的条件。

第二种战略是我所提倡的,即把表征系统的句法作为心理学理论的组成部分加以阐发,于是在对上述问题作出功能主义回答时,我们可以运用这种心理学理论。这明显是一条可以继续走下去的正确的途径,因为它简直就是心理学理论应该从功能上加以阐释的普遍规律的一个例示。

让我对这种功能主义探讨所涉及到的东西再作一点阐述。如果有人要详细地写出假定了内在表征系统的理论 H ,那么这样一些谓词像“ x 是表达式-标记”,“ x 是句子-标记”,“ x 和 y 是同一类型的表达式-标记”,就会要么作为原词出现,要么可根据别的句法原词加以明确定义(在集合论或高阶逻辑之内)。当我们将 H “功能化”时,即是说, H 的实现是这样一些属性和关系的 n 项构成,它们既符合于该理论的句法谓词,又符合于更直接的心理学谓

词,那么这些句法原词就应包含在心理学原词之内。如果我们以这种方式将该理论功能化,那么对上述问题的回答便可以从理论 H 中顺利地解读出来。^⑨例如,为了简单起见,假定“x 和 y 是同一类型的标记”是 H 的一个原词。如果我们想知道,有机体 X 的两个内部事件在 t 时成为同一类型的标记意味着什么,那么答案很简单:c 和 d 是同一类型的标记,当且仅当有这样一种物理关系 R,它是 H 在 x 中于 t 时的一种单独实现的适当成分,而且 c 具有与 d 的 R。^⑩(如果“x 和 y 是同一类型的标记”不是一个原词,而是一个被定义项,答案则稍微复杂些;但仍然可以通过这个理论而得到解答。)要旨在于,在阐发心理学理论时,没有必要详细说明 R 是什么;我们可以把它留给未来的神经生理学。(此外,我们可以承认,在不同的有机体内,存在着不同的物理类型—同一的关系。)我们把这个问题的留给未来的神经生理学这一事实并不意味着我们的理论本身有任何不清晰之处,从某种意义上说,该理论无疑详细说明了:一个表征系统内的两个内在标记是同一类型的标记是什么意思。

关于这些假定了内在表征系统的理论还有最后一点须予以说明。我已说过,内在表征系统的语法应当在关于信念和愿望的心理学理论中予以明确表述。关于内在表征系统的语义学是否也应作为心理学理论的一部分加以陈述呢?这取决于我们要求心理学理论干什么。如果心理学的任务是要陈述:

(i)有机体在受到感性刺激的情况下其信念和愿望演化所根据的规律,以及

(ii)这些信念和愿望影响其躯体行为所根据的规律,

646 第五部分

那么,我认为,在陈述心理学规律时,我们显然没有运用表征系统的语义学的必要;就心理学而言,内在表征系统中的句子也许是毫无意义的。这不是观察心理学理论的唯一方法——关于信念-愿望心理学的一种更广泛的概念将在下一部分予以介绍,语义学观念在那里将有名副其实的作用。但是至少在目前有必要强调一下狭义的心理学的心理学。因为我们已知道,内在表征系统的句法和类型-同一性条件在功能上应看作是由它们出现于其中的心理学理论所描述的,而且这种理论可以说是狭义的心理学的心理学,即是说是并没有运用关于表征系统中的句子的任何语义描述的那种心理学。这一点很重要,因为它意味着,表征系统的句法和类型-同一性条件在原则上是可以不考虑那系统中的句子的意义而得到确定的。

V. 真值

关于内在表征系统的意义理论是什么样子呢?在第1部分和第3部分中,我曾暗示过我关于这个问题的看法的一个方面:关于内在表征系统的意义理论在某种程度上一定是由多少有点塔尔斯基式的真值理论语义学构成的。如果我们不为内在表征系统给出一种真值理论的话,那么我们就无法理解下述观点的意义,即我们的有些信念是对的,而有些则是错的;我认为我们的确希望能够弄清这一观点(基于已简述的理由)。而且,我所知道的唯一没有明显缺陷的真值理论就是塔尔斯基的那种理论。因此结论是,我们有必要为内在表征系统提供一种塔尔斯基式的语义学理论。G. 哈曼最近写了一篇文章,其中的观点我大部分赞同,但他对这一点

提出了异议：

关于我们以之思维的任何表征系统的意义复合理论，不论该系统是心灵语言还是英语，是没法找到理由的。（哈曼，1975，p. 286）

然而，据我推想，如果关于真值的概念有意义的话，那么真值与意义一定以这样的方式关联起来：“凯撒越过了卢比肯河”这句话的直值是来自于这句话的意义加上凯撒越过了卢比肯河这样的事实。从这个意义上说，意义理论一定包含着真值条件理论。而且就我看来，真值条件理论一定是类似于塔尔斯基使之声名远播的那种复合理论。

哈曼对复合语义学的批评是基于一个重要的洞见。他在文中指出，像戴维森那样的哲学家，都有一个严重的问题，即他们把关于真值条件的知识当作是语义学中必不可少的东西：

戴维森（大概）会说，说者理解[“雪是白的”这句话]是根据这样的事实，即他知道，当且仅当雪是白的，它就是真的。困难……在于[说者要知道这样的事情，他]必须通过某种方式向自己表征雪是白的。如果有关的说者用“雪是白的”这些字眼以有关的方式表征雪是白色的，……那么戴维森的[理论]就成了循环论证。而且，如果说者有一种有用的心灵语言，他们能以之表征雪是白色的，从而使该理论避免循环论证，那么仍然存在关于心灵语言的意义问题。（p. 286）

由此得出的启发是，关于真值条件的知识对内在表征系统的语义学并不重要。该理论必须把真值条件而不是关于真值条件的知识

648 第五部分

归之于英语或心灵语言的句子。因为如果它不把真值条件归之于这些句子,那么谈论我们的信念是正确的或是错误的就变得毫无意义了。

我曾假设,我们的确应该弄清下述观点的意思,即我们的某些信念是对的,而有些是错的;我还假定,我们应该理解这样的观点,即我们的有些信念是关于尤里乌斯·凯撒的,而有些则是关于夸克的。然而据我看,这里有一个严重的问题,即我们为什么想要澄清这些观点。我们这样做的愿望所根据的,除了在对世界的严格科学说明中没有任何地位的天真的形而上学之外,难道还有别的什么东西吗?

只要注意到我在上一节末尾提到的一个事实,就会发现这个问题并不是无理的问题。如果心理学的任务就是要陈述:

(i)有机体在受到感性刺激的情况下其信念和愿望的演化所根据的规律,以及

(ii)这些信念和愿望影响其躯体运动所根据的规律,
那么对信念和愿望的语义学描述与心理学就没有关系,即使不说出关于被相信和被期望的句子的任何意义,或它们的真值条件是什么,或它们的主观材料是什么,人们也能陈述那些规律。例如,我们可以设想一种包含着下述规律的极原始的心理学的:

在内部表征系统中存在一种联接词“—+”,以至于,对于系统中的所有句子 S_1 和 S_2 来说,当一个人相信「 S_1 —+ S_1 」且想望 S_2 时,那么他也想望 S_1 。

[编者注:在本选集中,“—+”用来代替蹄铁号 \supset (即蕴涵)和作为

实质条件句(“如果……那么”)的箭号(即 \rightarrow)]。满足这种规律的联接词“ \rightarrow ”也许表示“仅当”,即是说,它遵循的也许是条件句的真值表;而它遵循真值表这个事实并不是我们在陈述心理学规律时必须说明的东西。既然我们所想象的心理学是极原始的,那么它可能还包含以下规律:

表征系统中有一类特殊的句子,可称之为“观察语句”,它具有这样的属性,即其中的每个观察语句都与一种特定类型的感性刺激相联系。每当适当类型的感性刺激发生时,有机体便相信这个观察语句。

我们凭直觉可以推想,如果一特定观察语句以这种方式与典型地由附近的兔子引起的那类视网膜刺激相关联时,那么这个观察语句的意思就接近于“附近有兔子”;但即便真是这样,心理学理论也没有必要说这是真的。因此我们为什么还要从语义学上描述内部表征系统中的句子呢?为什么不直接说信念和愿望只是人与无意义的句子之间的关系呢?这样说将阻止我们把信念当成是真的或把它当成是关于兔子的,但是这是否遗漏了某些有科学价值的东西呢?

这里是对此问题的一种答案——我认为它不是奎因、戴维森及哈曼(也许还有刘易斯,1974)这些哲学家的作品中强调的全部内容,而是其中的部分。假设我们和一个外国人在一起,他的语言是我们所不懂的。一只兔子在外国人的视线中飞奔而过,外国人端起枪,朝兔子奔跑的方向瞄准——至此我们有充足的理由认为,他相信附近有只兔子。但我们能用前面讨论过的狭义心理学的词

650 第五部分

汇来予以述说吗？我们不能说“他相信那个句子……”（说出的名称），因为我们不懂他的语言，如果我们转而说，“他相信某个句子，它是与那种类型的感性刺激相连的观察语句……（根据规律它可作为句法上相关的句子的依据，……等等）”，那么又会怎样呢？困难显而易见：只有对我们这个外国人的心理学了解得一清二楚的人才能填上这些空。第三种可能性要好得多：

他相信他的语言中的某个句子，它在他的心理学中所起的作用相当于“附近有只兔子”这句话在我的心理学中所起的作用。

然而，可以说，这的确牵涉到语义学概念！因为它不正是这样一种转弯抹角的说法吗：

他相信他语言中的某个句子，这句话翻译成我的语言就是：“附近有只兔子”？

而翻译不是一个语义学概念吗？

这个回答并不能让我满意，因为它虽然明确激发了翻译概念的引入，但并没有带来任何非翻译的、语义学概念的引入；也就是说，它虽然推动了关于一种语言同另一种语言的关系的、多少^⑧有点语义学的概念的引入，但并未推动像“真”或“指称”这样的把语言同世界关联起来的概念的引入。据我看来，完全没有理由认为，另一个人有真实的或关于兔子的信念。我猜想，奎因式的回答是，我们需要像真值这样的、与我们自己的语言相关联的概念（如，陈述这样的概括即“每一个具有‘P 或非 P’形式的句子都是真的”）；^⑨接着我们就可能使用这样的翻译概念，它是我们基于别的

理由所一直推进的,以便让真值概念传递到外国语言中。在此我不想讨论这种奎因式的答案,但我觉得它相当脆弱,而且我想作出更好的回答。

我认为,我们之所以能够把诸如真值和指称之类的语义学概念应用于人们相信和期望的句子,是因为我们坚持这样的理论,即人们的信念在很多情况下是关于世界的可靠指示器。而陈述这种理论的唯一方法就是使用真值概念(有时大概还有指称概念)。而且这种理论不是一种毫无根由、能轻易予以抛弃的形而上学,它对我们获取关于客观世界的信息是至关重要的,因为我们在形成对世界的看法时,经常利用我们关于别人信念的看法。一个小孩相信他做了我不喜欢的事情这一事实(根据他的行为常能推论出来的事实)使我有充分的理由认为,他的确做了我不喜欢的事情;大多数物理学家相信有引力波这一事实(从阅读一些物理学书籍能推论出的事实)有充分理由使我相信,的确存在着引力波等等。这些推论很明显是根据某些可靠性原则进行的,这些原则说明了在何种情况下一个关于某些事物的信念可能是真的。陈述我们所需的原则绝非易事:毕竟,一个孩子相信圣诞老人这一事实并不足以让我也相信圣诞老人;某一宗教派别的大多数成员相信下个月在亚利桑那州的一个农场上将有飞碟着陆,这一事实也并不足以让我相信将有飞碟在那儿着陆。(我认为,这些推论的失败并不完全是由于与结论错误有关的孤立的论据。)显然我们的确有一套可靠性原则,尽管我们还不能清晰地予以阐述;可以设想总有一天它们将得到系统化,变成一种清晰的理论。从广泛的意义上说,我们甚至可以把这种“可靠性理论”看作心理学的组成部分。

652 第五部分

我不知道这种可靠性理论会是什么样子。我猜测它必须包括的不仅有真值概念,还有指称概念,因为我们在该理论中想说的是,有些人对物理学有可靠的信念,而对经济状况则有不可靠的信念等等。因此我们可以设想,“真”、“指称”等就是这一理论的原词。如果我们这样做了——而且如果我们设想,可靠性理论是如此详尽而明确地予以阐述的,以至于在该理论打算运用于其上的每一有机体中它得到了独特的实现——那么我们就可以运用这个理论对内部表征系统的真值和指称作出功能说明。这种功能说明当然是受欢迎的,因为功能说明总是受欢迎的:它承认了这样的可能性,即指称关系在不同有机体中是由不同的物理关系实现的。

VII. 再论内在表征语义学

关于内在表征系统的语义学,除了已有的真值理论语义学以外,还有其他的语义学吗?我认为有;而且我认为这使人对命题的可能世界分析产生了严重的怀疑,即使撇开本体论上的考虑。事实上,我认为,(撇开本体论的考虑)它让某些人怀疑:任何关于命题的意见是否可能。在解释这些问题时,我将先讨论口语和书面语的语义学,因为我想讨论一下奎因在这方面所提出的某些观点,但是我将把话题逐渐转向内在表征系统。

多年以前,奎因有这样的说法:

词汇学家可能关心一种语言中的语言形式同另一种语言中的语言形式之间的同义性或者……他也可能关心同一种语言中的语言形式之间的同义性。这两种情况如何令人满意地统一

于关于同义词概念的一个一般的公式化表述,是一个悬而未决的问题……(奎因,1963, p. 56)。

奎因这里提出的观点与命题问题直接相关:如果人们假定命题的存在,那么就等于对奎因的“悬而未决问题”假定了一个肯定的答案。因为如果句子意云的是命题,那么若两个句子意云的是同一个命题的话,则它们显然是同义的;而且这将是关于同义词的一个普遍的概念,它既能作语言内的应用,又能作语言间际的应用。

奎因的“悬而未决”问题果真是悬而未决的吗?至少这一点可以肯定:语言内部的同义词要比语言间际的同义词易于定义得多。(同样,说者内部的同义词要比不同说者之间的同义词易于定义,这说明同义词概念中也许还有进一步的差异。)奎因在(1960)第9节和第11节论证的观点是,至少是对奎因所谓的情境语句来说,语言内的同义词比语言间际的同义词要易于定义。他指出,对于特定的说者来说,“珠穆朗玛峰”和“高里三客峰”(Gaurisanker)(作为独词语句来看待)之间意义上的区别可由下述事实来说明,即不同的感性刺激促使他选择一个而不是另一个;在特定的语言共同体中,这些独词语句意义上的区别则由这样的事实来说明,即这些句子对这一共同体的大多数成员来说在主体内部是非同义的。据我看来,这种“促使选择”的说法太过于行为主义化了,并引起了奎因的不必要的担忧,如对“第二意向的刺激”(一种言语刺激,如“务必赞成以‘E’开头的一独词语句,否则我把你揍成肉饼”)的忧虑——试比较(1960), pp. 48-49。但我认为,奎因的基本观点是正确的:我们可以通过证据方面的考虑来解释语言内部的意义差

654 第五部分

异。我在其他地方论述过这一点,^③所用的是非行为主义的(但我承认是理想化的)证据概念;它不仅解决了奎因方法所碰到的诸如第二意向问题之类的令人讨厌的难题,而且使说明不必局限于情境语句。我认为,这种探讨的有趣之处在于,它为你说明了意义的差异,而这些差异是你从直观上所期盼的,但是根据可能世界的方案你又得不到这些差异(如果对不精确名称没有采用不合理假设)。例如,根据可能世界方案,下述两个句子在意义上就是等值的:

(10) 珠穆朗玛峰是高里三客峰

最终等值于:

(11) 珠穆朗玛峰是珠穆朗玛峰;以及

(12) 温度是平均分子动能,

最终等值于:

(13) 温度即温度。

但是即使没有查看我的说明的细节,再清楚不过的是,这些等值性将不会抓住关于说者内的同义词的足够复杂的证据标准,甚至对相信句子(10)和(12)的说者来说也是如此。

在我看来,这种证据上的考虑(再加上真值理论的思考,如果后者不是多余的话)为说者内的同义词提供了很自然的说明。然而,对说者间际的同义词而言,情况则大不一样,因为对两个意义上不同的单词提出证据标准,在这里是相当困难的。困难的根源很清楚:你和我对什么可以看作是一个句子的证据可能各持一端,这不是因为这一句子对你而言与对我而言是不同的某东西,而是因为我们的信念的其他方面存在着差异。人们也许会想办法寻找

下述形式的规则,即“如果在你的句子和我的句子之间存在着如此这般的证据上的不同,那么它们在意义上也不同”,这是奎因在“设法消除附带信息效应”时所提出的一项任务。奎因对完成这一任务的可能性提出了许多值得注意的疑问(1960, § 9)。而我认为,如果根据我在菲尔德(1977)中提出的正式模式来看待这一问题的话,那么你就会发现奎因的怀疑是值得强调的。

在我看来,我们实际上所用的语言间际同义词的标准几乎毫无例外地是由真值理论语义学提供的。这不完全正确;如果在一种语言里有两个句子 S_1 和 S_2 , 而且就真值理论(或可能世界)语义学的观点来看,它们是等值的,但是在证据方面又彼此有别,如果在另一语言里有一个句子 S_3 , 它从真值理论上讲与 S_1 和 S_2 是等值的,但是事实上从证据上说,它与 S_1 比与 S_2 相似得多;那么我们认为把 S_3 翻译成 S_2 肯定是错误的——翻译成 S_1 似乎是必要的。但是除了在这样的情况下以外,如果证据作用的任何主体间的相似性是必要的,那么就几乎没有什么错误;如果火星人通过他们强有力的望远镜辨认出了珠穆朗玛峰即“高里三客峰”,并将之命名为“Schrdu”,我们就可以用“珠穆朗玛峰”或“高里三客峰”没有差异地翻译他们的名称“Schrdu”,不管我们多么想说“珠穆朗玛峰”和“高里三客峰”对我们来说在意义上如何不同。人们也许会以翻译是一种松散的语用学的概念为由来反对这一论证:真实情形(人们也许会说)是,“Schrdu”在意义上既不同于“珠穆朗玛峰”,又不同于“高里三客峰”;我们没有区别地用“珠穆朗玛峰”或“高里三客峰”来翻译这一名称,是因为这些与“Schrdu”近乎于等同。我对这一回答表示同情,只有一点例外:那就是假设有这样一

656 第五部分

种有点清晰的想法,它是关于指称同一事物的单词之间在意义上存在的语言间际差异的,这是我认为有必要确立的假设。

因此,我的观点是,真值理论语义学——即没有可能世界的可能世界语义学——几乎足够了,但是也存在某种它不能解释的意义上的非常细微的区别。这些意义上的细微区别看起来只有在语言内部才可以清晰地区分开,这一事实似乎使关于命题的那个看法疑窦丛生。

对公共语言中的句子间的同义词的这些说明显然适用于内在表征系统中的句子间的同义词。同样明显的是,为了界定同一内在表征系统中的句子间的证据上的差异,人们无需运用狭义意义上的、心理学之外的任何概念。因而,在努力把这种证据上的思考引入关于内在表征系统的语义理论的过程中,似乎没有任何特殊的问题,就此而言,在努力引入真值理论概念时似乎存在着一个特殊的问题。

我想通过提出一种我认为悬而未决的问题来结束这篇论文:在内在表征系统的意义理论与口头或书面语言的意义理论之间存在什么样的联系呢?根据一种对于口头和书面语言的意义理论的颇有影响的探讨,^⑤此类语言中的句子的意义应该根据通常与这个句子有关联的信念(或愿望等)来解释。大致地说,了解“凯撒是自私的”的意义就是了解:这一句子通常与相信凯撒是自私的这一信念相关联;人们的公共语言里的每一其他句子都是如此。(当然,常见的关联必须循环不断地加以阐述。)这种方案预先假设:即使完全不求助于人们口语或书面语言中的“凯撒是自私的”这样的句子的语义特征,人们也可以解释:相信凯撒是自私的意味着什

么:因为如果人们在对信念的说明中,依赖于口语或书面语句子的语义特征,那么根据信念来解释句子的意义就会陷入循环论证。因此关键问题是,那一预先假设是否正确。

用内在表征来表述这一前假设,那就是:即使不借助于对下述事实的解释,即公共语言中的某些词或词组表示的是凯撒,或者说公共语言中的某些其他的词或词组表示的是自私这一属性,那么人们也能解释:内在表征系统中意云凯撒是自私的这一句子是什么意思。如果这个前假设成立的话,那么上述对意义理论的探讨方案就是相当有吸引力的:如果加以阐发,那么它将把所有关于口语或书面语的语义学问题还原为关于内在表征系统的语义学的相应问题;进而那些问题即使不再求助于语言也能予以回答。然而对于这个前假设的正确性我还心存疑惑。我猜想,在典型的事例中,使我的内在表征系统中的一个符号成为表示凯撒的一个符号的东西部分是:由于我在公共语言中得到了表示凯撒的名称,因此这个符号在我的表征系统中就有它的作用。如果诸如此类的东西成立的话,那么上述关于公共语言意义理论的探讨方案就似乎被摧毁了。但我相信,现在我们所讨论的问题太复杂了,还有待更进一步的研究。

与这种意义理论对立的方案就是试图将内在表征系统的语义学还原为关于公共语言的语义学。这样做就等于假定:虽然狗、黑猩猩等动物可能具有内在表征系统,这些系统对于解释它们的行为有着重要的作用,但这些表征系统并不是意义和真值之类概念所适用的系统:这类语义学概念只适用于有口语和书面语的有机体的表征系统。对我而言,这个结论似乎是不合理的。虽然如此,

658 第五部分

这种方案中仍有某种可取的东西,因为意义的某些方面能更直接地用公共语言予以解释,而用内在表征系统则不然。这里的主张不是说,意义的这些方面不用信念和愿望也能得到解释,原因有二。首先,当你解释公共词汇“凯撒”指称凯撒这一事实时,你可以自由选择某种信念和愿望,如无语言的有机体所能具有的那些信念和愿望;因为我们已承认,这种信念和愿望可以用一种独立于语言的方式来说明。其次,有这样一种重要的意义,在此意义上,人们在关于公共语言的语义理论中一样可以诉诸于别的信念和愿望;我们可以诉诸于被诠释为对内在句子的态度的、我们都喜爱的信念和愿望。唯一的限制在于:我们应诉诸于那些内在句子的哪些语义特征。

我无意于从这最后的讨论中抽引出任何确定的结论。我想说的只有一点,即有必要阐发的是内在语言的语义学理论和公共语言的语义学理论,而不是独立于一类语义学的另一类语义学,不是把一种还原于另一种。还原主义的战略(尤其是试图将公共语义学还原为内在语义学的战略)值得探寻,但是替代性战略也值得探寻。

注 释

① 例如可看阿姆斯特朗(1968)第3章,普特南(1975b)。

② 齐硕姆(1957,第11章)讨论过布伦塔诺问题,有关布伦塔诺的资料可从中查寻。

③ 见斯托纳克(1976)。刘易斯对功能主义的评论散见于各种著作中。见刘易斯(1970),(1971),(1972b),(1974)和(1975)。

④ 见奎因(1960)第45节。

⑤ 相信*不同于相信为真:一个外国的说话人可能相信一个英语句子为真,但并不知道其意思是什么,例如假如他在《纽约时报》上看到了一个标题。相信句子S为真(即相信S是真的),如果(1)成立,就等于相信*一个句子S*,S*意味着S是真的。相信是真的与真值概念有关,而相信*则没有关系。

⑥ 巴那克-塔尔斯基公理说的大致是:一实心球体可分解为有限的不连续的片断,而这些片断可重新组合成两个与原来实心球体大小相同的球体。更明确的陈述及证据,详见杰克,1973,pp.3-6。

⑦ 对于学究而言,我指的是冯·诺伊曼的集合论,包括选择的公理;它被有限公理化了。

⑧ 刘易斯(1972a),pp.182ff。

⑨ 丹尼特(1975b),p.410,丹尼特似乎将核心信念概念的引入作为一种特殊手段,从而使句子存贮方法免于陷入要求无限多的句子被存贮的荒谬性,但我认为下述讨论将表明:此概念可能是所与的、强有力的、独立的驱动力。

⑩ 为了更充分地说明这一点,我们首先必须说明信念(比如与愿望相反)在推论性思考等中的典型的作用。——姑且称之为作用B。这一来,我所倡导的唯物主义的相信*观大致是这样的,在每一个相信者中,存在着句子的某种贮存库,以至于事物以那种方式贮存,其明显的结果就是那些具有作用B的事物。

⑪ “几乎”一词在此似乎必不可少,因为“是……的明显结果”是非及物的。

⑫ 在这里接受所述的不确定性不过是对工具主义关于日常信念话语的观点的一种小小的让步。

⑬ 例如可参阅福多(1975)和哈曼(1973),第4章,第2节。

⑭ 菲尔德(1972)。

⑮ 戴维森(1967)。

⑯ 在这段中,我假定专名总是严格的指示符。然而有些人也许否认这一点,他们相信,当我们使用名称“伯特兰·罗素”时,我们是把该名称与属性H联系起来,我们认为,H是我们从罗素身上单独分辨出来的;当我们讲“伯

660 第五部分

特兰·罗素没有头发”时,我们所说的在世界 W 中是真的,当且仅当在 W 中事实上有一在 W 中具有属性 H 的事物,他在 W 中没有头发。

⑰ 我的意思其实不是假设在句子类似物中的原始符号是名词或谓词;谁知道它们是什么呢? 我所提倡的说明假定的是:这种符号所引出的一切(现实世界)指称问题都可以以唯物主义可接受的方式加以解决。

⑱ 我应补充的是,在构建关于人与命题之间关系的替代说明时,困难尤为严重,如果人们把命题当作精致的命题的话。

⑲ 刘易斯肯定相信这一点,我猜测斯托纳克也是如此。

⑳ 这甚至适用于非唯物主义的功能主义者,他们也要求心理属性的实现,他们只是承认:这种实现是非物理的。(只有工具主义的功能主义者才全盘否定实现的必要性)。

㉑ 刘易斯(1970),(1971)和(1972)。

㉒ 我并没有强加上刘易斯的下述要求,即 H 是一种常识心理学理论。但你如果想增加这一要求,你可以这样做,本文这节的论证将不受影响。

㉓ 刘易斯对实现是什么的说明,严格说来是不准确的,因为他没有采纳我曾提出过的预防措施,按照他的说明,仅当某事物在所有有机体、所有时间实现疼痛时,此事物才用来实现疼痛,这一来便剥夺了功能主义在此问题上的发言权(试比较哈曼 1973,第3章第4节)。但很显然,我所提倡的(相当于哈曼所提倡的)其实就是刘易斯所追求的。

㉔ 这里我可予以简化:我们可以设想, T_1, \dots, T_n , 包括该理论所需的非心理学术语,而且我们还可以设想,该理论所需的某些心理学术语,不包括在 T_1, \dots, T_n 之内,因为它们所指独立于 H 被固定了。

㉕ 基于与本文无关的理由,刘易斯觉得下述设想很省事,即该理论已被重写,以至于所有的心理学术语都是名词。我不这样认为。

㉖ “我把状态看成一种特殊的属性:事物有时具备的属性”(1861, p. 165)。当刘易斯说“疼痛”指的是一种状态时,他当然不是说,它指的是可被称作状态-标记的东西,即个体的内在事件。我本人认为,“状态”一词最好专门用来表示我称之为状态-标记和状态-类型的东西。

㉗ 严格地说,在这里省去“物理的”一词可能更好,这一来,如果唯物主义仅仅偶然是正确的话,那么我们就可以承认,有机体存在于那些可能世界

中,在那里 H 是由具有功能属性的、不可还原的心理属性所实现的。

⑳ 据刘易斯的说法,“在疼痛中这一属性”与“疼痛”指的是不同的属性:前者指一种功能属性,后者指在 X 中于 t 时实现了功能属性的非功能属性。在“疼痛”是指功能属性还是指它的实现这一问题上,我仍想保持中立态度。

㉑ 由于承认这一可能:即被相信的句子在真实语言中是真实的句子,而不是句子类似物,因而我承认这一点。

㉒ 这是在假设:句法理论是以这样一种方式阐述的,即所有的句法谓词是标记的谓词。假设这一点并非假设句法能够以唯名论的即不量化集合这样的抽象的实在的方式被给予。一旦实现了这一点,以所需要的方式阐述句法理论这一任务就没有什么困难了。

㉓ 有时,R 可能完全是一种析取关系,比如说“x 和 y 都是左脑中的事件,x 具有与 y 的 R_1 ,或者 x 和 y 都是右脑中的事件,x 具有与 y 的 R_2 ,或者一个是左脑中的事件,另一个是右脑中的事件,左脑中的事件具有与右脑中的事件的 R_3 。”析取实现对其他功能理论是必需的,例如,由于析取属性如“C-纤维的兴奋或 X-神经元的激活”,疼痛在特定时间的特定有机体中完全可能实现。

㉔ 我认为,以刚才所解释的非常松散的、语用学的方式所用的翻译概念是否应被当作是一个真正的语义学概念,这是有某种问题的;因此才说“或多或少”。

㉕ 试比较奎因(1970),p. 11。

㉖ 菲尔德(1977)。

㉗ 见希夫尔(1972)。刘易斯(1975)的探讨很相似。哈曼(1968)的下述建议看起来也有异曲同工之妙:“意义的 2 阶理论”以“意义的 1 阶理论”为前提。

参考文献

D. 阿姆斯特朗(1968):《唯物主义的心灵理论》,伦敦:劳特利奇和基根·保罗。

H. 普特南(1975b):“论属性”,《哲学论文集》,第 1 卷,剑桥大学出版社。

R. 齐硕姆(1957):《知觉的哲学研究》,康乃尔大学出版社。

662 第五部分

R. 斯托纳克(1976):“命题”,载 A. 麦凯等编:《语言哲学争鸣集》,耶鲁大学出版社。

D. 刘易斯(1970):“怎样定义理论术语”,《哲学杂志》,67:427-446。

——(1971):“对同一论的论证”,载 D. 罗森塔尔编:《唯物主义与心身问题》。

——(1972a):“普通语义学”,载 D. 戴维森等编:《自然语言的语义学》,多雷特赫特:里德尔。

——(1972b):“心理物理学与理论的同一”,《澳大利亚哲学杂志》,50:249-58”。

W. 奎因(1960):《语词与对象》,麻省理工学院出版社。

——(1970):《逻辑哲学》,恩格尔伍德·克里福斯,NJ:普伦太斯-霍尔。

T. 杰克(1973):《选择的公理》,阿姆斯特丹:北荷兰。

D. 丹尼特(1975):“大脑写作与心灵阅读”,《明尼苏达科学哲学研究》,7。

G. 哈曼(1968):“意义的三个层次”,《哲学杂志》,65:590-602。

J. 福多(1975):《思想的语言》,纽约:托玛斯,Y. 克罗威尔(1973):《思维》,普林斯顿大学出版社。

H. 菲尔德(1972):“塔尔斯基的真值理论”,《哲学杂志》,69:47-75。

——(1977):“逻辑、意义与概念作用”,《哲学杂志》,74:379-409。

D. 戴维森(1967):“真值与意义”,《综合》,17:304-23。

S. 希夫尔(1972):《意义》,牛津大学出版社。

第六部分

意向性与心理内容理论

心灵能做什么：非意向世界中的意向性*

普莱尔·雅各布

心理符号语义学研究的 主要问题与方向

本书研究的是语义性或意向性问题——语义性或意向性怎样统一于非语义的、非意向的世界的问题。意向性是心灵的一个重要特征——即使不是别的心灵的重要特征，也是人类心灵的重要特征。它是那使人类的某些心灵状态——所谓的命题态度（如信念和愿望）——“关于”（或表征）某些现实的、可能的和不可能非心理的、心理的事物和事态的东西。换言之，如果个体的心灵状态有意向性或处在表征状态，那么它就有语义属性。尤其是，个体的信念有真值条件：它们可能是真的，也可能是假的（像实际所是的那样）。在当代哲学中，对意向性有两种一般性的探讨方案：即所谓的上-下方案和下-上方案。

* 节译自 P. 雅各布：《心灵能做什么》（剑桥大学出版社，1997）一书的“导言”、第2、5章，小标题略有改动，第一个小标题为译者根据内容所加。

666 第六部分

我所说的上-下方案具体体现在戴维森的著作中。其计划是:从生物——人类——开始说明意向性,展示其是具有丰富的命题态度的系统,拥有讲说自然语言的能力和把命题态度归之于别的生物的能力。从这一由上到下的观点来看,个体具有丰富的命题态度,其最引人注目的东西就是它们的整体论特征。正如戴维森所说(1982,p.473):“一个信念离不开多个信念,而多个信念又离不开别的像意图、愿望和……口才之类的基本态度。”上-下方案在对下述观点所作的“超验”论证中达到了登峰造极的地步,这观点是,“一生物如果不是另一生物的言语的解释者就不可能有思维”;如果他不拥有语言和关于信念的概念,他就不可能有信念(戴维森,1975,p.157,170)。这一论证之所以令人神往,理由在于:如果它是正确的,那么它就证明了一种极强形式的反个体主义:任何生物如果不是一个社会共同体的一员,就不可能有思维。

相比之下,下-上方案则是从更诚实的生物出发的,常识和认知科学家往往把表征状态或认知映象归之于它们。即使细菌、青蛙、猫和狗既不能讲出人类的自然语言,又据说不能把心理状态归之于别的生物,但它们仍能够进入具有起码的意向性的状态,即进入表征了它们的环境的某些方面的状态。据推测,即使它们不具有任何关于图像的概念,但它们仍能建立关于它们的环境方面的图像。有自然主义思想倾向的哲学家(如德雷特斯克、福多、米利肯和帕皮诺)往往强调,更简单的物理、生物系统与具有丰满的命题态度集合、掌握了人类语言、并有把心理状态归之于他物的生物之间存在着连续性。像倾向于自然主义的哲学家所说明的那样,下-上方案假定:如果我们想理解意向性的某些令人困惑的特征

(如错误之可能性或错误表征的能力),那么我们最好的步骤就是从最简单的、最纯粹的实例出发,因为它们完全没有得到充分发展的命题态度系统的那些复杂性和诡秘性。根据他们的观点,我们不应让人类认知更高翱翔所取得的成就牵着我们的注意力转。因此我们应首先考察细菌、青蛙、猫和狗,因为使用人类语言和把心理状态归于他物的能力离它们显然很遥远。

我想,人们不妨姑且承认上下方案的辩护者的观点(借用马尔科姆的一个著名例子):当一只狗追逐一只猫时,狗关于猫爬上橡树的表征(正确或不正确)是不同于我关于猫爬上橡树的信念的。因为第一,关于猫和橡树我有几个不同的、即便是相对真的信念——诸如猫是其脚上长着软毛和爪子的猫科哺乳动物;橡树有根、枝和嫩枝,在秋天,其叶子会落下来等等,毫无疑问,我也有许多虚假的信念。人类不仅有虚假的信念,而且甚至对猫和树有宗教的或迷信的信念,逻辑实证主义者一直认为它们的“认知意义”是暧昧的。对于猫或树,狗大概没有这样的信念。我也不知道狗的神经系统的状态怎样表征猫或树。关于人类对一事物是有毒的、真实的信念与老鼠逃避有毒食物的倾向之间的令人有趣的差异,埃文斯(1981, p. 131)曾作过如下的描述:

老鼠仅用一种方式表现“信念”——不是通过吃——而是不局限于下述方式,即关于某物有毒的普通信念由以表现出来的方式。被试可能表现有此信念,方式即是使另外的某人不吃该食物,或将它给予自己所恨的敌人,或用其实现自杀的目的。这些变化都是来自于不同的信念可以与之相互作用的计谋,当然同样的变化也源于一信念与别的信念的结合。例如

668 第六部分

当它与这样的信念结合时,即食用小剂量的毒品会使一个人免除其毒效,那么它大概会使一个人每天都吃少量的食物。它有利于许多不同的计谋,它借别的信念的中介作用而影响任何计谋,这正好是信念状态的本质。

毋庸置疑,老鼠的倾向比人类的信念所服从的目的要少。不过从它们的多目的性比人类信念的少或单一的目的比人类的多这一事实推不出下述结论,即狗的神经系统中的状态或者老鼠的逃避倾向缺乏关于性、意向性或语义属性。狗不像人类,不能具有关于猫或树的宗教信仰,不过,在马尔科姆的狗的神经系统中的状态,则像人类的信念一样,是可能犯错误的:例如当猫事实上是在邻近的枫树上时,狗能错误地把它的位置表征为在橡树上。马尔科姆的狗没有关于信念的概念,更不用说有关于真的概念,这一事实并不妨碍它错误表征它的环境的特征。

丹尼特(1983b, p. 69)——他自认为是下-上方案的辩护者——曾在稍微有点不同的语境中以普世性的口吻写道:“如果你把这(两种方案之间的差别)比作修建一条跨大陆铁路,你就会从两端开始,计划在其中间的某处相会。”我不知道中间位于两种方案之间的什么地方。即使我遵循从下往上的战略,试图发现意向性的自然主义基础,即使我不相信这样的主张:如果一生物没有关于思维或信念的概念,那么它就不可能有思维,我也仍然认为:在由戴维森所倡导的上-下方案中仍存在着重要的洞见。用非戴维森式的术语说,该洞见就是:支撑具有成熟的人类命题态度的基本系统是两类不同的人类认知能力:获得人类语言的能力和形成关于信念的信念、关于愿望的信念、关于信念的愿望、关于愿望的愿望

的能力,一言以蔽之,把命题态度归之于别的东西的能力。因此其核心思想就是:获得具有言语和(如我将称呼的)“元表征”能力(即形成关于命题态度的命题态度的能力)的大脑结构,在能进入具有意向性(或语义属性)的状态的生物系统的种系进化中,就一定是一个转折点。

如果一个人是物理主义者(像我一样),那么就会产生两个问题,第一,物理系统的状态——大脑细胞的碎片——怎么可能有意向性?一系统的非语义属性的哪一个允许其有语义属性?第二,一系统的行为的哪一方面可以用它所具有的语义属性而加以解释?一系统的语义属性的因果作用可能是什么?由于本书探讨的是上述两个互补的问题,因此自然有两个部分。第一部分(第1—4章)考察一个体的命题态度的语义属性是否能从个体的心灵的非语义属性和关系中派生出来。在当代心灵哲学中,这常常被认为是把意向性自然化的任务。第二部分(第5—8章)试图回答下述问题:物理主义者在对个体的意向行为的因果解释中能承认个体命题态度的语义属性有一种作用吗?用丹尼特的术语来说,个体的命题态度能够成为“语义发动机”吗?如果不能,那么物理主义者就得承诺意向性的副现象论——或者说关于个体命题态度的语义属性的副现象论。这就是心理因果性问题。

在第1章中,我描述了一种立场,我将其称为意向实在论,我还证明,接受它就得承诺:把意向性自然化以及表现那种意向性这一双重计划可能有因果上的有效性。在第2—4章中,我对将个体的命题态度的语义属性从人的心灵的非语义属性与关系中派生出来这一任务所作的基于信息的目的论语义学探讨,作了论证。这

670 第六部分

一探讨方案假定:信息是语义性的关键性组成部分。第2章陈述了信息语义学的主要观点。在第3章和第4章中,从假定信息语义学观点出发,我讨论了个体的命题态度的语义属性的三个特点,它们应当借助于意向性的信息探讨而得到说明:信念内容对于它们的信息源的不敏感性;内涵性问题(或指称的不透明性)以及错误表征问题。进而我区分了文献中常常混淆的两个问题:不完全相关的问题和具有自然法则普遍效力的依赖性的及物性问题。在第4章中,我论证了对后两个问题的目的论语义学探讨,这相似于德雷特斯克近来的观点,而不同于米利肯的纯目的论语义学方案,有信息作为其基础。

第5章基于思想的语言假说讨论了“关于心灵的计算表征理论”,在过去20年中,思想的语言假说的主要倡导者是福多。它有两个目的,一是它旨在解决个体命题态度的语义属性的构成性问题。二是想解决心理因果性问题:心理符号被设想具有语义和句法双重属性。心理符号的句法属性而非语义属性被预期具有因果上的效力。在考察由思想的语言假说所提出的概念争论之后,我论证的是:它并没有证明下述论点,即个体的命题态度的语义属性是因果上有效的。在第6章中,我的主张与福多的近来主张相反,认为:语义原子论和语义整体论之间的抉择方案并不是排他的,因而意向实在论者不一定要接受语义原子论。在第7-8章中,对于个体的命题态度的语义属性为什么缺乏因果效力这一问题,我区分了两种理由。因此我把两种副现象论的威胁即先占(preemption)威胁和外部主义威胁区分开来了。前者——我在第7章中予以讨论——之所以产生,是因为个体命题态度的语义属性可能是

由个体大脑的更基本的物理(化学或生物)属性所预先占有的。而后者——我在第8章中讨论——之出现则是由于这样的事实,即个体命题态度的语义属性并不是个体大脑的局部的属性。我认为,对外部主义的威胁,德雷特斯克近来已提供了回击的基本材料,这就是认为,个体的命题态度是语义发动机,而且对下述论点的证明作出了概述,这一论点是:个体命题态度的语义属性在因果上是有效力的。不过在第8章中,我解释了我为什么不同意他在个体发生中所形成的表征的语义属性作用与种系发生中所形成的表征的语义属性作用之间所作的明确的区分。最后在第9章中,对于我就语义性所作的信息和目的论语义学的综合解释与福多的纯信息解释之间的差异,我作出了反思。我论证说:前者不同于后者,并不能用来沟通语义学和心理学。

信息语义学:下—上战略

在第1章中,我曾断言,意向实在论面对的两个任务之一就是说明,个体的命题态度的语义属性即使不是来自于完全的非语义的、非意向的属性和关系,但怎么可能来自于几乎没有语义的、也几乎没有意向的属性和关系。该任务就等于是要分析“关于”或表征的概念。什么样的基本的、非语义属性和关系可以授予一系统以表征别的事物和事态的能力?这就是意向性之自然化意味着什么问题。

对于这一任务有三种主要的探讨方案,其中有两个依赖于信息探讨方案,一个是纯目的论的。即使我相信:信息方案要想为意

672 第六部分

向性的某些令人困惑的问题提供令人满意的解答,就必须用目的论概念加以补充,但我坚定地认为:信息是意向性的真正的构成要素。我主张:融合了信息和目的论的方案能解决意向性的某些难题,但对该主张的辩护则要等到第3章和第4章。

在这一章中,我将陈述信息探讨的基本观点。信息语义学——我将如此称呼该探讨——一直得到了许多哲学家的阐发——包括巴怀斯和佩里(1983),德雷特斯克(1981),埃文斯(1982),福多(1984;1987b),伊斯雷尔和佩里(1990),斯托纳克(1984)和斯坦伯(1977)。信息语义学的占主导地位的观点是:个体命题态度的语义属性首先来自于个体心灵与他或她的环境之间的信息关系。即使信息语义学与关于指称的因果理论(如克里普克和普特南的)有某种姻缘关系,但如我将要简略说明的,它并不拘泥于实际的因果关系。由于接受了意向实在论的观点,因此我将假定:信息语义学的目的不像(克里普克和普特南所倡导的)指称的因果理论那样,而在本质上指向了思维的内容——仅仅是派生而来的、语言符号的内容。这是非常重要的,因为该任务就是试图把握思维的语义属性,而这些属性并不以语言和言语能力、更不用说言语共同体的存在为前提条件。我认为:所有信息语义学家所共有的一个自然主义特征就是:他们按照从下到上的程序开展工作。由于这个理由,我将称之为下—上之人。他们试图理解,有心灵的造物的微妙特征怎样一致于、吻合于我们关于别的物理和生物系统的观点,怎样才能与它们溶为一体。

根据第1章(§6)所作的区分,信息语义学者可以被称之为“超验的”外在主义者:^①基于他们的信息探讨,他们可以接受这样

的观点,即个体的思维以来自于他或她的环境中的对象、属性和关系为前提条件。个体的内在状态并不能看作是真正的思维,除非个体的心灵处在与来自于他或她的环境中的对象、属性和关系的信息关系中。如果他或她不具有与来自于其环境中的对象、属性和关系的各别的信息关系,那么个体就不可能思维。此外,信息语义学者可能是反个体主义者,并假定:一个体的思维的许多、大多数或全部内容依赖于个体的社会(或语言)共同体的别的成员所思维的东西。但是信息方案本身并没有什么东西使他们承诺(我以前所说的)“超验的”反个体主义,即承诺下述观点:如果个体不是社会(或语言)共同体的成员,他或她就不能思维——不能从事思维活动,或者说孤独的鲁滨逊就不能形成真正的思想。

因此信息语义学对于个体命题态度的语义属性之下—上探讨方案与戴维森(1975:15)的解释战略之间在侧重点上存在着重要的差异,这一战略包含在一生物如果不能解释另一生物的言语就不能有思维这一著名的命题之中。信息语义学者大概会反对这种强“超验的”、反个体主义命题。不过他可能赞成下述的弱观点,即只有对思维能形成高阶思维的生物才能从事解释工作。换言之,他能够接受——实即欢迎——这样的观点:元表征能力的获得在能进入具有语义属性的状态的系统之种系进化阶梯上是一个真正的起点。

在信息语义学中处于至关重要地位的信息概念就是在许多科学研究领域中都要涉及的那个概念,正像人们常说的:一只脚印或手印携带着拥有该脚印或手印的个体的信息。在这个意义上也可以说:一生物化石携带着遥远过去的有机体的信息。树干上的年

674 第六部分

轮数目携带着该树的年龄的信息。一张染上了石蕊的纸在与一特定的液体接触时会红,这一事实携带着关于下述事实的信息,即该液体中存在着某种酸。也是在此意义上,分子生物学家谈论由一片 DNA 所携带的遗传信息,它复制成 RNA,并成为细胞中蛋白质合成的样本。例如—X—射线携带着关于断腿的信息。温度计的玻璃管中的水银柱上的高度携带着关于温度的信息。汽车上的油表或速度计量器分别携带着油缸中的汽油量和车轮的速度的信息。一只时钟携带着关于时间的信息。罗盘针携带着关于地磁北极的信息。在所有这类例子中,可以合理地假定:信息关系存在于指示器与它所指示的、独立于具有命题态度的行动者而存在的东西(或信息源)之间——即使在下述人造物的事例中也是如此,在那里,行动者事先安排好了那人造物。—当信息关系形成了,它就决定了某种信息加工装置是否挑选那种信息。

有必要强调的是——例如正如埃文斯(1982, pp. 123—124)和皮科克(1983, pp. 6—7)所述——在信息术语的上述科学使用中起着基础作用的、关于信息的信息—理论性概念与内在地关联于个体的命题态度内容的日常信息概念之间存在着差异。我所谓的日常信息概念实际上根源于弗雷格关于命题内容的观点。

为了弄清语言哲学中所做的工作的差异,让我们考察一下话语(S)这样的简单句,其主语包含着一个单称术语:

(S)长庚星闪闪发亮。

在当代语言哲学中,对于下述问题是存有争论的,即话语(S)所表述的命题是什么?根据弗雷格的看法,名称“长庚星”对于话语(S)所表述的命题所起的作用就是成为表达名称的指称、名称

的意义(or sinn)的方式。因此在他看来,由话语(S)所表述的命题包含了一种表达行星即长庚星的方式。根据反弗雷格的所谓“直接指称理论”,由话语(S)所表述的命题并不包含表达“长庚星”的指称的任何方式;它包含的只是那行星。弗雷格的观点把由话语(S)所表述的命题等同于话语所表述的思想或信念的对象,而直接指称理论则把它等同于这样的事态,即那话语的真值条件。弗雷格的追随者在支持他们的观点、遵循弗雷格的路线时指出:一个有理性的能说英语的人坚持(S)为真,也同样会坚持(S')为假,尽管等式(I)为真:

(S')金星闪闪发亮。

(I)长庚星 = 金星。

弗雷格等人断言:话语(S)与话语(S')必定有不同的认知意义或不同的信息价值:它们携带着不同的信息。根据直接指称理论,用“金星”取代(S)中的“长庚星”并不会改变(S)的内容或真值条件:已知“长庚星”和“金星”是共指称的,如果话语(S)为真,那么话语(S')也为真。同样,根据信息语义学:用共指称的术语替换单称术语并不会改变话语的信息内容。弗雷格关于命题内容的观点以之为基础的信息概念根据的又是这样的认知概念,即它反映的是具有命题态度的行动者所能学会的东西。如果一个有理性的能说英语的人不知道(I)为真,那么他或她也许赞成话语(S),而不赞成话语(S')。他或她不可能怀疑(I')的真:

(I')长庚星 = 长庚星。

在正常情况下,他或她可能发现它根本就没有信息。然而他或她则有可能发现(I)之真是有信息的。因此,弗雷格关于命题内

容的观点所根据的信息认知概念在流行的自然主义计划中不可能无循环地加以使用。在严格的语言哲学的范围内,弗雷格在信息概念和具有命题态度的行动者发现其有信息的东西之间所建立的链条并不是唯一有用的观点,因为反弗雷格的直接指称理论往往也强调这样的事实,即两种具有同一真值条件的不同的话语——诸如(S)和(S')——在“信息”^②的相关意义上并未传递相同的信息。在科学中所用的信息-理论性概念不同于弗雷格的认知概念,而类似于它的反弗雷格对立面,有望成为与具有命题态度的行动者之存在无关的概念。

我对于信息语义学的陈述依据的是德雷特斯克(1981)的经典著作,其重要特征之一就是它根植于认识论中的关于知识的确实性(reliability)理论。认识论中的确实主义(reliabialism,为阿姆斯特朗、德雷特斯克、戈德曼和诺齐克所阐发)认为:关于环境的信念状态具有对信念所关于的环境条件的法则依赖性,而这种依赖性就是信念之确证的关键性因素,也即是使信念状态成为一种知识状态的东西。信息语义学为了把握信念的内容即它的语义属性,而不是把握信念的确证,用了认识论确实主义的许多工具。个体的一切信念绝不能被证明是或者说没有资格成为知识状态。不过,个体的所有信念都有某种内容——某种语义属性——或别的内容。因此信息语义学的任务就比确实主义的认识论更根本,因为它旨在从自然主义方面解释所有信念——被证明或未被证明——所具有的一种属性。此外,它从自然主义角度还要解释:不仅是信念而且是所有别的命题态度所具有的一种属性。在本章及下一章,我想澄清的是,信息语义学能把握信念的语义属性的哪些特

征。即使信息语义学的确实主义成分是有争议的(且一直有争议),^③但我将在第4章中证明:它可基于进化的(适者生存主义的)根据而得到辩护。

在第一章开头,我曾述及这样的事实,即命题态度并不是我们人类心灵中的唯一的居住者(或公民):有意识的经验、感觉或感受性质也是心理的状态。它们也有内容——即使是非概念性的内容。经验的非概念性内容是为当今哲学家广泛讨论的一个课题。我认为,这个课题能从信息语义学中得到好处。考虑到有意识经验的存在,我曾说过:表征主义——即主张心灵首先是一个表征系统——面临着这样的挑战:证明表征主义不仅能够认识命题态度而且也能够认识有意识经验,这是表征主义义不容辞的责任。

在陈述了信息语义学的基础、尝试着消除下述担忧即它可能会偷偷摸摸地把意向性从后门再引进来之后,我还想说的是:信息语义学的用处之一就是它有助于我们迎接表征主义所面临的挑战。其实,我认为:信息语义学有进一步确证一个更强的主张的资源,这个主张是:一系统只有拥有真正的信念才可能有经验。正如我所理解的那样,信息是意向性的一个真正的构成部分这一论点得到了下述事实的支持,即信息语义学有助于理解有意识经验因而可为表征主义的战略提供辩护。尤其是,我在第4章中将倡导以信息为基础的、关于意向性的目的论语义学方案,以反对诸如由米利肯(1984;1993)所倡导的纯目的论方案,我这样做有两个理由,其中之一是:在我看来,信息语义学有助于解释有意识经验的某些特征。另一理由是:信息语义学为信念内容的意向性提供了精彩的说明。

关于心灵的计算表征理论

在这一章中,我打算讨论一种有影响的观点,(按照流行的作法)我将称之为关于心灵的计算表征理论(简称为 CRTM)。在过去的岁月里,其杰出的倡导者是福多。不同于关于心灵表征主义理论的别的非计算阐释,CRTM 对于心理(或认知)过程提出了相当强的假定:心理过程就是计算过程,即对于符号的形式操作。就我所知,计算表征理论的目的具有两重性,一方面,它想为命题态度的构成性所引起的问题提供一种解答。另一方面,它想为意向实在论的因果命题的强解释作出辩护,根据这种解释,不仅命题态度是原因,^④而且它们的语义属性在因果上也有效。因此计算表征理论依赖于(或承诺)两个根本假定,其一是思想的语言假说(LOTH),其二是这样的观点,即心理学解释既是意向的,又是法则学的,即是说,心理学解释涉及到类法则的普遍原则,而这些普遍原则又关系到或从量上限定着命题态度的内容(或语义属性)。我将把后者称之为心理因果性的具有法则效力的意向特征。

正如上面两段所述,把意向性自然化涉及到对四个问题的讨论;不过在第 3、4 章中,我只讨论过内涵性、错误表征和及物性问题,而未涉及构成性问题。^⑤现在我想予以讨论。试考察我的信念 r ,即相信对象 a 有属性 F 。按照原先的假定, r 是关于对象 a 有 F 这一事实的表征。 r 据以表征事实(a 是 F)的语义属性是复合的,因为它是由关于对象 a 的表征的语义属性与关于属性 F 的表征的语义属性复合而成的。个体命题态度的语义属性如同语言符号的

语义属性,正好在下述意义上是构成性的,即复合心理状态(或语言表达式)的语义属性依赖于它们的部分(或构成要素)的语义属性及结合的原则。根据我归之于意向实在论的第一个命题(见第1章),个体命题态度的语义属性先于别的任何有语义属性的东西(尤其是公共语言符号)的语义属性。^⑥因此个体命题态度的语义属性的构成性不可能依赖于语言符号语义属性的构成性。如果计算表征理论对命题态度的构成性提供了一种解决方案,那么它就为意向性之自然化作出了贡献。

与此同时,CRTM的目的就是想通过为强因果命题提供辩护而为意向实在论的第二个任务作出贡献。请记住,在第一章中我已论证过:弱因果命题能够为标记物理主义所接受。因此在讨论以思想语言假说为基础的、关于构成性的CRTM解决方案时,我将对强因果命题的CRTM辩护作出探讨。

使CRTM富有魅力的东西是:它通过作出心理过程是计算过程这样的假定(根据思想语言假说),有望同时澄清下述相互联系的三个问题,它们分别是:复合的命题态度怎么可能基于它们构成因素的更简单的语义属性而具有复合的语义属性;命题态度怎么可能产生别的命题态度;命题态度怎么可能卷入意向行为的产生之中。如果你根据数字计算机的机器-语言的模型设想个体的思想的语言,那么关于心灵的计算机模型就有可能解释语义属性的组合与命题态度的因果属性。用福多(1987b,p.17)的话说就是:

计算机是一种对于符号的因果属性和语义属性之间的媒介问题的解答。因此如果心灵是一种计算机,那么我们就得开始考虑:你怎么可能有关于心理过程的、已取得成功的理论……

680 第六部分

这种理论可解释:在有因果关联的思想中怎么可能有非随意的内容关系……在计算机设计中,因果作用借助于符号的句法学与语义学之间的平行论而进入了具有内容的短语中。但是如果没有心理符号,这一观点对心灵理论就没有任何好处。

下面我就转向关于构成性的问题。

接受思想语言假说(LOTH)的最好的理由之一显然就是:它为个体命题态度的语义属性构成性提供了一个合理的、与意向实在论的假定相一致的解釋。思想语言假说有两个组成部分:语义学还原论的部分和句法学的(或计算的)部分。就语义学方面而言,人们可以把LOTH当作一种深一层的所谓还原,即把个体命题态度的语义属性还原为心理符号——(一种假设的)心灵语言符号——的语义属性。就意向实在论的主张而言,自然语言符号的语义属性(以及别的有语义属性的实在)要还原为个体命题态度的语义属性。根据LOTH,后者反过来又要还原为思想语言中的符号的语义属性。这就是福多(1975, p. 75; 1987b, p. 17)所说的RTM(心理表征理论):

对于任何有机体O,以及任何关于命题P的态度A而言,存在着一种(“计算的”/“功能的”)关系R以及心理表征MP,以至于:

MP意味着P,以及O有A,当且仅当O具有对于MP的R。

根据LOTH,一个体有命题态度也就是他或她处在对于一种心理符号的特定的(计算)关系中。根据综合了意向实在论与LOTH的观点,个体命题态度的语义属性来自于(个体大脑中的)

心理符号的语义属性。因而根据这种观点,意向性的自然化(信息与目的论说明,这是第2-4章的课题)无疑是关于心理符号的语义属性的(所有别的语义属性都源于此)。

LOTH引发了两个问题,其中一个我已在第3章中提出过,它是这样一个问题,即它似乎想借助心理符号的观念,赞成关于概念的观念的模棱两可的用法。第二个问题就是:LOTH与意向实在论是否一致。

根据LOTH,人类大脑组织——神经元的集合——的标记构成了具有语义属性和句法属性的心理符号的标记。因此概念应当等同于这种心理符号。正如我在第3章中所说的那样,心理学家常常把概念称之为心理表征。根据他们的观点,一个概念就是一个媒介:它可以有物理和句法属性,并进入因果关系。而在源自弗雷格(1892)的哲学传统中,概念并不是媒介;它是有一定抽象性的某种东西,是媒介所表达的东西——因而不具有因果属性。如果我们想到一英文符号(或单词)——如“狗”——那么一个概念就是弗雷格称之为单词的意义(或 the sinn)的东西——例如单词“狗”所表达的东西,只能由个别的狗所例示的某种东西。也许,一个不同情LOTH的哲学家会论证说:假如根据LOTH,心理符号有本原的(非派生的)语义属性——最先例示的语义性,^⑦那么我们应该说的就不是心理符号有意义(或语义属性),而毋宁说它们就是意义。不过我认为:我们不当把心理符号等同于意义。如果心理符号就是(不同于“有”)意义,那么它们怎么可能还有物理属性就完全是神秘莫测的了。即使LOTH诉诸于心理符号,但它在内容的媒介和由媒介所表述的内容之间又陷入了含糊不清。在我看

来,这种模糊性就像一本书的概念所引起的模糊性那样无伤大雅,因为当我们谈到一本书时,我们谈的可能是具有因果属性的物理对象——一种媒介——而且我们也可以谈及由该媒介所表述的内容(或故事)。^⑧一本书所讲的故事对于书,就像心理符号的语义属性对于心理符号一样。正如我们不能通过对故事在书本上由以写出的墨水作化学分析来确定书中所叙述的故事内容一样,我们也不能通过考察心理符号的电学的和化学的属性来决定它的语义属性。然而作为媒介的书具有由字母构成的属性,字母又是用能从化学上加以分析的某种墨水写上去的。因此我们不能把心理符号当作是意义。

基于意向实在论关于个体命题态度的语义属性先于语言符号的语义属性这一论点,那么便会产生这样的问题,即 LOTH 是不是偷偷摸摸地重新引入了语言符号的语义属性对于命题态度的语义属性的优先性。换言之,会产生下述问题:第二个所谓的还原步骤——把个体命题态度的语义属性还原于心理符号的语义属性——与意向实在论保持一致了吗?即使这是一个困难的形而上学问题,但我认为下面的论述对保持意向实在论与 LOTH 之间的一致是有利的。首先,如果我们承认了自然的(或“公共的”)语言(可用之于口语交流)与思想的语言(或心灵语言)之间的差异,那么意向实在论就包含了这样的论点,即命题态度的语义属性先于属于自然或公共语言(用于口语交流目的)的符号的语义属性。因此,上面所接受的意向实在论还原对假设的思想语言(或心灵语言)的符号的潜在语义属性以及它们与命题态度的语义属性的关系保持了缄默。第二,在我看来,意向实在论者试图根据言语模式设想思

想、思维和思维过程是自相矛盾的。意向实在论根据默默的言语模式构想思维则未能保持连贯性。

据说,思想语言的惯用语句不仅有语义属性,而且还有句法属性或构成性结构。正如几位哲学家和语言学家(如乔姆斯基、戴维森,更不用说弗雷格)所指出的,自然语言有福多所说的产生性的属性,即具有构成无限多的(潜在的无限的)句子的属性,其中每一个都是由属于一部字典(或词典)的单词所组成的语法系列。即使一部字典由有限的单词和习语所构成,但由于它的循环的句法(或语法)规则,一种语言便包含着不定多的、符合规范的句子。这样一来,自然语言的构成性就可解释它的产生性:复合表达式的意义依赖于它们的部分的意义和句法规则。而根据意向实在论,个体命题态度也必定是产生性的:在个体的生命历程的任何一点,他或她接受的只是他或她可能接受的命题态度集合的子集。但是就意向实在论的第一个论点而言,个体命题态度的产生性并不能来自于他或她所讲的自然语言的构成性。因此就引出了现在的问题:怎样说明个体命题态度的产生性?

在我尝试着回答这个问题之前,我想考察一下自然语言的一种直接相关的属性,即福多(1987b)以及福多和莱波尔(1992)称之为系统性的属性,亦即这样的属性,如果一种语言包含了一种表述了复合命题P的句子S,那么它就一定包含了别的、表述了结构上(或句法上)相关的命题的句子。例如,如果一种语言包含了一种表述了命题Rab的句子S,那么它就一定包含了一种表述了命题Rba的句子。^⑨正如产生性一样,一种语言的系统性能由该语言的构成性结构加以解释:某些复合表达式的意义依赖于它们部分的

意义和合成的(或再合成的)句法规则。再者——根据意向实在论——一个人的命题态度也必定是系统性的,或者如埃文斯(1982, p. 75; pp. 100 – 105)所表述的那样,它们必定遵循他称之为普遍性强制的东西。它们一定有平行的属性,因为一个人如果思考 Fa 这样的想法,那么他或她就一定能思考结构上相关的想法,诸如 Fb 这样的想法(假如他或她有关于对象 b 的表征)或 Ga 这样的想法(假如他或她具有关于属性 G 的表征)。但是,根据意向实在论的第一个论点,个体命题态度的系统性不能来自于他或她所讲的自然(或公共)语言的构成性结构。因此两个直接有联系的问题便产生了:我们应怎样说明思想(或命题态度)的产生性与系统性?^⑩

试考察我的合取信念(或思想) p 与 q 。如果没有更简单的信念 p 和信念 q ,我就不可能有这个复合信念,因而两个更简单的信念就被认为是复合合取信念 p 与 q 的构成要素。迄今为止,信念的产生性常常是由这样的观点来解释的,即信念有构成性结构:复合信念有更简单的信念作为其构成要素。但是按照福多(1981, pp. 29 – 30, 1985, pp. 18 – 19)的观点,试考察我的析取信念 p 或 q ,或考察我的条件信念:如果 p 那么 q 。我相信 p 或 q 并不要求我:要么相信 p ,要么相信 q 。我相信如果 p 那么 q ,也是如此。我相信 p 或 q 大概正好以相信 p 为基础(或正好以相信 q 为基础)。因此信念 p 和信念 q 并没有必要是复合析取信念 p 或 q 的构成因素。析取信念 p 或 q 即使不像合取信念 p 与 q 那样,不可能有信念 p 与信念 q 作为组成因素,但是如果断言合取信念是具有构成性结构的信念,而析取信念则不如此,那么就是怪诞的。当然我们可以维护这样的主张,即析取信念有构成性的结构,条件是假定析

取信念的标记涉及到与心理符号标记的(计算)关系,而后者进又有构成性结构,即有意指 p 或 q 的心理符号。既然这种复合心理符号的意义可以依赖于它的部分(一符号意指 p ,一符号意指 q)的意义以及制约着析取符号形成的句法规则,因此我们可以诉诸于心理符号的构成性结构来说明命题态度的产生性和系统性。就我的理解而言,这是接受思想语言假说的令人信服的理由,至少在缺乏对于命题态度的构成性的更好说明的情况下是如此。

在这里,有两个问题应予以考究。一个就是:对于析取信念的构成性是否可找到下述替换性说明,它不求助于对于意指 p 或 q 的析取心理符号的计算关系。另一问题是福多(1987b, p. 152)的下述观点,即系统性是思想(或能容纳思想的心灵)的“一个偶然特征”——而不是基本的属性。在这一点上,福多关于系统性的观点与埃文斯关于他所谓的普遍性强制的观点似乎是背道而驰的。须知,福多对于思想的语言的论证是建立在析取信念的复合性的基础上的论证,或者说是基于由命题联系词所提供的命题态度的复合性的论证。它不求助于涉及到被量化的(或一般的)思想的内在结构的复杂性。单纯的命题思维与(第一阶)量化之间的差异可能会影响人们在第二个问题即系统性究竟是思维(或心灵)的偶然特征还是基本特征上所采取的观点。

库内(1995)不求助于思想语言中的析取符号的构成属性,而对析取信念的构成性提供了一个替代性说明。可以这么说,信念之类的命题态度有两个自由度或维度,其中一个它们的态度维度(或准诺),另一个是它们的命题内容。我们似乎可以先减去专属于每一态度的态度维度或准诺,再提出这样的问题,即难道没有

686 第六部分

这样的心理状态,它关联于特定的命题内容而又缺乏与不同命题态度相联系的特定的准诺吗?库内(1995)的设想是:存在着这样的状态,他称之为有能力接纳充满着特定命题内容的思想的状态。^⑩正如我在第6章中要论述的那样,相信 p 与接纳思想 p 之间的差异对福多是确实有用的。接纳—思想遵循的是下述原则:(I)–(V)。(I)接纳非 p 这一思想的任何人一定接纳了思想 p 。(II)接纳了思想 p 与 q 的人一定接纳了思想 p ,而且接纳了思想 q 。(III)接纳了思想 p 或 q 的任何人都一定接纳了思想 p 。(IV)相信非 p 的任何人接纳(或必定能接纳)思想 p 。(V)相信 p 或 q 的任何人接纳了(或必定能接纳)思想 p 。^⑪如果是这样,那么我们就接受下述关于析取信念的替代性说明:有析取信念 p 或 q 的必要条件就是:人们有接纳思想 p 的能力,或有接纳思想 q 的能力。

如果这种关于析取信念的构成性的说明被当作是对于复合信念的构成性的一般说明,那么它就关系到了系统性的地位问题。如上所述,福多把系统性当作是现实的人类思想(或心灵)的偶然的——而非构成的——属性。如果系统性是偶然的,那么就有可能存在着具有非系统思想的心灵(其思想没有系统的关联)。如果这是真的,那么就可能存在着能思考 Rab 这样的想法、不能思考 Rba 这样的想法的心灵。事实上,福多(1987b)称这种可能的心灵为点状的心灵。^⑫与福多不同,埃文斯(1982:75)设想:他的普遍性强制可以解释为思想的构成性属性:

……我们没法避免把(着重号为我所加)关于个别对象 x 的思想(意思是说它是 F)当作是两种可分离的能力的作用;一个

是思考 x 的能力,它在关于 x (意思是说它是 G 或 H)的思想中一无例外地起作用;另一个是形成关于成为 F 意味着什么的概念的,它能一无例外地在关于别的个体的思想(意思是说它们是 F)中起作用。

假设我们承认,上述根据接纳思想的能力的复合结构对析取信念的构成性的说明是根据接纳相关思想的能力的复合结构对任何复合信念的构成性的一般说明的一个例证。那么像库内(1995)所论证的那样,我们似乎有一个赞成系统性的基本特征(以及相关的反偶然性)的论证。如果某人相信埃菲尔铁塔高于世界贸易中心,那么他或她就一定相信:后者不可能高于前者,这大概是信念的逻辑的(概念的或基本的)属性。但是根据上述原则(IV,任何人只要把信念非 p 与接纳思想 p 的能力联系起来,相信世界贸易中心不可能高于埃菲尔铁塔,那么他就一定有接纳世界贸易中心高于埃菲尔铁塔这一思想的能力。由此可知,如果某人相信埃菲尔铁塔高于世界贸易中心,那么他就一定有能力接纳世界贸易中心高于埃菲尔铁塔这一思想。果真如此,就不能认为一个人相信埃菲尔铁塔高于世界贸易中心,除非他能接纳世界贸易中心高于埃菲尔铁塔这样的思想。因此系统性是思想(或心灵)的一个基本特征;非系统性就不是人类心灵所容许的一种经验的选择。

接下来的问题自然是,我们是否应当赞同上述依照接纳思想的能力的构成性对析取信念的构成性所作的说明。我想,思想语言假说的倡导者对上述关于析取信念的构成性的说明至少会提出三点诘难。

第一,上述说明的倡导者应该给我们作出这样的说明,即接纳

688 第六部分

或有能力接纳下述思想意味着什么,这种思想实际上并不具有与心理符号的计算关系。

第二,没有把握的是,库内关于接纳思想 p 或 q 的原则(Ⅲ)就是思想语言假说的真正替换品。从语义学的观点看, p 或 q 这一思想的内容在真值条件方面等价于这样的思想(既非非 p 又非非 q)。两个思想在同样的条件下既可能是真的,又可能是假的,就像它们的真值表所显示的那样。当然,它们的逻辑形式(或句法结构)彼此有别。于是,问题便出现了:一个人能通过接纳(既非非 p 又非非 q)这样的思想来接纳或 p 或 q 这样的思想吗?如果他不能,那么他大概只能以那种方式接纳 p 或 q 这样的思想。如果是这样,那么接纳思想 p 或 q 似乎就少不了这样的条件,如处在与心理符号的关系中。这一来,库内的说明看上去就不可能与思想语言假说区别开来。如果人们通过接纳后一思想而接纳前一思想,那么就会导致这样的问题,即当思想 p 或 q 被当作是思想(既非非 p 又非非 q)时,原则(Ⅲ)与接纳思想 p 或 q 是否一致的呢?换言之,原则(Ⅲ)可能是适于接纳 p 或 q 之类的思想的一种条件。即令原则(Ⅲ)不能使我们说明思想(既非非 p 又非非 q)的复杂性,但我们还是能通过诉诸于原则(Ⅰ)而说明它,因为根据这一原则,接纳思想非 p 的一个必要条件就是有接纳思想 p 的能力。果真如此,思想 p 或 q 与思想(既非非 p 又非非 q)之间的真值条件等值性对库内的说明就算不上是致命的打击。

第三,对系统性的基本的(非偶然的)本质的论证是以对库内的下述说明的基本原则为基础的,这个说明根据接纳有关思想的能力的复合结构对析取信念的复合性作出了解释。但是基本原则

本身似乎不适当地受制于与“高于”这样的非对称性概念有关的思想。人们不妨承认库内的说法:信念的逻辑(或概念)属性正好是:如果某人相信埃菲尔铁塔高于世界贸易中心,那么他或她就一定有接纳下述思想的能力,即世界贸易中心高于埃菲尔铁塔。但是关于非对称性关系的概念又怎样呢?像约翰爱玛丽这样的信念又怎样呢?说约翰爱玛丽这样的想法是系统性的就等于说:如果某人相信它,那么他或她就一定能接纳玛丽爱约翰这样的想法。然而,如果某人相信:约翰爱玛丽,那么对于玛丽是否爱约翰,就简直没有什么是他或她一定要相信的。因此接纳玛丽爱约翰这样的想法的能力并不是约翰爱玛丽这样的信念的逻辑的(或概念的)特征。上述关于系统性的基本本质的论证似乎只是毁了那项工作。

这为我们关于思想系统性的偶然性争论留下了什么呢?即使我认为,上述关于系统性的非偶然性的论证由于缺乏普遍性因此从根本上说是不成功的,可是我并不准备不加选择地接受下述观点:思想语言假说纯粹是偶然的。

首先,循着皮科克(1992, pp. 48 - 51)的路径,我愿意说:一生物是否把两个对象当作是对两个概念的例示(或归属),它能否把存在于它们之间的关系概念化,这是以经验为根据的、偶然的。但是如果它使概念上的区分适宜于把一种状况范畴化为“Rab”状态的一种例示,那么它是否有能力对“Rba”作出反应就不仅仅是偶然的。自然,如果系统性不只是偶然的,那么思想语言假说也不只是一种经验的假说。

第二,假如我们承认福多的观点:涉及到非对称性关系的思想的系统性是这些思想的偶然的属性,那么思维就有这样一种根本

690 第六部分

特征,即如果得到认可的话,似乎最好由思想语言假说并不是经验主义的假说这样的观点来把握的特征。我所说的思维的特征是我在第3章中已提及的那个特征(与超强内涵性有联系),即从个别思想到一般思想,例如从思想 Fa 到思想 $(\exists x)Fx$ 转化的能力。须知:如果思维离不开形成一般思想的能力,那么一个生物就可能有这样的心理符号,它们具有与在她的环境中而不是在思想中所例示的属性的信息关系,如果她没有从单称心理习语到一般心理习语转化的逻辑力量,即如果她不能形成一般思想的话。有人可能会抵制这样的主张,即思维是由形成一般的(或量化的)思想的能力所构成的,方式就是强调这是关于思维的一个太强的条件,进而主张应回到思维的弱条件,因此就思维而言,人们一定只有这样的能力,即通过用命题联系词把更简单的思想结合起来而形成复合的、逻辑上非构造的或非分析的思想。换言之,有命题演算的能力对于生物之思维来说是充分的。不过,如果假定:从个别到逻辑上的一般结构的转化的能力是由它要思维的东西构成的,如果假定思想的逻辑的、句法的结构就是能容许这种转化的东西,那么我想我们就接近于这样的认识了,即能在这种意义上思维的生物一定有思想的语言(或某种近似的東西)。这一观点非常适合于解决超强内涵性问题(第3章中所认可的),并使思想的语言成为思维从概念上说所必不可少的东西。

注 释

① 如我在第一章中所述,单纯的“非超验”外在主义者断言:即使不是个体所有的命题态度,至少是许多命题态度的语义属性依赖于个体与他或她的环境中的对象、属性及关系之间的关系。而“超验的”外在主义者则赞成更强

的观点:如果个体不处在对他或她的环境中的对象、属性及关系的关系中,他或她就根本不能进行思维活动。

② 关于语言哲学内的这样一种非弗雷格式的信息概念,可参阅卡普兰(1978),里查德(1990),萨蒙(1986)。

③ 尤其是米利肯(1989b,1990b)提出了争论。

④ 正如第一章所述,弱因果命题可为标记物理主义所接受。

⑤ 在第3、4章中,我区分了两个问题,一是析取问题(在信息结构中,它是作为不完全相关问题而出现的),一是合取问题(在信息结构中,它是作为具有法则效力的依赖性的及物性问题而出现的)。

⑥ 公共语言符号来自于具有音位学属性的自然语言的符号。“公共的”之附加物旨在包容这样的可能性,即来自于思想的语言——一种“私人的”语言——的符号的语义属性具有对于个体命题态度的语义属性的优先性。思想的语言的符号则没有音位学属性。

⑦ 参看第1章,第4节,关于本原的或原始的与派生的意向性之间的区分的证明。

⑧ 模糊性为什么是无伤大雅的,其理由之一就是:LOTH理论家对媒介和内容大概用了两个不同的术语——像我刚才论述过的那样。

⑨ 关于思维的系统性是否与福多(1987b)所说的语义学原子论相一致的争论,可参阅下一章。

⑩ 福多(1987b,pp.148)论证说:对于LOTH系统性的论证是不受下述特征影响的,可以设想,该特征弱化了产生性论证。不像前者那样,后者假定:个体的可能的思想(或命题态度)的集合是无限的集合(或是无限度的)。同样,关于个体的语言能力是产生性的这一假定,涉及到这样的假定,即他或她能理解无限长的句子。拒斥关于LOTH产生性论证的方式之一大概应是拒绝所涉及到的理想化描述。

⑪ 事实上,库内在接纳思想和有接纳思想的能力之间作出了区分,因为他假定:判断这样的心理事件和信念这样的心理状态之间是不同的。他设想:有接纳思想 p 的能力就是相信 p ,而接纳思想 p 就是对 p 作出判断。既然我不赞成事件和状态之间所作的区分,因此我只拟考察接纳思想 p 这样的状态。

692 第六部分

⑫ 换言之,相信 p 或 q 的任何人都接纳了(或必定接纳)思想 q 。

⑬ 在第 6 章中,我再来考察点状心灵的可能性。

参考文献

D. 戴维森(1975):“思想与话语”,载戴维森:《真理与解释探究》,牛津,克拉伦登出版社 1984。

——(1982):“有理性的动物”,载 E. 莱波尔和 B. 麦克劳格林编:《行动与事件》,牛津,布莱克韦尔,1985。

G. 埃文斯(1981):“对语义理论和不言而喻的知识的答复”,载 S.H. 霍尔茨曼等编:《维特根斯坦:遵守规则》,伦敦,劳特利奇和基根·保罗。

——(1982):《指称的种类》,牛津,克拉伦登出版社。

D. 丹尼特(1983b):“人工智能与心理学研究战略”,载 J. 米勒编:《心灵的状态:与心理学研究者的对话》,伦敦,1983。

F. 德雷特斯克(1981):《知识与意识流》,麻省理工学院出版社。

J. 福多(1975):《思想的语言》,纽约:克伦威尔。

——(1981):《表征:关于认知科学基础的哲学论文》,麻省理工学院出版社。

——(1984):“语义学,威斯康辛风格”,载福多(1990):《内容理论与其他论文》,麻省理工学院出版社。

——(1985):“福多对心理表征的导读”,载福多(1990)。

——(1987b):“使心灵更加物质”,载福多(1990)。

——(1990):《内容理论与其他论文》,麻省理工学院出版社。

J. 福多和 E. 莱波尔(1992):《整体论:购物者的向导》,牛津:布莱克韦尔。

J. 巴怀斯和 J. 佩里(1990):《条件与态度》,麻省理工出版社。

D. 伊斯雷尔和佩里(1990):“什么是信息?”载 P.P. 汉森编(1990):《信息、语言和认知》,牛津大学出版社。

R.C. 斯托纳克(1984):《探索》,麻省理工学院出版社。

D. 斯坦伯(1977):“论语言表征的因果理论”,载 P. 弗伦奇等编:《语言哲学之当代展望》,明尼苏达大学出版社。

C. 皮科克(1983):《感觉、内容、经验和思维及其关系》,牛津大学出版社。

——(1992):《概念研究》,麻省理工学院出版社。

R. 米利肯(1984):《语言、思想与其他生物学范畴》,麻省理工学院出版社。

W. 库内(1995):“思维、言语和‘思想语言’”,打印稿。

M. 里查德(1990):《命题态度》,剑桥大学出版社。

W. 萨蒙(1984):《弗雷格的难题》,麻省理工学院出版社。

D. 卡普兰(1987):“愿望命题态度”,载 P. 弗伦奇等编(1979):《语言哲学之当代展望》,明尼苏达大学出版社。

计算心理学与解释理论*

H. 普特南

我与我的朋友 Z. 皮利辛在饭后曾进行过一场辩论。它涉及的是皮利辛所提出的下述论断：“如果没有一个关于心理表征的内容同一性的精确规定的概念，那么认知心理学就是不可能的。”后来我想到，我拒绝这一论断的理由直接与 D. 戴维森对于意义理论和心灵哲学的众所周知的影响有联系，因而在皮利辛的允许和宽恕(但愿如此)下，我便决定在本文中主要阐述我反对他的论断的理由。

心理表征

当我们想到“那里有一棵树”，或想起关于常见自然事物的任何其他普通观念时，我们可考察一下：我们心理发生了什么。根据一种模型即心灵的计算模型，心灵有一套程序或一组规则，类似于控制计算机的规则，思维涉及到对单词和其他符号的操作(在能为计算机用语词表达的意义上，这种操作并非都是“有意识的”)。

* 译自 D. 罗森塔尔编：《心灵的本质》，牛津大学出版社 1991。

然而这种模型就现在的情形来看,几乎是空洞无物的(尽管它在那些不承认像计算机这样单纯的工具,能当作人类心灵这样特殊的东西的模型的人中也激起了热情)。其所以说它空洞,是因为它并未具体说明心理功能作用的程序或规则系统,而正是这种程序才构成了心理学理论。仅仅只说正确的心理学理论能被描述为计算机(或某种类似于计算机的东西)的程序(或类似于程序的东西),不管它可能是什么,这无异于空谈。因为事实上任何能由一组法则描述的系统至少能为计算机模拟。从弗洛伊德的深层心理学到斯金纳的行为主义都可以被描述为一种计算机程序。

然而,当今致力于“人工智能”研究的计算机科学家和那些思考指称、语义表征和语言运用等的认知心理学家,心中都有比主张心灵能为数字计算机模仿这类几乎是空洞的假说更具体一点的假说。(即使那假说由于包含有心理过程的因果结构之类的东西,暗示了这些心理过程是按决定论的或概率论的、按有限程序程序化的规则而发生的,因而不完全是空洞的。)但计算机研究工作者和认知心理学家都趋于一致的进一步假说则是:心灵借助于表征而思维。在现今关于“表征”的言论中,似乎出现了两种不同的观点。

第一种观点以试图为计算机编制模拟智能行为的程序的经验为基础,以为:思维不仅涉及对任意对象或符号的操作,而且也离不开对具有非常特殊的结构即一种形式化语言的结构的那些符号的操作。计算机人员的经验告诉我们:人工智能中最有趣和最成功的程序,从典型的方面来说,最终都涉及给予计算机以某种像形式化语言和一组支配那种形式化语言的规则(比方说语言中的“推理”)之类的东西。

696 第六部分

第二个与“表征”一词有联系的观点是：人的心灵通过建构关于它的环境的某种模型——关于世界的模型——而(部分地)进行它的思维。当然这种模型实际上自然不需要相似于世界。但有把握说：在表征系统中的项目和“外在那儿”的项目之间应当有某种非偶然的关系，因而心灵在它的表征系统中就能顺利地读出“外在那儿”发生的事情。

一旦给予一种形式化语言以指称定义，在那一语言中的一组语句就可看作是一种“表征系统”或“世界的模型”。

例如，假设我们想表达这样的事实，即巴黎市比维也纳市大。如果我们有一谓词，比如说 F ，它表达“比……大”的关系（即如果我们用形式化的记号写作 F_{xy} 的开型句子与任何两事物所处的关系有关，当且仅当它们是两城市，且第一个比第二个大——比如说，在人口方面），如果我们有“个别的定项”或专名，如 a 和 b ，它们分别表示巴黎和维也纳两城市（即按语言的指称定义， a 相关于巴黎， b 相关于维也纳），那么我们就能够通过把 Fab 这样的句子放入我们已有的句子项目（我们“关于世界的理论”）中而表达这样的事实，即巴黎是一比维也纳大的城市。同样，任何事态，不管它多么复杂，只要能用谓词、专名以及形式化语言的逻辑方法予以表述，那么便可断言：通过把表达了事态的公式整合到“关于世界的理论”中就能获得那事态。

当我们的“表征系统”本身是一理论时，当我们运用我们的表征系统的方法囊括了作出形式演绎时，我们就会明白：同一对象——形式化语言，包括演绎的规则——能够是这样的形式化语言，即计算机科学家一直把它设想为大脑或心灵（根据这种观点，差别

似乎不太大)的计算的媒介,同时是表征的媒介。心灵同时把形式化语言(或某种非常类似于形式化语言的东西)作为计算的媒介和表征的媒介,这可以被称之为当今认知心理学的工作假说。

在我看来,这种工作假说的一部分的确是正确的。我相信,如果不把某些心理项目当作表征,那么我们就根本不能说明思想和语言的功能作用。当我(正确地)想到“我面前有一棵树”时,单词“树”在我心里说出的句子中的出现就是一种有意义的事件,在单词“树”的出现的外延中的项目之一正好是我面前的树。而且,开型句子“x在我面前”与我面前的存在的关系属性有关联(在我的语言的严格的语义学中),根据这些相同的事实,完整的句子“我面前有一棵树”就是这样的句子,即当且仅当我面前有一棵树,它便是真的。

这里可能使心理学家们发生意见分歧的问题是:有多少心理项目是表征;为了解释有意识的思想和理智的行动,假定一庞大而复杂的无意识的表征系统有什么作用,等等。

关于“思想的语言”的证实主义语义学

至此,我所说的与皮利辛以及其他“命题主义”认知心理学家的见解毫无二致。为了这一论证起见,我将假定所有这一切都是正确的。当然,实际情况也许要复杂得多。心灵可以使用一种以上的形式化语言(或毋宁说形式化语言类似物)。大脑的不同部分可以以不同的媒介进行计算。句子类似物和映象类似物以及这两者之外的东西可以运用于实际的计算过程之中。但我们不妨假设

698 第六部分

对于皮利辛的观点来说是最好的情况,即以^①一种形式化语言完成它的一切计算的心灵。

心灵对自己计算媒介的理解是基于什么呢? 福多(1975)认为,说我们不应把“理解”一词用于“思想的语言”本身,这显然是不妥的。(“思想的语言”是一表示假设的、大脑中的形式化语言类似物的名称。)因为“思想的语言”和“大脑中的形式化语言”都只是隐喻。它们可以是科学上有用的宝贵隐喻,但作为隐喻,它们与理解这一概念又是不可分离的。某物确实不能是一种语言,除非它能被理解;某物不能是一种语言类似物,除非存在有适当的理解类似物。如果大脑中的某些表征是句子类似物和谓词类似物,那么相应的理解类似物又是什么呢?

我以为回答是这样的:大脑对它自己的计算和表征的媒介的理解就在于它拥有一种关于那种媒介的证实主义语义学,即有一种可计算的谓词,^①该谓词能表达可接受性或被证明的可肯定性或可信性。如果加以理想化,我们便可把语言当作(部分地)通过一组规则而被解释的东西,这组规则把确证度(即主观的概率)赋予既与经验输入有关,也与其他句子类似物有关的句子类似物。这些规则必定是可计算的;它们为心或脑或机器所拥有,其原因在于“它装有电线”,一定得遵循它们,或者说由于学习而开始遵循它们。(我并不假定思想的语言一定是天赋的,或者说它一定与说话者所获得的自然语言相分离。)

但是,为什么是一种证实主义的语义学,而不是戴维森所说的意义理论呢?

很明显,如果我们把思想的语言理解为“表征系统”,那么我们

就等于赋予谓词类似物以外延,赋予句子类似物以真值条件。但是包含了对思想的语言的特定解释的“意义理论”并不是心理学。事实上,如果我们像戴维森那样表示它,那么它独有的原始概念便是“真实的”,而“真实的”并不是心理学概念。明确地说:意义理论产生了这样的定理(“T-句子”),如(假定思想的语言是英语)“‘snow is white’(雪是白的)在思想的语言中是真实的,当且仅当雪是白的。”这里没有包含任何心理学词汇。

我们可能会说:“噢,理解就在于大脑知道那种意义理论的‘T-句子’”,但“知道”这一概念在亚人认知心理学中不可能是一个原始概念。^②

假如我们想说:心灵不借助于表征而理解了“雪是白的”是什么意思,而且它知道:“雪是白的”这一表征是真实的,当且仅当那种事态存在。这不仅把心灵当作“知道”事实的某东西,未把知道分解为更基本的、更少意向的过程,而且它还违背认知心理学的基本假定,即在思考事态时,理解它们是什么等等,不能离开表征而完成。从根本上说,那样一来我们就无法摆脱把表征直接比作未概念化的实在的神话。

另一方面,如果我们说:大脑或心灵使用“雪是白的”这一句子(或相应的句子类似物)足以证明这一解释:当且仅当雪是白的,“雪是白的”在思想的语言中才是真实的,这就是主张大脑(无疑地)“知道”T-句子这种说法所包含的意思,那么,“使用”究竟基于什么,对此我们无法提出任何理论。这就是证实主义语义学所要告诉我们的东西(而且,就我能理解的而言,只有证实主义语义学才如是说)。我以为,证实主义语义学是有利于功能主义(或认

700 第六部分

知)心理学的自然语义学,它包含有“信念”(或相信度)概念,而此概念正是使心理学成为认知的心理学的东西;同时它是可计算的语义学,而这又是使心理学成为功能主义的心理学东西。

我们当然想让语义学与行动联系起来。这意思是说:那个模型必须体现功利的功能和确证度功能。这种功能也必须是可计算的(或严格地讲,是部分可计算的)。这样的理想化自然是严肃的:我们假定:信念类似物(由确认度功能表达的)和抉择类似物(由功利功能所表达的)是完全一致的。信念和抉择(或信念表征和抉择表征)两种实际的(神经病学上表现出来的)类似物完全可能是不一致的,只要在实际的决定非得作出来时,有解释这种不一致性的程序就行了。用赖辛巴赫在另一情境下所使用的术语说,一致性可以是“虚构的”,而不必是“事实上的”。它对于心灵哲学的意义,我将在本文结尾简要地予以讨论。

现在的问题是:如果关于表征的媒介的大脑语义学是证实主义的,而不是真值条件的,那么心理表征的“内容”概念又会怎么办呢?

鲁利坦尼亚国*的两个小孩

假设地球的某地方有一被称为鲁利坦尼亚的国家。在这里,北方和南方所讲的方言只有微小的差别,其中之一就是:在北方方

* 虚构的充满风流的地方,据说那里的人奢侈放荡,浮光掠影,无须认真履行职责,狂热地献身于皇室。

言中,“grug”意即银子,而在南方则表示铝。假如有两个小孩,奥斯卡和埃尔默,都是在该国长大的。他们在遗传结构和环境方面非常相同,不同的是,奥斯卡在南方长大,而埃尔默在北方长大。再假设,在北方由于某些原因,壶和盘通常是由银子做成的,而在南方,一般由铝制成。因而北方的小孩一长大就知道壶和盘是由 grug 做成的,南方的小孩一长大也知道壶和盘是由 grug 做成的。

我们可以设想,奥斯卡和埃尔默有关于“grug”的相同的心理表征,有与 grug 相联系的相同信念,等等。其中一些信念即使在语词和心理表征上是同一的,但在意义上则自然有区别,例如当奥斯卡相信“我的妈妈有 grug 盘和壶”,埃尔默相信“我的妈妈有 grug 盘和壶”时,指示代词“我的”指称的就是不同的人,因而“我的妈妈”指称的是不同的妈妈。但是如果附带信息中的这种微小差别不足以构成“心理表征”的内容中的差别(倘若如此,符号的意义与我们在同符号相联系时所得到的附带信息之间的常见差别,似乎完全可以舍弃),^③那么我们似乎应当说:关于 grug 的心理表征的内容,对于处于生命这一阶段的奥斯卡和埃尔默来说实际上是相同的。

我无意于主张:单词 grug 在奥斯卡的个人习语中与在埃尔默的个人习语中具有相同的意义;在别的场合,我曾论证过,(1975a)我们应当注意:在两个共同体中,指称上的差异影响着个体的说话者的言语。具体地说,当奥斯卡试图确定 grug 是什么时,他最终不得不求助于“专家”。这些专家没有必要一定是科学家,他也许会直接请教他的父母(他们进而可能请教零售店店主甚或科学家)。但关键在于:既然 grug 的外延在两个共同体中事实上是不

702 第六部分

同的,根据我在其他地方辩护过的意义理论:外延上的差别构成了意义上的差别;既然外延是约定俗成的,而不是由个别人决定的,因而可得出这样的结论:在奥斯卡和埃尔默的个人习语中,grug一词的意义是不同的,虽然没有任何“心理的”东西,在他们头脑中也没有任何东西构成意义上的差别。意义不存在于头脑之中。如果是这样,grug一词的意义便有差别,意义存在于那个词的指称之中,它是经共同体的实践客观地形成的,而不在奥斯卡和埃尔默所拥有的关于grug的概念之中。

但是皮利辛对之感兴趣的、乔姆斯基^④对之也表示了关注的内容概念是这样的概念,它在外延中分化出了这类客观的差异。皮利辛寻求的是关于心理表征的内容概念,在这种心理表征中,地球上的“水”和孪生地球上的“水”,据说对于有相同的“水”概念的说者而言有相同的内容,即使有这样的情况:即地球上的“水”指的是 H_2O ,而孪生地球上的水指的是XYZ。至此我所述的不过是:不管根据哪一种关于内容的概念,在奥斯卡心中的grug与在埃尔默心中的grug有相同的内容。

那些词不仅有相同的内容,而且大脑在它的计算中可能会使用的、对应于语词单元grug的任何心理符号或谓词类似物在此阶段也都有同样的内容。但是根据皮利辛所倡导的那种理论,如果grug一词和存在于它后面的心理表征在这阶段都有同样的内容,那么它们何时出现内容上的不同呢?到奥斯卡和埃尔默都长成大人,都学会了外语以及其他东西时,他们便不再有关于grug的相同的概念。奥斯卡懂得:grug是英语中称之为“铝”的那种金属(我设想鲁利坦尼亚的每个人都在中学学过作为第二语言的英

语),埃尔默也晓得北鲁利坦尼亚语中被称作 grug 的金属是英语中所说的“银”。他们都知道用来区分银与铝的大量事实,并认识到;在南鲁利坦尼亚,grug 的意义不同于北鲁利坦尼亚所说的 grug。

然而,根据本文第一部分所述的那类证实主义模型,不存在这样的阶段,在此阶段,奥斯卡心中(或埃尔默心中)的 grug 一词或相应的心理表征改变了它的指称。从内在的方面说,把一个符号当作改变其指称来对待,事实上就是把它当成了不同的符号。这绝对是不可能的。根据内在的考察,会发生的只能是奥斯卡得到了关于 grug 的更多的信念。在某一时期,他只知:壶和盘由 grug 制成,grug 是一种金属;它有一定的颜色。后来他掌握了关于 grug 的附加事实(例如用英语说出的那些东西)。当一个词的运用为附加信息的连续获得所限定,而不考虑在任何阶段那个词被赋予的新的外延时,会发生的情况是:含有那词的不同句子的确证度发生了变化。而且含有此词的不同句子在确证度方面的变化是一种连续的变化。其所以是连续的,是因为它仅仅由先于附加信息的概率的条件化而产生。这一简单例子所要说明的是:在关于语言(甚或内在表征系统)运用的功能主义模型中,没有什么东西能自动地帮我们作出这样的决定,即我们何时应当说一种表征的“内容”发生了变化。我们能完整地描述心理符号的运用,而无须这样的标准,即把心理符号的内容方面的变化与附加信息方面的变化区分开来的标准。因而,我们碰到了关于心理符号的同义词的难题。

可能的解决办法

在“‘意义’的意义”一文中,我主张通过考察外延和有关的定型(stereotype)这两者去判定某些语词(自然类型的语词)是否是同义的。所谓定型,我指的是一组确定的信念或观念化信念,即在与那词发生关系时预计所有说话者都有的那类信念。例如,如果所有说话者都相信老虎是有条纹的(或想象中的老虎是有条纹的),那么条纹便是关于老虎的定型的一部分。定型的观念能推广到心理表征上吗?定型的同一能成为我们判断心理表征的同义的标准吗?为了探讨这一点,还是让我们设想一个奥斯卡,不过我这次只假设:奥斯卡生活在非常靠近北极的实验研究站。也许这是一个由在极端寒冷的条件下对找石油进行实验的科学家和工程师所组成的共同体。我设想:在该共同体中,由于某种原因,没有绿色植物,但有人工植物甚至人工草,它们是人们为了使那里更宜人而引进的。

现在就来考察一下奥斯卡和一个普通的美国小孩的个人习语中的语词“草”,我们把这个美国小孩叫做埃尔默。如果我们把“草”的定型简单地当作知觉原型,那么草的定型在奥斯卡心中与在埃尔默心中也许是完全一样的。但奥斯卡无论在什么意义上都不知道草是有生命的或生长的事物。我以为,在这种情况下,我们通常把假定奥斯卡和埃尔默心中的心理表征或“草”一词有相同的内容视之为错误的。

也许我们应当要求:除了有关于“草”的相同的知觉原型以外,

奥斯卡和埃尔默应有同样的“标示物”，例如他们应当相信“草是一种植物”，但这又直接会引起我称之为感染难题(infection problem)的问题。如果他们真的相信草是一种植物，那么为了使他们关于草是一种植物的信念相应恰当地成为相同的信念，他们关于植物的概念又必须怎样相同呢？我认为，在实际的解释中，我们的政策是不让感染扩散得太远，即为了确定一个奥斯卡和一个埃尔默有关于草的相同概念起见，我们可以要求他们都相信草是一种植物，而无须要求他们关于植物的概念实际上相同。但这等于接受了这样的观点，即解释实质上是一种日常生活中常见的、与利害有关的工作。

如果我们谋求一个关于心理表征的内容实际同一的概念，此概念正是皮利辛所要论证的，那么感染问题就变成了一无限倒退的问题。例如假定奥斯卡和埃尔默是生活在美国的两个小孩；他们都知道草是植物。但埃尔默知道植物是能用显微镜看到的，而奥斯卡的植物概念指的是有可见的大小以及绿色等的存在物。这一来，区分表征内容中有什么差别，以及在附加信息中有什么差别的问题在作为标示物的“植物”的层面上又重新出现。

关键之点在于：发现定型是否相同，这是一种有效的翻译方法论，但它丝毫不像是规则系统那样的东西。定型本身是以语词表现出来的信念。一旦我们同意对于某些词语或表征的翻译，尽管是尝试性的，那么定型理论便可以为我们怎样翻译其他语词或表征提供钥匙。但它不能用来定义关于语词或心理表征的内容同一的概念。

它也不能定义我们已提过的另一概念，即知觉原型的同一性

的概念。我们已认识到,即使两个词与同一的知觉原型有联系,但它们在意义上也可以是不同的。因为一词的意义不仅是有关的知觉原型的功能,而且也是或多或少有点抽象的不同信念的功能,人们在与语词(草是植物)发生联系时常有这类信念。当两个词(或两心理表征)与同一的知觉原型有联系时,说它们有相同的内容便是错误的;而且把知觉原型的同一性当作心理表征的内容的同一性的必要条件也是错误的。因为两个说话者可能赋予“单身汉”一词以相同的意义,即指已到结婚年龄而尚未结婚的成年男子,但却可以有与该词有联系的相当不同的知觉原型。根据语义内容的任何直观概念,把知觉原型上的每一差异当作语义内容上的差异,就如同把附带信息上的每一差异当作语义内容上的差异一样,是不可思议的。

乔姆斯基对我的“‘意义’的意义”一文所作的反应是:主张人们可以通过放弃“内涵决定外延”的原则去拯救弗雷格的“内涵概念”。根据乔姆斯基的建议,以我所说的孪生地球的例子为例,我们应当说:“水”一词在地球上和在孪生地球上有着相同的内涵,即使它在地球上与在孪生地球上有着不同的外延。然而根据已述的观点,这似乎滑向了错误的道路。因为如果我们用“‘意义’的意义”中所用的方法,把一术语的外延当作它的“意义向量”的组成部分之一,那么我们便有清楚的理由说:grug 在北鲁利坦尼亚与南鲁利坦尼亚有不同的意义。既然 grug 一词在鲁利坦尼亚人的两种方言中有不同的指称,因而它也就有不同的意义。一旦我们决定撇开指称(或毋宁说指称上的差异),去问 grug 一词在奥斯卡和埃尔默心中是否有相同的内容时,我们便是在着手从事一项不可能

完成的任务。我们所做的远远不是使判定表征是否同义这一件事更容易,而是使之不可能。事实上,我们用来接近下述原则的第一步就是注意外延,这个原则就是确定在实际翻译实践中语词是否有相同的意义的原则。“撇开”外延上的差异只会使按原则判定意义何时已有变化成为完全不可能的事情。

另一方面,在确定同义词时,不管是语词的同义还是心理表征的同义,如果我们执意要把外延当作确定同义词的一个因素,那么我们仍不会得到像规则系统之类的东西。因为正如确定定型涉及到确定其他语词的意义一样,确定一名词的外延总是与确定其他名词的外延有关。如果语词有不同的外延,那么就不能把它们当作是内容上相同的,这是一种在我们让那种解释计划生效之后、对于翻译有用的原则。但是只有在我们已开始解释一种语言时,我们才有关于一名词的外延是什么的观念。

这样一来,我们怎样着手进行解释呢?我所要辩护的答案就是已为戴维森和奎因有力地支持过的答案。(D. 威金斯曾说过:这种观点可在维科那里找到其雏形。)它是这样的观点,即认为解释在本质上是一种整体论的事业。解释一种语言(不拘它是公共语言还是思想的语言,假定思想的语言是存在的)涉及到找出一种翻译图式,一种“分析性假说”。而这种图式能被学到,能在我们的本国语言中产生与被翻译的表述等价的、垂手可得的表述,而且最重要的是,这种图式是这样的东西,当我们领会了说外国语的人的为这种翻译图式所说明的、说话者所要说出的意义时,我们就能“理解”他们的目的、信念和行为。正如维科所述,在解释中,我们就是努力把被翻译的东西的人性提高到最高限度。如果这是对

708 第六部分

的,那么 we 实际上所有的、用来判断任何符号或符号类似物的内容的唯一标准就是我们的成功解释的直观标准;把这些东西形式化涉及到把我们关于人是什么以及人的术语中什么是可理解的全部概念形式化(参阅普特南,1976)。

上述讨论并不是主张:解释理论是某种不能研究的东西。相反,解释是某种能以许多不同的方式,从许多不同的角度加以研究的东西。但是这一讨论要说的与皮利辛的见解不同,即解释理论和认知心理学研究的课题是迥异的,在其中一个课题上取得的成功在很大程度上不依赖于在其他课题上的成功。我已阐发的例子所要表明的是:即使不以任何方式去解决解释难题甚或保护指称翻译(reference-preserving translation,或外延赋值,assignment of extension)的难题,也能提供一种完善的功能主义心理学,包括关于思想的语言的完善的证实主义语义学。因为用功能主义术语描述一表征系统怎样起作用是一码事;而形成对于那表征系统的解释又完全是另一码事。

功能主义心理学和解释理论之间存在着差异,其部分原因是:功能主义心理学把人类心灵看作是计算机。它力图陈述计算的规则。而计算的规则又有这样的属性,即虽然这些规则的相互作用是错综复杂的、整体性的,但它们的作用在任何特定时刻又是局部的。机器常常是按照特定的或有限的指令,在有限的数据的基础上促使一个数字从一址到另一址。就此而言,解释绝非局部的。一种翻译图式,不管它对有限的语料库多么有效,但也不得基于附加的文本而经常作出修改。

让我们借助于似乎有点离谱的方式来阐述这一点。几年前,

H. 菲尔德(1972)提出了一种观点,认为指称是符号运用与事物(包括属性)之间的一种物理主义的关系。他在查珀尔希尔评论我的“语言与实在”(普特南,1975b)一文时,以某些例子又进一步说明了这一点。在考察一位使用了“引力”一词的牛顿式科学家这样的实例时,菲尔德认为:我们要搞清“引力”这一术语究竟指称哪一对象或属性,应以找到科学家当时事实上打算指称哪一对象为基础,如果他的理论作为整体是对真理的一种合理的、受欢迎的逼近的话。我之所以提及这一点,主要是因为:即使此处涉及的解释概念像能想象到的那样不同于皮利辛的查明(finding)“内容”这一概念,但同样的考虑也会出现。菲尔德所寻求的“物理主义关系”最终是这样的关系:它的定义蕴涵着把语言理解为整体,而不只是个体的符号(因为构成一种使之可被合理地接受的理论解释的东西依赖于对整体的理论的关注),同时“物理主义关系”还是这样的关系,它的定义就其规范的意义而言,包含着对合理性的分析。任何关系,它的定义包含着这些因素,是否严格地被称之为“物理主义的”,对此我不打算予以探讨。但重要之点很明确:每一种对解释概念的分析工程或迟或早要含有这样的清醒认识,即对解释概念的分析要么与对关于合理性的规范概念的分析不可分割,要么与对维科的“人性”之类的概念的分析不可分割。

企图用绝对精确和规范的方式完全成功地陈述解释的正确程序,显然是一种乌托邦式的空想。就其本身而言,这似乎不像是极端重要的观点。在陈述物理学规律的过程中,完全的成功毕竟是一种乌托邦式的计划。在描述人脑的功能结构方面的完全成功肯定也是如此。即使我们没有碰到完全的成功,但物理学可能如此,

710 第六部分

因为物理学规律能够连续不断地被接近;功能心理学也是可能的,因为(我们希望)人脑的功能结构能部分地被描述和接近。同样,人们可以期待,在描述解释的实践及程序的过程中,人们是能够取得部分的成功的。

我不怀疑这一点的真实性;而且我也的确不怀疑:这些课题的每一个能在多层面上(物理学毕竟不仅是基本粒子理论,而且还是磁流体动力学、固体状态物理学等等)得到研究,能在与其他许多需要不同的精确概念和许多不同的词汇的工程相互联系中得到研究。重要之点在于:在认知心理学中实在地可预见到的那种部分成功,与在解释理论中可实在地预见到的那类部分成功,是极为不同的。

在一定的意义下,我希望本文是对解释理论的一种贡献。我们在解释理论中可能得到的那些令人欣慰的成功更有可能是哲学的探讨,而不是技术的“结果”。但是即使我们密切注意到可能是来自于计算机科学和有关领域的对解释理论的各种各样的贡献,但像 M. 明斯基(1975)所指出的那样,这些成功也可能正好是由于它们的自然偏颇。明斯基强调:在我们把解释机械化时,我们典型地做的事情是把解释限定在一组确定的文本上,这组文本是关于特定的课题的,并具有共同的语汇,具有一组共同的计划,一组共同的经验或其他假定。解释理论中的局部成功总是有限制的,其理由在于:任何完全的成功,任何对句子——处在一种可变的语言、一种可变的理论中的、涉及不同课题和具有不同预设的句子——的解释程序都包含着对人性概念或合理性概念(依我看来,二者是密切相关的)的分析。把解释限定在特定的构架内的作用就

在于,用不着去处理那完全是乌托邦式的、对那些概念的算法分析工程。

然而,我们在认知心理学中可能取得的那种局部成功是相当不同的。这里,我们可能期待的是分辨产生了不同功能作用的各种物理状态。如果大脑用类似于形式化语言的东西进行计算这一假说是正确的,那么我们就可望把这些东西当作大脑中的谓词类似物和句子类似物,或者说是关于操作这些类似物的计算规则。我认为:任何人,只要对这两类局部成功可达到的境地有明晰的观念,只要不承认自己被下述十足的乌托邦式的梦想,即对人是什么(或什么是合乎理性)作出完全的数学分析冲昏头脑,都会立即明白:它们是独立的工程,的确不以其他东西为先决条件。

N. 古德曼(1978)提出了与一元论相对抗的多元论。根据这一术语,下述观点是一种一元论,即我们只应当瞄准认知心理学这一门科学,因为它一举解决了句法和语义学问题(并且在证实主义语义学的意义上以及在解释的意义上解决了语义学问题)。与之相反,我主张:我们应当让百花齐放,我们应当让百花以各自独立的方式,在不同的季节甚至在它们各自的花园里竞相怒放。

关于戴维森的心灵哲学

我以为,我所阐述的观点带有一定的“戴维森式”的性质。在这里,更明确地说明这是怎么回事也许是适当的。在“心理事件”一文中,戴维森(1970)提出了“异态一元论”的论点。它主张:在物理事件和心理事件(即用信念和愿望这样的词汇描述的事件)之间

712 第六部分

存在的是一种“标记－标记同一”，而不是“类型－类型同一”。他关于不存在“类型－类型同一”的论证现已广泛地被误解了。对于许多人来说，它似乎像一种“大笑话”。戴维森对“类型－类型同一”的非存在所作的论证仅仅根据这样的事实，即心理概念和物理概念服从于不同的标准。（既然这种论证妨碍了各种类型－类型同一，甚至“水是 H_2O ”这样的同一，或“光是电磁辐射”这样的同一，那么它便有点令人失望。因为在还原之前，“水”与“ H_2O ”的确从属于不同的标准，就像“光”与“电磁辐射”一样。）不过，这分明是一种误解；戴维森的确有一种论证，它既是一种难以捉摸的论证，又是一种饶有兴味的论证。

戴维森非常熟悉 A. 特夫斯基的工作（参阅特夫斯基和卡里曼，1975，1982）。因而他相当清楚：从真实的语言报告中是不可能读出抉择的。正如特夫斯基的非常细致的经验研究所表明的那样：人对于他们自己的抉择的真实的语言报告完全是不一致的。如果我们根据下述准则行事，即把人们（真诚地）说出的，他们所作的抉择归之于他们，那么我们便无法解释他们的行为。因为表达出来的抉择完全是矛盾的（例如他们非常严重地违背了选择及物性的逻辑属性）。也许这是戴维森应更明确地阐明的步骤。没有理由认为：使我们能直接分辨被试的思想语言的脑显微镜的有效性能使问题得到解决。真实的语言报告大概符合于呈现在被试大脑中的心理表征。也没有理由相信：病人的“思想语言”表征比他的真实语言报告更一以贯之。如果我们能“阅读”思想的语言，在特定的情况下，我们便可毫无疑问地发现编码在大脑本身中的这类报告，诸如“我宁愿 A 而非 B”，“我宁愿 B 而非 C”，“我宁愿 C 而

非 A”。我们在谈话中发现这种报告时,我们所做的事情就是观察被试的整个行为,并尽力根据他的全部行为,根据更多的语言资料、更进一步的“语料库”,确定什么是对于病人的行为和言论的最合理的重构。根据那些可能是作为他公开宣布的抉择的子集的抉择,或是与他所宣布的有点不同的抉择,一种合理的重构可以解释被试的行为或大部分行为,可以解释那些表达出来的、而我们确认病人的确没有的、作为大杂烩的抉择。例如我们可以判定,病人所表达的某些抉择是特定情境下的设想的产物。

某些人自然会说:在比病人的思想语言更深的层次上,必定存在着一个我们称之为柏拉图式的思想语言的层次,这种思想语言按照隐秘的规则,以神经病学上显著的表征的形式记录了被试的真正的抉择。但没有绝对的理由相信这一点。人可能正好是一部装有电线的计算机,因而在大多数时候,他们最终所做的事情可以容许某种合理的解释或其他的解释。作为最佳的对行为的合理重构的合理解释本身没有必要实际地、物理地编码在他们大脑的某处。

简言之,戴维森的观点是:适用于公开语言的东西肯定会适用于真实的心理表征或神经病学上的显著状态,从因果关系上说,这些表征或状态处于公开语言的背后。正如我们必须说:人的真实的抉择实际上并不相同于他所公开声称的抉择,我们同样必须说:人的“真实抉择”(即是在合理重构他的行为中最好应归之于他的抉择)并不相同于他编码于他的大脑表征中的东西。信念-愿望解释属于我所说的解释理论层次的东西。它像一切解释一样是整体性的,与利害相关联着的。心理学家常说,大脑中仿佛存在着概

714 第六部分

念。而我所论证的(以及我认为戴维森所论证的)观点则是:大脑中可能有句子类似物和谓词类似物,但没有概念。“心理表征”正像任何其他符号一样需要解释。

注 释

① 严格地说,是部分可计算的(部分循环的)谓词。在心理学中,与理想化的归纳逻辑相对立,我们不能要求:不同程度的信念总要得到定义。我要着重指出的是:在一种心理学模型中,并非所有学习都是归纳,即使在归纳逻辑学家广义地使用该词的意义,也是如此。我们正是通过别人谈论或出示那些东西而知道那些东西。对别人告诉我们或(在某种程度上)出示给我们的东西的相信必定归并到“确证度”函项中,不管人们是否承认这是好的“归纳逻辑”。

② 理由在于,正如福多和丹尼特所强调的,认知心理学就是计算心理学。让能作出推论等的小人作为对人的解释的组成部分就行了,假如那个小人最终“履行了职责”(正如丹尼特所述),即被搪塞为计算机的算法。但这意味着“知道”最终事实上是“被履行了”。

③ 我们不能把我们所拥有的伴随信息中的每一个差异当作是语词的意义上的差异,其理由在于,这样做就等于舍弃我们的概念与我们所具有的、包含着那些概念的信念之间的区别,正是这种区别才成了意义、同义词、分析性等直觉概念的基础。抛弃意义/信念的区别就等于赞成奎因的下述观点:我们最好是彻底地抛弃意义概念。这对我们的心理表征的“内容”与我们相关的信念之间的区别也是适用的:这不过是一幅画像,即关于“我们头脑中的词语”的画像。画像并非总是错误的:它们可能是有用的模型。但是如果关于心理表征及其“内容”的画像必须有其用途,那么“内容”在信念发生了某些变化的情况下一定会保持不变。

④ 在“‘意义’的意义”一文发表后,乔姆斯基给我写了一封长信,批评了这种观点。福多(1979)为这个计划作了辩护。

参考文献

D. 戴维森(1967):“真理与意义”,《综合》,X V II, 304-323。

——(1970):“心理事件”,载 L. 福斯特等编:《经验与理论》,伦敦,79 - 101。

H. 菲尔德(1972a):“塔尔斯基的真值理论”,《哲学杂志》,LXIX,347 - 375。

——(1972b):“理论变迁与指称的不确定性”,《哲学杂志》,LXX,462 - 481。

J. 福多(1975):《思想的语言》,纽约。

——(1979):“认知心理学中作为研究战略的方法论唯我论”,《行为与脑科学》,63 - 109。

N. 古德曼(1978):《创造世界的方式》,印第安那波利斯。

M. 明斯基(1975):“表征知识的构架”,是为他的《计算机视觉的心理学》一书的一部分,纽约。

H. 普特南(1975a):“‘意义’的意义”,重印于普特南(1975e),215 - 271。

——(1975b):“语言与实在”,载普特南(1975e),272 - 290

——(1975e):《心灵、语言与实在:哲学论文集》,第 2 卷,剑桥。

——(1976):“实在论与理由”,重印于普特南(1978),123 - 138。

——(1978):《意义与道德科学》,伦敦。

什么是心理表征理论*

斯蒂芬·斯蒂克

1. 引言

心理内容或心理表征理论是近来非常流行的理论。由于有许多时髦的产品,因此市场提供了一个令人眼花缭乱的选择范围。诸如因果协变理论、目的论理论、功能作用理论以及由关于指称的因果理论所引发的种种理论。还有单因素理论、多因素理论、狭义理论、广义理论以及关于这些论题的大量的、五花八门的变种。^①甚至,在该领域的重要的杂志上,似乎很难找到这样的一卷,它无论如何没有一篇文章对某人的心理表征理论提出赞成或(更有代表性地)反对的论证。而且这类文献的大部分都带有明显的、敦促它的腔调。探索一种充分的心理表征理论不只是一种时尚的追求;许多作者还坚持认为它具有极端的重要性。对此问题,几乎没有人埋怨福多轻描淡写地予以陈述,他告诉我们:建立关于心理内容的自然主义理论是为常识意向心理学辩护的一个不可或缺的步骤。“如果常识意向心理学真的被摧毁了,那么这将是我们的物种历

* 选自 S. 斯蒂克与 特德·A. 伍德菲尔德编:《心理表征文选》,布莱克韦尔出版公司,1984。

史上无可比拟的、最大的理智灾难。”(福多,1987,p. xii)F. 德雷特斯克也使用了同样的预言灾变的术语。他隐晦地指出:如果没有关于心理内容的适当的自然主义理论,我们最终将不得不“放弃把我们自己作为人类动因这一概念。”(1988,p. x)

尽管不乏关于心理内容的各种解释之优缺点的争论,但关于心理表征理论应该做什么的问题,即心理表征理论应该回答什么问题,什么才可算作正确的回答,几乎没有什么旗帜鲜明的讨论。而这些正是本文将集中关注的问题。在设法回答它们的过程中,从提出另一问题入手将是大有裨益的:为什么有那么多的人要一种关于心理表征的理论;是什么东西使建立这种理论的计划变得如此迫切?这是我在第2节准备回答的问题。

心理表征理论应研究什么这类问题常常被忽视,即使这是不幸的,但也不值得大惊小怪。这些问题所导致的那种方法论的自我意识在哲学中几乎不合时尚。结果就常常出现这样的情况,即哲学家们为压根就不明确的问题提供了周密的解决方案——或者用福多的话说,他们为不存在的疾病提供了治疗。因此我想作为哲学方法的一个基本原则加以强调的是,在我们开始考察答案之前,我们应花大量的时间澄清问题。当我们用这一战略去建立关于心理表征的理论时,某些非常令人震惊的结论便会接踵而至。

下面是我在本文中要辩护的结论的预览:

(1)一当我们开始思考心理表征理论应该干什么时,我们心理便很清楚,可以提供的答案实际上是非常不同的。这里的方案不是一个,而是多个。这些方案可划归两个不同的家族,即使仅在一个家族内就有值得注意的重要差异。

718 第六部分

(2)由于有一个(且有争议的)例外,因此我将陈述的方案并不能真的得到那些习惯于用文献中居于主导地位的、众所周知的哲学分析技术的哲学家们的追随。准确地说,它们在本质上是跨学科的方案,经验理论的建构与检验在其中起着关键作用。不过(还是有一些例外),必然会对这些方案作出重要推进的那类跨学科工作在文献中显然尚属阙如。

(3)上述事实可能已表明:人们实际上在探寻的方案不同于我将描述的那些方案——我简直无法设想那些探索心理表征理论的人正在做什么。但是即使不知道那些替补方案的某些细节,我仍倾向于提出一种模糊的结论:这个热闹非凡的领域中的大多数表演者并没有拿出一个能用他们常用的方法成功地加以探寻的、一以贯之的方案,这就是我的争鸣。

(4)即使我们撇开这些顾虑,我将概述的任何方案对回应下述担忧未必有很大的助益,这些担忧已使许多人感觉到:建立一种心理表征理论是刻不容缓的问题。

(5)但是我还要论证的是:这些担忧本身是极其错误的。

关于威胁和前提就说这么多。下面该言归正传了。

2. 我们为什么需要一种心理表征理论

毋庸置疑,人们想要心理表征理论的理由是很多的。但是在许多动因中,有一个格外突出。过去十年中,对取消论的关注已成了心灵哲学中的核心课题,进而建立心理表征理论就被视之为争论的关键步骤。即使取消论者在阐述其论点时几乎是不明确的或

不够谨慎的,但我认为,该理论充其量提出了两个本体论主张,其中一个比另一个更强硬、更不稳固。弱的主张是:常识心理学的表征状态——如信念和愿望——在关于人类行为原因的成熟理论中没有任何作用。如果我们用“认知科学”这一标签作为表示在解释人类行为中有着作用的各种科学的大杂烩,那么取消论者所倡导的就是:常识心理学所构想的意向状态并不是认知科学本体论的组成部分。强的主张是:这些常识心理状态甚至根本就不存在。不存在这类东西,就像不存在燃素或热的流质或巫婆之类的东西一样。赞同这两个观点的人通常假定:第一个能用来支持第二个,即使尚不清楚应怎样进行论证。许多人设想:心理表征理论在对付取消论的争论中可扮演重要的角色,这几乎是用不着怀疑的。但是这种作用实际上应是什么尚不明朗。在第3、4节中,我们将考察关于心理表征理论是什么的两种看法。当我们在这一课题上取得某种进展时,在第5节中,我们将回到下述问题,即心理表征理论怎样用来论证或反对取消论。

对于取消论的两个论点中的弱的一个——主张信念和愿望在成熟的认知科学中没有作用——有不同的论证。一类论证是强调民众心理学所描述的认知过程与机制的结构。它们坚持认为:这些结构与在一种或另一种所谓有前途的科学纲领中构想的结构是不相容的。第二组论证强调心理状态的语义或意向属性,就像在常识心理学中被解释的一样。第二类论证中的某些论证相当浮华琐碎,并充满专业性。它们阐述了随附性、个体主义和意义整体论之类的复杂深奥的概念。但正如福多所述,对于许多人而言,关于语义属性的最令人忧虑的事实就是他们的直觉本体论怪物。

720 第六部分

意向非实在论的最深层的动机不是来自于……对于个体主义和整体主义的相对技术性的忧虑……而是源于某种本体论直觉：在物理主义的世界观中没有意向范畴的位置；意向的东西不能被自然化。（1987,p.97）

这种忧虑对于解释加之于任何可接受的心理表征理论的被广泛承认的强制有很大的助益。这种理论必须是自然主义的。它必须说明：心理状态的表征属性怎么能用与由自然科学提供的、更广泛的物理主义自然观相一致的术语加以解释。尽管它被广泛接受，但我仍倾向于认为：自然主义强制有严重的错误。在第6节中，我将为我的疑虑阐述一点理由。

不过，即使人们接受了自然主义强制，但再明显不过的是，这种强制只能是下述故事的组成部分，这个故事讲的是修正心理表征理论意味着什么。因为基于对自然主义强制的任何合理的解除，谈论关于心理表征的大量的自然主义说明将是可能的，而且这些说明在许多方面将是彼此有别的。我们肯定不想说：所有这些说明都是正确的。因此现在让我们提出这样的问题，即是什么东西把好的说明与坏的说明区别开来了。怎样才算是修正了一种理论？

3. 对常识概念的描述：第一类方案

我们关于我们自己和他人的日常言论的主要特征就是我们通过提及心理状态的内容而实现对它们的分辨。例子俯拾即是：

布什相信：戈尔巴乔夫在莫斯科。

我认为下午会下雨。

我的妻子希望我晚餐时不迟到。

在这些以及别的大量例子中,内容的归因是易如反掌、不成问题的,无疑也是有用的。而且在典型的事例中,对于这些归因,主体间存在着广泛的一致。很显然,一定有某种复杂的心理机制,作为这种无处不在的实践的基础,而且似乎可以合理地假定:这里所说的机制包括关于在其之下适于描述像相信 p 或期望 p 之类的心理状态的条件的丰富的缄默的知识储备。如果我们在相当宽泛的意义上使用心理学中盛行的“概念”,那么这就等于假定:作为我们实践之基础的机制包含着关于心理表征的概念。心理内容理论的一个完全合理的目的应该是描述该概念。修正这种理论就是对作为我们日常实践基础的概念或缄默的知识主干给予准确的描述。^②

描述作为对内容判断之基础的概念结构,这样的方案至少大致类似于哲学和认知科学中一直被探寻的其他各种方案。在生成语言学中,有这样一个常见的假定,即说者的言语判断和实践是由缄默的语法知识的根本主干所使然,因而语言学家的任务就是对说者心照不宣地知道的东西作出明确的说明。在认知心理学中,有大量的工作旨在阐明作为各种社交和实践技能基础的概念和知识结构。遵循这些路线的一个最令人着迷的方案就是努力揭示“民众物理学”——当我们在物理世界中生活时,我们所利用的关于物理世界的信息系统——的概念和原则。使该研究特别令人感兴趣的东西是这样的发现,即许多人使用的是关于物理世界的错误的民众物理学,而且不只是细节上的。显然,这种引导着这些人的物理判断及行动的心照不宣的理论与其说接近于牛顿的物理学,倒不如说接近于中世纪的动力理论。^③这类发现可能使取消论

722 第六部分

的论点更加合理。如果人们能够根据一种有严重错误的物理学理论指导他们在这一世界中的活动,那么至少也有这样的可能性,即他们在描述、解释和预言人的行为时,他们也可能根据一种同样错误的心理学理论。

第三项工作极其相似于描述我们的心理表征的常识概念的计划,它属于自苏格拉底时代以来为哲学家们断断续续使用的那种概念分析类型。过去 2500 年中,游戏规则几乎没有什么变化。它常常是这样进行的:

S:(大概是苏格拉底)请告诉我,X 是什么?(在这里,X 可由“正义”或“慈悲”或“知识”或“原因”或“自由”……所置换。)

C:(也许是刻法洛斯,或齐硕姆)我很高兴地告诉你,某物要成为 X 的例示,就必须是 y 或 z。

S:但那不对。因为你肯定会假定:a 是 X,但它既不是 y,又不是 z。

C:你说得很对。让我再假定:某物要成为 X 的例示,它必须要么是 y、z,要么是 w。

S:恐怕那也不行,因为 b 是 w,但它显然不是 X。

当 S 用尽了反例或 C 用尽了定义时,游戏也就结束了。在这场游戏中,即使没有人作出精确的记录,但知道内情的打赌者常把赌注押在 S 上。

这场定义和反例的哲学游戏几乎是没有意义的,除非我们要分析的概念作出两个假定。第一,目标概念能够通过详细说明必要而充足的条件而加以描述或定义。为了胜一场,S 要么能够提出一个例子,它是该概念的例示,但不能为定义所把握,要么能

够提出这样一个例子,它符合那定义,但不是该概念的例示。而且,一般假定:定义是属性的一种布尔并置,或者说是关于那个论题的某种相对明确的变种。第二个假定是,游戏者得到了关于那个目标概念的足够信息,因而使他们能判断它是否适用于广泛的事例,包括以前从未想象过的真实、假设的事例。为了明白这一点,试考察下面两个众所周知的例子:

(I)如果某人请你保存他的武器,并要求:当他得了精神病后就归还,归还这些武器是正义的要求吗?

(II)假如史密斯刚在委托销售代理人陈列室签字购买福特牌汽车。即使史密斯不知道这一点,代理人对汽车没有明确的所有权,但是就在那之后不久并在很远的地方,G. 史密斯死了,他的旧福特车的所有权转到史密斯手上了。因此他相信他拥有一辆福特汽车,他的信念既公正又真实。史密斯知道他拥有一辆福特汽车吗?

我们希望人们怎样回答这类问题,或者我们为什么应严肃地对待他们的回答,这是很难搞清楚的,除非我们假设:他们已经不言而喻地知道了非常类似于我们正设法澄清的那组充分而必要条件的东西。

我们为什么一直较详细地进行着对概念分析的传统哲学探讨,这两个缘由。一是关于心理表征的大量哲学文献似乎正好受制于定义与反例的范式。关于心理表征本质的哲学理论一般对什么是下述形式的断言的充分必要条件提供了说明:

心理状态 M 有内容 P。

对这些理论的批评常常依赖于直观的反事实,即这样的事实,

724 第六部分

在其中,定义说 M 有内容 P,但直观否定了它,或者相反。^④第二个理由是,现有大量的证据表明:以这种传统哲学方案为基础的假设可能是完全错误的。如果是这样,那么在关于心理表征的哲学文献中起主导作用的方案就是严重缺乏基础的。

在心理学文献中,对支持传统哲学分析的假定最著名的挑战也许来自于 E. 罗施及其同僚的工作。^⑤根据罗施等人的观点,作为人的判断之基础的心理结构,在其将项目归入范畴时,并没有利用范畴从属关系或大致同等的任何东西的众所周知的充分必要条件。他们实际上所用的正好是在过去 15 年中促发了大量经验研究并继续为人们热情地予以探讨的一种争论。前不久罗施提出:范畴化依赖于原型,而原型则可以被视之为对大多数典型的或富有特征的范畴成员的理想化的描述。例如鸟的原型大概包括诸如会飞、长着羽毛、会唱等特征。在确定一特定的例示能否包括在该范畴内时,主体要评估原型与被范畴化的例示之间的相似性。不过,在原型中详述的特征与从属关系的必要或充分条件不可同日而语。因此,例如一动物可能缺乏作为原型的鸟的一个或多个特征,但仍被归类为鸟。鸱鸒尽管既不会飞又不会唱,但仍被归之为鸟类。取代原型理论的是这样一种假说,即范畴化是由样本(exemplars)所促成的。而样本可以被看作是对范畴的特定成员的细致的心理描述,这些成员是从事范畴化的人所熟悉的。根据这种解释,人们也是通过作出不言而喻的相似性判断而决定一个项目是否是一范畴中的成员。然而,根据样本理论,被归类的项目可比之于范畴的作样本的成员。^⑥

新近的研究表明,对于许多概念来说,原型和样本说明都不能

轻松自在地解释所有的材料。就某些概念而言,已有这样的建议,即主体的判断依据的是某种酷似众所周知的科学理论的东西。而在其他情况下,据说不存在作为范畴化判断之基础的永恒概念。更确切地说,据论证,主体“基于飞”而构造出各种不同种类的概念,以回应范畴化要求出现于其中的情境。^⑦

近年来,关于概念和范畴化即使已做了大量的工作,但对意向范畴——诸如相信 p , 或意欲 q 等范畴——则没有系统的经验研究。因此现在我们只能思考这种研究将有什么启示。也许最安全的赌注是,不管引起意向范畴化的心理机制可能是什么,它也不会利用“经典的”概念——能用一组充分必要条件加以定义的概念类别。这里的论证直接就是归纳式的:被研究的任何常识概念最终被证明都不能被分析为一组充分必要条件。甚至基于流行的有效证据,好像也不存在经典的概念。第二个似乎合理的考虑是:引起意向范畴化的概念或“知识结构”比在传统的哲学分析中所提供的那些概念更加复杂。我猜想:我们关于心理内容的“概念”将更像一种理论,而不像柏拉图式的定义。

假如这些考虑是对的,那么结果怎样呢? 最明显的结论就是,在寻求建立心理表征理论的过程中,传统的构想定义、寻找直观的反例的哲学方法将不得不放弃。该方法试图详细说明这样一组条件,它们是所有在直观上符合目标概念的事例都将满足的。但是如果关于一种状态是否有内容 p 的直观是由原型或样本或不言而喻的理论所引导的,或者说如果决定我们直观判断的心理结构在一定程度上是在回应直观判断所必需的环境时被建构的,那么就将不存在这样的条件。因此如果定义加反例的方法之运用是

该领域哲学理论的标志,如果心理表征的常识概念类似于从经验上被研究的每一别的概念,那么就有这样的看法,根据这种看法,绝不可能有关于内容的哲学理论。

然而,重要的是不曲解这一结论。因为即使哲学分析的传统方法可能非得予以放弃,但我们在建构关于心理表征的日常概念的描述性理论时则没有理由不运用别的方法。语言学家、认知心理学家和认知人类学家都已阐发了各式各样的探讨常识概念结构的方法,其中没有一个方法预先设定:这些概念具有能为一组充分必要条件把握的经典结构。这些方法中的一个或更多的方法由于有点独出心裁,因而不妨用来探究作为我们关于心理表征直觉判断基础的机制。

在本节中,我们一开头就提出了心理表征理论应做什么这样的问题。现在我们至少有一种似乎合理的答案的轮廓。据设想,心理表征理论就是要描述这样的概念或知识结构,它们是人们对信念、愿望和别的意向状态的内容的日常判断的基础。然而,如果这就是哲学家们在企图建立一种心理表征理论时想要的那种理论,那么它就是这样一种极好的赌注,即他们将不得不放弃“哲学研究”(如传统上所设想的),而代之以研究认知科学。

4. 作为自然现象的心理表征:

第二类方案

关于常识概念的描述,在不为对这种描述必须像什么的先验的哲学要求所妨碍时,就是一种完全合理的活动。但是它还不是

那些寻求心理表征理论的人心目中所想的唯一的计划。为了弄清可能有什么样的替代品,试考察关于疾病的概率。有一种基本的人类学文献,其目的是要描述用于各种不同文化中的疾病概念。^⑧如果你对人们怎样设想疾病有兴趣,那么这正好是看个究竟的地方。但是如果你对什么是疾病感兴趣,那么你应当研究的正好是生物学或医学,而不是认知人类学。对于金子或空间或质量或遗传也可以提出完全类似的观点。如果你想知道人们怎样设想它们,那么对常识概念或知识结构的描述就是应予探讨的方案。但是如果你想知道金子或空间或质量或遗传真的像什么,那么你就应该研究化学或物理学或遗传学。

有关科学对于它怎样设想那有趣的项目有时会直言相陈。《物理学和化学手册》将告诉你关于金子你想知道的一切,而且可能还远不止于此。但是在许多别的事例中,科学在不为那概念提供完全清楚明白的或哲学上令人满意的解释时也能相当成功地使用该概念。在这些情况下,科学哲学家常常插手有争议的概念,并设法更明确地予以表示。近年来,对适当性(fitness)、合乎语法性和时-空以及其他广泛的概念进行过富有启发性的研究。^⑨这项工作部分可看作是直接的概念描述——设法呵护科学概念和理论,就像语言学家、认知心理学家和认知人类学家一直设法呵护常识概念与理论一样。更有甚者,近年来许多科学哲学家在分析科学时已开始运用认知科学的技术,并常常产生有趣的结果。不过,哲学家所找到的概念,以及概念在其中起着作用的理论,有时又是含糊不清的或者是阐发不够的。在这种情况下,科学哲学家设法对他们描述的概念和理论作出改进就是顺理成章的。说描述止于

728 第六部分

何处,建构开始于何处,这常常是不重要的,而且对大多数目的而言,它几乎是无关紧要的。

对于心理表征理论应做什么的问题,我们现在给予第二种相当不同的回答看来是时候了。根据这第二种说明,心理表征理论不太关注心理表征的常识概念。大街上的那个男人或女人的直觉或缄默的知识完全没有什么关系。该理论想说的是:心理表征实际上是什么,民众心理学认为它不是什么。为了做到这一点,它就必须描述进而也许是匆忙拼凑心理表征的概念,就像我们可得到的、最好的认知科学所使用的概念那样。因此根据这种说明,心理表征理论一开始就是认知科学的认知心理学的组成部分,即使它最终也能成为它试图描述的那门科学的概念基础的组成部分。

在关于心理表征的大量文献中,我只知道一个作者,他旗帜鲜明地从事着我所陈述的那项工作。他就是 R. 卡明斯,在他的《意义与心理表征》一书中,他对心理表征的概念提出了详细的说明。但是他特别强调,他所关心的概念不是我们关于意向描述的日常语言以之为基础的民众心理学概念(p. 26)。毋宁说,他的目的就是対认知科学中的一种值得尊重的并仍有生命力的研究传统所使用的心理表征概念给以说明,这种传统试图建立卡明斯所谓的关于认知的“正统”计算理论。它“假定认知系统是自动地被解释的形式系统”(p. 13),而且过去 20 年所做的关于问题解决、计划、语言加工和高水平的视觉加工的大量工作完全可放入正统的计算范式之内。

卡明斯的方案的一个必要的部分就是阐明关于认知计算理论的解释战略。他说明了这些理论试图解释什么,在这种范式中成

功的解释必须做什么。既然在这些计算理论中所用的表征概念必须澄清被提供的解释的意义,那么这种解释结构便对关于心理表征的说明施加了很强的限制。下面便是卡明斯对他的研究方案的描述。

首先要确定表征在某种特殊的、诉诸于表征的科学理论或理论框架中起着什么样的解释作用;然后再追问:表征必须是什么——它应当怎样被说明——如果它真的有那种作用的话。(p. 145)

即使卡明斯的目标是认知的计算理论中所用的表征概念,但是他认识到:这不是认知科学中唯一有前途的研究传统。“在当今认知科学的赛跑中,存在着许多不同的构架”(p. 26),包括“正统的计算主义、联结主义、神经科学”(p. 12)以及其他各种传统。相同的方案极有可能用于别的一些传统所利用的表征概念,即使结果最终可能完全不同。“假设……(别的研究传统)都使用了相同的表征概念,这种假设似乎是天真的”(p. 12)。如果我们问:这些构架的每一个是怎样理解心理表征的,“那么我们不可能得到齐一的回答”(p. 26)。

这种多元论图景是我极力支持的一种观点。它为我本节中所叙述的心理表征理论之说明,增加了一个重要维度。因为如果认知科学之内的不同范式用了不同的表征概念,那么就将有我们一直在讨论的那类心理表征理论,而会出现许多理论,而且追问这些理论中的哪一个是正确的也没有意义,因为它们并不是相互对抗的。每种理论都旨在描述认知科学某一支中所使用的表征

730 第六部分

概念。如果认知科学的不同分支用了不同的表征概念,那么对心理表征将有许多正确的说明。自然也可以这样认为:认知科学的不同分支本身是对立的,而且关于心理表征的正确理论就是这样一种理论,它描述了由正确的认知科学所用的心理表征概念。但是我没有理由假定:在认知科学中,存在着关于那些理论的独一无二的正确构架。有许多要解释的现象,进而有许多层面,富有启发性的、科学上可接受的解释可以在其之上被给予。因此在这个范围内,我不如做一个多元主义者。

这不是详细讨论卡明斯对心理表征概念的说明的地方,因为它常用于关于认知的计算理论中。但是卡明斯的工作中有一些课题是我想进一步探究的,因为它们将把我们带回到这样的问题上,即心理表征理论在与取消论的论战中应怎样发挥作用。

5. 心理表征理论与取消论的论战^⑩

正如卡明斯所注意到的那样,他想描述的心理表征概念是从那个系统的历史和“一个系统的普通环境中的实际项目”里抽象出来的(p. 81),“根据计算主义,认知系统是由它们的计算属性而得以个体化的。”(p. 82)。两个系统即使在其历史、环境甚至在物理构成上彼此有别,但仍可以具有相同的计算属性。由这种心理表征概念所产生的分类学是个体主义的——如果两有机体或系统具有相同的物理构成,那么它们的表征状态表征的就是同样的事物。不过,卡明斯紧步普特南、伯奇等人的后尘,也主张:民众心理学所用的关于意向状态的分类学则是反个体主义的——“信念、愿望不

能以独立于环境”或历史的“方式而加以具体说明”。卡明斯由此所得出的结论是：信念、愿望以及常识心理学的别的意向状态都不属于认知的计算理论所认识的项目之列。“根据关于[认知的计算理论的]观点，普特南和伯奇的反个体主义论证所证明的就是：信念和愿望在 CTC(即认知的计算理论——译者)感兴趣的‘心理状态’的意义上并不是心理状态”(p. 140)。

看来，我们这里所拥有的正好是赞成取消论的论证的开端，在这一论证中我们所描述的两类心理表征理论正好起着一种作用。概括地说，论证是这样进行的：首先是描述常识的心理学和认知的计算理论所使用的心理表征概念。接着，对两者作出比较。如果它们有重大的差异，那么常识心理学的表征状态就不是计算理论的本体论的组成部分。如果这类论证不受制于认知的计算理论，那么它就不能证明取消论者是有理的，因为正如我们已注意到的那样，当代认知科学是一种斑驳纷呈的学科，而且别的各式各样的研究传统在那里争奇斗艳。因此要对取消论作出合理的辩护，这种论证就必须能为认知科学的竞争场所中每一种可行的研究传统所重复。作为这种理由套理由的方案替代物，取消论者可能试图压缩这一过程，以替换这种证据接证据的方案，方法是说明存在着某些特征，它们在任何科学上可接受的心理表征概念不得不具有的，进而再论证这些特征不能由内在于民众心理学中的、对心理表征的说明所认可。^①

然而，即使所有这一切都有利于取消论者，但是他无疑还没有证明自己是合理的。为了知其所以然，让我们还是回到卡明斯对常识心理学所解释的信念与认知计算理论中的表征状态所作的比

732 第六部分

较。根据卡明斯的看法,常识心理学从反个体主义的角度看待信念——它们不能以独立于环境的方式而得到具体说明。相比较而言,认知的计算理论所断定的心理状态从个体主义角度来看,则能被个体化——它们能独立于环境而被具体说明。由此出发,卡明斯似乎得出结论说:常识心理学的本体论不同于认知的计算理论的本体论。两种理论谈论的是不同的东西。而且以此方式进行推论并不只卡明斯一人。我本人像某些别的作者一样,在以前的大量的出版物中曾提出了类似的论证。^⑩尽管有如此不同的辩护者,但论证的前提是否支持其结论则一点也不清楚。从前提可衍推出的是:民众理论和计算理论对它们所谈论的状态提出了不同的、难以调和的断言。但是这肯定不足以表明:它们谈论了不同的东西。如果是这样,那么理论家们分道扬镳就几乎是不可能的。难道不可能有这样的情况吗,即民众心理学和计算理论谈论的实际上是相同的東西,而错误的正好是民众心理学?

其实,这里争论的不过是这样的问题,即是什么决定了一种理论所用的那些术语的指称。那些支持取消论者的人常常这样写作,仿佛他们已接受了指称的摹状理论的某种观点。但这是取消论者所采纳的一种极端可疑的理论。危险之一是,摹状理论的朴素的观点往往使取消论丧失其重要性。如果对于心理状态,常识所说的和认知科学所说的之间存在的微妙的不一致足以表明:常识和认知科学假定了不同的实在,那么取消论自然是正确的。但是有谁注意呢?从没有人这样认为:常识心理学最终被证明是一贯正确的。而且,如果我们假定:理论上的微妙的差异总会导致不同的本体论,如果我们设想后一类理论常常比前一类理论更接近

于真理,那么我们最后就有一个相当疯狂的观点——一种泛取消论。因为肯定有这样的可能,即在下一个世纪,我们现在所接受的每一种理论将有某些改进。如果这足以表明:流行的理论所假定的实在并不存在,那么我们现在所相信的一切东西都不存在!

关于指称的摹状理论的第二个关注点就是:即使是该理论的那些极其精致的阐释也完全没有触及到实质。如果它们是这样,那么从民众心理学和认知科学在心理状态上针锋相对这一事实就不可能引出令人感兴趣的本体论结论。清楚地注意到这一点的一个哲学家就是 W. 利康。下面就是他对该问题的看法:

我倾向于抛弃刘易斯关于理论术语指称的卡尔纳普式的或赖尔式的群集(cluster)理论,而接近于普特南的因果历史理论。像在普特南的“水”“老虎”等例子中一样,我想让日常使用的“信念”(作为民众心理学的理论术语)模糊地指向我们尚未充分把握的、只能由成熟的心理学昭示的自然种类。我预期:“信念”最终会指称有知觉能力的存在的、携带着信息的某种内在状态……但是它所指称的那类状态大概只有通常由常识赋予信念的一些属性。因而,我认为:我们分辨信念和愿望的常见方法在分辨自然中的真实实在时也是有效的,不过它也许无助于分辨常识主张它能予以分辨的实在。(1988, p. 32)

正如利康所强调的,这种观点的一个结论就是,我们的常识理论关于信念和别的表征心理状态的本质的看法最终也许是极其错误的:

我完全情愿放弃我们关于信念或愿望(或其他任何东西)的众

734 第六部分

多常识的或陈腐的理论,并断定:如果不作出这样的结论,即我们不再谈论信念或愿望,那么我们正好陷入了一系列的错误而不能自拔。(1988, pp. 31-32)

不幸的是,当争论集中在取消论时,利康的路线像摹状理论的朴素观点一样有着同样的缺陷——即使是在对立的方向上。因为根据利康的看法,难以澄明的是:怎么可能说明民众心理学的假定不是认知科学已知分支的本体论部分。甚至,在利康看来,我们为什么要说燃素真的不存在,这是极不清楚的。它就是我们现在称之为“氧气”的材料,而早期的理论家“正好陷入了一系列的错误而不能自拔。”因此似乎是,如果我们接受利康所赞成的指称理论,或者接受朴素的摹状理论,那么取消论者的主张将是不屑一顾的。根据摹状理论,取消论则有点真实性;而根据因果-历史理论,取消论则有一点错误。

这一切把我们带到了什么地方呢?如果我们想把取消论当作一种有趣的学说,而不是有点真或有点假的学说,那么我们对指称的说明与朴素的摹状理论相比就得少些限制,与因果-历史理论相比则要多些限制。很显然,我们的确想把取消论当作有趣的学说——或者说我过去常常这样想。因此当我开始进行我已述及的论证时,我最初的反应就是到处搜寻关于指称的更有前途的陈述。然而,我现在认为:那是一个严重的错误。

问题不在于:对指称的替换性说明难以找到;恰恰相反,构建似乎正好能说明我们想说明的东西的关于指称的说明是相对容易的。它们比因果-历史理论有更多的限制,比摹状理论有更少的限制。但这又引发了迄今似乎非常熟悉的问题:这些指称理论的

哪一个是正确的呢？是什么东西使那种理论正确？而且，根据指称概念与心理表征概念之间的直接联系，再清楚不过的是，第3、4节中关于后者所说的，能以相同的合理性在关于前者的说明中得到重复。如果是这样，那么对指称就不存在任何唯一正确的说明。毋宁说，将有这样一种说明，它描述了我们关于指称的常识概念（或者说如果不只一个常识概念在起作用，那么有几种说明），而且还将有另一种说明，它描述了在心理学或语言学或认识论或某种别的学科中对一种或另一种计划有用途的似指称的概念。

如果这都是对的，那么某种令人震惊的结论便会随之而来。首先，取消论不能被视之作为一种单一的命题，甚至也不能像在第2节中那样把它当作是两个命题。为了弄清这一点，试考察先前区分过的两个取消论论点中较弱的——它主张：常识心理学的假定并不是认知科学本体论的组成部分。如果我们最近的思考是对头的，那么这个论点就失去了意义——它没有明确的真值条件——除非它与指称的某种特定的说明有联系。因此如果对指称有许多完全正确的说明，那么关于“弱”取消论的命题就有许多不同的解释。而且可以合理地假定，基于其中的一些解释，取消论者的断言最终会被证明是真实的，而基于另一些，它将被证明是虚假的。但是如果这是正确的，那么就不能肯定：关于取消论的命题的任何解释不过是那种使一种断言令人感兴趣的东西。在我们意识到关于该学说的任何可理解的观点不得不相关于或隶属于一种指称理论之前，也许可以合理地主张：如果取消论是真的，那么某种严重的理智灾难就将接踵而至。毫无疑问，这更加使该学说妙趣横生。但是一当我们明白有把该学说隶属于某种特殊的指称理论

的必要时,情况就将是另一番景象。因为肯定没有人愿意说:不管取消论被相关于什么样的指称理论,它也将摧毁理智的屋脊。自然,人们仍可能认为:存在着某种特殊的指称理论,因而如果取消论由于隶属于该理论而成为真实的,那么令人不安的结论便将随之而来。这大概是对的,谁知道呢。但如果它是对的,那么肯定也是不明显的;这正好是一个需要论证的主张。这个论证该怎样进行,我没有一点头绪。因此只有当我们得到了关于该问题的某种领悟时,我认为才有理由怀疑:取消论的益处被过分夸大了。

6. 取消论与自然主义强制^⑬

大家还记得,福多有这样的观点,即取消论或“意向非实在论”的“最深刻的动因”就是这样的推测,即“有意向的东西是不能被自然化的”(1987, p. 97)。福多所想到的对于非实在论的含蓄的论证大概有前一节所描述的结构:一个概念要想在有关的科学理论中得到利用,就必须能够被自然化。因此如果意向概念不能被自然化,那么它们在任何可敬的科学理论中就不能被利用。福多本人的内容理论在很大程度上是出于这样的期盼,即这种推测可能被埋葬。在这最后一节中,我想考察与此有关的两个问题。第一,缓解“有意向的东西不能被自然化”这一忧虑意味着什么,我们所考察的那一种心理表征理论在这一过程中怎样发挥作用? 第二,如果该计划失败了,而且我们发现我们不能将意向概念自然化,那么将有怎样的恶果?

对第一个问题的回答似乎是很清楚的,至少就轮廓而言是如

此。为了不再受意向的东西不能被自然化这一担忧的困扰,我们不得不对下述心理表征概念作出自然主义的说明,它用于(或可能用于)认知科学之中,而且心理表征的常识概念可能与之相同一。不过,这一回答引出了两个问题。第一个简直就是我们前一节全力对付的问题的一种翻版:证明两个理论概念的跨理论同一意味着什么?在我看来,对此问题不可能有确定的回答。但是在前面我已说过我对此问题的一点想法,这里用不着再讨论这个问题。第二个问题是从本文一开始就潜伏着的问题,即当将心理表征“自然化”的争论开始提出来时便有这样的问題:说心理表征说明是自然主义的说明意味着什么?即使我知道没有人对此问题作出过详尽的回答,但文献令人信服地表明:那些想对心理表征作自然主义说明的人想得到的是像定义一样的东西——一组必要而充分的条件——用可为物理或生物科学确凿无疑地接受的术语予以表达。

能否为心理表征提供恰如其分的自然主义的说明,这自然是一个悬而未决的问题。我自己的猜想,不管其价值如何微不足道,就是:该计划是完全没有希望的。不过,与福多以及其他许多人相比,我倾向于认为,用不着死死抱住那个问题不放。福多主张,如果我们不能对心理表征作出自然主义说明,那么该概念在严肃的科学中就没有地位。果真如此,取消论者就将取得重大的战果。更有甚者,他们也许会取得胜利。但这一主张给我以固执己见的印象。要知道为什么如此,我们只要考察几个例子就够了。我们先看音素的概念。成为一个/p/或一个/b/意味着什么?如果你想要自然主义的回答,即用物理学或生物学术语提供充分必要条件的回答,那么恐怕你会大失所望。因为尽管有多年深入的研究,但

现今尚无有效的自然主义回答。^④当然,这一情况可能发生变化。语音学家可能对一连串是/p/的声音意味什么作出自然主义说明。但是常见的情况大概还没有变化。如果没有变,那么成为一个语音取消论者,进而否定语音的存在,这肯定是没有理由的。关于科学上毋庸置疑的有用之物的其他概念也完全可以如是说。在灵长目动物的性格发生学中,对照料行为不可能作出自然主义说明。在刺鱼的性格发生学中也不可能对攻击行为的自然主义说明。但是因为我们不能用物理学和生物学语言去定义照料行为,而否定它的存在则显然是有悖常情的。受过适当训练的观察者可能以极高的主体间的可靠性觉察到照料行为(音素)。而且我想论证的是,这足以使那些概念成为经验上可接受的。想要更多的——尤其是想要所讨论的概念被自然化——似乎是动机不明的,而且是愚蠢的。心理表征的情形看上去是完全类似的。做一个取消论者也许有充分理由。但心理表征不能被自然化这一事实不是这些理由中的一个理由。

注 释

① 关于因果协变理论,请看德雷特斯克(1988),福多(1990b)。关于目的论理论,请看福多(1990c),米利肯(1984),帕皮诺(1987)。关于功能作用理论,请看布洛克(1986),菲尔德(1977),洛尔(1982)。关于因果指称理论所引发的理论,请看德维特等(1987)。关于单因素理论,请看哈曼(1986)。关于多因素理论,请看麦金(1982)。关于狭义理论,请看福多(1987),德维特(1990)。关于广义理论,请看伯奇(1979)。

② 关于作为我们通过描述其内容对心理状态作出归因的能力之基础的机制,还有一种相当不同的说法,可参阅戈登(1986),戈德曼(1989)。关于戈登/戈德曼的观点的拓展性批评,可参阅斯蒂克和尼科尔斯(1993)。

③ 参阅麦克洛斯基等(1980),麦克洛斯基(1983)。

④ 参阅布洛克(1986),p. 660;菲尔德(1986a),p. 444;琼斯等(1991),§ 4.2;洛伊尔(1987),p. 296。

值得注意的是,福多在几种场合阐述了这样的主张,即他满足于“世界的一部分关于(表达、表征或符合于)另一部分”的充足条件,即使它们不是必要条件。(福多,1987,p. 98,还可参看福多,1990a,p. 52 ff.)

但是正如琼斯等(1991)所说明的那样,如果我们从字面上去理解他,那么很难相信:这就是福多真的想要的东西。因为提供仅仅是充分的条件正好也是易如反掌的事情。

如果 x 是福多最近关于“玛丽亚·卡拉斯”(1923-1977,生于美国的希腊女高音歌唱家——译者)所说出的话语(或者说,如果 x 是引起那话语的概念),那么 x 表征了玛丽亚·卡拉斯。

如果 y 是福多最近关于“意义整体论是一种怪诞的学说”所说出的话语(或引起它的思想),那么 y 就是关于意义整体论的,并且是真的,当且仅当意义整体论是一种怪诞的学说。

这是两个充分条件,而且我还可以不费太大代价说出许多来。

⑤ 像罗施经常述及的那样,她在该领域的工作是由维特根斯坦的《哲学研究》所促发的。

⑥ 关于原型与样本理论的文献的杰出的评论,可参阅史密斯等(1981)。

⑦ 参阅墨菲等(1985);巴萨卢(1987);里普斯(1989)。

⑧ 例如可参阅,默多克(1980)。

⑨ 关于适当性,参看索伯(1984b);关于合乎语法性,参看福多(1981c);关于时-空,参看斯克拉(1974)。

⑩ 本节部分出自于斯蒂克(1991)。

⑪ 事实上,这就是我所尝试过的战略(1978)。另一类战略仍有利于取消论者的下述论证:认知科学中的一种或另一种竞争的研究传统并不是可怕的竞争者,因而用不着考虑其功过。

⑫ 参阅斯蒂克(1979),(1983),第二部分,以及斯塔克(未发表)。

⑬ 关于这节所讨论的争论的更详细的看法,可参阅斯蒂克等(1994)。

⑭ 至于有价值的观点,可参阅弗赖伊(1979)和皮克特(1980)。

740 第六部分

参考文献

- F. 德雷特斯克(1983):“知识概略与信息流”,《行为与脑科学》,6,pp. 55-63。
- J. 福多(1987):《心理语义学》,麻省理工学院出版社。
- (1990a):“信息与表征”,载 P. 汉森编:《信息、语言与认知》,英吉利哥伦比亚大学出版社。
- (1990b):《内容理论与其他论文》,麻省理工学院出版社。
- (1990c):《心理语义学》,载利康(1990)。
- E. 史密斯等(1981):《范畴与概念》,哈佛大学出版社。
- G. 墨菲等(1985):“理论在概念一致性中的作用”,《心理学评论》,92。
- L. 里普斯(1989):“相似性、典型性与范畴化”,载 S. 沃伊斯尼多等编:《相似性、类比与思想》,纽约,剑桥大学出版社。
- R. 米利肯(1984):《语言、思维和其他生物学范畴》,麻省理工学院出版社。
- D. 帕皮诺(1987):《实在与表征》,布莱克韦尔。
- H. 菲尔德(1977):“逻辑、意义与概念作用”,《哲学杂志》,74:379-409。
- B. 洛尔(1981):《心灵与意义》,剑桥大学出版社。
- M. 德维特等(1987):《语言与实在》,麻省理工学院出版社。
- G. 哈曼(1986):“广义的功能主义”,载 R. 哈尼什等编:《知识表征与信念》,亚利桑那大学出版社。
- C. 麦金(1982):“内容的结构”,载 A. 伍德菲尔德编:《思维的对象》,牛津大学出版社。
- T. 伯奇(1979):“个体主义与心理的东西”,《中西部哲学研究》,4:73-121。
- R. 戈登(1986):“作为模拟试验的民众心理学”,《心灵与语言》,2:158-171。
- A. 戈德曼(1989):“心理学化的解释”,《心灵与语言》,4:161-185。
- S. 斯蒂克(1991):“真正的相信者存在吗?”《亚里士多德学会论文汇编增刊》,65:229-244。

S. 斯蒂克等(1993):“民众心理学:模拟试验还是不言而喻的理论”,《心灵语言》,7:112,35-71。

M. 麦克洛斯基等(1980):“无外力作用的曲线运动:关于对象运动的朴素信念”,《科学》,210。

T. 琼斯等(1991):“驱逐灾变:J. 福多的心理语义学的批判性评注”,《心灵与语言》,6:58-82。

B. 洛伊尔(1987):“信息与信念”,《行为与脑科学》,5:79-80。

G. 默多克(1980):《关于疾病的理论》,匹兹堡大学出版社。

E. 索伯(1984b):《选择的本质》,麻省理工学院出版社。

D. 弗赖伊(1979):《关于言语的物理学》,剑桥大学出版社。

内 容 理 论*

J. 福多

I . 问 题

引 言

在哲学中,有这样两个习以为常的看法:(I)意向/语义谓项形成一种封闭的循环,(II)意向状态在本质上是整体性的。(I)可以解释为:“用包含了别的意向/语义谓项的词汇系统表述满足某些意向/语义谓项的充分条件也许是可能的;但用纯属非意向/语义的词汇表述这些条件则是不可能的。”(II)可以解释为:“任何东西都不能表现意向属性,除非它表现了许多意向属性;一事物处于任何意向状态的形而上学的必要条件包含它处在许多别的意向状态中。”(I)据说排除了从物理主义方面说明意向归因的真值的充分条件的可能性;(II)据说排除了点状心灵的可能性。

分别地、全面地加以考察,(I)和(II)有助于为关于意向解释的哲学怀疑主义提供坚实基础。例如,(I)似乎排除了关于心理学的物理主义本体论,因为如果心理状态是物理的,那么对于它们

* 译自J. 福多:《内容理论与其他论文》,麻省理工学院出版社1990,第3-4章(第3章为“I:问题”,第4章为“II:理论”),有删节。

的例示来说,就肯定存在着可从物理主义方面加以详述的充分条件。^①但是问题在于,如果心理学的本体论不是物理主义的,那么就没有这样的科学。

比较而言之,(Ⅱ)真的能与物理主义相一致;为什么就没有既是物理主义的又是整体论的属性呢?不过(Ⅱ)仍有可能排斥具有科学地位的意向心理学。心理学规律获得普遍性的一个重要方式就是用量词限定所有处在特定心理状态中的有机体(如相信 P,或意欲 Q 或别的什么的所有有机体)。但整体论意味的则是:如果拥有许多心理状态中的任何一种,那么就一定拥有其全部。所以,心灵越是整体论性质的,两个有机体(或同一有机体的两个时段)的心理生活就越相似,以致它们都从属于相同的心理学规律。在整体论的范围内,只有当两个心灵具有它们的全部意向状态时,它们才具有这些意向状态中的任何一个。既然两个心灵根本不可能具有它们的所有意向状态,那么(Ⅱ)越真,意向心理学假定的普遍化事实上就越不可能普遍化。^②(一旦接受了关于意向内容的整体论观点,是否有可能在什么地方停下来,以免走向极端,这是一个饶有兴味的问题。我倾向于认为:严肃地承认不存在分析/综合区分的任何人都不得以否定的方式回答这个问题。不过我不想在这里论证这一点。)

概而言之,意向实在论值得付出的代价——至少为着科学的目的——就是对意向状态作出物理主义和原子论的说明。正如我所述,不可能得到这样的说明,这是哲学中普通的常识。

不过也有声势与日俱增的少数派,他们与这种多数人的意见分道扬镳。尤其是“信息”语义学的最新发展预示了这样的可能

744 第六部分

性,即可能出现关于谓词与其所表达的属性之间关系的自然主义和原子论理论。这样一种理论自然不是对意向性的一种完善的理解,只能算作一个好的策略。但它有助于引出怀疑论者的利爪,因为他的路线是:不可还原性和整体论是意向性和语义可估值性所固有的。假如在意向的怪圈中有适当的原子论和自然主义的突破,那么就可合理地断言:关于意向性的主要哲学问题已被解决了。剩下的则是这样一项工作,即或多或少的经验理论的重建,或者是多少有点普通的工作。

随之而来的也许是写一篇回顾性的文章;近来,“自然化语义学”^③方面的工作进展是如此之快,以致依我看来,一般性的考察也可能是有用的。这正好是我这里打算要做的。第3章中,我将就过去20年关于自然化问题的探讨怎样进化作出概述。(既然我首先想做的事情是澄清对自然化问题结构的流行看法,那么我的态度就将是辩证的、论辩性的,我将满足于使用我的具有历史准确性的、平常的聪明因子。)在第4章*中,我将说明:什么是关于以信息为基础的语义理论的、在我看来是有前途的解释:它将采取谓词表述属性的物理主义的、原子论的和假定的充分条件的形式。对于我的朋友同道与我一起成功地虚构的这个条件所提出的反例与反论证,我将仔细予以讨论。我想让你(以及我和灰猫)相信:这些反例和反论证都是无济于事的,或者说它们不管用什么方式注定是没有用的。

不过,即使我认为它们没有用这一点是对的,但某人明天又一定会找到一个有用的反例。因此这篇论文提出的主张不是说:不

* 即本中译的下一部分:“内容理论之二:理论”。

再会有关于意向性的哲学问题。毋宁说,我倾向于作出的——但愿我能使你心悦诚服加以严肃地接受的——主张就是:曾使自然主义语义学建构似乎是绝对无望的大量问题现在比过去显得要好对付得多。因此当人们在语言哲学和心灵哲学中开展自己的工作时,对承认意向非实在论持审慎态度是很明智的;在最低限度上,采取比过去 40 多年的哲学风格更审慎的态度是明智的。

1. 背景:德雷特斯克

我认为:F1 说出了德雷特斯克的理论的基本观点。

F1. 如果“Ps 引起 Ss”这一概括是反事实的辩护,S-事件(例如符号的标记)便表述了 P 属性。

例如,“狗”的标记表述了狗的属性,因为“狗引起‘狗’的标记”这一概括是反事实辩护。

我欣赏如此陈述德雷特斯克主张的方式,因为它澄清了他的图式与斯金纳图式的连续性。不过用德雷特斯克自己的表述说,基本的语义关系是“携带信息”(而不是“表述属性”)。对携带信息的初步解释是由 F2 提供的。

F2. 如果“Ps 引起 Ss”是一规律,那么 S-事件便携带着关于 P-事件的信息。^④

不过,F2 也是德雷特斯克难以接受的。例如根据他的理论,Ss 携带关于 Ps 的信息,当且仅当任意的 S 由 P 引起的概率始终是 1;而事实上,德雷特斯克只要求:“Ps 且只有 Ps 引起 Ss”是一条规律。

他对这种极强条件的主要论证是这样的:假如我们承认:即使

Ss 由 P 引起的概率是低于 1 的某种 P, Ss 也携带着关于 Ps 的信息。这样我们可得到一种情境, 在其之下, Ss 携带关于 Ps 的信息, Rs 携带关于 Qs 的信息。但是 S&Rs 并不携带关于 P&Qs 的信息(即是说, 因为 P&Q 以 S&R 为前提的概率低于 p)。^⑤

但我认为这个论证是有毛病的。一种语义学理论为什么非要独立地把信息内容归属于一种符号系统中的每一表述, 这完全是没有理由的。如果内容只归属于原子表达式, 关于分子符号的语义学是借助某些常用于真值定义建构中的技术建立起来的, 那么它才会如此归属。在下面, 我要假定的是: 把表征自然化的问题可还原于用原子符号(换言之, 如果被自然化的正好是心理表征, 那么它们就是原子心理状态)把表征自然化的问题。

F1 和 F2 的关系比其表面的关系更为直接, 因为我们可以假定: 只有当“Ps 引起 Ss”是一条规律时, 它才是反事实辩护。因此, F2 中表现明显的、信息与法则学性质之间的联系在 F1 中则是隐晦的。因为关于规律和反事实辩护的观念与斯金纳、德雷特斯克两人的语义学核心观点有如此紧密的联系, 因此那些理论具有一个重要特征, 它对我们后面的讨论非常重要: 两者都意味着: 你的词语的意思是什么完全取决于你标记它们的倾向, 它们的标记的实际历史从语义学上说是没有关系的。

这个原则——实际的历史在语义学上无关——源于信息语义学的下述基本观念, 即一种符号的内容唯一地由它的具有自然法则普遍效力的关系所决定。粗略而直观地加以表述, 一事物服从什么规律关系到它的虚拟进程; 关系到它做了(或曾做了)什么, 如果环境是(或曾经是)如此这般的话。比较而言, 一事物的实际历

史不仅依赖于它所从属的规律,而且也依赖于它碰巧遭遇到的环境。斯金纳与德雷特斯克假设自然化语义学否认了实际历史而赞成纯虚拟偶然性,这是否正确,是我们在第4章中将讨论的问题。直到那时,我们将严格地坚持这样的原则,即只有具有自然法则普遍效力的联系和他们所特许的虚拟式才是有意义的。

2. 错误和选言问题

你们一定在什么地方出了差错,而至今我们又没有办法找到它。事实上,对此看上去存在着一种二难推理。粗略地说,假设“狗”意指狗(在它的外延中,有狗且只有狗)是因为狗引起“狗”是一条规律。那么就有两种可能性:

第一种可能性。只有狗才引起“狗”。果真如此,那么只有在“狗”的外延中的东西才引起它被标记。因此“狗”的所有标记似乎必定是真的。

第二种可能性。某些非狗引起“狗”。例如,假设不是狗就是黑夜中的猫(类)充分地引起了“狗”的标记。事实上,F2说的是,那些符号表示的是这样的属性,它们的例示对于它们的标记从法则上说是充分的。所以“狗”表示的不是作为狗的属性就是作为黑夜的猫的属性。因此“狗”的外延是狗和黑夜中的猫的并集。被黑夜中的猫引起的关于“狗”的标记是真的,而且我们对虚妄和错误仍一无所知。

如果F2是一种关于内容的因果理论所能达到的最高成就,那么那些理论好像不能在意指选言的某东西的符号的真实标记与意指非选言的某东西的符号的虚假标记之间作出区分。信息语义

学的文献一直把这称作“选言问题”。

情况实际上会是怎样呢？似乎有这种可能，即在一个符号的虚假标记中你最不希望有的东西是：它被不在那符号外延中的某事物所引起。但是这恰恰是 F2 陷入困境的一个条件。因为：

(I) 一符号的每一标记(包括虚假的)由这样的某物所引起，它具有足以引起该符号标记的某种属性，这是自明之理。而且

(II) 根据 F2, 任何属性, 其例示足以引起符号的标记, 是由此而为那符号所表示的。

既然一符号的外延正好是这样的一组事物, 它们具有由那符号表示的属性, 那么由 (I) 和 (II) 似乎可得出结论说: 一种符号的每一标记是由属于其外延的某物所引起的; 因此一种符号的任何标记绝不可能是虚假的。说得婉转一些, 这是不令人满意的。

更有甚者, 它如此不令人满意, 以至于自然主义语义学是否可能这一问题近来事实上被等同于选言问题能否在自然主义框架内得到解决这一问题。因而本论文的其余部分将讨论近来寻找这种解决方案的种种尝试。

除了我后面要详细讨论的之外, 所有解决选言问题的标准尝试都表现出了某种家族相似性。基本的观点就是区分两类情境, 以至于有规律的协变在一类情境下决定了意义, 而在另一类情境则相反。事实上, 得到修补的理论则说: 一种符号表述一种属性, 条件是那属性的例示对一类情境下的符号例示从法则上说是充分的。既然发生在一类情境下的符号标记就事实本身而言是由在符号的外延中的事物引起的, 那么便可得出结论说: 如此这般的所有标记都是真实的。然而, 这样一些属性, 其例示(只)在第二类

情境下引起符号标记,并不由此而为那符号所表述;所以在第二类情境下出现的符号标记就事实本身而言并不是由在那符号的外延中的事物所引起的;因此剩下的只能是:这些标记也许是错误的。

这一来,那得到修补的理论的战略就是要通过限制选言问题而对之作出解决。大家都承认:第一类情境中的符号标记事实上是真实的;^⑥而且由此还得承认:如果一种符号的标记是由第一类情境中的一种以上的事物所引起的,那么那符号的意义就是选言的。但是根据新的说明,并非所有类型的情境都享有这种传递不可错性的特权;例如第二类情境就不能如此。因此新的说明就为说明错误的可能性提供了条件,正如我们已知的那样,旧的理论则不能做到这一点。

思考第一和第二类情境之间差异的方式尽管有趋同性,但也有微妙的区别。哲学家可能会合理地碰到这样的问题:“我们对思维、信念等的规范的说明是借助于运用那些(明显地)表示了实际或可能的事态的——内在于 that 从句中的——短语而完成的,这是为什么?例如我们为什么总是通过使用‘天要下雨’这一表达式而领会天要下雨这一想法。使这种实践成为可能的思想、事态又怎么样呢?”(帕皮诺,1988,大约 p. 88,提出了这类问题,洛尔也是如此,1981)。这与下面要展开的、我相信首先是由 D. 戴维森提出的问题有密切的关系,这问题是:我们应怎样理解下述事实,即能表现为独立式直陈句的表达式也能以命题态度的动词补足语的形式出现吗?

对此,所有信息说明在本质上都是千篇一律的;它们认为:会发生的就是一种原因论上的确认。当我们用“天要下雨”陈述天要

750 第六部分

下雨这一思想的意向对象时,我们不过是参照在某些环境下引起其被持有的事态而领会那个想法。这很像由酒精中毒所引起的恍惚;你是参照使该状态发生的那类事物而说明该状态。

至此都不错;但一般来说,既然一意向状态的标记可能有各种不同的原因(顺便说一下,不像由酒精中毒所引起的恍惚标记),问题便来了,即在哪一些环境下思想的原因就事实本身而言同一于它的意向对象。回答是:根据定义,这种巧合是在第一类情境下获得的。事实在于:选言问题是促使信息语义学家尝试作出第一类和第二类情境区分的一个原因,但不是唯一的原因。其他的哲学兴趣也有相同的迫切要求。

因此一切都很正常;对于类型一与类型二的区分,我们所需要的是一种令人信服的——自然是自然主义的——解释,我们将在关于内容的信息解释的框架内理解下述两个问题,即错误怎么可能,用我们常用的方法使意向状态个体化怎么可能。然而,对这种区分的令人信服的自然主义解释原来是建立在一种脆弱的基础之上的。

3. 德雷特斯克对错误的描述(略)

4. 目的论/功能主义方案

解决选言问题的基本观念已如此被限定在第一类情境中:

(I)如果在第一类情境中, P_s 引起 S -标记是一条规律,那么 S 意指的就是 P (如果 P 是选言的,那么也是如此);而且

(II) S 在其之下得到标记的一切情境并非都有资格作为第一类情境,所以发生在其他类型情境下的 S 的标记据事实而言就不

是虚假的。

第一类情境似乎不可能是学习的情境;但这里还有一种可供选择的建议。正常的情境就是我们所需要的那类情境。下面我们花一点时间来考察这种建议。

乍一看,这种观点是颇有吸引力的;它源于下述似乎合理的直觉,即错误就是某事在其中出差错的情况:“在信念虚幻的地方……我们也就有望解释对规则的背离:要么是环境中的异常,就像在光学幻觉或别的误导性证据中那样,要么是在内在信念形成机制中的异常,就像在一厢情愿的思维或错误的记忆中那样”(斯托纳克,1984,p.19)。相反,正常的情境大概正好是这样的情境,即在其中一切都无偏差。无论在哪种情况下——既然每当一切都无偏差时,你所相信的就可能为真——如果 S- 标记在正常情境下都为真,那么它也许就是对的。

如果在正常情境下,意义与真值的条件最终是相同的,那么它也可能是对的。正常的——至少当它如此被运用时——是一种规范的概念,而且真也是一种规范的概念,因此如果用前一概念重构后一概念,那大概也没有什么奇怪的。乍一看,人至少都是喜欢争辩的。

诚然,如果意向循环应诉诸于符号标记的正常情境而予以打破,那么我们最好是有某种自然主义的描述,以澄清:情境在有关方面为正常意味着什么。这样一种描述可能是什么样子呢?粗略地说,其设想就是:正常性应当是由于(自然的)目的论而莫明其妙地实现的;即由于某种关于生物学机制的或多或少的达尔文式的/历史的概念而实现的,正是这些机制做了它们被选择做的事情。

752 第六部分

因此,那个描述的梗概是这样的:有机体的心理状态标记是由这样的事件如发生在有机体的居住环境中的事件所引起的。当然,机制——典型地说是中枢机制——也是存在的,它们在这些因果交互作用中起着中介作用。这些机制大概经历了一种进化的历史。它们大概是选择过程的产物,而且很有可能的是它们选择的东西明显就是它们在联结心理状态标记中所起的作用。因此存在着这样一些认知机制,也存在着这样一些认知状态;前者的功能就是在适当的环境下产生后者的示例。

严格地讲,自然不能由此说:认知状态本身——像相信 P 或意欲 Q 或怀疑逃跑的人将搬回布鲁克林之类的状态——有正常的功能;事实上不能推论说:它们具有全部一切功能。(你完全可能拥有一部机器,它的功能就是产生本身无功能的东西。在消费者协会,你可能得到大量这样的东西。)有一种目的论陈述,据说它描述了信念形成的机制,这一假定并不意味着存在一种描述了信念的目的论陈述,它更没有这样的含义:信念(或换言之,别的意向状态)能借助它们的功能而个体化。这一点很重要,因为信念形成机制(例如神经系统)有功能比信念有功能更加直观。后一观念的不合理性不应当损害前一观念的合理性。

对选言问题的目的论解决方案并不要求意向状态能在功能上被个体化。解决选言问题所需要的是:在获得一种信念的正常和异常环境之间作出区分(因此在获得一种信念的第一类环境与别的环境之间作出区分)。只要存在着像信念形成的正常功能机制这样的东西,那么即使不存在正常地发挥功能作用的信念,也还会有这样的区分。正因为如此,关于选言问题的目的论方案就不要

求:对“7 是素数这一信念是什么”之类的问题必须作出达尔文式的(或任何)回答。

.....

迄今为止,我们一直追随着丹尼特和米利肯等人,假定:成为达尔文主义者对于目的论语义学来说是必不可少的。但是,对于把内容还原于选择的历史,人们可能深感绝望,进而追求一种关于内容的非历史理论;该理论认为,内容不是由实际上主宰着心理状态进化的选择压力所决定的,而是由在某些反事实为真的情况下所应用的选择压力所使然。例如扑捕苍蝇和扑捕或蜜蜂或苍蝇,在这个世界都是同样有利的。但是青蛙扑捕的意向对象是苍蝇,而不是或苍蝇或蜜蜂,因为扑捕苍蝇在邻近的世界已被选择了,在那个世界有苍蝇,不管那里是否有蜜蜂,但是扑捕或苍蝇或蜜蜂在邻近的世界则没有被选择,在那个世界存在着蜜蜂,除非那里也有苍蝇。事实上有这样一个问题,即在方位模糊不清的两种属性中究竟哪一种被偶然选中;所以人们就用跨越反事实世界的差异方法以消除混淆。而诉诸于反事实就特许了一个关于选择的内涵(带一个“s”)概念;它把选择实际上关注的结果(捕获苍蝇)与那些仅仅是偶然获得的结果(捕获蜜蜂)区别开来了。

但是问题又来了,即使似乎有这样的可能,即这些反事实实际上是在确定功能时所关联到的东西,但它们为什么关系到内容的确定呢?试考察下面的例子:我假设,喜欢甜的功能就是得到外面的糖块(从而得到热量),我还假设,摄取糖精是非功能的。如此构想对于关于功能的反事实探讨是有好处的:在所有甜的东西都是糖的世界里,对甜的喜欢是好事情,而在所有甜东西都是糖精的世

754 第六部分

界里,这则缺乏生存价值。但是麻烦在于:在这类例子中,功能与内容是分离的。嗜甜的功能就是使你摄取糖;但是它的意向对象——不是糖而是甜;这就是糖精为什么能满足需要的缘故。须注意,糖精也能满足对甜的需要;它恰好不是使那个需要消逝。^⑦

在我看来,关于内容的进化路线似乎犯了两个错误,其中任何一个都是致命的:一方面,它假定:你可以对功能作出历史的/选择的分析(因此一状态的功能就是它实际上被选择的东西),而关于功能你所需要的东西显然是某种反事实分析(一状态的功能就是这样的东西,它就是它被选择的东西,即使……)。另一方面,它假定:如果你给予一种状态以功能,你也就因此而给了它以意向对象,而嗜甜的例子足以说明:事实并非如此。

依我看来,你这里所得到的是一种贫乏而无用的理论。

在停止全面清算这种无用的理论之前,我先来说最后一点。你被求助于达尔文以建立意向性之基础的价值弄得迷迷糊糊,途径之一就是你让自己根据“自然之母心中所想到的”、用近乎于严肃的方式谈论进化目的论。这样做之所以令人迷惑,理由在于我上面强调的一点:我们借助人的目的、意图等严肃地、非形而上学地解释事物所用的表达式,对于共外延的谓词替换物来说,远不及进化解释所用的表达式那么透明。

就我所知,只要我们严肃地对待达尔文式的(即历史的)解释,下述主张就没有什么意义,即在 F 和 G 都必然共外延的情况下,一种状态选择的是 F 而不是 G。^⑧事实上,达尔文式的解释稳妥地把共外延的表征当作是同义的;而心理学解释则不然。因此如果你习惯于用诉诸于幕后的设计师的模式设想进化解释,那么你就

可能认为：当这些解释开始把内容个体化时，它们便超越其事实上的所能而为你做了大量的好事。

如果格兰妮要制造一机器蛙，那么她可能想到：她的蛙应当会扑捕苍蝇，而不会想到：她的蛙会扑捕那或者是苍蝇或者是蜜蜂的东西。因此她的机器蛙是一只苍蝇扑捕器，而不是或苍蝇或蜜蜂扑捕器，不管所有本地的或苍蝇或蜜蜂多么可靠地是苍蝇。（这正像丹尼特的“双钳牙”一样，即使这种情况与我们的直觉明显不一致。根据我的看法而不是根据他的看法，如果我制造了一机器，每当我放进贰角伍分钱进去时，我就想让它进入状态 S，那么该机器就是贰角伍分钱的接受器，即使在那森林的别的地方还存在着物理上酷似贰角伍分钱的墨西哥卢比，进而如果它碰到任何东西，它就使机器进入状态 S。）与“自然之母心中所想到的”那种说明不同，（所谓）“派生的意向性”的归因对稳定地共外延的表达式的替代物来说，显然是不透明的。尤其是，它们可以在扑蝇与扑或蝇或蜂之间作出区分。

因此派生的意向性不存在选言问题。我们在什么地方有这样的事物，其状态有派生的意向性（值得顺便一提的是，格兰妮迄今所造的一切人工造物的意向性），我们就能在那里对它们的状态的内容作出细致的区分。这是因为，我们能对我们的状态之内容作出细致的区分，而派生的意向性正好就是源于我们的意向性。把派生的意向对象归之于格兰妮的蛙就能在稳定地共外延的内容之间作出区分，因为把心理状态归因于格兰妮就可以在稳定地共外延的内容之间作出区分。在机械捕蝇器与机械扑蝇或蜂器之间的确存在着差异，因为在想让他们在一处捕食的那些格兰妮与

想让他们在另一处捕食的那些格兰妮之间的确有着不同。

不过,诉诸于功利选择的目的论解释的逻辑似乎是非常不同的。我们知道,有一点极不清楚,即诉诸于“自然之母心里所想到的”能使在稳定地等值的内容归因之间所作的区分合理化。而且你可能会如此表述布伦塔诺的命题:自然之母与格兰妮之间的不同显然在于:格兰妮重视的只是意向的差异,而自然之母则不然。我不是说:格兰妮比自然之母更精明;我强调的不过是:她非常精确。

因此,下述说法是非常非常错误的,即既然“……就有机体而言……(内容)……并不独立于自然之母的意图和目的,因而(它)正好像……(人工造物)中的意义一样是派生出来的”(丹尼特,p.305)。^⑨假设的类比使派生的意向性归因错谬百出,因为它对内容之间的差异估计不足,因此这种归因与那些把内容归之于自然之母的归因同出一辙。而且——也许更为荒唐——它从根本上误解了达尔文的纲领,而该纲领显然净化了关于任何具有各种解释之原理的东西的生物学,这些解释由于有一个“t”因而是有意图的。事实上(与形而上学相反),达尔文式的解释与把目的归之于自然之母风马牛不相及。与丹尼特所说的相反,达尔文的观点不是说:“……我们是由自然选择所设计的人工造物……”(p.300)。达尔文的观点更加深刻,更加美妙,显然也更令人震惊:在岩土是由风化而造成的人工造物的意义上,我们是由选择所设计的人工造物;这也就是说:我们不是人工造物,没有什么东西设计我们。我们是、并且一如既往地完全是我们自己。

自然,达尔文对布伦塔诺不屑一顾;达尔文事业的全部目的在

于让生物学摆脱布伦塔诺的研究路线。

这并不是关于选言问题的进化/目的论方案的全部错误。在前面,我已说明了作为目的论描述之基础的直观的自然天性,这种直观就是:错误就是出差错时所发生的东西。但是为了批准对选言问题的目的论方案,你所需的不只是这一点;你还需要:当事情是正确的时候——更具体地说,当事情是正常的时候——就事实本身而言,引起一符号被标记的一切东西就存在于那符号的外延中。正是这关于正常状态的目的论描述与关于内容的因果描述拴在一起了。目的论要限定的是在其中为正常的任何情境的类别;但是它也是这样的假定:正常地被引起的符号事实上适用于它们的、使语义学得以产生的原因。尤其是,正是这种假定准许把正常的情境与因果关系在其中促成内容的情境等同起来。

.....

基于所有这些,我倾向于认为:对内容的目的论说明是没有希望的。一方面,求助于目的论上的正常条件并没有为意向内容提供一个单义的概念;尤其是,它并没有解决选言问题。另一方面,第一类情境不可能与目的论上的正常条件相等同;正常地被引起的意向状态本身并不真的就意味着引起它们的一切东西。因此我们需要的是对选言问题的非目的论解答。就这样吧。

II. 理论

“……求助于目的论上的正常条件并没有为意向内容提供一个单义的概念……正常地被引起的意向状态本身并不真的就意味

758 第六部分

着引起它们的一切东西。因此我们需要的是对选言问题的非目的论解答。就这样吧。”这是第一部分讨论的结论。但是那毋宁说是以假定为武器来对付这样的人,他主张:由于不存在意向状态因而不会找到关于析取问题的解答,因此也用不着去管关于它们的意向内容的选言判断或别的东西的事实。武断一点说,你需要的只是一样东西;而你可能得到的则完全是另一种东西。一种关于内容的自然主义和非目的论理论到底是什么样子呢?

该论文的其余部分将探讨和拓展对于选言问题的一种方案,对此,我首先在《心理语义学》(1987)、然后又在《信息与表征》(即将发表)中作过概述。这一解答广泛地包含在对于内容的信息探讨的传统之中,^⑩它又没有把符号所意指的东西与其标记携带的信息等同起来;它也不是想通过区分第一类情境(即这样的情境,在其之下,引起一符号被标记的一切东西事实上就存在于它的外延之中)和第二类情境(即这样的情境,在其之下,符号据说是由它们所不适用的事物所引起的)而解决选言问题。^⑪其次,它至少有别于我在文献中所看到的、对于析取问题的其他的解决方案。

一开始我就必须承认:似乎有一种惊人一致的看法——也许在关注此问题的六或八人中——即认为我对于选言问题的处理方法是行不通的。但是格兰妮说:我不应当郁郁寡欢;她说:罗马不会在一天中被毁灭。因而我想扫描一下我所听到的以及我能想象到的、各种对于我关于选言问题解决方法的多或少的反对意见。这样做会部分地向你表明:我并没有忧郁不乐;它还想部分地让你相信:我的陈述事实上妥善地处理了公认的反事实;在一定程度上,它还为完善和深化该理论提供了条件。

1. 不对称的依赖性 (以及几乎是最后的目的论)

谬误引出了选言问题,但选言问题就其核心而言其实并不是关于错误的问题。选言问题真正涉及到的核心是意义与信息之间的差异。我们的讨论就由此开始吧。

信息与原因论相关联,而在这一方面,意义则不然。如果一符号的标记有两类原因论,那么由此可以说:那符号的标记携带的信息就有两类。(如果某些“奶牛”标记是由奶牛引起的,而某些“奶牛”标记则不是,那么由此可以说某些“奶牛”标记携带着关于奶牛的信息,而某些“奶牛”标记则没有这种信息)。比较而言之,一符号的意义就是它的所有标记共有的东西中的一种,不管它们怎么碰巧被引起。所有“奶牛”标记意指的是奶牛;如果它们不是,那么它们就不是“奶牛”标记。

因此信息伴随着原因论,而意义则不然,如果你把一符号的意义与它的标记所携带的信息等同起来,那么你由此就要碰到选言问题,这就是其原因。错误不过是一种提示;之所以陷入选言问题,仅仅是因为一符号的虚假标记可能有不同于真实标记的因果历史(进而携带着不同的信息)。但是正如我们在第3章中所看到的那样,还有另一类的原因论上的异质性事例(包括思维中的表征),它们也产生选言问题。

换言之,解决选言问题需要的不是关于错误的理论,而是关于意义的理论;如果一种意义理论足够完善,那么选言意义的条件就

会作为一种特例而出现(参阅我即将发表的论文中的讨论)。如果一个人同情斯金纳—德雷特斯克传统,那么构造这种理论的用意就在于解释:一种符号的意义怎么可能对它的标记的(实际或可能的)原因之异质性不敏感,即使一方面,意义据假定莫名其妙地被归结为信息,而另一方面,信息随着原因论的变化而变化。

你马上就会明白,关于内容的目的论理论究竟有什么错误。一种目的论理论的核心就是这样的观念:在“正常环境下”,一符号的标记只能有一类原因——即是说,决定意义的那类原因。(在正常情况下,只有奶牛引起“奶牛”,像目的论描述所说的那样。)但是这肯定低估了人们可能会说的意义的凝固性(robustness):其实“奶牛”标记是以多种方式被引起的,它们因此而都意指奶牛。解决选言问题,澄清一符号的意义对它的标记原因中的多变性怎么可能如此不敏感,这是描述同一任务的两种不同方式。如果要有关于内容的因果理论,那么就必须有某种方法,它能从一符号标记所能进入的所有别的因果关系中分辨出语义上相关的因果关系。我们最好不要通过对凝固性的不清红皂白的否定——即通过理想化地描述原因论异质性的背景——去做这样的事情。

那么我们该怎么办呢?首先我赞成的第一个近乎于建议的主张是:奶牛引起“奶牛”标记,(进而让我们假定)猫引起“奶牛”标记。但是“奶牛”意指的是奶牛,而不是猫,也不是或猫或奶牛,因为猫引起“奶牛”标记依赖于奶牛引起“奶牛”标记,而不能是别的。“奶牛”意指奶牛,因为像我后面要表述的那样,由非奶牛引起的“奶牛”标记非对称地依赖于由奶牛引起的“奶牛”标记。“奶牛”意指奶牛,仅仅是因为“奶牛”标记携带着关于奶牛的信息,它们不会

携带关于任何别的东西的信息。

请注意,这类陈述具有下述称心如愿的性质,即并未假定存在着第一类情境。尤其是没有假定存在着这样的环境——法则学上可能的,可从自然主义和其他方面顺理成章地说明的环境——在这类环境中,只有奶牛在语义上才必然引起“奶牛”。这绝不是说:存在着可顺理成章地加以说明的环境,在此环境中,所有奶牛在语义上都必然引起“奶牛”。^⑫根据现在的解释,“奶牛”意指奶牛所必须的东西是:某些“奶牛”标记被奶牛所引起(更明确地说,它们携带着关于奶牛的信息),由非奶牛引起的“奶牛”标记非对称性地依赖于这些奶牛。

目的论理论宣称:虚假标记的特别之处在于:在环境是正常的情况下,它们不可能发生;如果假定:事情实际上在某时是正常的(如果该理论是历史的/达尔文式的,那么就一定如此),因此便可以说:某时所说的(或所想的)就只能是真的。对比之下,我推销的理论宣扬的则是:无论何时,虚假的标记都可能发生;只有当它们发生了时,其他类型的标记也必定如此:没有被奶牛引起的“奶牛”标记,也就没有由非奶牛引起的“奶牛”标记;虚假的标记从形而上学上来说说是依赖于真实的标记的。^⑬既然非对称性的依赖性条件之满足与“奶牛”标记的因果历史中的任何异质性都是相容的,那么这种解决选言问题的方式与断定意义为凝固的这一点也是相容的。

在第3章所讨论的自然主义的意义上,这种说明也有自然主义的称心如愿的性质。它是原子主义的(在一个不可能出现别的非对称性依赖性的世界中,“奶牛”标记可能非对称地依赖于奶

762 第六部分

牛),同时它也是物理主义的(你不诉诸于意向的或语义学的术语,就能够说非对称依赖性是什么)。^④然而,尽管它有如此多的称心如意的性质,但我所概述的那个建议则仅仅是一个粗略的梗概。事实上有许多相反的意见。不过,在我们正式关注这些问题之前,我有三个我想强调的导言性* 的说明:短一点的一个你可称之为“泛语义主义”的学说,长一点的一个是关于本体论的,而更短的一个说的是谁有给出证据的重负。

泛语义主义。你的直觉中有这样的冲突:

一方面:……符号和心理状态都有表征内容。别的东西都没有这种从属于因果次序的内容:岩石,风暴,树,盘旋的星云都没有……语言哲学和心灵哲学的主要的、共同的任务就是表征问题……某事物怎么可能设法关于某事物;为什么只有思维和符号才能成功地做到这一点?(参阅我的《心理语义学》,1987,p. xi)

另一方面:乌云意味着下雨。某种斑点意味着麻疹……在所有如此这般的事例中,都存在着把一类情境与另一类情境联系起来的似规律的或法则学的规则性。这些规则性的例示就是这样的事例,即在其中,一种情境意味着某事情或者说携带着关于别的东西的信息,如果是这样,就没有必要有心灵,也没有必要有为心灵所用的符号。(伊斯雷尔,1987,p. 3. 着重号为原作者所加)

* 此处原文为 *prefecatory*,查无此词。现据 S. 斯蒂克等编《心理表征》一书所选该文对应处“*prefatory*”译出。

事实上,意义无处不在的观念是从对内容的信息分析所引出的一个自然而然的结论。如果意义真的(或多或少)还原于可靠的因果协变,那么由于到处都有大量的可靠的因果协变,因而就有望得出结论说:到处必然有大量的意义。在“‘烟’意指烟”,“烟意指火”这样的事例中,“意义”是单义的——而且意义携带着有关的信息——这样的直观就接近于以信息为基础的语义学的核心。

但是这可能是不正确的。如果是这样,那么(由于“携带有关的信息”是及物的)便可推论说“烟”意指火;其实不然。对比之下,根据非对称依赖性说明,这类事例最终又是对的。“烟”的标记携带着关于火的信息(当它们由火引起的烟所引起时)。但是它们并不指火。因为它们对火的依赖性非对称地依赖于它们对于烟的依赖性。撇开从火到烟的联系,从烟到“烟”的联系仍完好无损;在有火的情境下,我们使用“烟”一词并不依赖于烟所携带的关于火的信息。但是撇开从烟到“烟”的联系,从火到烟的联系也会随之消失;我们在有火的情境下使用“烟”正好依赖于“烟”所携带的关于烟的信息。

总之,世界上的意义比信息少得多。这是因为要有信息,你需要的是可靠的因果协变,而要有意义,你(至少)还需要非对称性依赖性。信息是无处不在的而非凝固的;意义是凝固的而非无处不在的。关于泛语义主义就说这么多。

本体论。我在第3章已说过,我假定:如果 X_s 引起 Y_s 这一普遍原则有反事实的支持,那么就存在着一条“涵盖的”规律,它把成为 X 的属性与成为 Y_s 的一个原因这一属性关联起来:反事实支持的因果普遍原则(要么等同于因果规律要么)是由因果规律所支

764 第六部分

撑的,而规律就是属性之间的关系。因此关于非对称依赖性的说明所传递的意思就是:“奶牛”意指的是奶牛,(I)如果在作为奶牛的属性与作为“奶牛”标记的一个原因的属性之间存在着具有自然法则普遍效力的关系;(II)如果在另一些属性和在为“奶牛”标记的一个原因的属性之间存在着具有自然法则普遍效力的关系,那么后一种具有普遍效力的关系就非对称性地依赖于前一种关系。

从本体论上讲,我倾向于相信这样的基本原则,即世界包含着属性及其具有自然法则普遍效力的关系;也就是说,关于属性中的具有自然法则普遍效力的关系的真理比关于个体的反事实真理要深刻得多——因而不应根据后者而加以分析。然而,从认识论来说,我又坚决肯定能知道属性中存在着一种具有自然法则普遍效力的关系,但我没有这样的看法,即反事实由于那关系所依据的事实而是真实的。因此从方法论来说,要求根据属性中具有自然法则普遍效力的关系而清晰地表达出来的哲学分析,必须借助由个体中的反事实关系而清晰表达的分析——用人们在做生意时的说法——来加以“兑现”,这可能是一个糟糕的主意。

这种方法论的观点是我感受相当强烈的一种观点,以至于我打算屈从离题的要求。这里想用几段篇幅,谈谈哲学家怎样由于假定关于规律的真理需由反事实真理来分析或分析为反事实真理而陷入了困境。

下面说的是克里普克对遵守规则的倾向说明的批评性讨论(1982)。根据该说明,加了“+”号的意义可根据一种倾向来加以分析,就如同当被问及“ $m+n$ 等于多少”之类的问题时“用两数码的总数”来回答一样。但是克里普克说:这类分析不可能完成,因

为我们没有这种倾向：我们的计算能力是有限的；我们还会犯错误，等等。对此，他设想他的对方会回答说：“困惑不过是来自于关于倾向的过于粗糙的概念，除了粗糙的、字面上的概念，余者皆同的倾向概念则是在哲学和科学中被标准地使用的概念。”因此被设想的东西其实是关于遵守规则的倾向性描述，它可通过在执行式和完成式之间作出区分来予以支持。

但是，根据克里普克的看法，那也是行不通的。因为“……我们应怎样使余者皆同这样的从句为真呢？也许（诉诸于反事实）是这样的：如果我的大脑是由极其特殊的材料构成的，足以把握大数字，如果它被给予了足够的能力，能完成这种大数加法……（等）……，那么假如相加问题涉及到两个大数字 m 和 n ，我将用它们的总数作出回答……但我们对之怎么可能有信心呢？在这个世界上，如果我的大脑是由特殊的大脑材料构成的，我怎么能判断什么会发生呢？这类思考无疑应留给科幻作家和未来学家。对于这种实验有什么结果，我们一无所知。它们可能会使我精神错乱……（等等）。”

克里普克显然是假定：我们没有理由赞成说，理想化条件所限定的普遍原则是有规律的，除非我们能说明在有理想化条件时为真的反事实。然而不可思议的是，人们为什么严肃地接受了这种方法论。例如如果分子和容器确实能满足由理想的大气规律所给定的条件（分子是完全活跃的，容器是无限地不可透入的；等），那么只有上帝才知道会发生什么；就我所知，如果这些事情都是真实的，那么世界就走向了末日。毕竟这些条件的满足在物理上是不可能的，然而谁又知道在这个物理上不可能的世界里又会发生什

766 第六部分

么呢?

理想的大气规律要想在科学上享有良好的声誉,没有必要要求:如果真的有理想的气体,我们就一定知道将会发生的一切。所必需的仅仅在于:我们必须知道(例如)如果有理想的大气,那么余者皆同,它们的体积就会随着加之于它们的压力而按反比变化。该理论告诉我们的这一反事实是真实的。^⑤

同样,如果存在着心理学的规律,它们能对无限的工作记忆作理想化解释,那么它们为了享有科学上良好的声誉,不必要求我们知道:如果工作记忆真的是无限的,可能发生的会是什么。我们有必要知道的不过是:如果我们确有无限的记忆,那么在条件相同的情况下,我们能对任意的 m 和 n 计算出 $m+n$ 的值。^⑥这一理论本身告诉我们的反事实是真实的。

同样,即使我们不知道属性之间的具有自然法则普遍效力的关系中的非对称依赖性使哪一些反事实为真,我们也能知道这些关系中存在着非对称依赖性。我们有必要知道的不过是:如果在 P_1 和 P_2 之间的具有普遍性的关系是非对称地依赖于 P_3 和 P_4 之间的具有普遍性的关系,那么在条件相同的情况下,割裂 P_3 和 P_4 之间的关系,也就会割裂 P_1 和 P_2 之间的关系。该理论告诉我们的反事实是真实的。根据同上。

为了倾吐我的心里话,我将随大流,不时谈谈反事实,舍此再没有更好的办法。既然大家都假定:关于反事实的话语可翻译为关于可能性的话语,那么我有时也会把“成为 Y 的属性与成为 X_s 的原因的属性之间存在着具有自然法则普遍效力的依赖性”与“ Y_s 在一切(近乎于? 详下)法则学上可能的世界中引起 X_s ”等同

起来。但是对其任何一点,我都不满意;在我看来,这类还原运动总是有损哲学家的面子的。我特别犹疑的是,我们将审视的某些难题不是来自于下述主张中的错误,即内容依赖于法则学关系中的非对称依赖性,而是根源于下述假定中的错误,即关于法则学关系的主张要求反事实世界/可能世界的转换。^⑦

谁有提供证据的重负。我要提倡的意义理论主要是用虚拟条件句加以阐述的。在这一点上,它与所有关于意义的信息理论毫无二致;这些理论的基本主张是:一符号意味着如此这般的东西,因为如果有如此之类的东西的例示,那么它们就会引起那符号的标记。因此,在其展开过程中,你自然可能提出下述反对意见:“我为什么应当相信所诉诸的反事实是真实的?我为什么要相信:如果实在地存在着如此这般的东西,它们就一定以你的理论所要求的方式实在地引起符号标记?”

回答是:不要忘记,这种东西被认为就是哲学。尤其是它是解决布伦塔诺问题的一种尝试,它旨在说明:对于有意向内容的物理状态来说,存在着可从自然主义方面说明的原子论的充分条件。在这样的背景下,我着手对反事实作出规定。如果我能澄清:当如此这般的东西就是事实时,“X”便意指某物,那就够了。不过我没有义务论证说:既然X意指某物,如此这般的东西就是事实。即说,解决布伦塔诺问题需要为意向性提供充分条件,而不是必要充分条件。因此如果你想与本文的形而上学结论论战,那么你就得建立这样一个世界,在那里我的反事实都是适用的,但“X”在那里并不意指我认为它要意指的东西。够了;让我们看其中的一点。

好,言归正传。

我们首先要看的不是一种责难,而倒像一个扑朔迷离的难题:即使你能得到足以妥善处理那些实例的理论,我也弄不明白:该理论为什么就是真的;非对称依赖性为什么在信息和内容之间作出了区分。

我们不妨先忘掉自然化问题(一些段落之后,我们再回过头来予以讨论)。我想阐明的是:当语义现象被常识认为游离于关于还原的形而上学争论之外时,非对称依赖性可能就深深扎根于对语义现象的分析之中。为了尽可能使事实公开,在此刻,与其谈论心理表征,倒不如谈一下语言学。

我假定:关于我们所用的语言表达式,我们有大量的实践。而且我还假定:这些实践似乎并不都在一个水平;其中一些以另一些为前提条件,因为一些起作用只是由于另一些适得其所。举一个陈旧的例子,如呼叫人名。它是这样进行的:某人呼喊“约翰”,如果是对的,那么约翰就会来。为什么是约翰?我意思是说:为什么你叫到的正好就是“约翰”?噢,因为实际是:叫出来的那个家伙就是其名字被呼唤为那个声音的家伙。这肯定是没有争议的。

须知,为了具体说明呼叫人名的实践,你必须求助于命名的实践。因此呼叫的实践是寄生于命名的实践之上的;如果没有后者,你就不可能有前者。但是我想:不能反过来吗?即使没有通过称呼其名而呼叫的实践,难道你不能有由命名构成的实践吗(例如有这样的惯例:如果一个人,其名字叫“约翰”,白肤色,那么“约翰是白肤色”就是真的)?我承认这样说是合理的,因此我也认为:呼叫非对称地依赖于命名也是合理的。

毋庸置疑,我可以与我的狗建立一种约定,据此,当我吹口哨

时,它就会来。即使我向我的狗吹口哨发出的声音不是我的狗的名字,它也会如此。在这里,学习语言的游戏的确就是训练。因为在我和我的狗之间有一种预先的约定,因此吹口哨才有作用;当我发出那种声音时,我是在教狗该做什么。相比之下,我能通过呼喊约翰的名字而呼叫他,而用不着这类预先约定。当称呼名字的惯例确定下来了,就有条件进行那完全是抽象的呼叫实践:任何有名字的人都能通过称呼其名而被人呼叫。

这些考察说明了非对称依赖性在语义学分析中站稳脚根的一种方式:我们的某些语言实践以我们的某些别的实践为前提条件,运用词项(即表示它们的承载者的名称,表示它们的外延中的事物的谓词)的实践似乎是那大厦的基础。^⑩但是是什么使这一切与凝固性,与信息、内容之间的关系发生联系呢?基本看法是:即使要得到板子的“板子”标记没有携带关于板子的信息(无论如何它们携带了关于要求的信息;即是说想得到板子的信息),“板子”的某些标记仍有可能带有关于板子的信息(尤其是,被用来表示板子的状态的标记会携带这样的信息);但是因为存在带有关于板子信息的“板子”标记,因此我不能通过运用“板子”一词去呼叫它而得到板子。于是,在一定意义上,我的“板子”要求在因果上依赖于板子,即使在它们的因果历史中没有板子。但是它们自然不像(根据信息语义学)我的“板子”断言因果地依赖于板子那样因果地依赖于板子。因此“板子”标记能够因果地依赖于与它们意指的板子一致的板子,就有两种语义学上相关的方式,一是通过由板子引起的“板子”标记,一是通过这样的“板子”标记,它们非对称地依赖于由板子所引起的“板子”标记。换言之,一是通过携带着关于板子的

770 第六部分

信息的“板子”标记,一是通过非对称地依赖于携带着关于板子的信息的“板子”标记的“板子”标记。

到目前为止,一切都不错;我们知道,在我们语言实践中的非对称依赖性怎么可能解释:即使一“板子”的标记只是一个要求而不是引起标记的板子,正像在想得到板子的例子中那样,一“板子”标记是如何意指板子的;即使“约翰”的标记常常用来称呼约翰,它不是由约翰而是由他的不在场所引起的,但“约翰”的标记怎么可能意指约翰。这就是说,我们已能看清非对称依赖性和凝固性之间的某种联系。

但是事实上,这对还原主义者自然没有任何作用。因为在这些例子中,我们是诉诸于语言实践中的非对称依赖性来解释凝固性的。语言实践依赖于语言政策;我称呼我的名字的非对称依赖性可归结为我的所作所为,即只要条件相同,当我想让其来到的人正好是用“约翰”之名予以称呼的人时,我就会发出“约翰”的叫声;如此等等。然而既然执行一项政策就是处在一种意向状态中,那么语言实践中的非对称依赖性怎么可能襄助自然化问题之解释呢?

首先要说的是:恰恰是因为单词的使用者试图履行某种或别的语言政策,单词就不可能有其意义。或者说正因为任何纯粹的心理现象,才有纯粹“在你头脑”中发生的东西。你试图把约翰称作“约翰”,本身并没有使“约翰”成为约翰的名字。它是怎么可能的呢?对于“约翰”之成为约翰的名字来说,在名字与名字的享有者之间必然有某种真实的关系;意向不可能独自建立起真实的关系。这自然是因为意向(仅仅)是意向的;你的意思可能是让“约翰”与约翰之间有某种关系,然而也许并不存在这种关系。更何况

你可能想让“约翰”与约翰之间有某种语义关系——例如一个应当命名另一个——然而不可能有这种关系。所能做的只能是把虚无与虚无联系起来；“意向关系”是一种矛盾修辞方法。由于“约翰”与约翰之间存在一种关系，某事情才一定在那个世界发生了。这就是那使语义关系的因果解释观念富有如此魔力的部分原因。（我认为这也就是维特根斯坦的“私人语言”论证中正确的东西。然而根据我对原文的理解，他把它与无关紧要的认识论混淆在一起了。因为“约翰”意指约翰，某事情一定在那个世界中发生了。这不是说：因为“约翰”意指约翰，某人就一定可以说出那事情已发生了。）

语言政策并不促成语义关系，但它们也许会促成因果关系，而因果关系也许会促成语义关系。无论如何，这也许是信息语义学由此而有生命力的一种希望。我奉行的是这样的政策，据此，我用“是一板子”断言板子的状态，它还是这样的政策，据此，我用“给我一块板子”请求得到板子，它也是这样的政策，据此，这些实践的第二步非对称地依赖于第一步。我奉行这些政策，就是我处在某种复杂的心理状态中，而我在那种心理状态中又有因果推论，尤其是有这样的推论，即在板子和我的“是一板子”的标记之间有某种因果关系模式；而且在板子和我的“给一块板子”的标记之间有某种（非常不同的）因果关系模式；第二种因果关系模式非对称地依赖于第一种。

现在我们大概可以过河拆桥了。政策本身对语义学也许无关紧要；而有关的东西也许是奉行政策所引起的因果依赖性模式。“板子”和板子之间的一类因果关系非对称性地依赖于“板子”和板

子之间的另一类因果关系,这可能足以解释“板子”标记的凝固性,不管这种关系是不是被维持着。(试比较,第3章所引的斯金纳倡导的学说:语义学依赖于符号和它们的所指之间的功能关系——一种具有自然法则普遍效力的依赖性关系。这种关系怎样起中介作用——即它是在神经学上起中介作用还是在心理学上起中介作用——则不是语义学描述的内容了。)

基本观点是:如果关于凝固性的非对称性依赖性描述能正好根据符号-世界因果关系而说出来,那么我们甚至能在计划是自然化的背景中予以说明。毫无疑问,正是说话人的语言政策引起了非对称性因果依赖性,根据这种依赖性,凝固性的条件得到了定义;但是,凝固性的条件又在量上限制了起中介作用的机制,因此无须借助于政策就能予以陈述;基于此,它们便可与自然主义和谐共存了。

从最低限度上来说,独立地承认语义关系向因果关系还原的人对于以这种方式接受关于凝固性的事实,没有一个会感到惊诧。例如信息理论正是以这样的方式定义“信息”的,即诉诸于从量上限定因果机制时的稳定的协变,通过这些因果机制,协变才得以维持。以这种方式,它们便对信息(而且是同一信息)为什么能在如此众多、各不相同的通道上传输作出了解释。

同样,如果有价值的是因果模式本身,而不是其作用引起这些模式的机制,那么我们的心理表征也许正是由于我们的概念所进入的因果模式中的非对称依赖性而成为凝固的。^⑨即是说,也许可能有这样的机制,它们维持心理表征与世界之间的关系中的非对称依赖性,即使对于我们的心理表征的标记,我们明显没有什么政

策。如果是这样,那么语言表达式的凝固性的条件和心理表征的凝固性的条件就有可能是同一的,即使机制是非常非常不同的,由于这些机制的作用,两类符号便满足了凝固性的条件。某些比赛是由帆船取胜的,某些是由汽轮取胜的,其在两类比赛中最终取胜的作用机制是非常不同的。但是取胜的条件在量上限制了那些机制,并且对两类比赛是相同的;不管你多么迫切,你取胜所要做的就是首先得加入进去(肯定是准时地)。

.....

2. 非证实主义者的表演

我们来到了一个重要的分界线。如果我们赞成迄今所阐发的理论,那么我们就是承诺了一种形式的证实主义。因为根据该理论,语义的真值(它以下述方式来自于语义关系的本质)正好是:

P:除非你有可能在法则学上把 X-例示与任何别的属性的例示区分开来,否则你就不可能有表达了 X 属性的符号(或概念)。

或者稍作变化予以表述:

P':如果“X”至少表述了 X,如果有一个 Y,你把 X 与 Y 区分开来在法则学上是不可能的,那么“X”既表述了 X,又表述了 Y(即它表述了选言属性:X 或 Y。)②

现在,对于 P,我不知道有没有很清楚的反例(P. 博格森已针锋相对地、果断地提出了一个,但我不信他能如愿以偿)。③不过在

774 第六部分

另一方面,我又不明白:为什么 P 或 P' 一定是真的。有一个意指 X 而不是 Y 的词为什么在原则上依赖于有能力对 Xs 和 Ys 作出分辨?(根据假定)作为 X 毕竟是一种不同于作为 Y 的事态,即使(根据假定)Xs 和 Ys 之间的差异在其中显露出来的世界离我们太远了,以致无法到达。但是如果作为 X 和作为 Y 之间的差异是真实的,那么作为 X 与作为(X 或 Y)之间的差异也肯定如此。如果作为 X 与作为(X 或 Y)之间的差异是真实的,那么我们为什么不能以顾及那种差异的方式谈论(与思考)呢?

我不知道,怎样才能使你相信可以找到那条思路;在这一点上,我也不敢肯定怎样才能使我相信可以找到它。在最低程度上,可以这样加以表述,即证实主义者的哲学化成就多少年来实际上并没有什么轰动效应。如果只是作为一种尝试,那么对于我们一直讲述的故事知道怎样去修改,也就不错了,假如我们不想让它衍推出 P 或 P' 的话。

我认为:答案再清楚也不过了。迄今的陈述有两个部分:存在着一种“信息”条件(粗略地说,“X”表述 X,当且仅当 Xs 作为 Xs 引起“X”s);以及存在着一种非对称依赖性条件,这是用来对付“凝固性”这样的情况,在那里,“X”s 由 Xs 以外的事物所引起。^②正是其中的第一个——信息条件——导致了证实主义。相应地,摆脱证实主义的代价就是一种语义理论,就现在所知的而言,该理论还不是纯粹的信息理论。我打算陈述有关的地理学,把它交给你,让你决定:以此为代价作一个证实主义者值不值得。

你大概还记得,在第 3 章中,当我们讨论语义学中的斯金纳—德雷特斯克传统时,我曾提到,下述主张几乎触及到了其实质:你

的词语(及思想)所意指的东西完全取决于你标记它们的倾向(依赖于它们的标记的、我所谓的“虚拟历史”),而它们的标记的实际历史在语义上是无关紧要的。

迄今的讨论没有偏离斯金纳与德雷特斯克。因为它假定:这学说是正确的——正如我马上要说的,语义关系纯粹是信息的^②——而且证实主义怎样仿效也很清楚了。试考察孪生人案例。对于水的概念与水₂的概念区别,人们首先强调的往往是:前者而不是后者形成于具有 H₂O 的环境。但是(纯粹的)信息理论并不承认这个解释。只有当概念的标记是由不同规律控制时,这些理论才在概念之间作出区别。因此只有当不同的反事实符合于它们的标记时,进而只有当有这样的(可能的)环境,在其中一个概念被促使得得到标记,别的概念则不然时才如此。因此如果你想让水的概念区别于水₂的概念,如果你想让纯信息语义学规律起作用,那么你就得假定一个世界,在那里水取决于 H₂O,而不取决于 XYZ,^③即一个 H₂O 与 XYZ 在其中泾渭分明的世界,(不容置疑,是一个 H₂O 与 XYZ 在其中可加以区分的世界)。这就是你从信息语义学到达证实主义的途径。

相应地,你避免证实主义的方法就是:你放宽下述要求,即语义关系只能由虚拟条件句加以表述;而且你把标记的实际历史也考虑进去。紧接着就是描述这类混合的理论。我设想:(符号)“X”和(属性)X 之间的关系有三个条件,以至于当它们同时被满足时,“X”就表述 X。我的主张大概就是如此。接着我将简要地对引发每一条件的各种原因作出评论。然后我再讨论这类理论难以相容的是什么类型的事实。我还将稍作梳理并引出启迪。最后

776 第六部分

——你就会高兴地听到——我的建议到此为止。

我断言：“X”意指 X，如果：

1. “Xs 引起 ‘X’s”是一条规律。
2. 某些“X”s 实际上是由 Xs 引起的。
3. 对于所有不等于 X 的 Y 而言，如果作为 Ys 的 Ys 实际上引起了“X”s，那么 Ys 引起“X”s 就非对称地依赖于 Xs 引起“X”s。

评注：

条件 1：“X→‘X’”是一规律。

* 这正好来自于德雷特斯克。它担保：“X”s 至少携带着关于 Xs 的信息(但注意，它们不只是携带了关于 Xs 的信息)。

* 它也可解释：“马”为什么意指马，而不意指小马(即使小马引起了“马”，如果说马那样引起的话)。其基本观点是：当小马引起“马”时，涵盖性的规律就是：马→“马”，而不是小马→“马”(同上)。

* 须知，条件 1 无论如何没有排除“马”意指马或(黑夜里的奶牛)，因为根据作为黑夜里的奶牛的黑夜里的奶牛有时被错误地当作是马这一关键假定，在成为“马”之标记的属性与成为黑夜里的奶牛的例示的属性之间的联系(不像成为马的标记的属性与成为小马的例示的属性之间的联系)，据说具有自然法则的普遍效力。即是说，信息这样的必要条件并没有完全解决虚拟问题。迄今为止，这已不是什么新闻。

条件 2：某些“X”s 实际上是由 Xs 引起的。

* 这诉诸的是“X”标记的实际历史，将其作为“X”的意义的构件，因此违反了纯信息理论的假定。

* 它排除了“‘马’意指孪生马”，“‘水’意指 XYZ”，如此等等。

* 它也承认这样的直观，即第一个被非自动机引起的“猫”（在贝克的例子中）是虚假的，假如这就是你觉得应具有的直观。（无论如何，它不需要这样的直观。如果你没有它，你就用不着论证说：为了语义学目的，只是包含了 X_s 的因果历史应视之为包含了 X_s 和 Y_s ，与此同时，对那 Y_s 的排除则是偶然的；在这样的事例中，直观应当是：最先由猫引起的“猫”意指猫或自动机，因而是真实的。）

条件 3：非对称依赖性。

* 对于混合的理论来说，这是解决选言问题之核心所在，因为该理论的纯粹形式原本是：例如假定 (a)：黑夜中的某些奶牛实际上引起了“马”，(b) 关于反事实的常见假定（即黑夜中的奶牛并未引起“马”，只能是马引起“马”），因而它否认“马”意指马或黑夜中的奶牛。

* 注意，我们这样做并不依赖于条件 2。假定“X”的实际历史必然包含 X_s ，以至于“马”不能意指孪生马，这是一码事。而假设它必然只包含 X_s （以至于，如果黑夜中的某些奶牛引起了马，那么“马”意指马或黑夜中的奶牛），这又是相当不同的另一码事。在那一理论中，接受条件 2 就等于承认标记的实际历史是由符号的语义属性构成的；而条件 3 则承认符号对于它们的标记的实际历史就像对于它们的反事实历史一样是凝固的。即是说，它承认一符号的标记实际上是由不在它的外延中的东西引起的。

* 条件 3 也必须否认“‘马’意指马的图画”，以便能说明“马”的隐喻式使用依赖于它的字面上的使用，等等。不要忘记：就事实

而言,并不是所有不是由马引起的“马”是虚假的。

总评:混合的理论正好有点证实主义的味道,但这只是就其肯定没有倾向性这一点而言的。我们常常不得不说:“X”s 意指 X 离不开把 X 与下述属性区别开来的法则学上的可能性,即如果它被例示,那么它就会引起“X”s。(因此我们不得不说:“水”意指某种虚拟的东西,除非有法则学上可能的世界,在其中 H_2O 不同于 XYZ,等等)既然我们所必需的东西是:存在着这样的法则学的可能性,即把 X 与任何在“X”s 的实际历史中实在地被例示的属性区别开来。(没有实在地引起“X”s 的任何属性,就事实而言是不能满足条件 2 的;根据现在的说明,这就是“水”不能意指 XYZ 的理由)。之所以说这种理论残留着证实主义,只是由于它假定:如果黑夜中的奶牛实际上引起“马”,那么“马”要么意指某种虚拟的东西,要么在法则学上可能把与黑夜中的奶牛区别开来。(即是说,残留的证实主义是必要的,因此由黑夜中的奶牛引起的“马”的标记可以服从于条件 3。)

其实,我认为:这种证实主义大概可被整合到关于内容本身的理论之中。这一来,你就得让实际的因果历史与关于类概念的语义学关联起来,如让像“水”之类的术语意指一切具有与局部样本的同类关系的某东西。根据假定:在局部样本中没有 XYZ,这就使 XYZ 不在“水”的外延中。比较而言,如果水的标记实际上是无差别地由 H_2O 和 XYZ 引起的,那么你要从“水”的外延中排除出 XYZ 就不能诉诸于实际历史。如果在任何从法则学上可理解的世界里,它们是无差别地由 H_2O 和 XYZ 引起的(即是说,如果对它们的区分在法则学上是不可能的),那么为了把 XYZ 从“水”的

外延中排除出去,你就不能求助于虚拟的因果条件句。因此唯一的出路似乎只能是:把 XYZ 放逐于“水”的外延之外,以与下述假定保持一致,即“水”所意指的东西一定与它的标记的因果关系有某种联系。我的忠告是:如果这就是证实主义,那么不妨相信它好了。

顺便值得注意的是:根据混合的观点,青蛙飞捕黑点点而不是苍蝇,这仍是真实的。因为某些青蛙的飞捕是由黑点点引起的(黑点点满足条件 2);不存在与青蛙心理学相一致的世界,在那里青蛙飞捕的是苍蝇而不是黑点点(苍蝇不能满足条件 3)。相反,麦克白* 的匕首概念表述的正好是匕首——而不是匕首幻象,因为即使匕首和匕首幻象在这个世界都能引起匕首标记,但仍有可能的世界存在,在那里麦克白能把它们区分开来。即使你不想太多的证实主义,为了对付幻觉论证的语义学翻版,你大概得要一点点证实主义。

碰巧,在我们对语义属性常常作出因果规律说明的地方,我们现在就有一种诉诸于因果规律和实际因果历史的说明。作为结果的陈述只能是最低限度的证实主义,有理由说,它是一个好东西。但是剩下的自然再没有什么可口的东西了。纯粹的信息理论并不是没有理由的;在一些方面,它们所做的胜过于混合理论。尤其是,它们长于谈论独角兽。

纯信息理论能像它们处理“桌子”和“椅子”那样处理“独角

* Macbeth(? - 1057), 苏格兰国王,其生平故事构成莎士比亚《麦克白》一剧的基本情节。

兽”。因为根据这种理论,语义关系所需要的不过是属性中适当具有自然法则普遍效力的联系,而且你在未被例示的属性中可能得到这种联系,因而“独角兽”意指独角兽所必需的就仅仅是一个法则学上可能的世界,在其中,前者由后者引起,并且满足通常的非对称性依赖性强制。根据纯信息理论,未被例示并不是属性中的语义学上值得关注的属性。

相比之下,混合信息理论对未被例示采取了相当严肃的观点;尤其是“独角兽”并不能借助对条件 1-3 的满足而意指独角兽,因为它异乎寻常地未能满足 2。结论是:纯理论能够把独角兽当作一原始的概念,而混合理论不得不把它当作一个事实上节略的描述。混合理论事实上必须说:表述了未被例示的属性的概念,其实是出自于表述了已被例示的属性的概念的构造;毋庸置疑,这里存在着某种古怪的罗素式的东西。无论如何,它也许不是一个悲剧。纯理论甚至不得不说:“方的圆”不可能是原始的,因为自然没有这样的法则学上可能的世界,在其中,“方的圆”是由方圆的例示引起的。因此如果包含了条件 1-3 的混合理论对于内容来说不可能是必要的,省略了条件 2 的对应的纯理论也是如此。

纯理论和混合理论两者都得承认原始的与派生的是一种类别划分。而且纯理论能够比混合理论更直接地承认语义学上是原始的和句法学上是简单的这两者之间的联系。我过去以为(参阅“关于天赋争论的流行态度”,福多,1981c):“原始的概念”意味的几乎就是“词典上的概念”(即大概是由英语的句法简单的谓词表述的概念)。而现在我则倾向于认为:它意味的几乎就是“词典上的、未被例示的概念”。从外延上说,这大概接近于没有差别,因为未被

例示的词典上的属性是非常非常罕见的。如此罕见,以致人们可能会冒险推断:它们的稀罕性决非偶然。也许已被例示的词典上的概念构成了一种语义学上的自然种类。

3. 小结:神怎样知道你正在想什么?

“即使是神也不能仅通过内窥你的大脑而分辨你的中枢状态的意向内容。”^⑥这是对“外在主义”的内容观的一种概述。它也是拒斥“功能作用”语义学的一种方式,因为根据功能作用理论,当你知道关于一种心理状态的心内的因果关系的事实时,你就知道其内容所随附的事实。

凝固性预下了赌注。如果像我所设想的那样,一种意向状态的标记的原因论在实践中能够以任意多样的形式一致于具有相同内容的所有标记,那么由此大概可以说:即使是神也不能借助注意一种心理状态的实际因果关系而分辨其内容是什么。这似乎是不令人满意的,因为真正的因果的——对立于信息的——理论常常认为:一种心理状态标记的实际因果关系就是那决定其内容的东西。作为这种理论之核心的直观是:它一定是某种像由猫所引起的东西,而正是猫又使某一思想成为一种关于猫的思想。这类直观与关于凝固性的事实之间的争执一直是我的讨论中的一个主题;而且关于内容的“混合的”陈述的确是一种公正对待因果理论和法则理论的尝试。

在你头脑中的东西并不决定内容,实际的因果关系也不决定内容,但是如果神来观察你的心理状态的实际因果关系和反事实

782 第六部分

因果关系的周围空间,那么他就能分辨你的状态的内容。一种状态的内容随附于它的实际的因果关系以及某些反事实。我的主张大致如此。

如果的确如此,那么(除了我们将考虑的某些解释)它就解决了有无可能为内容提供自然主义说明这一布伦塔诺问题。因此如果这是真的,那么它就是重要的。为了使该主张生动分明,我打算匆匆考察一个例子,它将说明:如果接受该理论,那么上帝会怎样查阅一种心理状态的实际因果关系以及有关的反事实,以解析选言问题的一个简单案例。这也许可以看作是我一直加以阐发的学说之主要部分的概括。

为简明起见,我假定:上帝在内窥你的大脑时所看到的是许多电灯泡,每一个上面都有一个字母。(高技术大脑上有 LCDs——即液晶显示器——译者)当你处在一种状态中,灯泡就在你头脑中发亮了,通过说明哪些灯泡发亮,一心理状态类型就得到了说明。而一心理状态类型中的一标记就是一时空区域,在其中,相应的一系列灯泡就会发亮。我想,这接近于神经学的合理性,同样也必然符合于严肃的哲学目的。

在上帝关注你头脑中的东西与世界中的东西之间的关系时,他所看到的就是:有许多实例,在其中灯泡相关于比如说局部末梢环境的特征而有规律地开和关。他能判断这些关系是有规律的,因为他能判断:它们是有反事实支持的。而且他之能分辨它们有反事实的支持,是因为根据假定,他不仅能看到实际的因果关系,而且还能看到环绕它们的反事实空间。

我们假设:在一种特殊的情况下,也就是说在你的特殊情况

下,这就是上帝所能做的。有一标号为 c 的灯泡,当有猫在场的时候,它有规律地发亮;还有一标号为 s 的灯泡,在周围有鞋的时候,它就有规律地发亮。我们可以假定:正确的陈述是: $c's$ 发亮意味着猫(即它构成了你对关于猫的概念的标记之持有), $s's$ 发亮意味着鞋。

但是上帝并不能假定这一点;至少迄今还未如此。他不能的理由是:他碰到了关于凝固性的问题。结果便是:即使你头脑中的某些 c 标记是由猫引起的,但你的某些 c 标记由鞋引起也会是真的。而且像猫 $\rightarrow c$ 因果模式一样,鞋 $\rightarrow c$ 因果模式也支持反事实;存在着这样的环境,在那里鞋可靠地引起 cs 。(我假定:统计数字无关紧要,即是说对于 c -状态的意向内容而言,不管鞋引起的 cs 对于猫引起的 cs 的相对频率最终是什么,这都没有关系。上帝不会用意向归因掷骰子。)

上帝也碰到了孪生地球-猫的麻烦。它们是自动机,因此既不是猫,也不是鞋。但是由于孪生地球-猫和真实的猫之间有相似性,因此它们也会开启 c 灯泡,而且由于孪生地球-猫和真实的鞋之间有相似性,因此它们也会开启 s 灯泡。既然上帝能看到反事实,那么他也就能够知道:所有这一切都是真实的。

因为上帝有凝固性与孪生物的如此之类的麻烦,因此他也有选言问题。据设想,这个问题是这样产生的:孪生地球-猫所引起的 cs 像鞋所引起或将会引起的 cs 一样,从语义学上说正好相似于猫所引起的 cs ,即是说它们都意指猫。鞋在其中引起 cs 的情境就是鞋在其中被错误地当作猫的情境;孪生地球-猫在其中引起 cs 的情境就是孪生地球-猫被错误地当作猫的情境,如果那里有

784 第六部分

猫的话。

然而上帝不能采取常说的那种理所当然的方式。上帝之爱要求他兼顾替代性的假说,即 c 是两可的,某些 c 标记意指鞋,某些则意指孪生地球-猫。^⑥这就是上帝对上述二难问题的解决方式。他自问:“实际的因果模式是什么样子?”“有一个世界,除了在反事实世界中 cs 不能被猫所引起以外,它大致相似于真实的世界,在这个世界中,因果模式是什么样子?”

对前一问题的回答否认了孪生物,在 c 标记的实际因果历史中根本就不存在孪生地球-猫,因此 c 标记并不意指孪生地球-猫。^⑦对后一问题的回答据说否认了鞋。这里有两种相互关联的可能性:

一个是:即使猫 $\rightarrow c$ 关联一直不在其适当的位置上,你仍得到了鞋引起的 c 标记。但是那些由鞋引起的 c 标记甚至不能意指猫。因为任何符号都不意指猫,除非它携带着关于猫的信息。但是没有符号携带着关于猫的信息,除非它的标记鬼使神差地、普遍必然地依赖于猫。然而根据现在的假定,鞋引起的 c 标记并不普遍必然地依赖于猫;即使是在猫 $\rightarrow c$ 关联被打破的世界中,你仍能得到这些标记。基本观点是:如果鞋引起 cs 不是莫名其妙地、普遍必然地依赖于猫引起 cs ,如果上帝准备放弃基于信息的语义学的基本原则,即是说放弃符号的内容莫名其妙地依赖于其标记所从属的合规律的因果关系,那么上帝就只能让鞋引起的 cs 意指猫。我猜想,上帝并不打算弃之不顾。

另一有关的可能性是:如果猫不引起 cs ,那么鞋也不会引起 cs 。如果这个反事实是真的,那么上帝就会修正下述假定,即 c 意

指猫：一方面，存在着不是由猫所引起的 cs （凝固性），另一方面，存在着这样的基本直观，即一种符号基于其标记指向猫之类的稳固的因果联系而意指猫。如果鞋引起的 cs 从因果关系上说依赖于猫——在猫如果不引起 cs 那么鞋也如此的意义上——那么上帝便会把一个 c 标记当作意指猫，即使它被鞋所引起，这也是对的。

因此上帝能借助观察你的中枢状态的实际因果关系和有关的事实现而分清你的意向内容。事实上，他能运用差异方法，正像别的有理性的动因一样。因此你的中枢状态的意向内容是什么，这是一个事实问题。上帝没有必要为布伦塔诺问题而忧心忡忡。我们也是如此。

4. 结论：我们解决了布伦塔诺问题吗？

假设本文中的一切都是对的。那么我们所做的就是对以一种词汇表达的语义关系（即句法上原始的谓词与由它表述的属性之间的语义关系）作出说明，这种词汇只涉及到自然主义的（尤其是因果性的）表达式和指称内涵的、只带一个 s 的对象的表情式（尤其是指称规律和属性的表达式）。它源于下述论述，即符号既是凝固的，又是信息的，并与具有法则性的依赖性的创造符号关系是吻合的。更有甚者，既然说明完全是原子论式的，因此意向性和整体论之间的联系就不是内在的，尽管许许多多赶时髦的哲学家近来走向了其对立面。

因此，这解决了布伦塔诺问题吗？换言之，信息加凝固性等于内容吗？ 信息与凝固性就是你说明意向性所需要的全部吗？

786 第六部分

我不知道怎样回答这个问题。对于把内容等同于信息的标准的责难就是选言问题。相应地,我在这篇文章中一直心照不宣地假定:如果你能得到一种内容理论,它能使下述两种直观协调一致,前一直观是:只有当“X”标记携带着关于X-例示的信息时,“X”才意指X,后一直观是:只有当你能够得到不由Xs引起的X-标记时,“X”也意指X,那么你才完成了解决布伦塔诺问题所必须完成的工作。不过关于内容,除了信息和凝固性之外,也许还有理由要求别的东西。这些理由可能会是什么呢?

例如,有一些人认为:你必须加上某种意识。然而,面对从弗洛伊德到乔姆斯基的、以否认其中之一为基础的成功的研究纲领,一意强调内容与意识之间的内在联系,这在我看来就有点卢德派的味道了。^⑧因此严肃地说,我不打算接受这种观点;但是我确实承认:如果我是错误的,而它又是一个严肃的观点,那么意向性问题大概是无望解决的,因为意识问题大概无望解决。

另一可能性是:你不得不搬进某种规范性。我有点同情这一主张。凝固性赢得了这样的观点,即使用符号的某些方法在本体论上寄生于别的方法之上。但是我们的要求肯定不止于此;我们希望:某些使用符号的方法最终是错误的。^⑨在我们建构的表征图景中,认为存在着错误的表征,认为它们是应予避免的东西,这样的观点根据何在呢?

人们可能会设法让规范性关系从本体论的关系中派生出来,但是进一步思考,这似乎又是不可能的。使用符号的一种方法非对称地依赖于另一种,这一事实意味着:我们宁可用第二种方法,也不用第一种,为什么是这样,是找不到明显的理由的。不止在此

处,而且在一般情况下,本体论的优先性从规范性上说似乎是中立的,尽管柏拉图反其道而行之。该怎么办呢?

到此,经历了整个令人厌倦的论证过程的读者,对于第3章所反对的关于内容的目的论说明可能会感受到一种怀旧的忧伤。正如我那时曾说明过的,关于功能的话语直接导致了(至少一种)规范性的话语;哪里有功能,哪里也就有出现错误功能和功能失常的逻辑空间。因此可以说:目的论理论在协调自然主义的要求与意向归因的规范性、以及与真实地加以评估的语义赋值之间的关系方面做了不少工作。而糟糕的目的论理论在解决意向不确定性时则显得如此软弱无力。

然而我们使之一分为二不是没有可能的。在我看来,第3章的论证已相当肯定地表明:你不可能从心理状态的生物功能中派生出它的意向内容(至少可以说,如果你对它的生物学功能的说明以它的选择历史为基础,那么就没有可能)。但是尝试着绕走另一条路也许是明智的:基于对于内容的一种独立的、非目的论的和自然主义的说明(例如像我们一直在探索的那种说明),你大概可以根据心理状态所表征的东西而解释该心理状态的功能。例如,信念P的功能就是把世界表征为:在它有时(某时)作为P的实例的情况下如此这般的P。假如表征的概念是独立地被定义的,以这种方式谈论就没有违反自然主义的顾忌。

而且,这种分析的一个论点就是:它为意向状态的功能所提供的说明可能是真实的。我假定,(至少是更高级的)有机体就属于决策理论机的种类;稍微改变一下说法,如果(除了凭运气,当且仅当)他们的信念是真的,那么它们就以最大限度地增加他们的功利

的方式行动。所以信念为了在这种机器中执行其功能,必要的条件就是:它应是真实的。因此虚假的信念就事实而言在其限度内是不能执行它们的功能的。根据这种观点,结论可能是:表征本身没有规范性的内涵。表征正好就是一种因果关系——正好是信息加非对称依赖性——进而既不是好东西也不是坏东西。当表征状态有根据它们的内容而定义的功能时(当如此这般地表征世界的一种状态具有把世界表征为某某东西的功能时),赋值便得到了。在这些情况下,错误表征就是功能缺陷,而且本身是很可惜的。

然而,这是极其繁复难解的;如果这类说明要让人信服,那么还有很多要做的事情。例如,如果弗洛伊德是对的,那么某些虚假的信念通过掩饰难以容忍的真实性也可执行一种功能。但它们由此完成了它们的功能吗?如果是这样,那么虚假的信念看上去就可能是有功能的,因此语义赋值和功能赋值便土崩瓦解了。这就使下述流行的看法成为可疑的了,它事实上主张:错误表征是坏事,因为它是一种功能失灵。我不知道人们会怎样严肃地理解这些例子,在这里我也不打算再进一步探讨这一争论。也许我们说到这个分上就够了:如果你受着关于意向归因的规范性的目的论说明这一观念的驱使,那么选择就仍是开放的,即使你认为(我认为你应当如此)关于内容的目的论说明是没有前途的。

假设我们能用类似于我们一直在讨论的方式巧妙地解决关于规范性的争端,那么断言已经解决了布伦塔诺问题是合理的吗?有一种想法,它是用来安抚那些坚持哲学问题永远不可能得到解决这一原则的哲学家的:即使意向性真的等于信息加凝固性,那么也没有必要得出结论说:信息加凝固性足以说明智能。处在具有

意向内容的状态的充分条件不一定非得是具有信念或愿望的充分条件,或者说没有必要是在别的心理条件的充分条件。

例如,可以论证说:信念不只是有内容的状态;它们是有内容的状态,而且它们的因果关系服从某种合理的决策理论的原则,以及某种合理的推理理论的原则,等等。我从未听说有论据表明:不满足别的一些意向性条件,你就不能满足成为一种信念的意向性条件。(关于内容的功能-作用理论可以衍推这一点,因为它们一般把内容与“最低程度的合理性”联系起来了;它们的坏的方面就说到这里。)如果内容等于信息加凝固性,那么一种好的内容理论就可能准许实实在在地把(非派生的)意向性归之于温度计、恒温器之类;即是说,根据这种好的内容理论,这类装置的某些状态就是可从语义学上加以估值的。我认为:不应把这看作是一种还原,即使(在我看来)把信念和愿望归之于温度计或恒温器的确是不可能的。

总之,结论大概是:有意向的东西是心理的东西的巨大超集;而且只要它不是一种关于心理的东西的疯狂的超集(例如只要它不是包罗万象的),那么就是可接受的。当你研究你的直观、寻找关于解决布伦塔诺问题的公认方案的反事实时,记住这一点大有裨益;人们不会去驳斥这样的理论,它仅通过说明状态S不是一种命题态度便衍推说:状态S具有某某内容。如果你对解决布伦塔诺问题的可能性常常忧心忡忡,那么记住有意向的东西可能是心理的东西的巨大超集也是有好处的。解决布伦塔诺问题终究可能使大多数心灵哲学仍陷在原来常见的困境中;因此技术性失业势必不会出现。

790 第六部分

最后的话。假设我们有关于内容的自然主义的充分条件。我们自然不会由此说：我们的任何中枢状态或我们的任何公共符号都有它们所表达的内容，因为它们满足了那些削价出售的条件。而且仅根据存在着内容的充分条件也不能推论说：宇宙中的任何东西实际上都有这种结果。“P 蕴涵 Q”对于 Q 是中立的。上帝能够接受任何真实的假言支的结论，只要他知道其前提不是虚假的；但我们做不到。

另一方面，如果存在着关于内容的自然主义充分条件，如果我们不知道这些条件没有被满足，那么我们至少有可能断言：例如“猫”可能意指猫，尽管我们知道还有相反的情况。这大概是心灵哲学（或语言哲学，或别的什么）最终所到达的一种令人满意的状况。因为“猫”意指猫，其表面的合理性毕竟是相当真实的。对于你们这些家伙，我一无所知，但是当其他行业的朋友问我：哲学家们近来进展如何时，我会告诉他们：哲学家这些天已形成了确实很深奥的主张——在第一类概念体系中，你知道吗？——其实，“猫”并不意指猫……他们嘲笑我。我觉得那很令人难堪。

注 释

① 这是正确的，即使像功能主义者所假定的：对处在心理状态的必要、充分条件的物理主义阐述一般不是类规律式的。

② 某些意向规律限制了特定时间下的特定有机体的状态中的关系。这些规律甚至可以泛化到通常没有心理状态的有机体上。

③ 为避免重复，我将把这当作表示一种内容理论的技术术语，这种理论既是物理主义的又是原子论的，即（Ⅰ）（Ⅱ）据以都为假的理论。

④ 事实上，德雷特斯克有两种关于内容的说法。在文中我叙述了其中一个，它把具有自然法则普遍效力的联系概念当作基本的；还有一个是用条

件概率加以阐述的。我认为:条件概率一说掉进了死胡同,而把内容与具有自然法则普遍效力的关系联系起来则是《知识与信息流》一书中的一个令人感兴趣的观点。不管怎样,为了这一讨论的目的,我建议阅读德雷特斯基的论著。

⑤ 一附带的论点是:必须保护静电复印(xerox)原则。

⑥ 这种对选言问题的探讨与认识论中的“范例”论证有某种姻缘关系。两者都假定:有这样的情境,因此一种符号适用于处在这些情境中的一类事物这样的事实是由它具有的符号意义所构成的。“‘狗’不可能只适用于罗弗,因为‘狗’的意义是这样构成的,即罗弗是人们说‘狗’一词所关于的那类事物中的一个范例。”

⑦ 我非常感谢 D. 罗森塔尔,与他的一席话使我茅塞顿开。顺便要说的是,糖精的例子并不罕见;关于性格形成学家所说的“超常”刺激的任何例子都有助于得出同样的结论。

⑧ 我迄今所论证的一直是:诉诸于选择历史不能把从 F 中表征事物的有机体与表征它们为有反事实支持(所有 F_s 且只有 F_s 是 G_s)的世界中的 G 的有机体区别开来。一种平行的论证路线得出了现在的结论:当 F 和 G 必然共外延时,求助于进化历史不能把对 F 的选择与对 G 的选择区别开来:如果你总是同时得到 F_s 和 G_s,那么选择其中一个的机制由此也会选择另一个,因此成为 F 和成为 G 的效用总是相同的。

⑨ 加以必要的修改即是:特迪的熊是人工的,但真实的熊也是人工的。我们装配出一种,而自然之母装配出另一种。哲学是充满着惊奇的。

⑩ 我在文章末尾将讨论一种理论,其变种严重地违背了信息语义学的措辞,即使对其精神的背叛不太严重。

⑪ 关于类型 1 与类型 2 的区分,参阅本书第 3 章(即中译 I)。

⑫ 试比较《心理语义学》(1987),在那里,我假定——现在我认为是错误的——基于信息的语义学必须详细说明这种环境。就我的分辨力所及而言,我假定这一点,因为我认为:任何关于内容的信息理论必然或多或少相近于下述名声不太好的表述:“所有奶牛且只有奶牛引起‘奶牛’”。这是严肃地接受语义凝固性的一个失误。

⑬ 就像一符号的其他“非标示”用法一样。参看福多(将出版)。

792 第六部分

⑭ 即使不是,也自然一样,用不着求助于意向(带一个s的)习语。非对称依赖性一说深深陷入了关于属性、关系、规律和别的抽象的实在论之中。这类实在论是否妨碍了语义学理论的论断成为物理主义的论断——如果妨碍了,语义学理论是否是物理主义的,要不要紧——这是一些有趣的问题。但是没有一个是我这里打算予以处理的。只需说:就我对该词的理解而言,自然主义用不着包含唯物主义,如果后者被认为否定了抽象实体的本体论地位的话。

⑮ 正如 G. 雷伊所述:“余者皆同从句的可行性不是依赖于对理想化的实际说明或可实现性,而是依赖于对它们所从属的规律的明显例外能否借助于独立的、可说明的冲突而加以解释。它是由具有诸独立理论的银行所开的一张支票,这张支票几乎就是那些理论,它们的独立的证据能使之有效。因此问题不是:余者皆同从句是否能被替换,而是所有的错误能否作为独立的冲突得到解释?”有必要阐述雷伊的观点所隐含的思想,即一般来说,为了知道满足一余者皆同条件的结果,我们不得不知道:如果没有“独立冲突”的源泉起作用,什么会发生。要知道它,我们就得在最低程度上知道“独立冲突”的源泉是什么,就得知别的什么规律能与正被考察的余者皆同规律相互作用。但是要知道它在这些条件下对科学理论构造实际有什么作用是绝对不可能的……。

⑯ 这个反事实自然不是空洞的。它事实上断言:我们做加法的能力只受我们的工作记忆的限制;尤其是受我们对于怎样加数字所知道的东西的限制。温和地说,这些断言常常是不明确的。例如时至今日,没有人知道这是不是真的,但是由于记忆的强制作用,正常的说英语的人都能解析该语言的每一句子。(“庭院小道”这类句子似乎就提供了反例。)结果,语言学中某些相当深刻的问题之解决就取决于这个问题。

⑰ 例如 S. 瓦格纳的“心理表征理论”(ms)就批评了我将倡导的、对该观点的一种阐释,他认为:“存在着微妙地相似于我们的世界的世界,在那里奶牛不可能被错误地当作马”,这“有不切实际的、不合理的结论”。事实上,我所坚持的只是:如果“奶牛”意指牛而不是马,那么把奶牛与马区分清楚在法则学上必定是可能的;这并不意味着一切都是不切实际的。你接受了瓦格纳所痛斥的结论,仅仅是因为你把我关于语义学的说明与下述关于模型的说明

搞混了,该模型是:如果 P 在法则学上是可能的,那么就有一个世界,其中,正好有作为 P 的实例。

⑬ 可以肯定:这不是非对称依赖性得以成立的唯一方法,例如,如我所主张的,如果导致错误标签和错误的预言等的语言符号之使用,非对称地依赖于它们的正确应用,那么该非对称性不可能根源于语言实践,就像称呼非对称地依赖于命名一样;存在着称呼的惯例,但没有标签的惯例。自然,对于我们的心理表征,我们没有语言实践(没有惯例)。

⑭ 我将模糊地使用“概念”;有时它指的是心理表征(遵循心理学用法),有时指心理表征的内涵(遵循哲学的用法)。上下文将说明哪一种理解是有问题的。当我想命名一个概念时,我将用加标记的对应英语单词;如“奶牛”代表奶牛的概念。

⑮ “即是”不如说“例如”,因为就现在的目的而言,我们打算在选言谓项和歧义谓项之间作出区分。

⑯ 对于我正建议的证实主义所包含的这类处理办法,博格森并不是唯一的批评者。

⑰ 信息与非对称依赖性条件是“X”表述 X 的充分条件中的项目;即是说它们是充分条件之满足中的必要的、联合的充分条件。

⑱ 德雷特斯克本人并不相信纯信息语义学;他关于解决选言问题的建议要求的是:关于获得一个概念的历史的事实是与对它的外延的确定连在一起的(参阅第 3 章)。关于这一点——没有不合理的东西——对于纯信息理论优于非纯信息理论,不存在先验的论证。但是正如我们已看到的那样,德雷特斯克在他的纯信息陈述之上再加上一个因果历史这一方法并没有把他带到他想去的地方。

⑲ 一个存在着 H_2O 但不存在 XYZ(或相反)的世界并没有考虑进来,因为即使 H_2O 在只有 H_2O 的世界唯一地控制了实际的水标记,而在那些假定 XYZ 与水标记的联系是法则联系的世界里,XYZ 则控制了反事实的水标记。记住独角兽:可能存在着关于未被例示的属性的规律。

⑳ 对比之下,他能通过注视你的大脑来判明你的哪一种心理状态在那个区域,你的心理过程属于什么范围。如果他是一个方法论的智者,那么他至少能够如此,我肯定他是这样的智者。

794 第六部分

②⑥ 对于歧义性和选言性之间的差异,我没有论述,因为我感觉到像这里的区分一样,“选言”问题实际上就是“选言或歧义性”问题,参阅注②。

②⑦ 严格地说,该观点只是主张:如果 c- 标记意指孪生物,它根据的必然是某些语义条件的满足而不是我们正在讨论的那一个。

②⑧ 塞尔论证说:意识一定介入了,因为别的东西不可能表明自己能把“派生的”意向性与真实的意向性区别开来。然而,如果现在的说明是正确的,那么这就不正确。粗略地说,如果 X's 的意向性依赖 X's 对条件 1-3 的满足,那么 X's 的意向性就是真实的。如果它是由 Y's 对条件 1-3 的满足派生的,Y 又不等于 X,那么它就是派生的。(中译注——这里所说的卢德派是指 19 世纪英国手工业者为破坏纺织机器而组成的集团,其成员一般带假面具,夜间行动。)

②⑨ 我要说的是,我(对博格森的论述)是有点同情的。麻烦在于:要求规范性是基础的,这就使人想到存在着更多的对自然化语义学的要求,即比它所提供的、对外延等概念的还原更多的要求。但是这“更多”等于什么?把一术语用于它的外延中的一事物就是正确地使用该术语。当你说那就是使桌子成为“桌子”的外延的东西,就肯定不再有这样的问題:用“桌子”于桌子为什么是正确的。因而似乎是,如果你有一个关于语义关系的还原理论,那么就没有为规范性奠定基础的工作要做了。

简言之,我不清楚:“悬而未决的问题”论证怎样——或者说是否——触及到现在的例子。我斗胆地说:克里普克对于意义的规范力量的担忧要么是没有意义的,要么是再次回到了还原问题;无论如何,它不是一个新问题。不过,在本文中,我已克服了这些忧虑。

参考文献

- F. 德雷特斯克(1981):《知识与信息流》,麻省理工学院出版社。
——(1983):“关于信念的认识论”,《综合》,55,3-19。
D. 帕皮诺(1988):《实在与表征》,牛津:巴兹尔·布莱克韦尔。
B. 洛尔(1981):《心灵与意义》,剑桥大学出版社。
R. 斯托纳克(1984):《探究》,麻省理工学院出版社。
R. 米利肯(1984):《语言、思想与其他生物学范畴》,麻省理工学院出版

社。

——(1986):“无规律的思想:无内容的认知科学”,《哲学评论》,95,47-80。

D. 伊斯雷尔(1987):“行动中信念的命题对象之作用”,CSLI 专题报告, No. CSLI-87-72, 斯坦福大学出版社。

D. 丹尼特(1978):《脑猝病》,剑桥。

——(1981):“真正的相信者”,载 A.F. 希思编:《科学解释》,牛津:克拉伦登出版社。

——(1987):《意向观》,麻省理工学院出版社。

J. 福多(1981c):《表征》,麻省理工学院出版社。

——(1987):《心理语义学》,麻省理工学院出版社。

S. 克里普克(1982):《维特根斯坦论规则和私人语言》,哈佛大学出版社。

P. 博格森(1989):“评 C. 麦金的《维特根斯坦论意义》”,《哲学评论》,1, 83-84。

R. 卡明斯(1989):“表征与协变”,载 S. 西尔弗斯编:《表征》,4, 哲学研究丛书,多德克特:克卢威尔。

G. 雷伊(未发表):“概念、模式和个体心理学。”

关于心理内容的种种理论*

桑福德·戈德堡 安德鲁·佩辛

在第二章中,我们探讨了卡明斯(1989)的所谓的“表征问题”。即是说,我们考察了各种不同的理论,它们是关于什么样的事物起着心理表征的作用的,而心理表征又反映在我们关于意向状态的日常民众心理学概念之中。根据笛卡尔的二元论,某种精神实体介入了;而在古典的同一论看来,该作用是由大脑状态类型所完成的;按照功能主义的观点,起作用的只是大脑状态的标记,而非类型;在行为主义看来,其实没有什么东西发挥那种作用,因为关于心理表征的话语正好是关于非表征的行为倾向的话语;而根据取消唯物主义的观点,也没有什么东西起那种作用,因为民众心理学在主张存在着心理表征时一开始就犯了根本性的错误。

在本章,我们将转向一个不同的问题,卡明斯把它称作“表征问题”。这里我们关心的不是:什么东西起着表征作用,而只关心:表征是什么——某物表征某物意味着什么。这是上述三种非取消论理论同样要面对的一个问题,因为不管是精神实体、大脑状态类型还是大脑状态标记完成表征行为,我们都必须解释:我们怎样或

* 节译自 S. 戈德堡与 A. 佩辛合著:《灰色物质:心灵哲学导论》,第三章, M. E. 夏普出版公司, 1997。序号为译者所加。

凭借什么才能表征。换言之,什么给予它们以不同的表征(即意向)内容?

这也是一个极其重要的问题。因为对于许多哲学家而言,意向性、表征这样的现象似乎是心理的东西的独特标志。就我们能谈论的而言,物理世界只是原子按不同方式排列的集合。这张桌子,这座山,这个水坑不过是:它们的任何东西、它们的任何状态什么也不能表征,什么也不能代表,也不能关于(*about*)任何东西。但是心灵的状态不知什么原因则不是这样:思维(包括信念、愿望等)都是关于事物的;它们有意向“内容”;甚至那就是使它们思维的东西。事实上,思维似乎只是宇宙中以这种方式拥有内容的唯一的项目。物理主义者的期望当然是:大脑将被证明也能表征,因为那时会有这样的机会,即把思维还原为中枢状态或过程。但是大脑能表征吗?对于表征问题的唯一合适的方案,即一种关于心理内容的理论,将让我们解决那个问题。

请记住前一段的一个重要限定。严格地讲,以为只有心理状态才能表征,这是错误的。许多物理对象和状态也能如此:例如这个句子表征了其产生者的一个信念,但它本身正好又书写在纸上。同样,一幅画可能表征它所描绘的场景,或某种更抽象的东西,树上的年轮表征其年龄,指纹是死者被谋杀的“记号”,等等。但是正如我们先前所看到的,这些都不是出自于其自身的表征;毋宁说,它们的每一个都能为具有心灵的生物用作表征。构成一个句子或一幅画的分子本身并不表征;正如普特南(1981b)所指出的那样,一只蚂蚁在沙滩上无意地划出的线条并不表征华盛顿、丘吉尔,即使我们能够如此去看待它。因而这些例子说明了一种“派生的”意

798 第六部分

向性,即它事实上是从我们派生出来的,而我们才享有一种“固有的”意向性:我们不是用我们的思维作为表征,它们正好就是表征。于是对前一段的限制就是:在探索关于心理内容的理论时,我们寻求的是对于固有的表征问题的解决方案。(关于这一问题更多的讨论,可参看塞尔,1983。)

但是请注意:当我们借助固有表征去“解决”派生表征的问题时,同样的策略并不能解决固有表征本身的问题。即我们可以说,句子从心灵派生出了它们的意向性,但我们不能说,心灵从(别的)心灵派生出了它们的意向性。这类说明不仅剥夺了心灵的“固有的”意向性,而且当我们需要探寻他心的意向性时也会使人迷失方向。我们寻求的是这样的说明,即它完全能用不诉诸于心灵或心理状态的术语解释心理状态的意向性:关于意向性的“自然主义的”(即非心理的)解释。因此在本章中,我们将按照自然主义的这种限制,考察各种企图解决表征问题的方案。

根据第2章所描述的心灵的计算理论进行下面的讨论常常是很省事的。按照该理论,思考一想法就是拥有非常近似于句子的某东西,它们以适当的方式实现于你的大脑中。正如一自然语言句子一样,这个心理语句将由单词或术语所构成,而那些单词或术语本身又是对象、属性等的(固有的)表征。而且这种“思想的语言”(LOT)假说的一个直接应用可看作是对于自然语言的派生意向性的一种解释:(例如)既然我们把我们的英语单词映射到我们的心理单词上,英语单词就有表征作用。因此我们能以下述方式陈述表征问题:一思想的语言(或心灵语言)的术语凭借什么而意指它所意指的东西?即是说,它从哪里派生出它的表征(即心理)

内容?^①

我们这里将考察对这一问题的四种回答,每一种都把内容置于不同的非心理属性的基础之上:相似,(因果)协变,适应作用或目的论,概念作用。首先,我们简要地描述每一个回答意味着什么,然后用大部分篇幅详细讨论其具体内容。不过我们得先简要讨论关于表征内容的非自然主义方案,它源于丹尼特;最后对贯穿于本章的下述问题作出一般的反思,即心理内容对行动者大脑之外的因素敏感到什么程度?

相似性涉及到这样的观念,即表征根源于某种形式的相似性:某物凭借相似于所表征的事物或属性而意谓它所谓的東西。这种说明首先与下述不再流行的观点有联系,即思维在很大程度上涉及到心理意象(因此,LOT 倡导者对此已无好感)。

协变是这样的观念,即一术语意指 x ,因为它与 x 's“协变”。即是说(文雅一点),当且仅当 x 出现,该术语就被例示或出现(“被标记”)。典型地说, x 是该标记的原因的组成部分。协变理论的典型应用就是对视觉表征的理论化:比如说,青蛙的某种神经模式在苍蝇出现时就意指苍蝇,当且仅当苍蝇出现在视域中。(不同于相似理论,神经模式在任何方面都用不着相似于苍蝇。)

目的论要解释这样的事例,在那里,我们认为某物意指 x ,哪怕它并不单独与 x 's 协变(可能与非 x 's 协变)。例如,蜜蜂的“花舞”实际上至少与花场以外的东西协变(例如吃蜂的鸟不在场),然而我们仍把它们当作花舞。适应作用理论通过把内容建立在生物学的(进化的)功能之上而说明这一点:这些是花舞,因为它们对于行为的、指向花的作用就是进化为它们所选择的東西,即使它们实

际上与非花协变。

根据概念作用观，一术语有其表征内容，不仅是由于引起它的东西，而且也由于它本身所引起的东西，即是说由于它对于个体的所有因果作用。这里强调的是所有，包含的不止是那些构成了该术语的生物学功能的作用。正如我们将看到的，这有助于对内容的极细致的归因，甚至因果作用上的细微差异也可看作是内容上的差异。解决由这一整体论方案所提出的问题也许是概念作用理论家们的重要任务。

下面让我们转向丹尼特关于内容的非自然主义方案。

1. 丹尼特的“意向观”

丹尼特提出了本章主要问题的一个更宽泛的变种。他不问某物怎样拥有它所具有的表征内容，而问：一完整的系统怎样才算是具有一般意义上的表征状态？他不关心特定状态或“符号”的内容，换言之，他重视的是有这些状态的系统。他的回答饶有兴味但又颇有争议：如果一系统的行为可借助于丹尼特所谓的“意向战略”或“意向观”而得到可靠的、广泛的预言，那么它就可看作是意向系统，即具有表征状态的系统。（这一节我们主要讨论他 1981 所阐述的观点。）

意向战略是有利于预言系统的行为的几个主要战略中的一个。试考察“物理战略”或“观点”：预言一物理系统（例如人，至少是人的躯体）的行为就是确定它的物理构造和自然规律。如果这样做实际上能避免繁复性，那么它在原则上总是有用的。

不过,采取对于系统的“设计观”有时更为省事。在这里,你可以忽略该系统的许多物理细节,假如它有某种设计,那么便可预言它将很好地按所设计的行动。设计观是大多数使用者对于计算机的一种观点。人们用不着有关于机器内部怎样进行物理运作的知识,也能预言:当一个人输入了不同的命令,什么会发生。

诚然,你运用设计观——你毋须知道关于电子学、物理学等的任何东西,也没有必要去确定任何原子的物理状态——所得到的东西是,你将专注于可靠性。毕竟,物理观在原则上把握提供完善的预言;相反,设计观在系统有故障时或让它在未考虑到的环境中运行时就会失败。把水倒进计算机中,设计观就会束手无策。但是既然事物一般是按它们被设计的那样起作用,且条件常常是正常的,因此一般来说,设计观比物理观更有用。

不过,对于人来说,设计观是没有用的;我们不是以计算机被设计的方式而被设计的。因为预言人的行为,常用的是意向战略。正如丹尼特所描述的:

首先,你决定把其行为应予预言的对象当作是一有理性的行动者;接着你基于行动者在世界中的地位和目的,去推论那行动者应有什么样的信念。然后你根据同样的理由,推论它应有什么愿望,最后你预言:这个有理性的行动者将根据它的信念采取行动以接近它的目的。根据所选择的信念和愿望的集合所作的小型实践推论,在许多例子中——但不是全部——将产生关于行动者应做什么的决定;这就是你预言行动者时将做的事情。(1981,p.18)

802 第六部分

丹尼特进一步设想,这肯定是我们通常预言他人行为的方法,他还概述了我们在“推论”行动者“应”有什么信念和愿望时应遵循的一些一般规则。从根本上说,这意味着我们把人当作有理性的生物,并由此去预言他们的行为。因此,例如如果你的朋友说:他将于 7:30 在剧院门前与你会面,你(也许无保留地)把不同的愿望和信念归之于他,诸如他希望你成为朋友,相信严守时间会增进友好的关系,进而你就会预言:在 7:30,他肯定会准时在剧院门前。你作出了这个预言,而并没有关于他在物理上怎样构成的信息,尽管他不是一个“被设计”的系统。

意向观自然也不是完美无缺的预言战略。人并非完全是理性的;你的朋友为了在 7:30 赶到剧场,在必须离开时可能作出了错误的计算。他还有可能正好忘记了。或者你并不像你所想的那样了解他。也许他被雇用,诱骗你离开你的房间以便行窃。因此意向观并不总是有用的。但它常常是有用的。而且它肯定比物理观更容易运用,因为后者试图根据物理学的规律预言你的朋友的行为!

为什么丹尼特的方案是有意义的?这不在于意向观的观念本身,因为上述描述反映了我们相互怎样对行为作出预言,这一点是没有争议的。丹尼特的方案之所以有意义,是因为上述关键的主张:一系统,如果其行为借助意向观能“可靠地广泛地被预言”,那么才可算作“真正的相信者”,算作真有意向状态。简言之,丹尼特主张:一系统之被认作具有表征状态的系统,必不可少的就是:它的行为可通过意向战略这样的方式加以预言。意向性是通过这种可预言性而合法地表现出来的。

“可靠与广泛”这样的限定自然是必不可少的。任何简单对象的行为毕竟也能借助意向观而被预言。我可以把这样的信念归之于我的桌子,即它相信它位于宇宙的中心,还可认为它有呆在那里的意图,进而便可准确地预言:它将呆在那里。但我们并不想说我的桌子真的可算作是意向系统。它的行为也很简单;我的预言尽管可靠,但并没有构成意向性所必需的充分的“广泛性”。

丹尼特的理论有说服力吗?首先,它是非自然主义的。即是说它并不打算把不同的意向概念(信念、愿望、关于)分析成任何非心理的术语。成为一意向系统也就是成为这样的系统,其行为可通过把它当作是有意向的而加以预言;仅此而已。它并没有解释什么是理性或是有理性的;丹尼特的理论在主张对一系统采取意向观离不开把它当作是有理性的时,认为这些概念是理所当然的。这些术语也是非自然主义的,因为它们一般被理解为强制性因素,基于这些,它才使各种信念和别的意向状态“具有意义”;因而它们含蓄地指称意向状态。一般而言,即使他的理论的非自然主义特征并未削弱其价值,但它意味着:它不可能成为对于我们所探寻的意向性的完满的解释。

还有另一个更特殊的问题。根据丹尼特的理论,一系统是否应算作有特殊意向的状态,正是一个解释和程度的问题。试想:人们在把心理状态归之于他人和他物时是多么的不一致。除了你,弗洛伊德也会把不同的意向状态归之于你的母亲。但是你也能像他一样预言她的行为。(她毕竟是你的母亲。)你们都承认:她将为你买一份丰盛的生日礼物,但你是这样作出这个预言的,即认为她有这样的信念:你是母亲的一个最可爱的孩子,而弗洛伊德则是这

804 第六部分

样作出预言的,即认为她有促使你更依赖于她的(无意识)欲望。谁是对的呢?她究竟有哪一种信念?丹尼特的理论没有告诉我们该怎样作出回答。

丹尼特的理论还难以说明:一系统何时才可算是具有意向状态的。黑猩猩、猫、鸟、蛇和昆虫有意向状态吗?它们的行为似乎完全可以通过那个意向战略而得到不同程度的预言,但它是一个程度问题。一生物在哪一点上跨越那个界线而获得真正的意向性?一黄蜂可以作好细致的准备,以保证它的子蜂有足够的食物。但它具有这样的复杂性以致拥有保证子蜂有足够的食物这样的愿望吗?黑猩猩从行为方面说可能就是人类,但它突然干了非人的事情就不然了。那是非理性的吗?或者说,认为它完全有意向状态是错误的吗?

某人是否相信 p (p 在这里指任何命题),这会是一个程度问题吗?我们说的不是深信 p 的程度问题;你可以轻易地相信:尼克松对“水门事件”的内疚比你的朋友更强烈、更可信或更可靠。我们追问的是:你所拥有的某种信念在或大或小的程度上是否可当作是关于尼克松感到内疚这一信念。首先,正是信念 p 具有不同的确信度;其次才是 p 被相信的程度。信念以这种方式涉及到程度,无论如何是不大可能的。我必须确实知道我有过的是哪一信念,才能判断我拥有它所具有的真实性或可信性!然而丹尼特的理论就其整体论的和解释性方案而言,使人想到的似乎正好是这样的问題。

最后,对于某些实际起作用的系统在其发挥作用时为什么要采取意向观,丹尼特的理论并没有作出解释。在他看来,意向状态

以广泛的行为模式为基础,而与任何内在于行动者的东西无关。我们不能以相应的方式把意向观用之于分立的、内在的状态;只有在有行为模式的时候,它才能应用,而且分立的内在状态不像人那样能表现出适当的行为模式。因此丹尼特并不能说明:是什么东西使我们的内在状态成为意向状态的。因此他不能解释意向观的下述成果,即它通过指出对那些意向状态有益的任何内在状态而把意向状态归之于行动者。但是如果你不能指出作为行动者的意向状态起作用的内在状态,那么你就不能解释:为什么把意向状态归之于行动者能使你如此成功地预言她的行为。于是我们得放弃丹尼特的理论,把我们的注意力转向关于表征内容的四种自然主义理论。

2. 关于内容的自然主义理论

2.1. 相似性

17-18 世纪流行的知觉理论认为:我们实在地知觉到的东西不是世界本身,而是心理映象(image)或“观念”,它们通常是外部对象使其出现在我们心灵之中的。有不同的论证支持这一观点。例如,我们做梦时,我们似乎看到了许多常见的对象,但是其实我们并没有看到这些对象;毕竟,我们的眼睛是闭着的。因此我们实际所看到的就是这些对象的心理肖像。既然梦中的“看”(据说)在质上正相像于实在的看,因此实在的看也必定不涉及到对象,而只与对象的心理映象有关。

806 第六部分

对我们来说,有意义的是,这一理论演变成了一种关于思维的理论:如果在知觉 x 时,我们得到了关于 x 的心理映象,那么在思维 x 时,我们不知何故把同一个映象或许是它的微弱的复制品提取出来了。如果这就是思维由以构成的东西,那么我们对表征问题就得到了一个自然的回答。使一心理对象成为关于 x 的表征的东西正好是:它是 x 的一种映象——尤其是它相似于 x 。

相似性虽然可能成为派生的意向性的有用基础,但在今天它显然不能成为我们自己固有的意向性的基础。一个一般性的问题是:把相似理论与物理主义调和起来是很困难的。个体的大脑及其状态除了相似于别的大脑及其状态以外,显然不再相似于任何东西。如果思维涉及到相似性,那么大脑除了思考别的大脑之外不能思考任何东西。但是既然我们思考了别的东西,因此要么是相似理论错了,要么是我们必须诉诸于与大脑没有明显联系的纯心理映象。今日占主导地位的物理主义显然倾向于前者。

而且,即使是对于不是物理主义者的人来说,相似理论也有问题。因为相似性是一个难以捉摸的概念。如果相像就是具有许多或大多数内在的属性,那么事实上映象除了相似于别的映象以外,并不相似于任何东西。试把猫的映象(如图像)与一个真实的猫加以比较。真猫是哺乳动物,有皮毛,活泼,18 英寸长,由血肉构成,而映象不是哺乳动物,没有毛,是死的,5 英寸长,由纸和柯达化学制品构成。它们几乎没有共同之处。心理映象与真实的项目甚至没有共同之处。但是根据相似理论,猫的映象并不能当作是猫的象征。

当然这里也有某种意义的相似性:相片(或心理映象)看上去

像猫,或至少对我们来说是如此。即是说,在它们之间存在着可知觉到的相似性,即使没有实际的相似性。因此主张相似理论的人也许会说:我们的思想(心理映象)是关于事物的,不是因为存在着实际的相似性,而是因为在它们与它们所表征的对象之间有可知觉到的相似性。但是如果相似论者想使之得到改进,那么就将面临新的问题。

两事物实际上能够独立于任何心灵即通过享有它们的大多数固有属性(如一只猫可能相似于另一只猫)而彼此相似。^②但是假如两事物在许多属性方面各不相同(就像在我们的映像与猫的例子中那样),那么相对于某种能知觉它的心灵而言,它们之间将只有一种可知觉到的相似性;换言之,“可知觉到的相似性”这一概念以心理活动为前提。然而,这意思是说:用我们的心理映象与对象之间的可知觉的相似性解释我们自己固有的意向性就是对我们上述的自然主义限制的背离。因为一般来说,我们试图解释的正好是我们进入意向状态之中的能力;因此我们不能根据“可知觉到的相似性”解释我们思想的意向性,因为知觉是我们要解释的一种有意向的心理活动。既然对于我们怎样知觉那些可知觉的相似性未能作出说明,那么我们的解释就是有缺陷的。

相似理论还碰到了更深层的问题。例如它与我们以抽象的、一般的方式思维的能力是势不两立的。我们能思维诸如真理是一种效力或所有三角形有三条边这类思想,这是怎么回事?心理映象怎么能相似于“真理”或“效力”?正如大主教贝克莱在两个世纪前关于三角形所说的,你所诉诸的任何心理映象都是关于特殊三角形的(直角或非直角,等腰或非等腰,大或小等);特定的映象绝

808 第六部分

不可能相似于所有三角形。但是这一来,相似理论怎样解释我们思维所有的三角形的能力呢?

最后,正如福多(1975)所指出的,在解释我们思维的命题本质时存在着一个真正的难题(pp. 179ff)。试考察像弗雷德长着短头发这样的思想。关于弗雷德的任何映象将表征他的头发之外的许多方面。它可能再现:他多高或他多重或他的眉毛像什么,他是坐着还是站着的等等。如果像相似理论所主张的那样,思考弗雷德就是诉诸于关于弗雷德的映象,那么什么映象能在弗雷德有短头发这一思想与弗雷德高个、很重等这一思想之间作出区分?简言之,既然映象是模糊的,而一般来说,思想则不然,那么映象似乎就不适合成为思维的媒介。因此以映象为基础的相似理论家便陷入了困境。

总之,相似理论家主张:思维 x 就是诉诸于相似于 x 的心理映象。然而这一主张很难与物理主义调和。而且心理映象实际上并不相似于任何物理对象。为了求助于他们的可知觉的相似性,便预设知觉的心理活动,进而便背离了我们的自然主义约定。最后相似理论竭力解释抽象和一般思维,设法涵盖我们思维的命题本质。

当对这些问题中的一些可以作出有效的回答时,相似理论便不再被当作是一般内容理论的切实可行的基础了。不过在进一步讨论之前,我们应注意:我们的意思绝不是说心理意象(imagery)在认知中没有什么重要的作用。关于此争论的别的细节,可参阅后面的“进一步阅读书目”。

2.2. 协变

协变理论的主要倡导者 J. 福多(1990)认为,运用协变作为意义的基础这一现代观念可以追溯到行为主义者 B.F. 斯金纳的 50 年代的语义理论中的可资利用的组成部分(参看第 2 章关于行为主义的讨论)。该观点是,英语单词“奶牛”意指奶牛,因为语词反应“奶牛”是在奶牛刺激的控制之下。把这个观念应用于思维,我们便得到了心理内容协变理论的内核:如果一表征(心灵语言中的一术语)在奶牛的控制之下,即是说如果在奶牛的出现与行动者头脑中的那个表征的标记之间存在着某种有规则的联系,以致表征与奶牛“协变”,那么该表征就意指奶牛。(由于它依赖于因果联系,那么该理论有时也被称之为关于内容的“因果”理论。还要注意:该理论无需这样的观念,即表征必须相似于它所表征的东西。)

在阐述这一理论之前,我们必须注意两个不得不解决的一般问题。第一,我们想让心理表征“奶牛”表征所有的奶牛,而不仅仅是行动者实际所碰到的特定的奶牛。^③第二,我们想要让“奶牛”意指奶牛,尽管在奶牛与“奶牛”的标记之间的因果链上有许多中间环节,包括奶牛的外表,视网膜映象和中枢活动。根据 K. 斯特尼(1990, p. 113)的观点,我们把这些分别称之为“广度”和“深度”问题。其意思是说,如果协变理论要求助于作为意义基础的因果联系,那么它就必须足够广泛地(以至于“奶牛”的意义中包括了所有的奶牛)、并以适当的深度(包括奶牛,但又排除奶牛的外表、视网膜映象等)解释因果链。

沿用我们的例子,协变理论的一种粗浅的形式可以表述如下:

810 第六部分

“奶牛”这一表征意指奶牛,当且仅当奶牛引起“奶牛”被标记。不过我们马上就会明白:该理论太粗浅了,以致不可能成功。首先当我们想解决上述广度问题时,我们就不能要求:所有奶牛实际上引起“奶牛”被标记。毕竟,任何特定个体在一生中只能碰到所有存在或曾存在过的奶牛中的极小部分。第二,这一粗浅的理论不能解释:我们是怎样思维不存在的事物的,如独角兽。根据该理论,要有意指独角兽的表征,必须有一个与不存在的独角兽——不可能的怪物——有因果联系的表征。第三,我们不能要求:只有奶牛才引起“奶牛”的标记,因为我们允许我们有犯错误的可能性。即是说,我们必须承认:一匹马(远距离的、黑夜中)有时也会错误地引起我们的“奶牛”标记。因此这一粗糙的理论需作出某种修改。

处理前两个问题也许不会太难。正如福多(1990c)在阐述协变理论时所说,有关的因果关系从本质上说并不在个体的奶牛与“奶牛”之间,而毋宁说是属性中的更一般的“具有自然法则普遍效力的关系”(p.100)。即是说,使“奶牛”意指奶牛的关键的部分就是存在着把两者关联起来的自然规律,粗略地说是表现为“奶牛引起‘奶牛’”这样的形式的规律。更明确地说,存在着把成为奶牛的属性与成为“奶牛”的属性关联起来的因果规律,基于这一规律,奶牛的例示(即标记)引起“奶牛”的标记。这一改进是怎样对付前两个问题的呢?

规律的一个重要特征就是:它们能支持反事实。一个反事实就是一个以虚拟时态表现出来的“如果—那么”句子;一个例子可能是(C)“如果有P,那么它就引起Q”。因此,说“ P_s 引起 Q_s ”这一自然规律支持反事实不过是说:基于这一规律,它的对应的反事实

(C)为真;或者说,如果“Ps 引起 Qs”将成为一条规律,那么(C)必定为真。但是应注意,(C)完全可能是真的,即使实际上没有任何Ps! 反过来,这意思是说:即使没有Ps,“Ps 引起 Qs”仍能成为一个完全有效的自然规律(即是说只要(C)是真的)。但是独角兽引起“独角兽”,完全有可能是一个自然规律——从而是我们有意指独角兽的表征的根据——即使没有符合于那规律的任何独角兽。简言之,这意思是说:以福多的方式阐述的协变理论可能涵盖我们思维不存在的对象的能力。

怎样回答第一个问题呢? 协变理论借助“奶牛”标记与奶牛的似规律的联系解释了“奶牛”标记的意义。但是科学要发现那些细节,那么就极有这样的可能,即那规律对知觉的原因如此敏感,以致支持像这样的反事实:“如果奶牛在适当的知觉条件(好的光线、靠近等)下出现,那么它就引起‘奶牛’标记。”但是即使大多数奶牛在适当的条件下不出现,因而在任何特定个体的头脑中实际上并未引起“奶牛”标记,那么该反事实(以及它的相应的规律)仍可能是真的。如果这是正确的,那么广度问题就得到了令人满意的解决。^④

这又带来了第三个问题。我们必须为我们犯错误的可能性提供解释;有时非奶牛引起我们标记“奶牛”。事实上,这个问题变得相当深奥了。如果(比如说)马像奶牛一样能促使我们标记“奶牛”,至少在某些环境下是如此,那么“奶牛”实际上就没有与奶牛这一属性协变,而完全是与不同的属性即关于是一头牛或一匹马的属性协变。即使周围没有奶牛标记的例示,“奶牛”有时也会出现,但周围倘若没有(让我们假设)一个或奶牛或马的标记,那么它

812 第六部分

绝不会出现。(只有当某物或是奶牛或是马或者同时是两者,那么它才是一个或奶牛或马)。但是如果是这样,那么根据协变理论,“奶牛”意指的就不是奶牛,而应是或奶牛或马。后一属性被称之为“选言”属性,这是从逻辑学中借用的一个标签(在那里,“A 或 B”这样的陈述形式被称之为“选言判断”),进而这个问题——协变理论怎样看待我们的错误——便作为“选言问题”而为人所熟知。

也许,协变理论家解决选言问题的最初尝试是由 F. 德雷特斯克(1981)作出的。他的基本观点是区分开两个时期,即人们学习概念的时期和随后的简单运用概念的时期。以我们教一个小孩运用不同语词所用的方式作为范例,他指出:在学习时期,我们强化奶牛与孩子对“奶牛”标记的运用之间的严格关联(即是说奖赏正确的用法,惩罚错误的用法)。德雷特斯克论证说,这足以使那小孩让“奶牛”意指奶牛;进而“奶牛”便会保持其意义,不管后来什么东西引起对它的错误运用。因此后来的马所引起的“奶牛”便不会意指或奶牛或马。

正如福多所说(1984b, 1987),这是一有趣的观点,但也是一面临着大量难题的观点。因为一方面,就我们的许多或大多数概念来说,不可能明确区分开学习时期和使用时期。而且当学习语词和学习概念之间存在着毋庸置疑的重要联系时,德雷特斯克的战略似乎充其量只能用之于语词意义,而不能用之于心理表征的意义。毕竟,为了“强化”所需的关联,教师有什么办法到达我们的心理表征呢?为了使这一点更明确,试想某个没有语言而长大的人。尽管他缺少许多概念,但他(似乎)仍能有许多别的概念,包括奶

牛。倘若如此,德雷特斯克战略怎样解决选言问题呢?

更糟的是,这一战略忽略了重要的反事实。假设,在你的学习时期之后,在C环境中的一匹马使你说出了“奶牛”这样的标记。下述反事实似乎是真的:如果你在学习时期,碰到了C中的马,那么你就会正好以相同的方式作出“奶牛”的标记,即使你的老师先前帮你建立起了完善的奶牛-“奶牛”关联,也是如此。但是这意味着你的奶牛-“奶牛”关联其实只是一个幻觉:事实上关联始终存在于“奶牛”与C中的或牛或马之间。因此你的“奶牛”标记意指的不是奶牛,而是C中的或牛或马——这一来选言问题依然未能解决。

德雷特斯克可能会通过强调学习与后学习时期之间的关键区分而作出回答:如果C中的马早就出现了,老师就会纠正错误的“奶牛”标记。然而,这种区分不能用来解决选言问题。为什么不能?因为这等于说:那使“奶牛”意指奶牛的东西(不是C中的或牛或马)是老师的一种行动,该行动又依赖于他的意向状态(他必定相信错误发生了,并意欲予以纠正等)。事实上,德雷特斯克的內容理论以意向状态为前提条件,因此是根据(别的)心理状态解释内容——进而违背了我们的自然主义约定。这样一来,德雷特斯克对选言问题的最初解答似乎未能如愿。认识到这一点之后,他完全放弃了协变理论,转而探讨关于内容的目的论方案(参阅德雷特斯克,1988)。与此同时,福多发展了自己对于该问题的解答。

福多注意到(1987),为了使“奶牛”意指奶牛,尽管它有时是由马引起的,我们必须找到某种自然主义的方法,以打破奶牛正确地引起“奶牛”标记的情境(根据假设)与马错误地引起“奶牛”标记的

814 第六部分

情境之间的明显的对称性(p. 106)。他设想,做好这一工作的途径就是注意到,错误的标记由以依赖于正确的标记的非对称的方式。

其基本思想是:我们不会在看到了一匹马时便作出“奶牛”的标记,除非“奶牛”意指奶牛已是事实;错误地把马认作奶牛正好是(根据协变理论)对典型地由奶牛所引起的一个符号作出标记。但是反过来就不行了:基于看一头奶牛而标记“奶牛”无论如何不依赖于这样的事实,即马也可能引起“奶牛”。更详细地说,如果事实不是奶牛引起“奶牛”,那么那匹马就不会引起“奶牛”(因此是被错误地当作奶牛);但是即使事实不是这样,即马有时引起了“奶牛”,那么奶牛仍会引起“奶牛”。换言之,如果奶牛没有引起“奶牛”,那么马也不会引起“奶牛”,但是即使马没有引起奶牛,而奶牛仍可引起“奶牛”;正如福多所述,马所引起的“奶牛”非对称地依赖于奶牛所引起的“奶牛”。因此,“奶牛”意指奶牛,而不是意指或牛或马,即使马有时也可引起“奶牛”。

值得注意的是,福多对选言问题的解答似乎避免了德雷特斯克所陷入的困境:它用不着区分学习概念与使用概念两个时期,它毫无困难地适用于心理表征和语词表征,而且它用不着提及任何人的行动或意向状态。所涉及的问题不是符号何时或怎样被学会,而明显是“错误的”标记是否非对称地依赖于正确的标记。因而福多坚持了下述基本观点,让意义根植于因果关系之中,而同时又有处理错误的机制。

因此,福多的协变理论,包括它对选言问题的“非对称的依赖性”的解决方案是可行的吗?值得注意的是:他绞尽脑汁,试图证明它是可行的。在一篇论文(1990c)的结尾,他陈述了 11 条以上

的不同反对意见,从相对简单的(如上述讨论的非存在的对象问题)到复杂的都有。接下来,我们按照福多的回答,简要陈述两种反对意见;在许多细节方面,我们参照的是福多的作品(1990c)。

第一个反对意见。在本节开头,我们已注意到任何协变理论都必须解决的两个一般性问题:广度和深度问题。即使福多讨论了前者,但正如我们所看到的,他对选言问题的回答似乎与后者有冲突。在“奶牛”与奶牛之间,存在着一长串因果媒介,就我们所知,包括奶牛的外表、网膜映象和视觉神经活动。我们想让“奶牛”意指奶牛,但有这样的问题:“奶牛”与这些中介的任何一个之间比“奶牛”与奶牛之间似乎存在着更好的关联。用我们的新词汇加以表述就是,奶牛所引起的“奶牛”本身似乎非对称地依赖于(比如说)奶牛形的网膜映象所引起的“奶牛”——在此情况下,“奶牛”意指的不应是奶牛,而是奶牛形的网膜映象。福多怎样解决这个问题呢?

他写道:“没有任何理由假定:知觉状态对末梢对象的因果依赖性非对称性地依赖于邻近刺激的特定排列对末梢对象的因果依赖性”(1990c, p. 109)。即是说他否认任何特定的中介(邻近刺激)对可靠的“奶牛”-奶牛联系之存在(知觉状态对末梢对象的依赖性)是必不可少的。他的这一主张的根据是我们的“知觉的恒常不变性”这一普遍接受的事实,即是说,我们知觉机制的输出(世界怎样向我们显现)比它们的输入(亦即在我们的视网膜上什么发生了)更稳定。^⑤例如试想:当你考察一个对象时,什么会发生:当你的视网膜上的映象变化和转换时,你不可能知觉到那对象在变化和运动。再回到我们的例子上来,这意思是说:即使是在所介入的

816 第六部分

网膜排列处在流变之中,我们仍设法让我们的“奶牛”符号得到标记。但这只不过是说:“奶牛”与奶牛的联系比任何“奶牛”与奶牛形的网膜映象的联系更加可靠:即使任何已知的奶牛形的网膜映象没有引起“奶牛”,奶牛依然会引起“奶牛”。于是奶牛所引起的“奶牛”并不是非对称地依赖于任何特定的奶牛形的网膜映象所引起的“奶牛”,因此“奶牛”意指的是奶牛,而不是奶牛形的网膜映象。至少在此事例中,福多似乎解决了深度问题。^⑥

第二个反对意见。大致说来,证实主义是这样的一种学说,它主张一术语的意义就是它的证实方法的一种功能,因此知道“奶牛”意指什么(例如)也就是知道怎样去证实:某特定的事物是不是奶牛。在 20 世纪初期,这可能有某种合理性,进而是一流行的观点,但现今它广泛地被视之为一种谬论(参阅进一步阅读书目)。因此第二个反对意见便是:福多的理论可衍推出这种错误的证实主义,因此他对选言问题的解决是无效的。

怎么是这样呢?试想存在着我们不能区分的两种不同实在的可能性。以我们在本章最后一节将详加分析的那个例子为例,试想:宇宙中的某个地方,存在着一个极像地球的一个星球,只有一点不同,即流淌在河里、湖里的清澈的液体不是 H_2O ,而是别的某物(比如说 XYZ)。再设想:我们没有把 H_2O 与 XYZ 区分开来的专门知识。在这种情况下,似乎存在着把每种物质与我们关于“水”的标记联系起来的同样可靠的因果规律;^⑦因此我们的“水”最终意指的就不是水,而是或水或 XYZ,即使我们在地球上从未接触过 XYZ。在孪生地球上,我们的对应的人也是这样:他们的孪生英语单词(比如说)“twater”意指的不是 XYZ,而是或 XYZ 或

水,尽管他们从未与水发生过相互作用。简言之,我们承认“水”(water)与“twater”有不同的意义,但是根据福多的协变理论,两种物质的不可区分性——它给予两词以相同的证实方法——确保它们最终意指相同的东西。因而便陷入了证实主义。^③

为了弄清这种证实主义怎样使福多对错误的说明变得无效,现在让我们转到 R. 卡明斯(1989)的分析上(pp. 58ff)。(同样的讨论在斯特尼 1990 和贝克 1989 的论著中也可看到。)改变一下例子,试回忆,根据福多的说明,“老鼠”尽管有时是由鼯鼯引起的,但是如果由鼯鼯引起的“老鼠”标记非对称性地依赖于由老鼠引起的“老鼠”标记,那么它意指的就是老鼠(而不是或老鼠或鼯鼯)。这种情况常常出现在下述反事实的相关真理和谬误之中。

(1)如果老鼠并没有引起“老鼠”,那么鼯鼯也不会。(T)

(2)如果鼯鼯没有引起“老鼠”,那么老鼠也不会。(F)

卡明斯说,假如现在不能把鼯鼯与老鼠区别开来,当我们开始问:在什么条件下,(1)是真的而(2)是假的时,问题便随之而起。

卡明斯从(2)开始分析。他注意到:鼯鼯不引起“老鼠”似乎有两种方式,要么是:如果(a)它看上去像一只老鼠,便不会引起“老鼠”标记,要么是:如果(b)鼯鼯看上去不像老鼠。如果(a)是事实,那么老鼠(看上去像老鼠)无论如何不会引起“老鼠”标记,而陈述(2)就是真的,这与我们所期望的恰恰相反。如果(2)是假的,就必定是这样的情况,即(b)是真的。进而便是:如果鼯鼯看上去不像老鼠,那么即使鼯鼯没有引起“老鼠”,老鼠仍会引起它们——这样一来,(2)就是假的。

迄今一切都不错。但是试考察这样的条件,在其之下,(1)是

818 第六部分

真的。老鼠不引起“老鼠”似乎也有两种方式,要么:如果(a)看上去像只老鼠但并未引起“老鼠”,要么:如果(c)老鼠以前看上去不像老鼠(即是说,就像老鼠现在的样子)。如果(a)是真的,那么老鼠和鼯鼠都没有引起“老鼠”,陈述(1)就如愿是真的。然而,我们在上一段已看到:如果(a)是真的,那么陈述(2)也是真的,与我们所期望的相反。如果我们假设(c)是真的,又会怎么样呢?如果老鼠以前不像现在,那么它们就没有引起“老鼠”。但那不会影响鼯鼠(它像老鼠现在看上去的那样仍继续像老鼠)和“老鼠”之间的联系——因此,与我们所期望的相反,陈述(1)结果又是假的。

简言之,假如老鼠和鼯鼠之间有不可区分性,那么就没有如此安排事物的方式,以至于陈述(1)是真的,而陈述(2)是假的:使(1)为真的世界状态也使(2)为真。这意味着我们不可能证明:由鼯鼠所引起的“老鼠”非对称地依赖于由老鼠引起的“老鼠”;因而根据福多的理论,我们不能让“老鼠”意指老鼠(不同于或老鼠或鼯鼠)。不仅福多的理论是证实主义的理论,而且至少在某些情况下,它对他化解错误即解决选言问题也是极其有害的。^⑨

让我们回忆一下我们已述的。我们描述了关于表征内容的因果协变解释的基本观点。我们知道:“粗糙的”理论面对着三个问题:广度问题、不存在的独角兽以及关于错误的问题(选言问题)。福多的“属性中具有自然法则普遍效力的关系”的方案成功地解决了前两个问题。接着我们陈述了德雷特斯克对选言问题的回答,它区分了学习时期与使用时期,基于几方面的根据,我们对之作出了批评。在考察了福多对该问题的“非对称依赖性”回答之后,我们从整体上对他的理论作出了简要的评价。我们注意到,它似乎

妥善地处理了深度问题,但有证实主义所具有的麻烦。不过关于此课题既然有许多要说的,而福多对这些诘难已作了回答(试比较:1990c, pp. 199ff),因此是换新的话题的时候了。(在本章最后一节,将从不同角度对这些问题中的一些作出交待。)

2.3. 目的论

关于内容的目的论方案与许多哲学家的名字联系在一起,包括德雷特斯克、D. 帕皮诺、P. 戈弗雷·史密斯以及最活跃的 R. 米利肯。由于所有这些作者从不同方面作了阐发,因此该理论极其深奥、难以捉摸和复杂难解。在本节中,我们只陈述、考察它的相当简明和一致的观点。

其基本观点是:在确定意义时,不是简单地考察什么引起表征,而是要考察表征的全部的因果作用(即它的常见的原因及其结果)。(概念作用语义学对这个观点作了不同的阐发,下一节我们再加以说明。)尤其是,人们关心表征的生物学功能或目的(因而有“目的论”的标签,即“关于目标或目的”);或者说,人们可能关心形成表征的机制的功能。一事物的生物学功能从根本上说就是该事物这样的作用,即该事物由进化过程世代所“选择的”作用。于是我们可以说:我们的眼睛有看的功能,因为它们在我们的看的过程中被选择了它们的作用:我们祖先中那些能看的人享有远远超越于不能看的那些祖先的生存的、再生的优越性,所有别的作用概不例外。因此把它应用于表征,我们的简单的目的论理论就会说:在生物 C 身上,一特定的心理符号比如“老虎”意指老虎,因为那符号有表征老虎的功能——而且它之所以有那种功能,是因为

820 第六部分

在 C 的一些祖先中,“老虎”曾引起了他们与老虎有关的适当行为(诸如跑开),他们在进化上就优于 C 的祖先中那些没有这类行为的人。

这一方案事实上为目的论理论解决关于错误的问题提供了自然的方式。因为某事物偶尔甚或经常不能完成一种生物学功能,这与具有那一种功能没有一点矛盾。眼睛的作用是看,心脏的作用是泵血,即使对某些生物来说,这些器官有时失灵;或者用米利肯的一个例子,使卵细胞受精是精子的功能,即使大多数精子从未做过这件事情。但是如果一事物即使在它没有执行那功能的时候仍有那功能,如果意义是一个功能问题,那么一事物在它机能失常时仍有其意义。这一来就有对错误的回答:“老虎”的符号有时是由美洲狮错误地引起的,但仍会意指老虎,只要它的功能——它被选择的东西——是表征老虎就行了。

还值得注意的是,目的论理论怎么可能避免(至少在某种程度上)证实主义,而后者曾使协变理论深陷窘境。试回忆一下上述的孪生地球方案。如果两种物质是不可区分的,那么协变理论就只能说:我们的“水”的符号意指或水或 XYZ。但是既然我们的现实的进化历史涉及到的只是与水的相互作用,而不是与 XYZ 的相互作用,那么目的论理论家就能够说:“水”被选择了它的对我们行为的、与水有关的作用,因此意指的是水,而不是 XYZ,也不是或 XYZ 或水。与此相仿,我们的孪生地球上的对应物“twater”(水)就意指 XYZ。(请记住,福多把这一观点融合到他的最终的、“混合”的理论中去了。)

但是证实主义只是在一定程度上得到了避免。例如,目的论

理论只在有限的(而且未必可能的)条件下处理了老鼠-鼯鼠案例,比如说,我们的进化历史只处理了两者中的一个,或者说至少就大的方面来说是如此。也许该理论由于也承认了人的历史的作用因而可能有一点帮助,即使我们的物种也许世代与两类生物打交道,假设弗雷德实际上只碰到了老鼠——他的“老鼠”意指的只是老鼠,而不是或老鼠或鼯鼠。(将其扩展到人的历史肯定是相当自然的,尽管进化的含义有所改变。)不过,问题依然故我。如果弗雷德无意中也碰到了两种生物,又会怎样呢?存在着这样的直觉,即当他没有看到他所碰到的老鼠与鼯鼠之间的差异时,“老鼠”和“鼯鼠”仍有不同的意义,目的论理论怎样同化这一直觉呢?^⑩

按照同样的步骤,福多(1990c)把他对目的论理论的批评主要集中在作为意义之基础的生物学功能这一概念的运用上。试看下面众所周知的例子。在青蛙的视觉系统中,存在着专门对细小的、黑暗的、快速运动的对象(如苍蝇)作出反应的细胞,并且它能够通过发动各种追捕、取食的行为(如舌头冲出)而做出这种反应。如果我们把这些反应细胞当作是符号的例示,那么便可合情合理地说:符号的意义是像苍蝇或那里有苍蝇飞过之类的东西。这种直观的观点为目的论理论所确认,因为如果那些细胞在青蛙的实际的进化历史中有因果作用,那么它们的功能作用就可能是表征苍蝇。这些细胞(进而青蛙)很容易被欺骗——如青蛙会对非苍蝇(如飞过的蜜蜂)作出反应,去捕捉并吃掉它们——该现象并不能摧毁这一结论,正如我们在上面讨论错误时所看到的那样。

这就是福多发起攻击的地方。因为他注意到:只要刚说过的功能故事是合理的,一个替换性的说明也同样如此。根据这一故

822 第六部分

事,这些细胞的功能不是分辨苍蝇,而只是分辨细小的、运动的、漆黑的点点。碰巧得很,有对细小-运动-漆黑的点点的分辨器对于青蛙来说就有重要的进化上的优越性,在它们的通常的环境下,大量飞过的黑点点对它们来说常常是苍蝇(而不是蜜蜂)。但是后一事实并不要求青蛙的有关细胞就是苍蝇分辨器。毕竟,它们作为对细小-运动-漆黑的点点的分辨器更加可靠。正如福多所指出的,自然选择并不管青蛙是在哪一种描述(“苍蝇”与“运动的黑点”)之下捕捉到它的苍蝇的,只要青蛙在捕食它们时实际上成功了。

但是,如果说自然选择不管这些话,那么对意义感兴趣的那些东西则不然。因此福多的观点实际上是说:就其描述方面的中立性而言,生物学功能在本质上是不确定的。不过,意义则不然:在世界上,苍蝇的概念与运动的黑点的概念不可同日而语。因而生物学功能不能作为意义的基础。事实上,福多已指出了不确定性的两种不同的种类。其一,青蛙细胞的功能是表征苍蝇还是只表征漆黑的点点(通常由苍蝇所引起的网膜映象);其二,它们的功能是表征苍蝇,还是表征关于或苍蝇或蜜蜂这样的对象的更一般的范畴(福多把该范畴缩写成蝇蜂)。把这两者区分开来可能有其必要。

第一个不确定性使人想到的是协变理论的“深度问题”,即这样的指责:奶牛与(比如说)奶牛形的网膜映象之间的协变比奶牛与“奶牛”之间的协变更合适。但是如果福多现在能够把同一个问题从根本上强加给目的论理论,那么随之会出现这样的问题,即目的论者为什么不能接受福多对该问题的回答。(该回答诉诸于知

觉恒常不变性来否定那个指责。)即使福多没有提到这一点,一个可能的回答也会油然而生。

也许在奶牛与苍蝇的例子之间存在着重要的不同,它妨碍了目的论者使用那一回答。奶牛毕竟比苍蝇大得多、复杂得多,因此能够而且事实上作为奶牛和“奶牛”之间的各种中介的网膜映象很可能比作为苍蝇和“苍蝇”之间的中介的那些网膜映象多得多。这一来,当知觉恒常不变性承认特定映象对于标记“奶牛”并非必要时(就我们所知),这对于标记“苍蝇”便不适用;于是,“奶牛”-奶牛关联,就“苍蝇”-苍蝇关联不大于任何“苍蝇”-映象关联而言,肯定大于任何“奶牛”-映象关联。如果是这样,那么目的论者就不能再使用福多对深度问题的回答。

然而,如果这番考察是对的,那么福多在苍蝇的事例中似乎也没有解决深度问题,因为他也需要更大的“苍蝇”-苍蝇关联。同样地,这些考察允许目的论者解决奶牛例子中的深度问题,就像福多所能做的那样:他可能简单地论证说:基于知觉的恒常不变性,有关细胞的功能显然就是表征奶牛,而不是表征奶牛形的网膜映象。事实上,解释一生物指向奶牛的行为的唯一的办法就是基于这个假定。^①如果是这样,自然选择就肯定会注意生物的行为发生于其下的描述,至少在奶牛的例子和所有同样的例子中是这样。

结果,福多与目的论者在这一点上似乎打了一个平手。如果奶牛的例子与苍蝇的例子之间没有不同,那么福多与目的论者便有同样的结局:知觉恒常不变性为目的论者解决了深度问题,正如我们原先假定它为福多也做了同样的事情一样。但是如果存在着不同,那么他们便要碰到同样的尴尬:两者在奶牛事例中都解决了

824 第六部分

深度问题,而在苍蝇事例中则相反。

第二个不确定性,即青蛙的细胞表征的是苍蝇还是蜂蝇,又怎样呢?斯特尼(1990)通过否定福多的指责回答道:青蛙的视觉系统表征的是苍蝇而不是表征逃跑的蜂蝇,他论证说,即使在青蛙的自然环境中,所有苍蝇碰巧成了蜂蝇(p. 127)。为了弄清其理由,试考察某些相关的反例。如果苍蝇随着时间而变化,以至于看上去不再像蜜蜂,那么自然选择就会导致青蛙不大可能去捕食蜜蜂。同样,如果青蛙的环境中既有苍蝇又有蜜蜂,那么自然选择就会使青蛙能在两者之间作出区分。斯特尼得出结论说:这些反事实的真实性就是相信青蛙的视觉系统适于甄别苍蝇而不适于甄别蜂蝇的理由。

这种对不确定性责难的回答能否为目的论者效劳尚不清楚。不过,求助于反事实似乎不仅与目的论方案的精神相悖,因为根据该精神,意义就是真实的(不是反事实的)历史的功能,而且更糟的是,它也背离了它的字面意义。正如福多(1990c)所注意到的(p. 75),如果条件不同,那么碰巧发生的事情就不可能提供真实的进化和选择上的利益,因此在对生物学功能的任何说明中也不可能有任何地位——如果是这样,它在对内容的目的论说明中也没有地位。因此斯特尼的回答是不合理的,不确定性责难是能成立的。

值得注意的是:根据纯粹的协变理论,青蛙的心理符号的意义最终是蜂蝇,因为“苍蝇”最主要是与或苍蝇或蜜蜂协变。如果你发现这一结果是反直观的(比如说根据斯特尼的回答),那么我们再一次打了个平手:协变理论和目的论理论都没有道出苍蝇这一意义。如果青蛙碰到的真的只是苍蝇,而不是蜜蜂,尽管有更好的

协变,那么福多的“混合型”的理论(参阅福多 1990c)为了说出苍蝇而不说出蜂蝇,当然会诉诸于真实的历史。但是,如果福多能够盯住真实的历史,以消除纯协变结果的歧义性,那么目的论者为什么不能消除生物学功能的歧义性呢?

但是当它的历史本质导致了目的论理论的不确定性时,卡明斯(1989)便指出了一个更一般的、成问题的结论(pp. 88ff):该理论将与倾向于关于内容的非历史概念的任何人势不两立。他通过设想一个在分子水平上复制有机体的机器给予了解释。试想你进入了这样的机器,并创造了关于你自己的实际的复制品。这样一来就有这样的问题:你的对应物与你具有相同的思想吗?

大多数人的第一个直觉就是:你的对应物肯定有你所有的一切概念和思想。该直觉也许可得到像在《星球跋涉》中所述的运输器那样的装置的支持,它似乎以相同的方式起作用:另一端的成功的“化身”与开端的“旅行者”似乎没有一点不同。而且它还得到了物理主义者所喜爱的东西的进一步支持:如果心理状态不知什么原因依赖于物理状态,那么物理的复制品也应是心理的复制品。事实上,它们怎么可能不是呢?

这就是目的论理论陷入困境的地方。根据该理论,内容是进化历史(或许是上述的个人历史)的一种功能。但是一复制品刚好是突然产生的;在任何显而易见的意义上,它没有任何历史,不管是进化的还是个人的。因此根据目的论理论,它的心理标记不可能有任何表征内容,从而它也不具有任何真正的概念。更一般地说,两种物理复制品根本用不着具有历史的属性;两种复制品的相同的心理标记最终可能有完全不同的内容。因此如果一行动者的

826 第六部分

心理内容和心理状态真的只依赖于她现在的物理状态(以及有关规律),那么目的论理论就陷入了困境。

再者,人们关于这种情况的直觉也许还有待更直接的考察。物理上同一的有机体的表征内容之间也许可能有重要的不同。正如丹尼特所说(1987, p. 166),他有关于他的妻子苏珊的许多信念。但是他的对应物则没有关于苏珊的任何信念;他甚至没有碰到过她。关于苏珊,他自然会有同丹尼特的思想(我们能够假定)在质上同一的思想,并不认为他从未碰到过她,但是他的意向状态仍不是关于苏珊的。如果你的直觉这样发生,那么你至少承认了那生物的个人历史在它的心理状态的个体化中起着某种关键的作用。一当你承认了历史,从而超越了纯粹的非历史的物理主义,那么就可能为目的论理论留下了某种地盘(关于这一争论更根本更详细的细节可参阅下一节的“内在主义与外在主义”)。

人们可能在不同程度上承认目的论对于内容的重要性。正如斯特尼(1990)所说的那样,诉诸于目的论可能是谨慎的也可能是狂热的。谨慎的目的论理论关注的只是具有相当内在和孤立的生物学结构的内容,正像在上述青蛙的例子中那样。而作为一种狂热的理论则试图说明更复杂的结构,如人类的意向状态。正如我们所看到的,福多甚至拒绝谨慎的方面;而斯特尼则支持谨慎的理论,拒绝狂热的理论。米利肯和其他的人则为狂热的理论作论证。关于目的论不管你此前持什么结论,我们仍可通过提出某些专门针对更狂热的努力的问题(从福多(1990c)那里吸纳的)而结束本节的讨论。

狂热的理论肯定这样的主张,即意向状态的内容根源于它们

的功能。为了更加明确,我们可以这样加以表述:某种特定的意向状态的内容根源于它的功能。但是一般来说,当我们的认知机制可以合理地认为有一种功能——有助于我们以提高我们生存和再生的能力的方式行动——时,它就是从那里向下述主张的一种大步跳跃:每一心理状态都有一种功能。

举例来说,我们的许多或大多数信念与愿望都没有那种必须给予它们以生物学功能的进化历史。它们是关于当前或现在的事件、处所和事物的,包括我们自己,或许还有直到现在尚无法设想的未来可能性。而且很难想象,许多信念的功能究竟是什么。正如福多所述,“相信 7 是素数这样的信念是什么?”(1990,p.65)。

愿望就更复杂了。因为在这里我们能像米利肯(1986,p.63)一样想象:愿望的功能就是使所期望的事情发生。可是问题是使这一观念一般地起作用。也许想吃的愿望的功能就是使吃的活动发生;为了证明这一主张,也许还有把拥有那愿望与实际的吃的行为联系起来的某种规律。但是想治愈癌症或想成为这个班的主席或只想得个 A 的愿望又怎么样呢?谅必没有把治愈癌症的愿望与此治疗之发生两者联系起来的规律。

充其量,大概只在治愈癌症的愿望和设法治愈之间存在着规律。但是,哎呀!提到这一点是没有用的,因为设法本身就是一种意向状态或行动,有一对象或内容。因此这一建议等于是说:愿望 D 的内容是由它与另一意向状态的内容的联系所决定的。结果它违背了我们先前的自然主义约定,按照这一约定,我们的全部目的就是根据非意向的属性去解释内容。

最后,有责难说,我们常常感到满足的远不及我们所期望的。

828 第六部分

我们也许期望确实昂贵的汽车,但只能将就买更能承受得了的汽车。有这样的情况,在那里愿望要么无因果效力,要么它促使其发生的不是它的对象而是别的与那对象有关的东西。但是下述说法似乎很古怪,愿望的功能就是使愿望的对象发生,而在极其常见的条件下并非如此。

让我们略作概述。我们先叙述了目的论关于内容的简要的理论,据此,意义根源于表征的生物学功能,接着指出:该理论能够解决关于错误和证实主义的问题,至少在一定程度上是如此。然后,我们考察了福多的论断:该理论是极其不确定的,它不能说明表征为什么可能意指苍蝇而不意指(比如说)黑点或蜂蝇。关于第一个不确定性,我们作了这样的论证:该理论事实上并不比福多自己的协变理论更糟;关于第二点,我们考察了斯特尼的“反事实”回答和福多对此的反应。紧接着我们探讨了卡明斯的观点:目的论理论的历史本质与非历史的物理主义倾向是相互冲突的。最后,我们考察了该理论的狂热的变种所面临的某些问题。

上述简要的说明由于突出了目的论理论的问题而非优点,因此对该理论的实际的奥秘像米利肯和别的人阐发的那样并未作出公正的评价。为了对该理论作出最公正的评价,我们必须再次提醒你注意原著。

2.4. 概念作用语义学与整体论^⑩

就像在目的论理论中一样,我们的下一理论的核心观念就是:为了确定符号的内容,人们不仅应注意符号的原因(像协变理论所做的那样),还要注意其在整个系统中的结果。不过像我们将看到

的那样,“概念作用语义学”(CRS)不大关心构成符号的生物学功能的结果,而更多的是考察符号的原因作用。全面地看,CRS 自然与功能主义(试回忆第 2 章的讨论)相结盟;其倡导者包括哲学家和认知科学家,如内德·布洛克,G. 哈曼,H. 菲尔德,C. 麦金,G. 米勒,P. 约翰逊-莱尔德等。在本节中,我们将根据布洛克 1986 的著名论文“关于心理学的语义学广告”所陈述的思想先简要介绍其关于 CRS 的阐释,然后详细讨论某些问题,尤其是关于整体论的争论。

布洛克事实上赞成“两因素”CRS,据此,意义根源于两种因素:一是完全“存在于头脑中”(源自普特南 1975 一文的短语)的概念作用因素,二是必然涉及到内在符号与其在世界中的所指之间的关系的的外在因素。我们现在更感兴趣的是意义的内在因素,即符号的“概念作用”,它将是符号在推理与思考中、在与别的符号相结合和相互作用中、在感性输入和行为输出之间的媒介作用中怎样运作的一种功能。^⑩应注意的是,如此被设想的概念作用是怎样抽掉了与推理之类的东西无关的那些符号因果作用。例如写出的单词的全部的原因作用就包括(比如说)新闻纸对你的手的影响;在你头脑中实现的心理符号的全部原因作用包括它对你的体重的奉献。但是既然这些作用与推理没有关系,那么它们就不是符号的概念作用的组成部分,结果,按照 CRS,它们也不是意义的组成部分。

不幸的是,被认为与推理之类的东西(以至于意义)“有关”的东西并不总是那么清楚明白。例如一符号在一个人的大脑中实现时比它在另一个人的大脑中实现时其发生作用可能更慢,但是那

830 第六部分

会造成意义上的差异吗？也许不会；但是速率上的差异可能正好会影响（即使是很微弱的）哪一些推理会延长，如果是这样，那么它就与意义有关了。甚至还有更不确定的情况。很显然，符号的哪些因果作用应算作是其概念作用的组成部分，对之加以区分对于 CRS 的任何变种来说都是一个重要的问题。

下面，看看布洛克两因素 CRS 怎样把某个（某些）信念“个体化”（即总合或分类）的两个例子。试比较布洛克通过说出下述两个句子所表达的信念：

1. 我处在被车辆碾过的危险之中。
2. 内德·布洛克处在被车辆碾过的危险之中。

布洛克强调，他的信念 2 对他的行为不可能有与信念 1 相同的避让汽车的作用，除非还有这样的事实，即他相信他就是内德·布洛克（既然他可能相信他是拿破仑）。如果有这些潜在不同的行为后果，那么我们就应说：至少在某种意义上，1 与 2 表达了不同的信念，即使在某种别的意义上，它们似乎表达的正是相同的信念。它们毕竟都是关于内德·布洛克的。

不妨把这种情况与下述关于两个不同的个体的例子加以比较，其中每个人都说“我口渴”。我们可能会说：在一种意义下，它们是关于同一信念的表达式：它们对两个个体的行为将有相同的解渴作用。在另一意义上，它们是关于不同信念的表达式，因为它们是关于不同的人的。

CRS 是这样调和那两类事实的。在第一种情况下，它说当 1 与 2 在概念作用方面不同时，它们具有它们的外在关系；或者更明确地说，它们的意义中的外在因素产生了相同的所指（关于“我”和

“内德·布洛克”),而它们的意义中的内在因素即它们在推理和行为中的功能作用则是不同的。在第二种情况下,与之相应,两个“我口渴”的句子的内在因素是相同的,而外在因素是不同的。引入某种术语,我们可以说:内在(概念作用)因素决定了一句子的(或信念的)“狭义”意义,而外在因素决定了它的“广义”意义。因此在第一种情况下,1与2有相同的广义意义、不同的狭义意义,而在第2种情况下,两句子有相同的狭义意义、不同的广义意义。这一术语现在便产生了一种将 CRS 与较早的内容理论加以比较的新的方法:协变和目的论理论对广义意义的关心超过了狭义意义,而 CRS 则更重视狭义意义。

或者说,先前的理论并不关心你关于 x 的构想,而只注意你与 x 的因果关系,而 CRS 由于把意义置于整个概念作用的基础之上,因而关心你的构想。例如在大多数情况下,狭义意义的差异可看作是构想方面的差异。如果汤姆与弗雷德对老虎有不同的信念,那么“老虎”对他们来说就有不同的狭义意义。一般来说,有此不同就是有关于老虎的不同的构想。构想(conceptions)方面的差异何时才相当于有不同的概念(concepts)? 请看下一段。

在下面的“内在主义与外在主义”一节中,我们将考察广义—狭义意义区分中所涉及到的各种争论。现在我们集中讨论由 CRS 的内在方面所引起的一个问题,事实上是一群问题,即“整体论”问题。该问题又是来自于另一个问题,即布洛克所谓的“伴随信息”问题。还是采用他的例子,试想:你无意中听说某人否定老虎是动物。即使在所有别的方面没有异议,你可能会认为:这个人确实没有理解老虎是什么,或者说他用“老虎”一词表达了不同于

832 第六部分

你的概念的概念。CRS 对此的解释相当精彩：既然你准备从老虎推知动物，而这个人则不然，“老虎”在你们两人中有不同的概念作用，因此意指某种不同的东西。但是，假如你现在听到另一个人谈论老虎，说了许多熟悉的事情（老虎有条纹，原本生长在非洲，有尾巴等），接着又补充了某种不熟悉的东西：老虎喜欢上等色拉而不喜欢可口的小羚羊。事实上，不仅这是不熟悉的，而且你也绝对不会相信它，因为它完全不同于你关于老虎所知道的一切。于是这个人打算从“老虎”推论出“食色拉者”，而你则不然。因此如上所述，“老虎”在你们两人中有不同的概念作用，从而意谓某种不同的东西。但是根据 CRS，我们似乎必须说：你用“老虎”一词时表达了不同的概念。

结果，许多人发现了反直觉的东西：单一信念中的差异尤其是相当琐碎的信念如动物有时是否吃草的差异并不意味着有关的概念是不同的。你们两人想到的毕竟都是老虎。然而，根据 CRS，构想上的任何差异事实上完全相当于有不同的概念。

CRS 抵制这种结果的一种可能的方式常常是寻找某种标准，以区分两类不同的信念范畴，即与决定一个行动者有哪一概念有关的那些信念，以及仅提供关于那概念的“伴随信息”而与其同一性无关的那些信念。这种标准如有用处，那么我们可能会同意说：关于老虎是不是动物的信念，对形成老虎概念是至关重要的，就此而言，关于老虎吃什么信念则不然；因此两个行动者尽管在他们用“老虎”一词表达的那个概念中没有不同，但在后一信念方面则彼此有别。不过这个建议对孤立的信念和概念而言尽管是可能的，但它作为一种普遍的解决方案似乎是没有希望的。因为其

一,不同的标准有必要针对每一单个的概念即一个极其困难的任务而加以阐发。其次,我们甚至不可能为大多数概念设计这样的标准:在我们没有任何明显的根据去作出判断的情况下,关于每一可能的信念是否会一个接一个地影响心理符号的意义(以及概念),我们就只好求助于我们的直觉。最后,几乎不可能达成一致的一致。因为毕竟还不能肯定:动物的信念对于拥有老虎的概念是必不可少的:试想一个小孩,他能轻易地认出老虎,但还没有形成明显相似于动物概念的任何东西。尽管如此,难道不能认为她有关于老虎的概念吗?(在与意义有关的信念和与意义无关的概念之间不可能作出清楚的区分,这一点事实上并不奇怪:它其实就是 W. 奎因(1950)在批评“分析的”与“综合的”句子之间的区别时所精辟地论证过的东西,正像我们简要地讨论的那样。)

如果我们放弃划出清楚的界线这样的希望,那么我们可能碰巧接受了下述反直觉的结论,即每一信念与意义有关,因为两个行动者在明显不重要的信念上存在着不同,就是他们在有关术语的概念作用上存在着不同,进而不仅在构想上不同,而且在他们用术语所要表达的那个概念上也有不同。这一观点就是著名的“(概念)整体论”,因为一行动者心中具有什么概念对有关术语出现于其中的全部信念网络是很敏感的。^④

概念整体论是一有争议的观点。它不仅是反直觉的,而且还面临着许多别的难题。然而我们相信:这些难题也许不是致命的。接下来,我们就来详细讨论其中的一些问题,并作出解答。除了偶然作点改造(及补充)之外,我们将主要按照 J. 福多和 E. 莱波儿在《整体论》一书中的第一章“顾客的向导”(1992)^⑤提出问题的顺

834 第六部分

序进行。我们关注的是四个问题,即人际间的理解、虚假信念和指称、概念的相似性与同一性、分析与综合的区分。最核心的主题是:整体论在原则上碰到了这些问题,但在实践上则设法予以回避。简言之,我们承认:福多和莱波尔正确地解析了整体论的某些含义,但同时断言:这些含义不仅不是致命性的,而且甚至也不是那样有害。^⑩

人际间的理解

问题。如果我心中所想到的概念是由我的相关的信念所决定的,那么当你我具有一切有关的信念时,我们将只具有概念。既然我们具有一切有关的信念是不可能的,既然成功的交流似乎需要我们用我们的语词表达相同的概念,那么根据整体论观点,成功的交流是怎样发生的呢?

回答。严格地讲,如果这里的“成功”需要概念的同一性,那么成功的交流即使有的话也很少见。不过所发生的是足够成功的交流,在这里,限定词大意是说“足以满足所有有关当事人的直接需要”。

假设特里西娅和维贾对苹果的某些信念彼此有别:特里西娅相信它们只生长在纽约,而维贾则相信它们是柑橘属水果。在早餐时,特里西娅说:“请给我一个苹果”,维贾答应了,此时他们的心中出现的是某种不同的东西,但此刻无妨。不过,如果他们旅行到另一可能的世界,在餐桌上传递关于对象的有关信息,那么就有问题了,因为他们最终接受了关于“苹果”的不同的“外延”(指称)。在另一个可能世界,特里西娅要一个苹果时,维贾就会递给他柑橘

或在加利福尼亚生长的某东西,对此特里西娅就会有这样的反应,“不,我要的是苹果。”问题在于,他们的确有不同、具有迥异外延的概念,但是那种差异在当前的这个世界中还不会妨碍他们交流的需要,因为在这个现实的世界里,有关的差异常常是无关紧要的。

因此一般来说(即使不是总是),当我们涉及到相对具体的对象或属性时,或当我们在这样的背景之下,比如说在其中,我们的术语的指称是清楚分明的,那么我们就能成功地交流。我们离这些环境越远,我们的交流就越差,因为我们没有直接的、共同的参照点。但是那几乎是不可能注意到的,因而也许几乎没有直接的关系。当概念上的差异被发现了时,所需要的就是来一点对话,直到它们得到澄清。在这里,“成功”与“不成功”的交流之间的差异因而只是一个程度的问题;在两者中,一个人所想到的确可由另一个人的信念来具体说明。

虚假的信念与指称问题

问题。如果一个人心中的概念由另一个人的信念来说明,那么(这完全是反直觉的)一个人就不可能有任何虚假的信念(哪怕是一种)。为什么不能呢?因为所有所谓虚假的信念最终都是关于另外的某物的真实的信念。例如,维贾关于苹果是柑橘这一信念就不是关于苹果的虚假的信念,而是关于另外的、在某些方面像苹果的、但也是柑橘的某实在的真实的信念(在另外的某个可能世界)。同样,如果两个人对 x 的信念不同,那么他们所想的其实不是同一个项目,就像刚才我们在所接受的外延概念中所看到的那

样。(因此,说特里西娅和维贾有关于同一事物即苹果的不同信念,并不十分准确。)这里的意思是说,在某一领域内的不同科学理论是完全不可比较的,因为它们事实上不是关于同一事物的。正如福多和莱波尔所述,“古希腊人关于星星……骑在类似玻璃的球体上绕着天空运行的观点与我们的下述观点实际上并无冲突,这观点是……[它们]没有……事实上严格地说……古希腊人以前没有关于星星的任何观点”(p. 12);根据 CRS,假如他们在信念上不同于我们,那么他们用“星星”一词所意指的就不是我们用该词所意指的。

回答。不错,严格地说,我们没有关于这类东西的虚假信念。不过,在现实生活中,我们说话很少那么严格。因此无虚假信念并没有引出什么问题。

就拿维贾关于苹果的观点来说。促使我们说他的柑橘信念是关于苹果的虚假信念的东西是:(例如)如果我们向他说明:他递给特里西娅的苹果事实上不是柑橘,那么他就可能放弃他的信念。现在他没有这个必要;他可以坚持认为:虚假的东西不是他的柑橘信念,而是他的这样一个“指代性的”信念,即这个对象是一个苹果。接着他可能指向另一个可能世界,在那里有似苹果的柑橘属水果,并说,“那些是苹果。”如果他固执己见,那么我们将有一个相当愚蠢的语义学争论。但是事实是,他可能让步,改变他已接受的关于苹果术语的外延(以及他由此要表达的意思)。因此即使严格地说,他用一个关于一物的新的真实信念(即相信实际的苹果不是柑橘的真实信念)取代了关于另外某物的另一真实的信念(即相信另一世界那些似苹果的柑橘属水果的确是柑橘的信念),而没有从

先前关于实际苹果的虚假信念转向关于实际苹果的真实信念,但我们仍可以假定他作了改变,因为为了实践的目的,我们关注的是早餐桌上的实际的东西。因此,严格地说,没有关于这类东西的虚假信念——而且关于“虚假信念”的话语也许能够被重构为关于“情愿改变某人所接受的外延”的话语,或至少可根据后一话语而加以解释。(在下面考察分析—综合的区分时,我们将碰到关于“虚假信念”问题的另一种探讨。)

在刚才的虚假的指代性信念的例子中也不难明白这一点。整体论的一个难以捉摸的含义就是:在标记某些概念时,我们心中总有多种属性(因为我们的相关信念在一定意义下就“包含”在我们的概念之中)。但是,每当我们标记那个概念时,它们用不着总是相同的属性。更明确地说,在用同一个单词时,我们心中常有不同的属性(严格地说不同的概念):许多不同的概念常常会被当作是同一个概念。而且我们可以对我们心中的属性给予不同的重视。因此,即使维贾在心中把某些属性揉在一起,并把它称之为“苹果”,但是在继续使用那个名称时,他更情愿放弃那些属性中的某一些。因而我们主张:在许多情况下,使用像“这个苹果”这样的指代性短语就是通过加强这些属性而明确地指称一般来说是最突出的、可观察到的属性(即使不总是如此)。

我们现在清楚了:在面对有关证据时,维贾为什么会放弃他的柑橘信念。那是因为,他说“这个”,就是强调柑橘属性之外的圆、红、离右手6英寸等属性。因此由于看不到与这些属性在一起的柑橘,他就会毫不犹豫地坚持用“苹果”一词表示现在的非柑橘的属性集合。总之,即使严格地说他现在所用的“苹果”一词表达了

838 第六部分

一个不同的概念(因为他放弃了他的某一个信念),即使他原先没有关于苹果的虚假信念,但他对诸属性的强调以及他随后省略某些属性的意愿(就像在指代性实例中所见的那样),就会引起下述不严谨的说法:他真的有过关于苹果的一个虚假信念。

福多和莱波尔关于我们的天文学与古希腊人的天文学并无冲突的主张又该怎样看呢?作出同样的思考是适用的。实际上,我们强调的是我们的信念。在我们指星星时,我们特别强调的是地球上可观察到的那些属性,就像古希腊人那样。因而当古希腊人主张而我们否定星星骑在类似玻璃的球体上旋转时,严格地说,我们肯定有不同的概念,因而讨论的不是同一个东西。但是我们的概念实际上有部分的巧合,比如说,在用指示代词时,我们不仅知道他们谈论的是**什么**,而且甚至能够不严格地说,它们是**错误的**,正如上面所说的那样。

概念相似性与同一性

不言而喻,我们迄今对整体论的辩护利用的是概念相似性这一观念。无论如何,人们其实很少有同一的概念,但他们的概念实际上又是“相似的”,因此我们可以对付着使用(试比较布洛克,1986b, p.629)。

问题。正如福多和莱波尔所描述的:

你不可能有关于相似的某物的完全的观念,除非你相应地有关于同一的某物的完全的观念,这似乎有某种合理性……如果整体论是对的,那么关于“同一(概念)类型的标记”的观念就只为下述实例所界定,其中每一(概念和)信念都被拥有

……如果对两个(概念)在信念系统未能完全一致的情况下是同一类型的标记这一主张没有说明,那么在这类情况下,我们应怎样构建关于两个作为几乎相同类型的标记的信念的观念呢?(p.19)

假如两个人不可能完全一致,那么我们怎么可能重构关于相似性概念的观念呢?换言之,我们怎么可能理解这样的观念,即两个人的不同概念被看作是相似的?

回答。这极像前面的回答。首先,有一更加初步的意见:一般说来,许多不同的概念(“不同的集合”)真的映射到了自然语言的单个词项上;那就是我们的最初的线索。映射到同一词项上的概念可能比映射到不同词项上的概念更相似:因而特里西娅和维贾用“苹果”所表达的两个不同的概念将被看作是相似的。即使是不可想象的,但在同一语言共同体内的不同行动者不可能把极其不同的概念映射到同一词项上。如果他们这样做了,那就可能很快在交流中被发现,并形成一致的意见。词项愈是具体,语言相互作用愈是频繁,极端的概念歧异性就可能愈少。

更明确一点说,基于这一点,并借用认知心理学中的许多实验材料,我们可能要求两个不同的行动者(或两个不同时间的同一个人)把他们将其与某一给定概念联系起来的所有属性列举出来。即使这些列举不完全相称(我们常假定它们不相称),但是相称的程度可以当作是概念相似性的有用的标尺。(即使予以强调,这也只是一个有用的标尺,就像假定概念的同一性对应于所列举的属性一样。在理论中,我们应当继续测量那些概念的相似性,而且要不断地进行下去。不过,在实践中,使回溯停下来的是上面所

840 第六部分

形成的那个观点,因为随着列举的拓展,我们便减少了后退。)换言之,我们把语言作为交流媒介来使用,为我们提供了关于概念相似性的初步意见,即使假定不存在着概念同一性。

再明确一点说:让我们回到关于被接受的外延的意见上。不管有什么重要的哲学屏障,我们仍可能要求特里西娅和维贾到每一可能世界去旅行,并根据他们在此现实世界中的苹果概念,把他们能挑出的苹果都挑出来。如果我们把它们分成并排的两堆,那么我们将看到什么呢?

假如仅在很少的世界中,特里西娅才有机会到达水果的产地,或维贾有机会得到关于柑橘的信息,那么那两堆水果在很大程度上就会是一致的。越是一致,他们的概念就越是相似;一致越少,相似性也越少。因此即使没有概念的同一性,我们仍能重构相似性。

当然,这也许没有重构相似性的所有方面。例如,特里西娅的葡萄水果的概念比她的V形8缸式发动机概念更相似于她的苹果概念,然而她所接受的外延在任何情况下很少是一致的。但是这种“异类词”的相似性并不是所讨论的那类相似性。即使你假定:特里西娅和维贾的苹果是同一的,但要讨论苹果与他们的葡萄水果的相似性毕竟还需要某种别的理论。(再者,也许还有借助所接受的外延解决这种相似性的方法。假如苹果和葡萄水果明显共有的属性达到很高的程度,那么基于这么多的工作,它们在具有V形8缸式发动机的世界之外的更多的世界中就合成一类了。

请注意,在我们的答复中,我们只用了基本名词。不过,我们认为,这种处理方法对于其他种类的、外延不太好界定的范畴和词

语也是有用的。不管哪种情况,在其中外延难以界定,都不可能见容于主张概念具有同一性的理论。

总之,在其中肯定没有概念同一性的条件下,我们是没关于概念相似的“健全的”观念的。但是我们其实不需要它,因为我们的实践——我们在主要是公共的现实世界中的公共的语言实践——提供了一个作为其替身的有用的相似性观念。

分析—综合的区分

20世纪70年代以来,分析—综合的区分已受到了哲学的高度关注。其基本观点简单明了。某些句子的真(或假)似乎严格依照有关语词的意义而确定,以至于可以说,知道了意义就足以知道句子是真还是假。经典的例子是:“单身汉是未婚的。”任何懂英语的人都立马知道:这是真的;如果某人不知道它是真的,那么某人就一定不知道句子中的某一单词的意义。现在试把这类句子与“单身汉有可能比已婚的男人活得更长”加以比较。这个句子是真还是假呢?一个人要搞清楚这一点,就必须作某种研究;知道单词的意义对于知道句子之真假是不够的。

第一类句子被称之为“分析的”,意思是说句子的谓项(或者至少是未婚这一对应的概念)在某种意义上已“包含”于主项之中(或至少是在概念“单身汉”之中)。于是仅通过“分析”主项概念我们便能发现谓项概念,进而能通过了解单词的意义(即通过把握有关的概念)而确定句子之真。后一类句子被称之为“综合的”,因为其谓项不包含在主项之中,而是加于主项之上的,旨在形成一个新的综合整体。既然谓项不包含在主项之中,那么仅把握主项概念对

842 第六部分

于知道句子是真还是假就是不够的;因而需要作出某种经验研究。

简言之,分析句据说是基于意义而为真(或假)的,而综合句是基于经验的事实而为真(或假)的。但是正当这种对句子类型的区分貌似合理时,本世纪最著名的哲学论文之一,即由 W. 奎因所写的“经验主义的两个教条”,则论证说:这种区分事实上是站不住脚的。该论证的技术细节,包括其精妙的注释和主要的意蕴超出了我们这里的范围。不过,其结论已向概念整体论提出了另一个问题。

问题。还是继续我们关于虚假信念的上述讨论,如果人的概念出现于其中的几乎所有信念对定义概念有所帮助(因而“包含于”其中),那么某种类型的(几乎)所有的直陈句就将是分析的。它们是真是假,只有借助于反思有关概念才可发现。但是这是很麻烦的,因为大多数直陈句似乎并不正好符合于这个模式。奎因的著名结论事实上是严格的对立物,他主张,所有这类句子都是综合的。在这里,该怎么办呢?

回答。所有断言实在之属性的直陈句都是分析的——从个人习语上说。所有这类句子也许真的都是综合的——从社会方言(或某种类似的东西)角度说。这是什么意思呢?

语言学和心理学之间存在着差别,而有关讨论则常常忽视了这一点。说一个语词意指什么是一回事;而说一行动者用一语词意指什么又是另一回事。后者是行动者的“个人习语”的组成部分,即是说是她自己的个人惯用语,而前者也许是更大的共同体的惯用语(或“社会方言”)的组成部分。行动者用一语词意指什么要由关于它的信念来说明,因此她诚实地表达的所有直陈句都为她

从分析上视之为真(或假)。也许她用一语词意指的东西并不是该语词所意指的东西,或更明确地说,不是她的语言共同体用该词所意指的东西。也许正如奎因所倡导的那样,社会的和科学的语言实践最好这样来加以描述,即把所有的直陈句当作是综合的。但是那与个人习语的讨论又没有关系。

有时似乎有这样的情况,即行动者从综合上把她的信念当作是真或假的。例如,她也许相信苹果是柑橘属水果为真,然而她似乎不是根据“苹果”一词的意义而如此相信的。但这是由语言的运用而产生的一个幻觉。根据所接受的外延和所强调的属性对之加以彻底改造(正如我们在讨论虚假信念时所说的),幻觉便会随之烟消云散。如果她在知道苹果不是柑橘属水果之后,因为强调属性而仍继续用“苹果”这同一个词表示冰箱里的那些食物,那么她想起的似乎是相同的东西,即使把不同的属性归之于它们,进而她似乎还会综合地对属性作出归属。不过,当我们注意到下述情况时,这种现象便会随之消失,那情况是,在懂得苹果是非柑橘属水果时,她也会改变她所理解的关于苹果的外延(从各种可能世界加以估价),而在发生这种改变时,那就正好是她所想起的东西。因此属性并不是综合地而是分析地得到归属的,尽管这有悖于偶然的发现。

还是回到先前的一个例子,对某个古希腊单词的最好翻译也许确实是“星星”,即使古希腊人关于星星的信念非常不同于我们的。根据最好的语义学理论,那个词所意指的也许真的是“星星”所意指的东西。但是那并不意味着某个古希腊人用那个词所意指的东西(他心中所想的)与我们用“星星”所意指的东西是相同的。

844 第六部分

他心里所想的东西肯定不存在于实在的世界之中,而我们心中所想的则存在——至少我们认为就是这样。

语言学和心理学的这种区分提供了另一种重构先前分析过的、关于虚假信念的观点的方法。在那里,我们把它当作是这样的某东西:在维贾碰到人们关于特里西娅和别人称之为“苹果”的东西的详细说明而可能改变其关于苹果的信念时,我们就说他那个信念是虚假的(因为有悖于关于另外的像苹果的某物——只是柑橘——的真实信念)。现在我们可以这样初步地予以设想:当他用“苹果”意指的东西不同于特里西娅和别人所意指的东西时,我们就说维贾的信念有时是虚假的。以这种方式,我们便能承认维贾的句子具有严格的、个人习语的分析真理,同时又符合这样的观点,即她有时说的是错误的。总之,关于分析—综合的区分的概念整体论所蕴涵的并非那么糟糕。换言之,它的分析的内涵不一定与我们的综合直觉相对立。

综述:在本节中,我们简要叙述了概念作用语义学,它试图把心理符号的意义建立在符号对行动者的认知过程所起的整个因果作用这一基础之上。接着,我们考察了“伴随信息问题”,即确定哪些信念与符号的意义有关(进而去定义由那符号表达的概念),哪些无关。由于难以提供任何可行标准,从而便导致了整体论的观念:人的(几乎)所有信念都与人的术语的意义有关。再接着,我们描述了概念整体论在同化四类问题时的颇有实用主义色彩的辩护。即使我们对这些问题的回答只是描述,但它们也许足以表明:那些问题也许不是致命性的。

这就完成了我们对四种主要的关于表征(即心理)内容的自然

主义理论的考察。我们还记得,这四种理论作出了共同的探讨,以用自然主义回答下述问题:是什么使我们得以处于意向状态之中?心理符号根据什么而派生出它的意义?由于他们把意义建立在不同属性的基础之上,如相似性、协变、目的论和概念作用,因而自然彼此有别。每种理论都有某些优点,但又有某些问题;很显然,它们试图回答的问题都是很困难的!

在本章最后一节,我们将转向对意义的本质的更一般的考察。其实,接下来的并不等于就是关于意义的替换理论,就像不是关于意义的理论化的严肃态度的争论一样;事实上,主要的问题在我们对上述四种理论的整个讨论中一直是隐晦不明的。它们于是就变成了下述令人困惑的问题:思想的心理内容在何种程度上对外在于行动者头脑的因素敏感?

3. 内在主义与外在主义

根据我们在第2章考察过的笛卡尔的实体二元论,心灵与身体是完全不同的两类实体。但是当此观点在哲学家中不再流行时,它的内涵则一直牢牢地支配着直觉。难以放弃的观点是,我们总是按从内到外的顺序接近世界:我们的心灵似乎包裹在我们的肉体之内,具有其自己的私人的心理“空间”,因此我们真正有权接近的就是我们自己的思想和经验。而且,我们似乎有优越的通道接近这些东西,因为我们能比他人更直接、更可靠地知道它们(参阅第5-6章)。因而,我们认识的主要目的似乎就是基于我们关于它的内在经验建立关于外在于我们的世界像什么的图像。

简言之,在某种意义上,我们的思想和经验与它们的环境似乎是隔绝的。这种意义当然不是因果上的意义;我们的心理状态常常是由不同的物理状态引起的,并对后者有因果作用。(请记住,实体二元论之所以遭拒绝,主要是由于它在解释心物因果关系时陷入了困境。)毋宁说,这里的意义有点像逻辑上的意义。更明确地说,正是我们思想的表征内容才似乎是纯内在的事件,完全在心灵之中,即使我们思想的因果关系不是这样。这一观点也许是笛卡尔二元论的重要遗产。

笛卡尔本人以怀疑论者的忧思的形式丰富了这一观念。他问道,有没有这样的主张,我们对它是如此肯定以致简直没有任何合理的方法予以怀疑?例如我们怎样知道:不存在某种强有力的、可恶的幽灵,其唯一的功能就是在一切事情上误导我们,它如此有力,以致我们所想到的一切可能都是虚妄的?我们怎样知道:世界事实上完全不同于我们所认识的样子,世界也许真的是不存在的?思维与它们的环境“在逻辑上隔绝”很有可能出现在这样的方案之中:我们的思维可能正像它们所是的样子,可能正好具有它们所具有的内容,即使在它们(对于笛卡尔来说,包括我们自己的身体和大脑)之外没有它们所表征的外部世界。

当代哲学家由于不是实体二元论者,因而感觉到,在这种方案中没有把大脑算进去的必要;据假定,大脑在某种意义上对拥有思想是必不可少的。因此直觉便受到了关于这类思想实验的更新的观点的检验,我们在上面批评目的论理论时曾陈述过它的一个变种。现在设想,有这样一个传说,你的大脑的分子对应于分子的复制品被创造出来了,并在装有营养物的大缸中存活下来了。一睿

智的神经科学家周期性地进来,用电极刺激那个大脑。以这种方式,他使你的大脑进入了不同的物理状态。它们实际上像你有特殊的思想和经验时你的大脑所进入的物理状态。假如在某种情况下,你有一思想,你描述说,“呵呵!有一卡车在我的右边开动了。”神经科学家又刺激大脑复制品,让它实在地进入你的大脑在那种情况下所进入的状态。问题随之而来:大脑复制品在那时也有这样的思想:呵呵!有一卡车在我的右边开动了!?(而且,它有全部经验——即将到来的关于卡车的视觉经验,包括所有末梢感受性质以及伴随而来的恐怖吗?)

大多数人都有这样的直觉,大脑及其复制品真的应当具有心理状态。大脑的不同的外部环境——一个大脑在你头上,而另一个只漂浮在一个大缸中——与大脑思维的内容似乎是没有关系的。^⑩但这也正好等于说:我们的心理状态与外部世界是隔绝的:我们的思想完全“在头脑中”(借用的还是普特南 1975 的著名的短语),只依赖于我们的大脑状态。

依赖关系究竟是什么,这是一个有争论的问题。例如在第 2 章中,我们探讨过心理状态是类型上还是标记上同一于大脑状态。事实上,由这些观点所衍推的另一有关的(且极为流行的)观点就是:心理状态随附于大脑状态。粗略地讲,说心理状态随附于大脑状态就是说:两个行动者(或同一行动者在两个不同的时间)的心理状态可能不会有什么差异,除非他们的大脑状态存在着某种差异,或者换句话说,如果两个行动者处于相同的大脑状态,那么他们必然地处于相同的心理状态。须知,随附性只涉及到这些关联,并不承诺心理与神经状态之间的任何同一性或特定的因果关系。

848 第六部分

在第4章中,我们将进一步讨论随附性;至于现在,我们只是借用它。因此我们认为,我们面前的观点即思想完全存在于大脑中,实即这样的主张:人的思想的内容随附于他的大脑状态。

这一学说的常见的名称是“内在主义”。它几乎类似于常识的观点。不过,大约在70年代以后,许多哲学家对此观点发起了攻击,而论证支持某种“外在主义”的观点。作为对立的学说,它认为外在环境(头脑之外的)一定与人的心理状态的个体化有关系。

事实上,在关于目的论的那一节中,我们已初步领教过外在主义。试回忆丹尼特的主张:丹尼特的复制品不可能有关于丹尼特的妻子苏珊的任何思想。也许,那复制品在质上具有同丹尼特的思想一样的思想,就其内部而言,它们“感觉起来”是相同的,但它们不是关于苏珊的,因为那复制品从未碰见过她。但这等于说:尽管他们处在相同的大脑状态,但并没有相同的思想内容——因此外在的历史环境关系到思想的个体化。在这种情况下,如果你与丹尼特有同样的直觉,那么你就有某些外在主义的倾向。

普特南(1981a)作为外在主义后面的原始的驱动力量之一,把这些倾向具体应用于“缸中之脑”的故事中。他论证说:这种大脑的思想不是关于外部世界的任何事情,而事实上只是关于它自己的输入机制的。他断言,那是因为,“关于”离不开处在与有关对象的正确的因果关系之中;外在因果关系对意向性是必不可少的。(因而丹尼特的复制品由于没有与苏珊的因果关系,就不可能有关于她的思想。)既然缸中之脑在因果上与世界相隔离,它的思想就不可能是关于世界上的任何事情,即使就大脑内部来说它们似乎是这样。我们大概可以说,“复制品的思想”不是真正的思想,就

像确凿无疑伪造出的伦勃朗* 绝不是真正的伦勃朗一样。

简言之,根据外在主义,物理复制品不一定没有心理状态;这完全取决于它们与在它们之外的世界的关系。因此只注意行动者的大脑状态在原则上肯定不足以确定那行动者在想什么;人们必须同时关注行动者的外在环境(包括他的过去历史)。

为了说明外在主义及其内涵,许多哲学家阐述了种种思想实验(参阅“进一步阅读书目”)。这里我们只介绍其中两个最著名的。在上面,我们事实上已简要讨论过第一个的一个变体,即普特南的孪生地球案例。我们还记得,孪生地球是一个实在地相似于地球的星球。但是我们现在强调的是单词“实在地”:孪生地球在每一细节上包括其居民都是地球的分子对分子的复制品。于是当你阅读这些词语时,同时还有你的一个复制品,也在阅读复制的M.E. 夏普公司所发表的这本书的另一拷贝中的这些词语。两个星球只有一点差别,即充斥于河、湖等中的清澈的液体不是 H_2O ,而是某种别的化合物,即XYZ,它在所有实际用途上与 H_2O 没有什么区别。我们的孪生人,由于是我们的分子上的复制品,因此身体的行动与我们的一模一样,进而把那化合物称之为“水”;但是它不是水,因为就我们所知,水是 H_2O 。

假设,你看到一个喷泉后便想,“水是多么的美呀!”在那一刻,一个孪生的你正注视着复制的喷泉,并想到了这样的词:“水多么的美呀!”问题是:你们有同样的思想吗?换言之,你们的思想有相同的内容吗?^⑧两种直观立即会发生冲突。一方面,孪生人是复制

* 伦勃朗(Rembrandt, 1607-1669),荷兰著名画家。

850 第六部分

品,实际上处于同一的物理(包括中枢)状态。(我们可以不管我们的身体包含 H_2O ,而孪生人的身体包含 XYZ。)基于我们对大脑在我们的心理过程中的作用所知的一切,这似乎等于赞成内在主义的下述观点,即孪生人一定有心理过程。但是另一方面,XYZ 并不是水,即令它与水没有什么区别也是如此。当你的思想是关于水时,那么你的孪生人的思想由于他的环境则是关于 XYZ 的。但是“关于”正是意向状态的本质:关于不同对象的思想就是不同的思想。因此即使普特南本人原先并没有区分孪生人的那些心理状态,但许多人已接受了那个教条,而且它已与普特南的名字紧紧地联在一起了。

在这里,支撑普特南的外在主义说教的东西主要是关于自然种类和自然种类术语的意义的独立的学说。从根本上说,自然种类就是关于对象的自然范畴;典型的例子主要有电子、基因、老虎等。至于指称自然种类的术语,普特南论证说:这些术语在它们的定义中有“索引”因素。而索引就是这样的语词,它们的意义在关键的方面依赖于说出它们的语境。例如像“这里”、“现在”、“我”等词:它们意指什么实际上取决于谁说出它们,以及它们在什么时候和地方被说出来。说自然种类术语是索引性的就等于说它们的意义同样依赖于说出它们的语境。就眼下的例子而言,“水”被用来意指这样的东西,即与周围的液体具有某种相似性的材料,这里的“周围”意指地球上的“这里”。孪生地球上的东西也是这样的,所不同的是“周围”意指孪生地球上的“这里”,在那里,液体是判然有别的。如果普特南是对的,那么英语中的“水”就有不同于孪生英语中的“水”的意义。如果这些异质的“水”有不同的意义,那么当

你说,“水多么的美呀”时,你所断言的东西就不同于你的孪生人在说:“水多么的美呀”时所断言的。因此这近乎于说:你们每个人用那些词表达的思想是不同的,于是我们借助语言哲学推导便到达了外在主义。(更详尽的细节,参阅后面第一个答复。)

但是当普特南的外在主义从这类对自然种类术语的考察中派生出来时,T. 伯奇(1979)则论证说:结论实际上可加以泛化。伯奇断言,物理环境上的差异(H_2O 与 XYZ)不仅能够造成意义上(进而心理内容上)的差异,而且还能造成语言共同体中的差异。

.....

我们再一次碰到了外在主义:两个奥斯卡尽管是物理的复制品,但处在不同的心理状态中。只是在这里,有关的差异不在于他们的物理环境,而在于他们的共同体的不同的语言实践。既然语言实践决定了每一个单词,因此外在主义的论证就不仅可适用于自然种类的思想,而且正好可适用于任何别的思想。(伯奇自己的例子有沙发、胸脯、拨弦古钢琴、契约、颜色等)外在主义显然是一种有说服力的理论!

值得强调的是,普特南和伯奇两人的思想实验及随后的结论是极有力量的。分子复制品可处于不同的心理状态中;一个人处在什么心理状态对他的物理环境和语言环境在许多方面是很敏感的。某些问题似乎会接踵而至,而其中许多则有点令人讨厌。不仅笛卡尔关于心灵是纯内在的机构的遗言要被放弃,而且到达我们自己的思想的优势通道以及一般地说关于它们的知识这样的观念(如前所述)也是如此:如果你的思想内容(以及你的心理状态)依赖于外在于你的因素,哪怕你对它们一无所知,那么就不能说你

852 第六部分

知道你自己的思想(关于自知的争论,请看第5章)。

关于心脑关系的问题也同样令人困惑。外在环境(尤其是语言环境)通过什么样的因果机制(如果有的话)对人的心理状态产生它们的决定作用?根据外在主义,大脑在决定人的心理状态时起什么样的作用呢?我们能搞清对有目的行为的常识解释的意义吗?常识把我们的行动解释为我们的意向状态尤其是这些状态的内容所引起的结果:我们喝药,因为我们期望身体好,并相信:药物会有疗效。同时我们的大脑对我们似乎会起相关的原因作用。但是如果是这样,如果这些内容依赖于外在于我们大脑的因素,那么我们怎么可能把我们的意向内容作为原因加以信赖呢?(关于心理因果关系更多的内容,请看第4章。)基于外在主义,有理由对神经心理学进行探讨吗?有希望为心理学提供自然主义的还原吗?

围绕着外在主义,正好产生了一些困难的问题。而且普特南和伯奇促发了旨在解决这类问题的全部工作,这些工作结果又与形而上学、语言哲学和科学哲学中的其他许多争论联系起来了。现在我们打算讨论一下许多哲学家用以回应外在主义挑战的方式中的一个小的样本。为了使行文简明扼要,我们将忽略普特南和伯奇之间的差别,主要集中在对普特南的答复上,我们将考察其中的五点。(关于内在主义与外在主义争论的主要论文以及关于该争论的某些评论,可参阅佩辛与戈德堡,1996)。

第一个答复是分析上面描述过的普特南关于索引性术语的概念。有人可能会论证说,当汤姆在纽约用“这里”一词,撒拉在洛杉矶用“这里”一词时,他们指的确实是不同的地方,但不能因此说:“这里”一词在他们那里有不同的意义;它只是有不同的指称或外

延,而指称是不同于意义的。如果自然种类的术语在这方面是索引性的,那么你和你的孪生地球上的对应者所说的“水”在外延上就是不同的(H_2O , XYZ),而在意义上则没有不同。如果它们在意义上没有区别,那么孪生人用了“水”一词的各种句子分别表达的思想在内容上就不会有差别——进而孪生人将有心理状态。

但是就日常语言的索引而言,如果这种答复可能是对的,那么普特南其实可论证说:“水”在两种话语的外延上的差异肯定与意义上的差异有联系。如果我们不在意说:被当作“这里”的地方完全随话语的语境变化而变化,那么我们同样不会说:被当作水的液体也这样变化。XYZ 在孪生地球上并不被认为是水,因为在那里根本就没有水;地球上的英语中的“水”一词并不适用于孪生地球上的任何物质。(不要因这样的事实而误入歧途,即孪生英语中有“水”的同音同形异义词!)如果讲英语的地球人(不懂孪生地球语)访问孪生地球,把湖里的那种液体说成是“水”,那么他就错了——这与下述情况是不同的,即当汤姆到洛杉矶旅行,一到达便说:“我到了这里!”

不管怎样,普特南自己的推理依然表明:他轻率地把自然种类词项比作日常索引大概是错误的。伯奇(1982)(用超外在主义的批评方法!)事实上相当清楚地阐述了这一点。他指出:如果“水”意指这样的某物,即与周围的液体有某种相似性的材料,那么在保留“这里”一词的索引性时,它的适用性就会发生变化,正像“这里”会变一样——访问孪生地球的地球旅行者在把 XYZ 说成“水”时就是对的。如果像许多人所认为的那样,该旅行者这样做不对,那么与普特南的观点相反,“水”(以及一般的自然种类术语)就不可

854 第六部分

能相同于日常的索引。

现在也许有可能对这个结论提出责难。例如人们可能会论证说:外延至少是日常索引的意义的组成部分,倘若如此,普特南把自然种类词项比作索引将有可能取得成功(即使在与上述方式相反的方向上)。不过我们不值得去探寻那条道路,因为未必要求助于那种同化本身:有意义的是普特南的下述观点,即“水”随着语境之不同而有意义的变化,不管日常索引是不是这样。

这里以及其他所有部分所贯穿的这一课题牵连许多复杂的问题,在下面的讨论中,我们只打算述及其中的两个。例如,我们怎样把语言个体化?地球英语应看作是不同于孪生地球英语的语言吗?我们怎样确定一个体所讲的是哪一种语言?你能够说,地球人一到达孪生地球便开始讲孪生地球语吗?(他肯定会流利地讲它!)如果不是这样,假设他生活在孪生地球上几个月、几年或几十年;如果可能的话,他在什么时候才会成为一个说孪生地球英语的人?在海外旅居2或3或4年的某人又怎么样呢?人能熟谙两种语言吗?有规律地在两个星球之间旅行的人又怎么样呢?如果地球上的某些湖、水坑等(或许多或大多数?)包含有XYZ,即使某个特定的个体从来没有碰到过这些湖和水坑,那么这又意味着什么呢?但是我们得放下我们现在的讨论。

对外在主义的下一个答复将在第一个答复完结时再接着说。“水”在地球和孪生地球上真的有不同的意义吗?促成这一主张的是它的几个相关的主张:

(a)“水”在两个星球上有不同的外延。

(b) H_2O 和 XYZ 属于不同的自然种类。

(c)说一物质是水就等于说它同我们的语言共同体的成员称之为“水”的东西具有相同的本质。

(a)事实上是坚持“水”在两个地球上意义不同这一主张的最直接的动因;(b)(c)则支持了(a)。现在的问题是:它们这样做是否能成功。换言之,它们为我们提供了相信(a)的理由吗?

试考察(c)。正如 E. 泽马齐(1976)所问的那样,谁可以算是“我们的”语言共同体的成员?如果该共同体包括所有讲英语的人,那么(他论证道)我们应包括孪生地球上的人。针对下述反对意见,即孪生地球人讲的是孪生地球英语,而不是英语,他指出:把孪生地球英语与英语非武断地区分开来的唯一方法就是看你是否已假定:他们称之为“水”的东西不是水,进而用未经证明的假定来论证;否则那两种语言就是同一的。但是如果语言共同体包括孪生地球上的居民,那么根据(c),称之为“水”的材料既不是 H_2O , 也不是 XYZ,而是选言物质:或 H_2O 或 XYZ。结果与(a)相反,两星球上“水”的外延是同一的。

即使我们假定:普特南找到了一种用有根据的假定来辩论的方法,以把孪生地球居民从我们的语言共同体中排除出去,但仍有麻烦。因为几乎所有的地球居民也会把孪生地球上的海、湖等中的物质称之为“水”。根据(c),而与(b)(a)相反,这意味着,XYZ 毕竟是水。事实上,泽马齐补充说,这是相当自然的,因为它反映了“水”和别的物质名称实际被使用的方式。试想想我们关于许多事物所说的话:纸也许有许多不同的化学结构,沙、布和头发等也是如此。再看氧气氛(D_2O)这一例子。当它在化学上不同于 H_2O 时,它常常被看作是一种水(并被如此命名:“重水”),就像其公式

856 第六部分

包括了 T_2O 、 HDO 、 HTO 和 DTO 的别的物质那样。在我们典型地称之为“水”的所有样本中,根本就不存在任何单一的化学实质或共同的本质,通常的水常常包含有一些 D_2O 、 H_2O_2 ,一些 H_2O_3 以及其他大量物质,而且硬水、软水、矿泉水、盐水和蒸馏水在化学上都是彼此不同的。如果我们接受(c),那么所能根据的就只能是这样的解释,即正讨论的共同“本质”是一种选言本质,以至于许多熟悉的物质包括 XYZ 大概都可看作是水。但是如果 XYZ 是一种水,那么再一次与(b)(a)相反,“水”在地球和孪生地球上有着相同的外延:水。

看起来,(c)至此正成为种种麻烦的根源:如果接受(c),我们最终便陷入了与(a)(b)的矛盾对立。如果我们放弃(c),又会怎样呢?这一建议的问题在于:如果我们放弃(c),那么我们首先便放弃了相信(b)(进而)相信(a)的主要理由。^④自然,这并未使(b)和(a)成为虚妄的,但它意味着:如果人们要接受关于孪生地球的外在主义教条,那么就应为它们提供某种独立的、合理的论证。因为如果“水”在两个星球上有相同的外延,那么我们就没有动力去主张,它在两个星球上有不同的意义。如果两种“水”在意义上没有不同,那么我们就应该说:两个孪生人在说出“水多么美”时,并未意指不同的东西。如果我们不能这样说,那么我们就没有理由主张:他们用那个句子表达了不同的思想。如果我们失去了这样的理由,那么我们就抛弃了外在主义:物理上同一的两个孪生人终于都有心理状态了。

在开始的答复中,我们考察过这样的主张:在两个星球上,自然种类的术语如日常的索引只是在外延上不同,而在意义上则没

有不同。在第二个答复中,我们探讨了它们在外延上是否是不同的(注意:如果它们没有不同,那么我们就没有理由认为:它们在意义上不同)。现在我们转到下一个答复上。

假设为了论证的方便,我们承认外在主义者的两个观点:“水”不同于普通的索引,而且它的外延在两个星球上也不一样。在这些情况下,我们一定要赞成说“水”在意义上(即在内容上)也不同吗?提问者是福多。

福多(1987)指出:在孪生地球出现之前,就已有有用的、合理的、难以丢弃的、据以可以将心理内容个体化的原则,即

A. 内容之同一与外延之同一相关联。

更明确地说,该观点是:外延随附于内容,以至于如果两术语在相关的内容方面(即概念上)没有某种差异,那么它们的外延也不会有任何差异,反之,除非外延同一,内容就不可能同一。^②

问题在于,还存在着另一也有用的、合理的、难以放弃的原则:

B. 心理状态随附于大脑状态。

问题之所以产生,是因为,正如我们已述的那样,孪生地球故事似乎使两个原则陷入矛盾之中。一方面,如果我们赞成(A),那么在假定外延存在着差异时(与上述第二个答复对应),我们就必须对两个孪生人的概念(因此还有思想)作出区分,这一来又与(B)相悖。另一方面,如果我们承认(B),那么我们就必须说:两个孪生人有同样的思想(进而有同样的思想内容和概念),尽管有不同的外延,这一来又违背了(A)。(我们不会忘记,普特南本人原来坚持这后一条路线,而与他的名字有联系的外在主义则采取前

858 第六部分

一路线。)如果人们要同时坚持 A 和 B,那么必不可少的是采取这样的方式,即让地球和孪生地球上的“水”成为同义词,而不管其外延上的差异(即使参看下述为另一选择而作的第五个答复也是如此)。

福多提出了这样的建议,即如果承认(B)为真,那么他指出,孪生地球就是一个难题,因为它似乎割裂了外延同一性与(A)所假定的内容同一性之间的联系。但事实上,他论证说:那是错误的,孪生地球案例并未破坏那个联系,而只是把它关联于语境。

“关联于语境”是什么意思呢?试考察一个类比。福多论证说,x 和 y 这两个对象的因果力是否相同也必须相对于语境而加以估计。例如如果你在不同的语境下检测你与你的朋友们的二头肌的力量——比如说,你们两人都试举椅子,但你的朋友举的是一把很重的椅子或固定在地板上的椅子——那么你就无法比较你们的二头肌的力量。因此我们可能会这样说:说 x 和 y 具有同一的因果力就等于说,无论 x 在特定的语境下能举起什么,y 在那语境下也能如此。

我们可以把这个观点用到思想内容上。在孪生地球之前,原则(A)可能是这样理解的:一词项的内容就像一数学函项,因为它(至少部分地)把你“从”该词项带“到”该词项的外延。两个孪生人之所以是一个问题,是因为(假如(B)),他们有相同的思想和概念,而根据关于(A)的函项观,他们的“水”就应映射于相同的外延之上——它们并没有如此。但是我们现在必须“把这些函项相关于语境”,进而把(A)修改成下面的样子:

A.* 一词项的内容把你从该词项及其语境带到那词项的外延上。

只有当两个内容把思想和语境中相同的函项实现于外延之上,换言之,只有当一个内容在特定语境下将那个词项映射到任何外延上,另一内容在那语境下也是如此,那么两个内容才会是同一的。

因此我们应怎样分析你和你的孪生人都想到“水”这样的事实呢?是的,你们有相同的大脑状态,你们有不同的外延: H_2O 与 XYZ。不过,如果你(与你的孪生地球人一道)住在孪生地球上,那么许多人的直觉便是:你们的外延是相同的(两者都是 XYZ);不过,如果你的孪生地球人(与你一道)住在地球上,那么他们还会有相同的直觉(两者都是 H_2O)。(即使你记得:这些直觉得到了下述事实的支持,即你和你的孪生地球人都不能分辨 H_2O 和 XYZ 之间的差异,从而在每一语境下都接受相同的外延,但这些直觉仍是相当有争议的。)在同样的语境下,即是说你和你的孪生地球人有相同的外延,因此将享有从词项、语境(星球)到外延的相同的映射。结果,我们就可以说,尽管外延上有不同,但你和你的孪生地球人都有内容(所以有概念)。于是我们得保留(B)和(A)(以及它的修改过的形式(A*))——而且因为复制的孪生地球人将有心理状态——因此我们得设法防止外在主义的危害。

现在,我们应该强调的是,在这里起作用的内容概念不知什么原因是不同于支配着我们本章大多数讨论的那个内容概念的。这个新的概念原来就是所谓的“狭义的”内容(或意义)。它与“广义的”或“宽泛的”内容形成了鲜明对比,后者其实就是在你详细说明

860 第六部分

了狭义的内容和语境之后所得到的东西。正如我们在关于概念作用语义学那节所述及的那样,该术语的核心意思是:狭义的内容完全取决于“头脑中的东西”;大脑状态的同一性保证了狭义内容的同一性。然而,由于外在的(或“广义的”)语境上的差异,两个孪生人的思想就可能有不同的外延,从而在广义内容上有着不同。因此福多的观点可以简化为:两个孪生人在广义内容上有区别,但他们共有狭义的内容——而且狭义的内容是心理学科学中所必需的东西,包括保护原则 B。^②

我们还应强调的是:福多的观点适用于所有类型的语词,包括自然种类的词项。于是它也为我们对普特南的早期观点提供了直接的回应,后者认为,既然 XYZ 不能看作是一种水,那么像“水”这样的词项在两个孪生星球中的意义就是不同的。我们完全赞成他的看法(如果我们愿意的话):XYZ 并不是一种水,但可作出补充说:它所衍涵的两种“水”在意义上的唯一的不同就是广义意义上的不同——因此一切都不错。

在这里,很多东西尚未述及。例如关于什么是心理学科学所必需的,就存在着大量的争论(例如可参阅伯奇,1986)。关于行为、因果力的准确的个体化,也有需予解决的困难问题,其中有两个对于原则(A*)和(B)都有重要的意义(例如可参阅福多 1991;以及本卷第四章)。关于狭义内容是否真的是内容的一种形式,也存在着繁难的问题。内容毕竟是心理状态的这样一种属性,据说它反映了心理状态的“关于性”,而且既然福多假定两个孪生人的状态是“关于”不同的事物(H₂O 与 XYZ)的,因此任何内容似乎一定也有不同。“狭义的内容”究竟怎样涉及到关于性,这一点的确

是很难弄清楚的。狭义的内容似乎是不可言喻的：你的“水”被认为具有与你的孪生地球人的“水”一样的狭义意义，但是共有的意义究竟是什么？单个的概念水在这个世界指称 H_2O ，在那个世界指称 XYZ，它究竟是什么？它似乎既未被 H_2O 所把握，又未为 XYZ 所把握，而且有根据说，（假如地球人未曾碰到过 XYZ，孪生地球人未曾碰到过 H_2O ）它也未被或者 H_2O 或者 XYZ 之类的人为构造所把握。这些再度成为困难的问题。关于更进一步的细致的讨论，我们提请你看福多的著作。

第二和第三个答复通过论证两种“水”的意义之同一性来向外在主义挑战：在第二个答复中，它们的外延有不同这一点遭到了否定，而在第三个答复中，其不同又得到了承认。在下一答复中，我们直接关注的是语词的意义与思想的个体化之间的实际关系，我们认为这是理当如此的事情。

与先前的答复相反，假如我们承认，“水”一词在两个星球上意指不同的东西。再让我们承认伯奇的主张：“关节炎”不仅在两个语言共同体中而且在两个奥斯卡说出该词时都意指不同的东西。我们的问题是：语词意义上的差异与心理内容有什么必然关系？对上两种情况的一个相当自然的内在主义的回答是：它们所说明的充其量是：在心理状态与我们用于一般和特指的心理状态时的语词之间也许存在着一种比我们以前所实现的关系更微妙的关系。于是它们理所当然地促使我们去考察这种关系。然而它们没有向我们表明的东西是：我们心理内容（进而我们的心理状态）的个体化依赖于外在于身体的因素。

语言的确是一个社会问题，而心理状态则是一种私人的、个人

862 第六部分

的问题;它们似乎很难匹配,至少难以充分匹配,因此如果不引进任何额外的社会装备,那么我们则无望把语词直接映射到心理状态之上。也许你的语词从不同于你自己的意图的因素中得到了其“法定”的意义(及外延);但这并不意味着它们的法定意义实际上就是你用那个词所意指的东西。借用 B. 汉南(1994)的一个例子,当 J.F. 肯尼迪发表其著名的“我是一个柏林人”演说时,他的语词对其德国听众的意义事实上是:“我是一免费的炸面饼圈”。不过,我们猜测,这不是他说那些词要传达的意思。他意谓的——以及话语后面的心理状态的内容——自然是“我是一个柏林人”。但是既然这个在语词和思想之间错误匹配的清楚例子已为某个未能完全掌握一门外语的人产生出来了,那么这个范例可以予以泛化:一个词在社会方言中所意指的东西也许不是一特定的个人用他的“个人习语”中的那个词所意指的东西。

在这里,普特南(1975)关于划分语言活动的观点是有重要意义的:语词(如“金子”和“水”)是通过共同体的专家的活动而获得其法定的意义的,进而非专业人员(他们对金子和水的实际本质也许一无所知)承袭这些意义。这里的断言不外是:这些社会习语的意义在很大程度上与个人习语的意义是无关的,因而与心理状态个体化也没有关系。

这类思考显然为 B. 洛尔(1985)所发展。他提出,关于心理内容的许多讨论,包括普特南和伯奇的那些讨论,是这样进行的:先提出关于内容的某个论点,接着对之提出质疑,其根据是该论点未能像对“that 从句”的正确归因那样将内容个体化。即是说,该论点对于内容在不同的情况下是否相同所作出的判断与我们所用

的 that 从句不统一。(that 从句是由 that 一词引导的句子,如“that 天要下雨”,它们是我们对思想内容作出归因的标准方法,就像在“汤姆认为天要下雨”中那样。)例如在孪生地球的例子中,内在主义受到了挑战,因为它主张,两个孪生人有同样的思想内容,即使(据论证)正确的 that 从句各不相同:汤姆认为 H_2O 是美的,而孪生地球的汤姆则认为 XYZ 是美的。不过,正如洛尔所强调的那样,这样的论证战略依赖于下述假定,即正确的 that 从句的归因之同一(或差异)隐含了心理内容(进而心理状态)的同一(或差异)——即是说,公共语言和私人的心理状态是紧密地交织在一起的,就像我们上面分析的那样。接着他提出了一些思想实验和论证以对这种假定提出挑战。我们将考察其中的一个。

洛尔对克里普克(1979)的例子作了改变,他要求我们设想一个名叫皮埃尔的人。皮埃尔生长在法国,只会讲法语,并最终相信(即是说基于他的保姆所讲的故事)有一个被称之为“伦敦”(Londres)的美丽城市。他用“伦敦是美丽的”之类的语词表达了这个信念。不久,他被带往伦敦(London)生活,并没有意识到这个城市与他过去听说的“伦敦”是相同的。结果他相信:伦敦(London)也是一个美丽的城市,这是他用“伦敦是美丽的”之类的词语所表达的信念。现在的问题是:这个“新”的信念与他儿时得到的信念是同一个信念吗?换言之,这些信念具有相同的内容吗?

洛尔论证说:这些是不同的信念。(确实可以断定,说皮埃尔最终相信:除了伦敦(Londres)以外,伦敦(London)也是一个美丽的城市,这是多么自然的事情。)例如它们非常不同地与皮埃尔的其他信念发生相互作用。假设,皮埃尔在法国曾相信:如果他在伦

864 第六部分

敦住过,那么他住过的那个城市与奥斯卡·瓦尔德曾住过的城市是同一个城市。然而基于他的“我住在伦敦”这一信念,他不会得出结论说:他现在就住在奥斯卡·瓦尔德曾住过的一个美丽的城市里。这一来,在皮埃尔的心目中,“伦敦”(Londres)与奥斯卡·瓦尔德有联系,而“伦敦”(London)则相反。我们可能会说,“London”和“Londres”对他有非常不同的概念作用,一个清楚的标志就是,包含它们的信念由于具有不同的内容因而是不同的信念。

即使你不相信概念作用由内容所决定,但你仍必须解释:同一个信念怎么可能享有两种如此不同的概念作用。(福多 1990a 实际上提出了一种解释方法,就是通过述及信念的不同“媒介”而予以解释。关于解决这一争论的替代性方案,也可参看下面对外在主义的最后答复。)

因此假设它们是不同的信念。洛尔论证说,有意义的事情是:信念归因的常见原则用的是关于它们的相同的 that 从句。因此把皮埃尔儿时的信念明确地描述为相信伦敦很美这一信念,就是相当自然的事情。(为了支持这一主张,如果你愿意的话,可假设:他的保姆曾在伦敦住过很多年,因此基于她自己的实际经验向他讲了许多真实的事情;而且她还向他展示了不同建筑物和纪念碑的图片。为了使这些东西合理化,再假设:当她想起她曾长时间忘记了的那些故事和图片时,仍保留着“Londres”的信念。)以为皮埃尔成年的信念不同于伦敦很美这一信念,这可能是不正常的。因此洛尔断言:我们有这样的情况,在那里,同一个 that 从句被用来对两个不同的信念进行归因。(接着,他进一步探讨了这样的情况,在那里,不同的 that 从句被用来归因同一个信念;请记住前几

段所述的内在主义对孪生地球案例的答复。)

启示何在? that 从句不可能是追溯心理内容的基础。正如洛尔进一步的分析所显示的那样, that 从句归因渗透着与个体心理状态几乎无关的各种各样的社会事务。这一来, 我们便进到了对外在主义的第四种答复的核心: 我们可以承认: 两个孪生人的语词有不同的法定意义, 进而也可同意说, 不同的 that 从句(对不同的社会习语至关重要)可应用于两个孪生人, 而用不着假定他们在心理状态方面彼此有别。我们用来描述他们的心理状态的那些语词可能是共同体所(或者相反, 外在地)敏感的, 换言之, 这并不意味着被描述的状态本身是如此敏感的。

这便把我们带到了我们对外在主义的最后答复, 对此, 我们可以非常简略。

我们可以同意说(与先前的答复相反), 法定的语词的意义对心理状态的个体化是关键, 但同时又否定, 与一开始就假定的相反, 心理状态是通过它们的内容而个体化的。因此这一转变认可的是: 同一心理状态与不同的内容可能有联系。如果是这样, 内在主义者便可假定说: 物理上同一的孪生人有不同的心理内容, 但同时又否定这样的主张, 即孪生人处在不同的心理状态之中。进而这便允许内在主义者坚持来自于上述第三个答复的原则(A)和(B), 同时又用不着修改; 它允许她假定(根据第四个答复) that 从句追溯的是心理内容, 而同时又否定: 皮埃尔的 Londres 和 London 信念是相同的心理状态。总之, 第五个答复承认外在主义思想实验的主要命题, 而不赞成他们的最有争议的内涵。

不过, 这一对外在主义的答复是极其复杂的, 具有其独特的、

我们这里不能予以讨论的困难。因此基于这一不完全的注解,我们便结束我们关于内在主义与外在主义争论的主要探讨。

注 释

① 自然,并不是每个哲学家都赞成 LOT 假说。而不赞成该假说的哲学家也就不会用那些术语去构想表征问题。为了与之调和,我们在下面的讨论中,有时在不同的构想方法之间转换,如有时集中在 LOT 术语的意义上,有时关注于自然语言语词的意义。

② “内在的”属性就是一对象在没有别的任何东西存在的情况下而能够具有的属性。例如即使世界上没有任何别的东西存在,猫子在原则上也有其长度。试比较“位于弗雷德的左边”这一属性。这是一种“外在的”属性,因为任何东西要具有它,那么另外的某物如弗雷德就必须存在。

③ 从这里开始,我们在引证时将用词语(如“奶牛”)指称表征本身。即使必要的时候将用“奶牛”来指英语单词,但一般应理解为指的是 LOT 语词。加着重号的词(如奶牛)有时将指概念(如关于牛的概念),有时指属性(如成为一奶牛的属性);当必须明确时,我们将尽力。

④ 不过我们应注意,这个答复被简化了,绕过了难题。例如,有这样一个严峻的问题,即如此之类的知觉规律(连同其反事实)对于严格的感觉概念(如红)之外的许多或大多数概念是否存在。即使它们存在,那些规律是否从根本上满足我们的自然主义约束,仍是有疑义的,因为它们正确的运作以推理这样的心理能力为前提条件。(试比较,卡明斯,1989,pp. 62-66。而且这类规律的存在似乎只会加剧深度问题。更详细的细节,可参看福多,1987,1990c)

⑤ 德雷特斯克(1981,pp. 162-165)为解决深度问题用了同样的方案。关于知觉恒常性的更多的内容,请参阅关于知觉心理学的任何教材,如霍克堡 1978,pp. 72ff。

⑥ 我们必须强调:关于该课题还有更多要讨论的问题。即使假定任何个别的奶牛形的“奶牛”-网膜印象联系并不比“奶牛”-奶牛联系更牢靠,但网膜印象的析取(即印象 1 或印象 2 或印象 3 或……)又怎样呢?当这里所说的中介是“奶牛的外表”而不是奶牛形的网膜印象时,福多的方案会有作用

吗?关于进一步的讨论,参阅福多 1990c, pp. 109 – 110, 斯特尼, 1990, pp. 120 – 121。

⑦ 我们这里所用的“水”,指的是英语单词。应用于心灵语言就相当简单了:只要把水这一概念理解为个体用英语单词“水”表示的东西就行了。(详见最后一节)还请记住我们先前的讨论:即使完全有效的因果规律实际上并未被“应用过”,但在适当的地方仍可能存在着这样的规律。

⑧ 我们应注意到:福多理论的全面阐释要解决的就是这个问题,方法是把标记的实际历史作为它们的意义的组成部分、作为协变的补充。于是,既然地球人的“水”在它们的实际因果历史中只有水,因此“水”也只意指水;类比的思考允许“twater”(孪生地球水)只意指 XYZ。然而,这种变换超越了纯粹的协变理论,因此福多称之为“混合型的理论”。在下一节中,我们更详细地考察了“实际的标记史”的作用。

⑨ 我们只讨论了由作为不可区分的类别的项目构成的案例,但卡明斯的分析事实上可应用于所有错误表征的案例。例如当作为一个类别的马与奶牛不是不可区分的时,它也许就是环境 C 中的马,即引起错误的“奶牛”标记的那些东西可能就是如此。

⑩ 人们自然可能会否定直觉。这样做所引起的争论之细节,请参看下节的“内在主义与外在主义”。

⑪ 关于下述观点的进一步发展,即对行为的(心理学)解释必须诉诸于头脑之外的因素——就目前的例子而言,诉诸于与奶牛形的网膜印象相对应的奶牛——可参看斯特尼,1990,第 5 章,及 p. 126,伯奇,1986。即使我们没有集中考察该观点本身,但某些与之有关的争论在下节“内在主义与外在主义”中还将被探讨。

⑫ 本节大量的材料是从佩辛 1995 中选取的,经哲学文献提供中心、《哲学研究杂志》出版者允许而重印的。

⑬ 这里被理解的概念作用就是整个句子(或整个思想或信念)在原来意义上所独有的某东西,因为推理就是我们用句子所做的某事。不过,我们可以把一语词(或一心灵语言术语)的概念作用理解为它对句子所起的概念作用。

⑭ 事实上,布洛克 1986 关于这一点的观点是完全不清楚的。一方面,

868 第六部分

他明确放弃了我们刚讨论过的关于分析—综合的区分(因此似乎是一个概念整体论者);另一方面,他在适当地地方所提出的又能被理解为一种与语境有关联的、分析—综合区分(就此而言,他可能不是一个概念整体论者。参阅布洛克,1986,p.629,福多和莱波尔,1992,p.184)。

⑮ 本书是对各种整体论的主要论证的一种批评,概念整体论是其必要的组成部分。即使我们强调了福多和莱波尔对各种整体论所提出的某些一般问题(不是他们对布洛尔的 CRS 的特殊的批评),但我们是按这些问题特殊地相关于概念整体论的方式予以探讨的。

⑯ 总注:我们大概可以说,整体论事实上依赖于在构想的差异与概念的差异之间作出的所期望的区分的意义。(因此尽管两个行动者有不同的信念,但如果在某些有限的语境下,“老虎”一词有关于它们相同的外延即指称,那么他们仍不会被认为有关于老虎的不同的概念,而只会被认为有不同的构想。)我们并未进一步阐发这个观点,而它可以用来回答下面的大多数问题,并确实贯穿于我们的整个讨论之中。

⑰ 在本节的大多数内容中,我们假定:意向状态是因它们的内容而个体化的;因此说两个个体处在具有相同内容的信念状态就等于说他们处在相同的信念状态,反之亦然。(不过在本节末尾,我们对此假定简要地提出了质疑。)这一来,将思想个体化的任何方法将也是把思想内容个体化的方法。

⑱ 两个思想要有相同的内容,它们就必须是由相同的概念构成的。为了表达的方便,下述大多数讨论将集中在自然语言词项的使用上。不过,与概念的联系是直接的:把概念当作是语词的意义,当作行动者心中所想的或用语词所表达的东西。所争论的问题可以如是被突出:“两个孪生人在说‘水很美’时意指的是相同的東西吗?”(在后面,我们要讨论语词的意义与指称或外延之间的区别。)

⑲ 试回忆一下我们上面关于“虚假信念与指称”的讨论。

⑳ 我们这里省去了一些源自于普特南对“语言活动之区分”的概念使用的重要而难解的方面。(试比较普特南 1975,泽马齐 1976 的更详细的讨论;还可看下面第四个答复。)

㉑ A 是由这样的观点引发的,即意向内容必然是关于性;说两个词项有不同的外延就是说它们关于某种不同的东西,从而在内容上相互有别。(或

在意义上。再者,我们大致在同义的意义上使用“意义”与“内容”:心灵语言词项的意义也即它的内容。)当我们下面集中在词项及其外延时,论证也可类比式地应用于整个思想(由词项所构成)及其真值条件;我们有时在两个范围之间转换。最后,我们应当注意:我们用我们的表述对福多的分析稍微作了一点修改。

② 即使福多对狭义内容有兴趣,但不要忘记:他拒绝概念作用语义学本身,他论证说:具有自然法则普遍效力的协变(整个概念作用)取决于狭义的内容。(参看上面关于协变的一节。这个观点,包括证实主义的观点,自然有问题。)从此,福多的观点围绕这些问题经历了重大的发展。

参考文献

- R. 卡明斯(1989):《意义与心理表征》,麻省理工学院出版社 1989。
- H. 普特南(1981):“缸中之脑”,载《理性、真理与历史》,剑桥大学出版社。
- (1975):“‘意义’的意义”,(1975),载 A. 佩辛与 S. 戈德堡编:《孪生地球编年史》,纽约:M.E. 夏普,1996。
- J. 塞尔(1983):《意向性》,剑桥大学出版社。
- D. 丹尼特(1981):“真正的相信者”,见《意向观》,麻省理工学院出版社 1987。
- J. 福多(1975):《思想的语言》,哈佛大学出版社。
- (1984):“重估观察”,《科学哲学》,1984,51:23-43。
- (1987):《心理语义学》,麻省理工学院出版社。
- (1990):《内容理论及其他论文》,麻省理工学院出版社。
- J. 福多与 E. 莱波尔(1992):《整体论》,牛津:巴兹尔·布莱克韦尔。
- K. 斯特尼(1990):《心灵的表征理论导论》,牛津:巴兹尔·布莱克韦尔。
- F. 德雷特斯克(1981):《知识与信息流》,麻省理工学院出版社。
- R. 米利肯(1986):“无规律的思想:无内容的认知科学”,《哲学评论》,95:47-80。
- W. 奎因(1950):“经验主义的两个教条”,载《从逻辑的观点看》,纽约,1961。

870 第六部分

N. 布洛克(1986):“关于心理学的语义学广告,”载 P. 弗伦奇等编:《中西部哲学研究》,第 10 卷,明尼苏达大学出版社。

T. 伯奇(1979):“个人主义与心理的东西”,载 P. 弗伦奇等编:《中西部哲学研究》,第 4 卷,明尼苏达大学出版社。

——(1986):“个人主义与心理学”,《哲学评论》,95,1:3-45。

E. 泽马齐(1976):“普特南关于指称物质名称的理论”,重印于 A. 佩辛和 S. 戈德堡(1996)。

J. 霍克堡(1978):《知觉》,第 2 版,新泽西:普伦蒂斯-霍尔。

A. 佩辛(1995):“概念整体论之辩护:对福多和莱波尔的答复”,《心理学研究杂志》,20:269-280。

第七部分

他 心 问 题

类 推*

伯特兰·罗素

迄今为止,我们所考察的这些公设对于关于物理世界的知识来说都是必要的。广义地说,它们使我们在一定程度上承认了关于物理世界的时空结构的知识,而让我们以完全不可知论者的态度对待它的质的特征。但关于他人,我们觉得我们所知道的不止这些;我们确信,他人在质上也有与我们极为相似思想和感情。我们并不满足于认为,我们只知道我们朋友心灵的时空结构,或者他们具有启动以我们的感觉为终端的因果链的能力。哲学家也许会声称,他只知道这个,但如果让他去冒犯他的妻子,你会看到,他并不把她只看作是个他知道其逻辑属性、而对其内在特征看都不看一眼的时空大厦。因此我们有理由推断,他的怀疑论与其说是出自本意倒不如说是职业性的。

我们所关心的是下面的问题。我们从自己身上观察到像回忆、推理、感到快乐、感到痛苦这一类事件。我们认为手杖和石头没有这些经验,而其他人却有。我们大多数人都不怀疑,高等动物能感觉快乐和痛苦,尽管有位渔夫却曾对我断言“鱼既没有感觉也

* 译自其专著《人类的知识》,第六部分,第8章,伦敦:乔治·艾伦—昂温出版公司,1948。

874 第七部分

没有情感”。我不知道他是怎样获得这一知识的。虽然大多数人也许都不同意他的说法,但他们对牡蛎和海星有感觉和会表示怀疑。不管这可能是什么样子,随着动物王国等级的下降,常识表现出的怀疑则在增加,但就人类来说,常识并没有产生任何怀疑。

显然,关于他心的信念需要物理学并不需要的某种公设,因为物理学只要有关于结构的知识就够了。我现在的目的就是要表明,进一步的公设是什么。

很明显,我们必须诉诸于某种可以比较合理地称作“类推”的东西。他人的行为在许多方面与我们自己的行为是相似的,因而我们假定,它必有相似的原因。人们所说的话是我们在有某些思想时所要说的,因此我们推论他们大概也有这些思想。他们向我们提供了我们在以后有时能予以证实的信息。他们的行为方式与我们在应当高兴(或不高兴)的外界条件下高兴(或不高兴)时的行为方式一样。我们也许会和朋友一起谈论一些我们共同经历过的事件,并发现他的回忆与我们的回忆是一致的;这一点在他忆及那些我们已经忘却但思绪又为他重新勾起的事情上更有说服力。再假设,你给你的孩子出了一道算术题,他凭运气得到了正确答案,这使你认为他有算术推理能力。总之,我对于刺激的反应与“无生命”的物质的反应在很多方面都不相同,而在所有这些方面别人却和我相似。由于我清楚地知道,那些支配我的行为的因果律一定与“思想”有关,所以我就很自然地推断,同样的因果律对于我的朋友们的类似行为也是适用的。

我们眼下所关心的推论不只是那种通过主张感觉有某种可以知晓的原因而使我们超越于唯我论之上的东西。这种能够满足物

理学要求的推论,我们已经考察过了。我们现在所关心的是一种特殊得多的推论,即与我们关于他人的思想和情感的知识有关的那种推论——假如我们有这种知识的话。很明显,这种知识或多或少是值得怀疑的。不仅有这样的一般性论证,即认为我们正在做梦,而且也有关于制作精良的自动机存在的可能性的论证。有比我们的学龄期儿子更会做计算题的计算机;有能完美地记录某某人在某某场合的话语的留声机唱片;在电影中,有这样的人,他们即使是真实的人的摹本,但本身并不存在。在事实上没有生命的地方,智谋在成功地产生生命的假相方面是没有理论上的限制的。

但是,你也许会说,在所有这些情况下,正是人类的思想产生了这些制作精良的机械装置。是的,但这一点你是怎么知道的呢?你怎么知道留声机不能“思考”呢?

首先,在可观察行为的因果律之间存在着不同。如果我对一个学生说:“给我写一篇关于笛卡尔相信物质存在的原因的论文”,如果他是一个勤奋的学生,那么我将会引起某种反应。留声机唱片可能会制造得如此之好,以至于能对这种刺激作出反应,也许比那个学生做得还好,但如果这样,他却不能告诉我关于其他任何一位哲学家的任何情况,即使我以拒绝向它授予学位相威胁。人类行为的一个显著特点就是改变对于一已知刺激的反应。一个有创造才能的人可能制造出一种自动装置,这种装置在听到他讲的笑话时总会发笑,不管这些笑话他已经听过多少遍;而一个人,在一个笑话听过几遍之后就会哈欠连天,并且最后会说:“第一次听到这个笑话时我是怎么笑的呢。”

876 第七部分

但是有生命与无生命物质的可观察行为之间的区别,并不足以证明,除了我自己的身体之外,“思想”还与别的有生命的身体相关联。借助纯物理因果律对生命躯体的行为进行理论上的说明大概是可能的,而仅根据外部观察驳斥唯物主义则大概是不可能的。如果我们真的相信,在我们之外还有思想和感情存在,那依据的一定是我们自己的思想和感情与之相关的推论,并且这种推论必然超越于物理学所需要的东西之外。

当然,我并不是在探讨我们何以相信他心存在的历史。我们发现自己在第一次开始思考时就相信它的存在;妈妈可能生气或高兴这一想法就是产生于孩提时期的一种想法。我所讨论的是建立一种公设的可能性,这一公设将在这一信念与事实之间,比如在“妈妈生气了”的信念与听到一大声叫喊之间建立一种合理的联系。

这种抽象图式似乎是下面这样的。根据对自己的观察,我们知道有一种具有“A引起B”的形式的因果律,这里A是一个思想而B是一个物理事件。然而有时候我们虽然观察到一个B,却不能观察到任何A;于是我们便推论一个未被观察到的A。例如,我知道,当我说“我渴了”的时候,我那么说,通常是因为我渴了。因此当我不渴的时候听到“我渴了”这个句子的时候,我便假定别的人渴了,尤其是当我面前站着这样一个热得受不了的、有气无力的躯体时,他不停地说:“我滴水未进,在这样的高温里,在沙漠中步行了二十英里”,更容易作出这种假定。显然,我对这种“推论”的信心随着事实的复杂程序的增加而增加,而且如果根据主观观察得出来的因果律的确定性足以说明那个事实的复杂性,那么我

的信心会随着该因果律的确定性的增加而增加。

显而易见,就多数原因都值得怀疑而言,我们所进行的这种推论是无效的。假使我们知道“A引起B”,也知道B已经发生;如果这被我们用来证明对A的推论,那么我们就必须知道只有A引起B。或者,如果我们只满足于推出A是或然的,那么只要我们知道在多数情况下正是A引起B就够了。如果你没有看见闪电却听到了雷鸣,那么你会满有把握地推断有闪电出现,因为你确信你所听到的那种声音极少是由闪电以外的东西引起的。正如这个例子所显示的,我们的原则不仅适用于证明他心的存在,而且在物理学中也常常被假定,尽管这种假定采取了一种不太具体的形式。我之所以说“不太具体的形式”,是因为未看见的闪电与看见的闪电只是抽象地相似,而我们设想的他心与我们自己的心的相似则绝不是纯粹抽象的。

如果他人的被观察到的行为的复杂性,通过像口渴这样的简单原因就能够给予充分的说明的话,那么通过减少其他原因的或然性,这种复杂性就可能增加所做推论的或然性。我认为,在理想的有利的条件下,这一论证可从形式上表述如下:

通过主观观察,我知道A引起B,其中A是一种思想或情感,B是一种身体行为,比如说一种陈述。我还知道,每当B是我自己身体的一个行为时,A就是它的原因。现在,我在我之外的身体上看到了一个B类行为,而且我此时没有A类思想或感情。但我在自我观察的基础上,仍然相信,只有A才能引起B。我据此推断,存在一个引起B的A,虽然不是我能够观察的A。以此为基础,我进一步推论,他人的身体与心灵是相联系的,这些心灵与我的心灵

878 第七部分

的相似取决于他们的身体行为与我自己的身体行为的相似。

在实践中,上述陈述的准确性和确定性一定会被弱化。因为我们不能肯定,在我们的主观经验中,A是B的唯一原因。即使在我们的经验中,A是B的唯一原因,我们又怎么能知道在我们的经验以外也是如此呢?我们没有必要确切地知道这一点,只要有很高的或然性就够了。在这些情况下,关于或然性的假定正好就是我们的公设。因此,这一公设可以表述如下:

每当我们能够观察A和B是否出现时,如果我们发现,凡有B的场合都有一个A作为其因果前件,那么大多数B都有A作为其因果前件,这一点就具有或然性,即使是在观察不能使我们确定A是否出现的情况下也是如此。

这一公设如果被接受的话,就可以为我们关于他心的推论及许多其他未加思考、仅凭常识就作出的推论提供辩护。

情感的类比*

斯图尔特·汉普希尔

1. 在本文中,我将只探讨与我们关于他心的知识有关的诸多困惑之一的一个根源。人们常说,不可能有充分的理由证明:关于他人情感和感觉的陈述是基于任何普通模式的归纳论证,即是说是基于从已观察到的东西到未观察到的东西这一熟悉的公认的推论形式。而我将论证的是,它们是可以这样予以证明的。我不否认,此类推论步履艰难。完全撇开哲学理论不说,谁都知道,这些推论是很棘手的。但我不赞成说,当把它们仅仅看作归纳论证时,它们在逻辑上是稀奇古怪或无效的。我相信,现代哲学家们已经在我们对他心的推论中发现了某种逻辑上稀奇古怪的和成问题的东西,甚至否认此类推论的可能性。之所以如此,至少部分原因是未能完全理解我们语言中的代词和其他语境表达式的功能,尤其是他们误解了这些表达式在与诸如“知道”、“肯定”、“证实”、“证据”之类的词语结合时的正确用法。如果我是对的,那么就更容易解释:唯我论者想说的东西为什么不能被正确地表达出来,为什么唯我论是语言学上荒谬的理论,同时,还不难解释:为什么它是一个吸引着那些将认识论区分与逻辑区分相混淆的人的理论。

* 译自《心灵》,LXI,241(1952,1):1-12。

2. 基于下文将阐明的理由,我拟引进两个准技术术语。我将把“我感觉头晕”、“你感觉头晕”、“他感觉头晕”以及含有动词“感觉”的其他例句当作这类句子的样本,这类句子的地位在正常运用中是有争议的。在我所创立的术语中,对于语句“我感觉头晕”的任何正常运用都将是一个自传性陈述的样本。在这里,这一短语只是一个以第一人称单数形式表达的、描述了某人的瞬间情感或感觉的陈述的速记形式。而“他感觉头晕”、“你感觉头晕”或“他们感觉头晕”这类句子的任何正常运用,都将是他传性陈述的样本——即是说描述了某人情感的、但不是以第一人称单数形式表达出来的陈述。在我看来,“我们感觉头晕”在正常使用时,则是一个半自传、半他传的陈述。有时人们可能不用第一人称单数、而用第三人称或一些假想的或其他的名字,来讲述自己的内部生活的故事;实际上,用第三人称,并放进旨在表现某人自己的情感和感受的描述,去撰写某人自己的讣告也不是不可能的。但在这种情况下,代词(或有词尾变化的语言中的动词的格)运用是错误的,并且是故意如此。单词“我”(或在有词尾变化的语言中的相应的动词的格)的一般功能是要清楚地表明,该陈述的作者同时也是该陈述的特指的主语。而前面提到的例外的、故意的错误运用则有意识地利用了这一事实。所谓“自传性陈述”,我指的是描述了某人情感和感觉的陈述,它以它的表达式的现在形式清楚地表明,该陈述的作者同时也是其特指的主语。一个陈述,例如小说中的、我们对之可能有争议的陈述,由于利用了该陈述语言形式自身之外的证据,不管它事实上是不是一个自传性陈述,因此在我的人为的、限定性的意义上就不是一种自传性的陈述。

人们常常注意到,这些关于情感和感觉的第一人称单数陈述有某些特异之处,尤其是当其主动词是现在时态时更是如此。这些特殊性已使一些哲学家将其说成是不可救药的陈述,而使另外一些哲学家干脆拒绝承认它们是“陈述”。这些特殊性出现于与这些语句结合在一起的词诸如“知道”、“相信”、“肯定”等的用法中,更准确地说,出现于它们不露面的用法中。就大多数陈述而言,“我想P是真的但也许我错了”和“我已证实P确定无疑是真的”是有正常用法的语句,不管P可能是什么。但通常没有这样的条件,在其中,人们会说“我认为我感觉头晕,但也许我错了”,或者说“我已确定无疑地证实我感觉头晕”,因此通常没有这样的条件,在其中人们可以恰当地说,“我绝对确信我感觉到头晕”。相形之下,“你感觉头晕”或“他感觉头晕”这类句子则常常出现于“我相信他感觉头晕但我不能肯定”或“众所周知他感觉头晕”等形式的陈述中。但另一方面,“他相信他感觉头晕”或“他确信他感觉头晕”则是不正常的用法。因此其必然结果是,如果作出关于瞬间情感和感觉的陈述的作者被清楚表明同时也是该陈述的特指的主语,那么当向这些作者提出诸如“你怎么知道”或“你的证据是什么”的问题时,这些问题对于那些陈述来说就是不恰当的。

根据这些事实可能作出的一个推论是,关于情感的他传性陈述绝不可能被直接知道为真。这里的“直接知道”意思是说,不存在这样的问题,即那陈述怎样认定为真,也不会有这样的问题,即支持这一陈述需要什么证据。但照现在的情形看,这是个明显虚假的结论,因为作为此种他传性陈述的特指主语的人一般都能够直接知道关于他的陈述是真还是假,而不需要任何证据。正确的

882 第七部分

结论只能是,在已述的那种意义上,此类他传性陈述的作者绝不可能直接知道他所作出的陈述是真还是假;作者总是能被恰当地问起他是如何知道、或者他根据什么相信他的传性陈述是真的;他被要求提供他的证据。因此,所谓的不对称性,就第一人称单数形式表达的陈述需要从第二人称或第三人称单数形式所表达的陈述中获取证据而言,并不是以第一人称单数形式表达的陈述本身有什么不同的问题。用第一人称单数所作的关于情感的描述和用第二、第三人称表达的关于情感的描述这两者,都可能由于间接证据的存在(如“我确信你在撒谎;你有撒谎的明显动机,你没有表现出你通常在感到头晕时所表现出的任何迹象”)或者对直接知识的恰当声明(如“我可以相当肯定地告诉你,我确实感到头晕,尽管证据是相反的”)而受到质疑。

这一点是很明显的,但当哲学家们在一般地谈论“关于他心的陈述”、进而探究适用于确证或反驳此类陈述的方法时,它却又易于受到可怕的轻视。也许有人会认为“关于他心的陈述”也就是我所说的他传性陈述,即描述了情感和感觉但不是用第一人称单数形式表达的陈述。而所谓的他心问题,有时被称作某类陈述如何可能得到检验的问题,并不附着于一类具有任何一种特殊形式的陈述。如果在听者的位置被认为取代了作者的位置这样的情况下,那么即使对于第一人称单数陈述来说,他心问题同样会出现。他心问题,准确地讲就是这样的问题,即对于任何一位事实上不是关于思想和情感的陈述的特指的主语的人来说,什么样的检验和证明是可能的。对于关于情感的任何陈述来说,不管这一陈述是以“我”、“你”还是以“我们”或者“他们”开始的,他心问题同样会发

生。

3. 对于这一如此重新表述的问题,常识性的回答似乎平淡无奇——确实平淡无奇,以至于仅仅给出这一回答可能不能令哲学家们感到满意;有必要进一步作的事情是对它为什么被认为是不充分的进行解释。这一常识性回答是:我们每个人有时是一个自传性陈述的特指主语,有时是一个他传性陈述的特指主语;我们每个人有时会作出或者说有可能作出关于情感的、不是推论性的且不需要支持证据的陈述,有时还会作出或可能作出关于情感的、推论性的且需要支持证据的陈述。为检验任何实际推论方法的有效性,所需要的不过是,我们每个人有时有可能将可疑的推论方法的结论与他所知道的、独立地适用于这里所说的那种推理方法的东西进行比较。就我们推论别人而非我们自己的情感的常见方法而言,由于我们每个人经常能够将这种推论的结果与他能直接而非推论地知道其为真的东西进行比较这一事实,因此我们每个人当然有这种可能。我们每个人,只要他是一个关于情感和感觉的陈述的特指主语,就都有作出这种检验比较的可能。我,汉普希尔在推测琼斯及史密斯的情感时,知道会被哪些标记误导,因为我清楚地注意到(尽管不可能确切表述),琼斯、史密斯和其他人在推测我的情感时一般会在哪里出差错。通过实验,我们在作为孩童时就直接学会了如何掩饰和欺骗,如何做作和抑制;正是在这一过程中,我们同时还学习了怎样识破他人的做作和抑制;我们直接弄清了掩饰的标记和场合,而且随着我们自己关于欺骗的形式和场合的直接经验的增加,我们又会不断地修订我们关于欺骗的准则。

这些就是所谓的常识的看法,乍一看,它们似乎允许我们将我

884 第七部分

们任何人所作的他传性陈述看作一个有效的归纳推论的结论。这一能用于任何特殊场合的推论方法的可靠性,原则上是可为我们每一个人检验的,如将其与我们的直接经验,即与关于这一特殊方法的成功和失败的非推论知识相对照,就可对其进行检验。而且我认为,第三个人的看法将确证第一个人的看法,这在处理这些问题时是很常见的。但是在往下深入之前,有要注意的是,哲学家们所说的类比论证是如何对待我所说的常识观点的,又是怎样将其曲解和过分简化的。“类比”有这样的一种意义,据此,通过将我正在使用的特殊推论方法与别人在讨论我的情感和感觉时对同一推论方法的其他使用相类比,我确实可以证实我所作的关于史密斯感到头晕的推论。凭借直接经验,我知道像头晕这样的感觉是如何被掩盖和暴露的。我和史密斯都 有可能检验那些我们在谈论别人时不时使用的关于头晕及类似感觉的间接推论方法的可靠性。正如哲学家们一般所表述的那样,类比论证之所以失败,仅仅是因为类比找错了地方。所需要的不是某种简单的类比,如在以我的感觉及外在表现为一方与以其他某人的外在表现以及其他某人的头晕的感觉为另一方之间的类比;必需且有效的是这样的类比,即不同的人在不同场合对同一论证方法的不同使用之间的类比。归纳论证,其可靠性应由我们每个人来检验,既可用之于“我感觉头晕”这一句子,也可用之于“你感觉头晕”、“他感觉头晕”这样的句子;它可用之于任何具有“X感觉头晕”形式的语句;任何听到或者使用此种形式的任何语句的人,任何需要检验在某一特定场合所表达的陈述的人,在其不需要此种归纳确证时,都能通过寻找与它运用的场合的类比,而得到这样的确证。对于那些

对关于情感和感觉的特殊推论方法的证明存有怀疑的任何人来说,有助于消除疑虑的类比就存在于这里所说的语句在不同场合的使用之间。因为在其中的某些场合,怀疑者,无论他是谁,都有可能非推论地知道这里所说的推论方法会导致一种正确或错误的结论。我们每个人都有可能从其自身经验中获悉,某种可推出“X感觉头晕”之类结论的推论方法一般是成功的。当然,如果我,汉普希尔,从来没有感觉过头晕,或者从未有过一点哪怕轻微的像头晕的感觉,那么当他人自传性地描述这种全然不为人所知的感觉时,我也会在一定程度上不知道他们是否说了真话。在某些极端的情况下,这种可检验性的彻底失败,因此交流的失败,确实会发生。如果是这样,那么我事实上愿意承认,我个人无法知道他人所说的是否纯属编造;我完全不知道他们在说什么。但是对于我作为其作者或者听众的、关于情感和感觉的、正常范围的陈述而言,我一般能指出这样的场合,在其之下,我是被讨论的特定陈述的主语,而他人则不得不运用这一现在仍成问题的推论方法。假如史密斯和我都怀疑对方在欺骗,并怂恿自己使用不可靠的推论方法。这还是一个可检验的、经验主义的怀疑,因为我们每个人都知道,当我们试图以这种方式行骗的时候,我们自己是怎麼做的。我们每个人都将自己的行骗策略建立在我们对他人关于我们的推论方法的观察之上。我们每人都知道,我们的不同行骗方法中有一些共有的东西,因为我们有时知道我们的骗术已经失败,因此从自己的经验中获悉,此种欺骗如何可能被识破。但对于那些表面上通人性并且敏感的生物来说,它们从未想到公开地用语词推测我们的情感,它们从未用语词承认过我们对它们的情感的成功推测,因

此要建立共同的心理语言是不可能的,因为用一种语言引导持另一语言的它们无异于在相互修正中兜圈子。对于完全沉默的人或者在推测我们时不表明自己心迹的人的情感,我们是没有充分的归纳根据予以推断的。完全不能交流和理解,这种情形是很罕见的,事实上我们一般都有可能通过与别人在谈论我们时的成功和失败的比较,消除我们对他人情感的推论方法的疑虑,这不是逻辑的事实,而仅仅是自然史的事实。

一开始就坚持下述自明之理是必不可少的,即所有关于情感和感觉的陈述,包括用第一人称单数形式表达的那种陈述,对于某些人来说都是“关于他心的陈述”,但对另一些人来说则不是“关于他心的陈述”。因为正是它们的这一特点,才使我们每个人都可以在直接经验中对自己在谈论他人的感觉时所运用的、众多的、特定的推论方法的可靠性进行检验。这一自明之理的重要性在类似的事例如“关于过去的陈述”中也能显现出来。哲学家们在创作时有时制造了混乱,好像我们能挑出一类陈述,作为“关于过去的陈述”,然后我们又能够询问此类陈述如何可能通过归纳论证被证明为真。因为有这样的¹问题,即在原则上我们如何能确证我们关于过去的推论的有效性?导致这一问题的错误与“他心”问题的情形相同。我们不可能挑出一类陈述来作为关于过去的陈述,除非我们指的只是以过去式表达的陈述。但时态,像代词和动词的格一样,可用来(在其他功能中间)将一个陈述与特定的语境或者与说出它或考虑它的场合关联起来。显然,同一个陈述,当其在不同的语境为不同的人所考虑、接受或反对的时候,可以是一个关于过去、现在或将来的陈述。同样,同一个陈述可以是他传性地作出

的,也可以是自传性地作出的。一个现在时态的陈述,即一个在这种人为的意义上的、关于现在的陈述,当其被证实或再认定的时候,可能会被再认定为一个关于过去的陈述;同样,一个关于将来的陈述,当其最终得到确证的时候,可能会被再认定为一个关于现在的陈述。正是确证这个概念,包含着对同一陈述之作出的不同语境进行比较的可能性。一般说来,它不存在于陈述自身之内,或者说不在它的表达式的语法形式之中,因为它是一个关于他心的陈述,或者说它是一个关于过去的陈述;这些是作出或考虑到这一陈述的环境的特征,亦即是部分为代词、时态和其他语境表达式所体现(而非陈述)的特征,无论这一陈述在什么时候由谁所作出、所重申或所否定。严格地讲,不可能有关于过去的、需要花费无谓的确证工作的陈述,就像不可能有一类过去的事件一样;同样,不可能有作为他心的一类心灵,或关于它们的一类陈述。语境习语如“其他”、“过去”与类-术语的这种混淆,根源于这些习语的未被注意到的双重使用,现在必须对此作出解释。

4. 经常有人提出,代词和其他语境表达式(“这”、“那”、“这儿”、“现在”等)的功能就是唯一地表示或指称某人、某物、某时、某地、某事等。斯特劳森先生为它们贴上了恰当的标签:“唯一指称表达式”(《心灵》,1950年10月);毫无疑问,代词和其他的语境表达式所用的方式之一就是这种唯一指称法——即是说,在一个特定的语境中,表示一个特定的人、物、事件。但这些语境表达式的特点就是,它们并不总是或很少用来唯一地指称或表示某个特定的人、物、事件;它们还有一种重要的被泛化了的用法,在这一用法中,它们并不涉及某一特定个体,而且可以在不涉及一切或任何特

888 第七部分

定言谈语境的情况下得到解释。以口号“现在就做”，或“绝不要把今天能做的事拖到明天”为例。在这一用法中，“现在”、“今天”、“明天”并不是唯一地指称，但在某些（而非所有）方面却有像变量一样的意义，而且可解释为：“现在，‘现在’可以指任何时刻”，或者“今天，‘今天’可以指任何一天”。再来看一个例子，“未来极不确定”：照现在来看，而且没有语境，这个句子的意思就相当模糊，可以用来作出两个截然不同的陈述，甚至两类截然不同的陈述。“将来”可以用于唯一指称的方式之中，那样的话，我们就需要知道话语的语境，以便确定哪段特定的历史被提及了，且被说成是不确定的；“将来”还可以以一种纯泛化的方式被运用——“无论在历史的哪一点上，将来总是不确定的”。这一常见的、泛化了的或者准变量的用法，在我们谈及“关于过去的陈述”、“关于月亮的另外一面的陈述”及“关于他心的陈述”时，就被转移到哲学中了。在一个论证中，当作出了从泛化的用法到唯一指称用法的转换，或者相反，但又未注意到这种转换时，这两类用法间的混淆就会发生；而这正是我们在论证关于“他心”的知识及系统阐述关于自我中心困境的过程中通常会发生的事情。怀疑唯我论的人可能不会以这种明确泛化的形式提出他的问题，但却会问：“我如何才能证实我关于你心里在想什么的推论呢？既然我对我自己作出的关于你的情感的推论没有独立的检验方法，那么这里也许有一个混乱：这里的“我”是指“我汉普希尔”吗？它是关于我汉普希尔与你史密斯特有的隔阂和对你的特有的不理解的抱怨吗？或者说“我”指的是“‘我’所指称的任何人吗？”而对于“你”，“你就是‘你’指称的任何人吗？”如果意想的是后者，并且代词是以一种泛化的方式被运用的，问题就

变成了：“既然我们之中的每个人，无论他是谁，都无法检验对其他自身之外的他人的情感所作的任何推论，那么我们中的每个人怎么可能证实对自身之外的某人的情感所作的每一次推论？”针对这一问题的泛化形式，常识回答又会提出：我们每个人，无论是谁，每次都有办法独立地检验他所用的推理方法的可靠性，尽管在需要使用某种特定的推论方法的那些场合，他自然不能独立地检验在同样场合的那个推论。当我汉普希尔在我自己的经验中检验我在谈论别人的情感时所运用的各种推论方法的可靠性时，我在这些检验的结论中所做的陈述，据推测还不是推论的结论本身；但是它们作为对我的推论方法的检验仍是有效的。除开这些混乱，唯我论问题现在可以被重新表述成：每当有人使用语句“我感觉头晕”时，就有一个人并且只有一个人有可能无需推论便直接知道所作的那个陈述是否是真实的；每当有人说“你感觉头晕”或“他感觉头晕”或“史密斯感觉头晕”时，就有一个人且只有一个人有可能无需推论便直接知道该陈述是否是真实的；每当有人说“我们俩都感到头晕”或者“他们感到头晕”时，就没有人可以无需推论便直接知道所作的那个混合陈述是否是真实的。因此，唯我论者也许能正确地说，与关于物理事件的陈述相反，关于情感的陈述的鲜明特性就是，最多只能有一个人能够正确地声称，在无需提供证据和证实的情况下，他能直接知道这类陈述是否真实。但是唯我论者的初衷是想在关于心灵的陈述之中，将关于他心的一类陈述作为值得怀疑的、有问题的陈述与被认为特许的、毋庸置疑的自传性陈述区分开来。正是这一区分是站不住脚的。

假如，在谈论我们的情感时，我们每个人都以唯我论的方式将

890 第七部分

自己限定在这样的陈述之上,即我们可以正确地声称,即使不诉诸于证据或推论方法,我们也能直接知道这些陈述是真的;我汉普希尔,将被允许说“我感觉头晕”,而你史密斯,将被允许说“我感觉头晕”;但是既然这一动词的其他格的各种用法离不开成问题的推论,那么我们将决不会被允许同意或不同意彼此的陈述,或者将自己置于一个听众的位置来讨论它们。在这些条件下,代词和动词的格就不再有别的功能,并且对谎言的所有论证和分辨都将被排除:我们的心理语言只能用来传达一系列无法讨论的表白。就这些话题所进行的一般意义上的交流将停止,因为交流必然涉及作者用以向现实或潜在的听众传达陈述的语句的用法,这一来,对于所作的任何陈述来说,这种语言的所有使用者在否定和确证时,都可能以这种方式由听者的位置转向作者的位置。再比较一下人称代词和时态的用法:因为指称大大早于或晚于说话时的事件的那些陈述,在作出时是相对地不确定的,因此可能有人会认为,只有现在时态的陈述才可当作是绝对可靠的。但如果我们不将“相同的陈述”理解为某种需要在不同语境中重新确认的东西,那么我们就要排除改正和否定陈述的最后可能性,与此同时我们还将排除所有关于它们的争论、对它们进行检验的可能性以及表述信念或怀疑的可能性;我们将因此而排除作出陈述的基本条件或观点;而我们可能已经这样做了,原因是未能认识到那些把相同的陈述与坚持该陈述的变化着的环境联系起来的手段的功能。“我有可能判断关于我自己而不是别人的情感的陈述之真实性”,这一常用的公式仅在使人误入歧途方面功不可没,因为表达式“我”和“其他”都可以以两种方式加以使用,而且常常神不知、鬼不觉地从一种用

法转换为另一种用法;正是这种转换引发了唯我论的结论——即只能确定地知道一个心灵的存在,只能确定地知道一组情感和感觉的发生。当然,如果该论证正确地予以陈述,那么从这一论证就不会得出这种关于一个心灵的结论。真正的自明之理是:“我们之中无论谁都不需要推论就能确信关于他自己的情感的陈述的真实性,但他绝不可能直接地、无需诉诸证据就能确信关于情感的陈述的真实性,他绝不可能直接地、无需诉诸证据就能确信关于别人的情感的陈述的真实性;”这一自明之理如果以这种形式、用一个准变量表达式作为主语予以陈述,那么就不可能成为任何唯我论结论的前提。

5.“知道”一词及其同源词的特性——它们与任何类型的陈述一起正确使用的条件,随话语的指示语境的变化而变化——并不局限于关于心灵和情感的谈论。这种特性充斥于“知道”、“确信”、“证实”这类语词的全部应用范围。而这些语词又是与任何类型的陈述结合在一起的。无论讨论的是什么课题,对知识或确定性的断言是否恰当,都一定取决于在什么时间、什么条件下是谁作出了这一断言。它绝不可能只是这个陈述自身或其话题的形式问题。任何经验陈述不管它是什么样的,都是它的用法或考虑的某些条件之下的一种不确定的推论。一个陈述,对某人来说也许是直接和确定的知识,而对另一个人来说却将永远是个不确定的推论,这一事实中无任何神秘可言;同样,同一陈述,在某一时刻可能被确定地认为是真的,而在另一时刻则一定是个不确定的推论,这一事实中也没有神秘可言。哲学家们(柏拉图、笛卡尔、罗素)通过创作已经发明了这一神秘性,仿佛所知的真和所具有的不确定性是陈

述的固有属性,这些属性以某种方式依附于那些陈述,独立于它们在其中被作出或被考虑的特定环境。那些研究句式转换模式的形式逻辑学家们,漠视那些把陈述与话语的语境关联起来的陈述的特征,这是正确且必要的。但是哲学家们关于用法与意义的问题以下述不同的语境为转移,在这些语境中,“知道”和“确信”之类的语词可以与具有不同形式和不同话题的语句同时出现。

6. 结论。“过去”、“现在”、“其他”不是类术语而是语境术语,而且不可能有作为过去事件的事件类别,不可能有作为他心的心灵类别,也不可能有作为关于它们的任何一个的陈述的陈述类别。“关于他心的陈述”,要么是个不完整的表达式,它要能为人所理解,就必须有关于它运用的特定环境的知识——如“我汉普希尔的心灵之外的心灵”;要么这一语境表达式可以在被泛化的意义上被运用,意指的是“关于作者之外的心灵的陈述,不管那作者是谁”。如果是后者,那么在提出关于他心的问题的时候,我们就是在探讨一种类比,这种类比可以使任何人将他在其中不求助于推论便知道的一个关于感觉的陈述为真的情境,与他在其中不知道其为真的情境相比较。而且当我们谈论检验任何实际推论的方法的可靠性时,所指的正是这一比较。

关于他心的知识*

诺尔曼·马尔科姆

I

我相信,关于他心存在的类比论证至今仍享有多于它所应有的声誉,因此本文的第一个目的就是要指出它是没有出路的。J.S. 密尔是接受这一论证的许多人中的一个,我将以他在这方面的陈述为代表。他向自己提出了这样的问题:“我根据什么证据知道,或者我基于何种考虑竟至相信,存在其他有知觉能力的生物;我所看到或听到的那些正在行走或讲话的人有思想和情感,或者换句话说,有心灵?”他的回答是这样的:

我之所以断言,他人也像我一样拥有情感,首先是因为,他们像我一样拥有身体,就我自己而言,我知道,我的身体是拥有情感的前提条件;其次是因为,他们表现出动作以及其他外在的标记,就我自己而言,我根据经验知道,它们是由情感引起的。我意识到,在我自身中有一个由统一次序所联结的事实系列。在该系列中,我体内的变化是开端,中间是情感,终端是外向行为。就他人的情况来说,我有关于这个系列首尾两

* 译自《哲学杂志》,LV,23(1958,11):869-78

894 第七部分

个环节的直接的证据,但没有关于中间环节的证据。然而我却发现,其他人的首尾两个环节之间的次序如同我自身中的次序一样是有规律的、恒常的。就我自己来说,我知道,第一个环节通过中间环节产生了终端环节,没有中间环节就不能产生终端环节。因此经验使我得出必然有中间环节的结论;这个中间环节在他人身上和在我身上或者相同,或者不相同。我必须或者相信他们是有生命的,或者相信他们是自动机。由于相信他们是有生命的,即由于设想这一环节与我对之有经验的本质是相同的,而且这一本质在所有其他方面也相似,因此我就将作为现象的别人归入下述同样的普遍原则之下,这些原则是我通过经验所知道的关于我自身存在的真实理论。^①

我将避开下述可能的反对意见,它认为,上述推理是一非常弱的归纳推理,即使它是基于对单个事例的观察。因为更饶有兴味的是下面一点:即假设这一推理能得出“那个人类形体(指的是其他某人而非自己)很可能有思想和情感”之类的结论,于是便有这样的问题,即这个结论对作出此推理的哲学家来说是否有意义,因为存在着“那个人类形体有思想和情感”这个语句对他来说是否有意义的问题。这一点何以会成为一个问题呢?这是因为,密尔由之出发的假定是,他缺乏对另一个“正在行走或讲话的形体”是否有思想和情感进行判定的标准。如果他有这个标准,他就可以运用它,有根据地肯定这个或者那个形体有或者没有情感(因为唯一可能的标准就在于我们能够公开看到的行为和环境),而用不着求助于充其量只能导致或然性的脆弱的类比推理。如果密尔没有对

自身之外存在的情感进行判定的标准,那么在这一意义上,他就不能理解“那个人类形体有情感”这一语句,因此他也就不能理解“那个人很可能有情感”这一语句。

作出下面的回答是司空见惯的倾向:“尽管我没有证实的标准,但我仍然能够理解,比如说‘他疼痛’这样的句子,因为我理解‘我疼痛’的意义,因此‘他疼痛’的意思就是他感受到了和我在疼痛时所感受到的相同的东西”。但这只是徒劳的花招。因为如果我不知道怎样认定某人^①有疼痛,那么我也就不知道怎样认定他有和我在疼痛时相同的东西。^②你不能指望依靠“相同”这一概念来增进我对“他疼痛”的理解,除非你能给我一个说某人具有与我相同的东西的标准。如果你能做到这一点,你就不需要类比论证;如果你不能做到这一点,那你就不能理解那一论证的假定的结论。我以为,自称信赖类比论证的哲学家都无法逃避这一两难困境。

为了修补类比论证,人们进行了多种尝试。S. 汉普希尔先生辩解说,^③作为一种推理方法,其有效性可以用这样的方式予以证明:他人有时能根据我的行为推论我感到头晕。汉普希尔说,我则有关于我本人情感的直接的、非推论性的知识。这样我就能根据事实,对别人作出的关于我的推论进行检验,进而检验这种推论“方法”的准确性。

检验任何事实推论方法的有效性所必需的不过是,我们每个人都有时有可能用他知道独立地适用于这里所说的推理方法的东西去检验可疑的推理方法的结论。由于我们每个人常常能够把这类推论的结果与他直接而非推论地知道为真的东西加以比较这一事实,就我们关于自己之外的他人的情感的常见

896 第七部分

的推论方法而言,我们每个人确实有这种可能:只要我们每个人是一个关于情感和感觉陈述的特指的主体,他就有可能作出这种检验性的比较。由于我汉普希尔清楚地注意到(尽管可能未予阐述),琼斯、史密斯以及其他人在推论我的情感时通常在什么地方出差错,因此,我知道,在推论琼斯和史密斯的情感时,由于什么样的标记我才可能犯错误。(pp.4-5)

据推测,我也能指出别人关于我的情感的推论什么时候不会出错。在确定了一些推理程序的可靠性之后,我自己就可以以一种谨慎的方式,带着对关于他人情感的结论的真实性的适度且合理的信心,运用这些程序,以得出那些结论。

我的第一个评论就是,汉普希尔明显地忘记了类比论证的目的,这一目的就是为“我所看到和听到的那个正在行走和说话的形体有感觉和思想”提供某种或然性(密尔)。因为他所描述的推理涉及到他人确实有思想和感觉这一假定:因为据假定他们根据对我的行为的观察作出了关于我的推论。但是关于他心存在的哲学问题是这样的问题,即自己之外的人类形体除了别的之外是否确实作出了观察、推论和断言。汉普希尔所设想的对类比论证的辩护是一种诡辩论证。

如果我们从汉普希尔所描述的推理中取消关于他人思想和感觉的全部假定,我们能得到的大概就是这样的东西:“如果我有如此这般的行为,我旁边的人就会说出‘他感觉头晕’的话。而一般说来,我当时确实感到头晕。因此当别人表现出同样的行为并且我说‘他感觉头晕’时,他很可能也确实感到了头晕”。但这里对来自他人身体的似句子的声音的提及却是无关紧要的,因为我必须

假定,那些声音表达的并不是推论。这一来,上述推理就完全变成了古典的类比论证:“当我有如此这般的行为时,我感到头晕,因此当别人有相同的行为时,他很可能也有相同的感受。”这一论证又陷入了关于相同的标准的二难困境之中。

关于类比推理,H.H. 普赖斯(1938)^④所提供的说法更有意思。他表明:“一个人关于他心存在的证据最初是来自其对语言的理解”。(p.429)他的意思是,如果别的物体发出像“有一辆汽车”这样的能为某人所理解的声音,并且这些声音给予他以新的信息,那么这就“为发出这声音的别的物体由一个像某人一样的心灵所驱动提供了某种证据……。假如我经常在它的附近,它不断地说出我能够理解并在我自己身上得到验证的话语。再假设这种情况发生在多种不同的条件下。因此我认为,我相信这一物体由一个像我的心灵一样的心灵所驱使的证据将变得非常强而有力”(p.430)。发出这些能传递信息的声音的物体不一定是人的身体。“如果橡树的树叶发出的沙沙声构成了向我传递新信息的可理解的语词,如果荆豆丛表现出了可理解的姿势,那么我就有证据说,橡树和荆豆丛为像我的智能一样的智能所驱使”(p.436)。即使这些能传递信息的可理解的声音不是由某个物体所发出的,它们仍能为(脱离身体的)心灵的存在提供证据(p.435)。

普赖斯所表述的这种推理虽然与古典的类比论证有很大的不同,但在形式上仍然是类比推理:通过内省我知道,当某种声音组合由我发出时,它们都是“自发的思维活动的符号”;因此当相似的声音组合不是由我发出时,它们也“很可能起着自发的思维活动的工具的作用,当然在这种情况下,它不是我本人的活动”(p.446)。

898 第七部分

普赖斯说,这一推理也可对其他一些我能理解但并未产生的神秘声音提供解释。他预料,可能有这样的驳难,即这一假说由于无法验证因而是无意义的。他说,“在我清楚地知道:如果这一假说是真的,那么世界将必定是什么样子——它里面一定有哪些实体、它里面一定发生了哪些事件——的意义上,这是一个完全可信的假说。通过内省我知道,思想和知觉活动是什么,我还知道,这些活动组合成单一心灵的统一体意味着什么……”(pp.446-447)。

我要提出的反普赖斯的观点是,橡树或厨房的桌子发出的任何可理解的声响,绝不意味着它们有感觉和思想。应提出的问题是:什么能说明树和桌子理解了它发出的声音?我们可以设想,有用的警告、真实的描述和预言、甚至对问题的“回答”,都出自于树,以至于它竟成了其主人的无价之宝。我们怎样认定它理解了那些语句呢?我们应该“问问”它吗?假如树“说”,它附近有个雌狐,我们“问”它,“什么是雌狐?”它“回答说”,“雌狐就是雌性的狐狸”。对于“雌性”和“狐狸”,它也许还能继续解答下去。这种表现可能使我们说,与它回答“我不知道”或干脆不回答这一可能情况不同,这棵树理解了这些词。但这能说明,在人能理解这些词的同样的意义上,那棵树理解了它们吗?对人来说,这一表现将导出一个假定——他能正确运用所说的这些词,但对树来说,则不行。为了弄清这一点,让我们想一想我们一般是怎么教小孩子识字的(如“勺子”、“狗”、“红的”),又是如何判断他是否理解了它们。在教学的最初阶段,人们不需要或不指望使用定义,而宁愿让他从蓝中挑出红、从猫中找出狗、从叉子中区分出勺子。这包括他的看、指、伸手去抓以及走近正确而非错误的东西等动作。小孩说“红”的时候,

将红色的东西放在他面前,说“蓝”的时候将蓝色的东西放在他面前,这说明只有与看、指、试拿、取、运等其他动作一起,才能掌握那些词。试想,他说出的是正确的词,却看着并伸手去拿错误的东西,我们还能说他掌握了那些词的用法吗?不,不能。词和行为之间的不一致将使我们说:他并不理解这些词。就树而言,在它的词与“行为”之间根本就不存在不一致性,因为在逻辑上它不可能有相应的行为。

由于没有像人的面孔和人的躯体之类的东西,因此说树或计算机正在看或指或取某物就没有什么意义。(当然人们总是可以为这些表达式发明一种意义。)进而说它理解或不理解以上这些词也是没有意义的。树和电脑既不能通过也不能不通过小孩所接受的那些测试。它们甚至根本就不能参加这些测试。一个物体是可理解的声响或其他信号(不管它们如何有序)的一个来源这一事实本身,并不足以证明它有思想或感觉。有信息内容的语句和有价值的预言如何出自荆豆丛,可能是一个严肃的科学问题,但绝不能这样解释,即荆豆丛有心灵。这不是解释而是胡说八道。

也许有人会认为,只有那些词,其意义有“知觉内容”才会有上述困难,比如,如果我们设想,我们的荆豆丛说出的只是纯数学命题,那么我们就有理由把思想归之于它,即使它没有感觉。但假如有一个不同寻常的“会计算的孩子”,虽然能给出算术问题的正确答案,却不能将数字运用于经验命题中的实在,也就是说,他不会数任何东西,那么我相信,任何人都不愿意说他理解了他说出的数学符号和真理的意义。而如果他能用正常方式数数,那么就不会有这种不情愿了。“用正常方式数数”涉及到看、指、接近、取等动

900 第七部分

作。也就是说,它需要人的面孔、人的躯体及人的行为——或者类似的东西。那些不具备人的形式的或没有像人的形式的东西,不仅不符合,而且不能符合思维的标准。这里,我只是想表达维特根斯坦下面这句话的部分意思:“我们只能说人类以及像人类的东西在思维”(《哲学研究》,§ 360),“人的身体是人的灵魂的最好图像”(同上,p. 178)。

到目前为止,我尚未论及类比论证的最根本错误。无论是在这一论证的古典形式(我的身体和他人身体之间的类比),还是在普赖斯的形式(我的语言与其他东西所发出的声音或信号之间的类比)中,这一错误均存在。人们根据自身的情况得知什么是思维、感觉、知觉,这是一个错误的假定。普赖斯在说“通过内省我知道了什么是思维和知觉活动……”时,给出了这一假定的表述(同前,p. 447)。这是哲学家所能作出的最自然的假定,而且起初似乎是唯一可能的选择。然而维特根斯坦已经使我们明白,这首先导致唯我论,其次导致无意义。下面我将尽可能简要地说明它是如何产生这些结果的。

相信人必须“从自我出发”才能知道什么是思维、恐惧或者疼痛的哲学家,却不相信应予观察的是人的行为,而宁愿相信某种“内在的”东西。他认为,行为是与内在状态、事件相联系的,而后者不过是伴随物,或可能是一种结果。他不能将行为当作心理现象的标准;因为他如果这样做的话,他就没有运用类比论证(如前所述),赋予“自身情况”的优先权也将没有意义。他相信,他在自身中注意到了某东西,并将其称作“思维”或“恐惧”或“疼痛”,然后设法推论说,别人身上也有同样的东西。接下来,他应予解决的问

题便是：关于他人身上相同的东西，他的标准是什么。而这一点他无法做到，因为拒绝将环境和行为作为判断他人心理现象的标准正是其观点的实质。除此之外还有什么可以作为标准呢？因此他应该作出的结论便是，关于他人思维、恐惧或疼痛的概念从一种重要的意义上来说是没有意义的。他不知道什么对其有利或者不利。^⑤他应该说：“在我之外还有思维和疼痛存在是不可理解的”。这是一种严格的唯我论，也是下述假定的一个恰当的结果，这假定是，人们只能从自我出发才能知道什么是心理现象。这其实等于说：“当我说‘我疼痛’时，这里的‘疼痛’指的是某种内部状态。当我说‘他疼痛’时，‘疼痛’指的是一种行为。我不能像我把疼痛归之于我自己那样把疼痛归之于别人。”

维特根斯坦之前的一些哲学家可能是由于“从自我出发”才导致唯我论的。但我相信他是说明了那一出发点是如何毁灭那出发点自身的第一个人。这一说明可以表述如下：人们设想，一个人在内部把某东西当作是思维或疼痛，后来一当它出现于灵魂中，便把它分辨出来了。但问题接踵而至：他作出了正确的分辨吗？“私人”分辨的支持者无言以对。他确信，他对灵魂中出现的事件的分辨是正确的；但确信并不能保证它是正确的。更有甚者，他并不清楚正确意味着什么。他不知道如何将事实上作了正确的分辨与印象中觉得作了正确分辨区分开来（参阅《哲学研究》，§ 258-259）。要是他把焦虑的情绪当成了疼痛的感觉呢？无论是他还是其他任何人都可能知道这一“错误”。也许他每次都错了！也许我们每个人都错了！此刻，我们应该明白我们是在胡说八道。我们并不知道错误是什么。我们没有与我们的内部认知相比较的标准、范

902 第七部分

例、习惯做法。内部识别没有击中靶心,也不会偏离它,因为根本就没有靶心。当我们认识到关于正确和错误的观念都不适用于假定的内部分辨时,后一概念的有意义的外观便丧失殆尽。它的瓦解也就同时颠覆了唯我论和类比论证。

II

类比论证的破灭也摧毁了据说已为之提供了解决方案的问题。哲学家觉得自己陷入了关于他心的困难之中,因为他假定,他首先是“从自我出发”熟悉了心理现象的。困扰他的问题是,如何实现由自身情况到他人情况的转变。当他的思想从自身情况的优先性的幻觉中解脱出来的时候,他就能正视熟悉的事实,承认他人的环境、行为及话语实际上是他判断他们心理状态存在的标准(不只是他的证据)。而这在先前似乎是不可能的。

但此刻他又有跳到对立的行为主义极端的危险,行为主义也是错误的,因为它相信通过对某人自己的环境、行为及话语的观察可以发现他在思维或生气。“从自我出发”的哲学和行为主义,尽管在某种意义上是势不两立的,但却有一个共同的假定,那就是第一人称现在时态的心理陈述可通过自我观察予以检验。根据“从自我出发”的哲学,自我观察不能由别人检验;而根据行为主义,自我观察可以由对所有人都有效的外在标准来检验。前者已是不可理解的;而后者至少对多种心理陈述来说是错误的。我们不得不断言,第一人称心理陈述不可能(或几乎不可能)由自我观察予以检验。由此我们可以说,它们根本没有任何检验;因为如果有的

话,就只能通过自我观察。

但如果像“我头痛”或者“我想知道她在哪儿”这样的句子表达的并不是观察,那么它们表达的是什么呢?如果我关于我头痛的陈述不是一个观察报告的话,那么在它与我头痛的事实之间有一种什么样的关系呢?作为批评的结果,关于他心存在的困惑最后就变成了某人自己关于自己的心理语句的意义的困惑。在我们的出发点上,正是语句“他头痛”惹出了麻烦;而现在却是“我头痛”在困扰着我们。

这一难题由以表述的一种方式可以换成这样的问题:“一个人怎么知道何时该说‘我头痛’这样的词语?”如果想一想一个人的这样一种奇异但又不是不可能的情况,即他已长至成年却从未体验过疼痛,那么就更有可能是提出上述问题。为了改变这种状况,便让他接受各种各样的注射,后来在接受注射的某一天,他跳起来并惊叫道:“我现在感到疼痛了!”我们不禁要问:“他是怎样认识到这种新的感觉就是疼痛的?”

请注意,如果这个人给出一种答案(比如,“因为我跳起来的方式,我知道那一定是疼痛”),那么这一事实正好表明他并未掌握话语“我感到疼痛”的正确用法。它们不能被用于陈述一个结论。当他把他的做法告诉我们时,他将认识到自己的误用。因此,“他如何识别他的感觉”这一问题要求给予的是不可能的答案。提出这一问题的倾向就是我们不能把握如下事实的证据:运用这一心理语句与识别或分辨或观察自己的状态完全没有关系。

上述假想的事例产生了一种特别强烈的诱惑,那就是提出“如何”问题,这一事实表明,我们有这样的观念,即为某人第一次出现

904 第七部分

的感觉安立恰当的名称一定更加困难。这意思是说,第一次之后就不那么难了。为什么会这样?我们会这样认为吗,即这个人那时会有一种关于疼痛的范式,他可以将他的感觉与其进行比较,从而有可能立刻知道某种感觉是或不是是一种疼痛?但是这一范式要么是某种“外在”的东西(行为),要么是某种“内在”的东西(也许是感觉的记忆印象)。如果是前者,那么他就是在误用第一人称语句;如果是后者,那么他是否能正确地把当下的感觉与关于疼痛的内在范式进行比较这一问题就是没有意义的。这一来,第一人称语句的运用可由范式决定的观点就必须予以放弃。它是我们所坚持的第一人称语句以某种方式依赖于对心理状态的识别这一错误概念的翻版。

这些荒谬的言论表明,我们必须以某种完全不同的方式来设想第一人称心理语句。维特根斯坦向我们提供了这样的建议(对此哲学家们尚未给予足够的注意),即应当把第一人称语句看作是类似于心理状态的自然的、非言语的、行为的表达式。例如,“我的腿受伤了”就应看作类似于叫喊、跛行、抱腿。这是个令人困惑的比拟,而且人们对此的第一个想法就是,没有比那两种东西更不同的东西了。通过说出这个语句,一个人便可作出一个陈述;它包含着矛盾;它或者是真的或者是假的;说这话时他或者是在撒谎或者是在陈述事实;等等。确切地讲,这些都不适用于叫喊、跛行、抱腿。因此怎么会有相似呢?但维特根斯坦知道这一点,因为他故意将这一语句比作对疼痛的“原始的、自然的表达式”,并说这是“新疼痛-行为”(同上,§ 244)。尽管我自身的局限性使我无法在这里完成这一点,但我仍然认为这一类比值得探讨。因为它至少

有两个重要优点:首先,它可使我们免受“一个人怎样知道在什么时候该说‘我的腿受伤了’”这个问题的困扰,因为根据类比,这 will 同“一个人怎样知道在什么时候该叫喊、跛行、抱腿”这一问题一样荒谬;其次,他解释了这样的问题,即他人说出的第一人称心理语句对我们怎么可能是重要的,因为尽管它不是一种识别,但根据类比,却与自然行为具有同样的重要性,而自然行为又可作为我们判断他人心理状态的前一语言标准。

注 释

① J.S. 密尔:《对 W. 汉密尔顿先生哲学的考察》,第 6 版,伦敦,1889, pp. 243-244。

② “假设他有疼痛实即假设他与我具有相同的东西,这样说并不是什么解释。因为语法的那个部分我是一清二楚的,即是说,如果某人说炉子处在疼痛之中,我处在疼痛之中,那么他说的就是,炉子的经验与我的经验是相同的”(维特根斯坦,《哲学研究》,纽约,1953, § 350)。

③ S. 汉普希尔:“情感的类推,”《心灵》,I (1952), pp. 1-22。

④ H. 普赖斯:“我们关于他心存在的证据”,《哲学》,第 13 卷,1938, pp. 425-456。

⑤ 正如维特根斯坦严肃指出的那样,哲学家们一般没有得出这个结论的一个理由可能是,他们假定,“在把一个事物与其身同一起来时”,他们有“关于同一的一个不可错的范式”(《哲学研究》, § 215)。

关于他心的类比推理*

亚历克·希斯洛普

如果我们都有他心问题,如果任何心灵理论既未设法回避这个问题,又未设法提供解决方案,如果没有办法避免要命的以己之心度他人之腹,如果科学的推论没有独立地证明我们对他的信念,那么就有望复兴对他心的(混合而成的)类比推理的兴趣。不管怎样,在哲学中我们常常凑合着利用我们所拥有的东西,而且一般来说,不作抉择在哲学中注定是富有魅力的。对于许多哲学家来说,科学推论即使没有把他们从怀疑论的无底深渊中拯救出来,但无论如何把他们从他心的类比推论中解救出来了。

这一章旨在消除他们的疑虑。它将对他的心的类比推理作出辩护,以抵御主张它是糟糕的归纳推理的种种攻击。在本章中,一切都围绕着这样的事实,即被辩护的类比推论是传统的类比推论与科学的推理的混血儿。其他章节再来回答下述种种断言,即认为它在一定程度上包含着概念上的不一致性或混乱。

对于作为归纳论证的他心类比推论,存在着两种经典的驳难。第一是,对其结论作出检验,在逻辑上是不可能的。另一个是,它

* 选自作者所著的《他心》(克卢威尔学术出版公司,1995)一书的第4章。这里只选译了前半部分。后半部分是对六种批评意见的反批评,未译。

是根据一个事例(常常将其作为一孤立的例子令人痛心加以利用)所作出的概括。如果两种驳难合在一起,那么具有逻辑必然性的事情是,该论证就将是根据一个事例所作的概括。然而,概括的贫乏的基础像它所具有的必然性之贫乏一样,对类比推理的反对者来说无关紧要。理应是无关紧要的。我对该推论的辩护将从考察下述责难开始,这个责难认为,这一推理是根据一个事例所作的轻率的概括。

对这一责难的标准的答案其实就是重申:如果不把这个概括当作是基于一个事例的概括,那么它就是不可理解的。这个答复显然与 A.J. 艾耶尔的名字联在一起。我要论证的是,这个答复是没有用的。

如果“一个事例”意味着,证明的基础就只是对与物理状态相关联的心理状态的一种近因的观察,那么上述责难就会因一虚假的前提而成为完全无效的。很显然,他心的类比推论在此意义上并非真的就是基于一个事例。它是以人们在自身的情况下所观察到的、心理与物理之间的多种多样的关联为基础的。另一方面,正如标准的答复所述的那样,如果“一个事例”指的只是这样的事实,即心理与物理之间的那些关联都是从某人自己的事例中引申出来的,那么上述责难就只注意到了下面的事实,即类比推论以之为基础的样本成员有共同的属性;亦即是说,只注意到了就个人自身而言的那些关联的事实。类比论证从具有共同属性的样本开始论证,这是相当普遍的情况。例如氢像所有的元素一样,有一种富有特性的光谱。该光谱的本质自然是由在地球上进行的实验所决定的。这一来就可观察到(实际上是被记住的):从太阳而来的光发

908 第七部分

出了这种光谱(与许许多多别的光谱一道),由此可以推论:氦存在于太阳上。在这个论证中,氦的所有样本都在地球上。

不过,标准的回答是不能接受的。就某个人自身而言,心理与物理行为之间的确存在着名副其实的、各种各样的关联。不过,严格说来,他心的类比推论事实上只是根据一种事例所作的类比论证,而不是根据具有共同属性的大量事例所作的类比论证。总之,两种不同的考虑表明:他心的类比推论可以正确地看作是基于一个事例的论证。

第一个考虑针对的是这样的观点,即认为出自某人的疼痛行为的每一个例示,而不是大多数或许多例示,是由他自己的一种疼痛所伴随的,这样的看法恰恰是错误的。这意思是说:如果我们把类比推论当作是根据人们自己身上所发生的心-物关联的许多事例所进行的推论,那么能够得到的唯一的结论就是:行为状态的许多事例都与心理状态有联系。这个结论正好是我们所不需要的结论,因为它与某些人的身体相吻合,这些身体的举动就像是其本身作出的,与心理状态没有任何联系。^①我们想要的结论是:对于每一完全像某人自己行动一样行动的躯体而言,其行为状态的许多事例确实与心理状态联系在一起。因此标准的回答是没有用的。

第二个考虑来自这样的事实,即我们相信:对于每一活生生的人体而言,至少有一个人,其躯体的行为状态为那个人的心理状态所伴随。^②正如我们将看到的那样,即使标准的回答成功了,它的结论即认为别人的躯体有与之相联系的心理状态,是没法支持下述主张的,即至少有一个与每一躯体联系在一起的人。唯一能说明这一点的类比论证就是根据下述事实所作的一种概括,这事实

是:对于每个人自己的躯体而言,至少有一个人(他自己),其躯体的行为状态是由那个人的心理状态所伴随的。如此一来,这就是且必然是一种根据一种事例所作的一种归纳概括;因此如果关于他心的类比推论真的承认:对于每一活生生的躯体而言,至少存在着一个人,其心理状态对应于那躯体的行为,那么该论证就必须视之为根据一个事例所进行的类比论证。

不过,R.I. 西科拉^③认为:标准的回答是“基本正确的”。但他又断言:“当严格地加以理解时,根据类比所作的论证本身还不是对他心的论证”。在他看来,到达他心须经过两个阶段。第一阶段即真正的类比推论,它肯定能够表明:“经验与他人的躯体是有联系的”,并可满足下述最低限度的要求,即对于每个人的躯体而言,存在着有联系的经验。在这个阶段,人们只是诉诸于“行为例示与经验之间的大量的关联”。这一阶段是多例类比论证。接着人们可以借助于假说推论表明:“一个人至少与每一别的、处在能活动的生存期间的躯体联系在一起”。这第二步不是类比推论,更不是根据单个事例的类比推论。

西科拉的答复是值得密切关注的,因为它是我所知的唯一的、可阻止下述令人讨厌的武断发生的尝试,这个武断是:他心的类比推论一定是根据一个事例的概括。我要论证的是:(Ⅰ)非多例的类比推论将支持经验与每一他人的躯体有联系这一结论,(Ⅱ)正如人们至此可以清楚地料到的那样,即使它是如此,但第二步的假说如果支持了所期盼的结论,仍还需要只能在个人自己的情况下才可获得的证据,因此并未避免孤立例示的问题。

我一直主张:关于他心类比推论的任何多例阐释都不能证实:

910 第七部分

每个他人的身体与心理状态联系在一起。下面是西科拉对我的主张所阐述的支持性论证(就他自己而言,他的愿望是驳斥这一主张):

(I)既然我只看到,我的疼痛行为的大多数表现形式为疼痛所伴随,我就只能这样推论:(II)行为状态的许多例示与心理状态有联系,而且据此不能得出结论说:(III)下述常识信念是正确的,即每个以相似于我自己的方式行事的躯体都有许多与其行为状态联系在一起的心理状态例示。(p.139)

西科拉的答复似乎是赞成说(III)不能从(II)推论出来。他的策略就是主张:

在那些情况下,当我表现疼痛行为但又没有疼痛时,我的疼痛行为仍与某种别的心理状态有联系。因而人们有权推论说:对于每一以酷似某人自己的方式行动的躯体来说,它的行为状态的许多例示都与心理状态有联系。(p.139)

不幸的是,这个策略一般来说是无效的。西科拉大概考虑到了伪装。但是当睡眠时,我们在一定程度上作出了行为,而且某些睡着的人做出了具有某种复杂性的行为。因此其实我们并未意识到行为的每一有关的细节,它也没有为心理状态所伴随。再者,当我们以某种方式行事比如说驾驶一辆卡车时,存在着“进入状态”这样的现象,在这些情况下,我们可能没有意识到任何伴随着发生的心理状态。假设司机处于昏睡或麻木状态,而且对于他们苏醒之前的一段时间来说,这类情况尤其会导致司机失去控制。这仅仅只是关于常见的、缺乏对自己行为的意识的一种戏剧性的例示。一

个人在重视或全神贯注于某项任务时,不可能充分注意到:为了提供西科拉的严格的关联,他们在怎样行动。如果是这样,那么就不清楚他们持续地处在有意识之中。因此我们确实不知道为心理状态所伴随的行为的每一有关细节。

在赞成我们相信所有人的身体在某时与特定类型的心理事件有联系的基础上,西科拉进一步考察了修改过的、专门讨论疼痛的论证:

1. 当我表现出疼痛行为时,我常常感觉到疼痛。
2. 因此疼痛行为的许多例示与疼痛是有联系的。
3. 尽管如此,仍有这样的可能,即当某些人的身体表现出疼痛行为时,这种疼痛行为从未伴随着疼痛。
4. 结论:根据类比所作的多例论证并未说明:所有人的身体在某时一定伴随着疼痛。(p. 139)

正是第二个前提才是至关重要的;疼痛行为的许多例示与疼痛有联系,但并不尽然。

西科拉赞成这个论证,但又主张:

它的结论对根据类比所作的多例论证并无特别的伤害。能够确认的是,我们不能肯定,对于每个单个的人体来说,其疼痛行为的某些例示与疼痛有联系,只要该可能性足够地微弱。可能有人反驳说:一般的人都不会赞成说,这类例外甚至有极其微弱的可能性。不过,如果这最终是另一种情况,在那里,我们只能期盼盖然的确定性,即使一般的人认为(或假设认为)我们有绝对的确定性,那么哲学家们就没有必要过度地沮

912 第七部分

丧。(pp. 139 – 140)

其实不然。麻烦不在于：我们注定得思考“盖然的确定性”与“绝对确定性”可能是什么意思。那个论证的致命的弱点在于西科拉所承认的东西，即能被肯定的只能是：疼痛行为的大多数（即使是压倒的多数）例示为疼痛所伴随。假如有这种不均匀的分布，那么这就有可能为真，即使人们常常碰到的几个身体有这样的举动，好像他们处在疼痛中，其实不然。观察的材料并不能肯定下述命题是足够确定的（即使不是绝对确定的），这个命题是：每一个作出了适当的举动的躯体至少有的时候为疼痛所伴随。如果没有进一步的证据，我们绝对没有办法证明下述断言：每一个这样的躯体是如此被伴随的。观察材料能够支持的不过是：某些躯体是如此被伴随的（足以对由疼痛所伴随的疼痛行为的例示提供所需的标准）。我们不能由此推论说：所有的躯体甚至有可能如此被伴随（不，自然是不可能的）。我们没有权利固守一种观点，不管是哪一种。因此基于这样的证据，我们就无权断言：“这种例外只有极小的可能性”。

关于任何特殊的躯体，我们也没有权利相信，正是它至少有时为疼痛所伴随。即使我们肯定大多数人的身体如此被伴随，但对于任何特定的身体，我们也不能因此而肯定它如此被伴随。在缺乏进一步的信息的情况下，我们没有条件说：是否有例外，如果有，它们会是什么。我们所掌握的证据，对于我们所碰到的特定的身体是否有可能成为一种例外，是中立的。确信这一点（盖然的或别的）对于我们每一个人来说都是至关重要的。

该观点可以这样予以阐释：以虱子在一地区的小学生头上蔓

延为例。据说,也许有 30% 的人将被殃及。不仅不能由此说,对于每个学校来说,有 30% 的人将碰到,众所周知,将不会这样。而且也不能说,30% 以上的人将碰到,但分布是不均衡的。同样,从说 80% 的行为状态为心理状态所伴随这样的事实,也不能推论说:对于每个人的身体来说,其行为状态的 80% 由心理状态所伴随,即使承认那个证据,这甚至也是不大可能的。真实的东西是:平均是 80%。如果没有进一步的证据,那么对于严重背离某些身体实例中的这个数字的可能性或不可能性来说,我们就没有发言权了。

因此我们甚至无权“盖然地”确信:对于每个人的身体而言,它的行为状态的某些例示与心理状态有联系。而且不难理解的是:推理超出了所有曾经存在和将存在的人的范围。如果有例外,不管多么少,那么也许对我个人不幸的是,我的某些最亲密的朋友正在提供这些例外,或者说我的近邻正在提供它们,或者说这个历史时期正在提供它们。也许以为外星人可能接管了我周围的那些人的身体这样的看法是有点令人不安的(至少我不是孤立的)。再者,有这种例外,不可能只有微弱的可能性。仅根据西科拉所利用的证据,是没有理由相信有或没有例外的。

后来,西科拉事实上已表明他意识到了这类困难,因为他接着考察的是那个被改造过的论证的“修订版”:

我发现,在我自己的疼痛与我自己的疼痛行为之间有许多相互关联。不过,我又注意到:有大量但却不太重要的时刻,我表现了疼痛行为而又没有感觉到疼痛。因而,每当我在另一个人身上看到疼痛行为时,与我自己的类比便促使我得出这

914 第七部分

样的结论,即有这样很少的但却很重要的时机,他不处在疼痛中。由此可知,如果我看到了现存的每个人的所有疼痛行为,统计上的理由会使下述事情有很高的概率,即有一些人体,其疼痛行为的例示并不为疼痛所伴随。因此我没有权利像我事实上有权利做的那样,根据基于类比的多例论证肯定,每个人的身体有时与疼痛有联系。(p.140)

他的答复就是主张:多例类比推论“能够借助大量的附加证据而支持那个幸存的结论”。

在这一段中,西科拉考察了明显比我的观点更强的观点,根据那种观点,所倡导的推论就是“有一些人的身体,其疼痛行为的例示不为疼痛所伴随,是极有可能的。”而我想论证的观点则是:根据所引的证据,我们无论如何没有可能提出一种意见。两种观点实际上都同样缺乏牢固的基础。既然西科拉对更强的主张的答复完全有可能是对准我的主张,因此如果它们真是如此,我就得对之作出考察。

这里所说的支持性的工作得这样进行,即清除必需的心理-物理关系中的例外。第一种支持有赖于这样的主张:“只存在某些环境,在其中,当我不处在疼痛中时,我才表现出疼痛行为”。果真如此,长时期的观察就会排除这些环境。但是正如人们——也许马尔科姆最为著名^④——常说的那样,这样的环境的目录是无限的。如果是这样,那么就不可能注意到它们,进而不可能注意到它们的不在场,最终不可能得出结论说,另一个躯体处在疼痛之中。

第二种也是最重要的支持涉及到疼痛刺激与疼痛之间的新增加的联系,

因为就某些种类的疼痛刺激而言,我所接受的刺激与感受到疼痛之间的联系比我表现疼痛行为与感受到疼痛之间的联系要大得多。(p. 140)

不幸的是,这个答复一般来说是不管用的,因为有许多心理状态,它们不可分辨地关联于刺激。尤其是,我们不知道哪一类刺激在发挥作用。既然西科拉注意的是引起严重疼痛的刺激,那么许多人仍有问题,因为不幸得很,他们从未碰到过它们。一当他放弃这些刺激,那么他关于那种关联更大的断言就变得越来越可疑了(众所周知,当内外各种环境改变了外部环境影响我们的方式时,疼痛阈就会发生变化)。

西科拉最后的支持,旨在提供可导致安全感的基础,就是祈求于中枢状态与疼痛感受之间的关联,这些关联甚至可能比疼痛与疼痛刺激之间的关联更紧密。我认为,随后还将出现的任何怀疑都是完全可以容许的。(p. 140)

人们可能会觉得:最温和的怀疑(关于“疼痛”的最温和的认识)如果连绵不绝,那么也会削弱人们的满足感,但是无论如何,流行的神经生理学似乎会拒斥这种主张。同一种类的心理事件可以关联于不同种类的中枢状态,即使在同一个体中也是如此。似乎不存在西科拉所假定(和所需要)的一一对应。^⑤不管怎样,这种动人的乞求实际上只对少得可怜的人有效(迄今不适用于任何人),甚至在原则上对大多数人是没有用的。

因此这里所展开的论证是不能接受的。总之,即使它是可以接受的,它也没法对付那种攻击。如果有关的联系有例外,那么我

916 第七部分

们便不能由此推论说：以适当方式行事的每个人的躯体至少有时为相应的心理状态所伴随(或者肯定为任何心理状态所伴随)。西科拉的论证甚至没有打算要产生或没有想到要产生那些必需的、没有例外的关联。

我所论证的是：西科拉的多例类比推论并不会支持他所必需的结论。因此他的双层方案的第一步失败了。我现在要说明的是，即使不是这样，西科拉在他的第二个步骤中对它的利用，如果不求助于一个事例即人们自己的事例，就不会取得成功。第二章已经阐明，事实将是如此。然而西科拉的第二步并未从一个特定的方面令人满意地补充和进一步证实更一般的论证。

西科拉的如意算盘首先就是对别的躯体、借助多例类比推论确定：

(I) 这些躯体在他们的整个生活中都与心理事件联系在一起，(II) 其中一些心理事件是相似于我自己的、可公开的记忆中的东西。……(p. 137)

——当这一点大功告成，他便断言，他能

证明这样的假说，即一个人至少与每一别的、处在能活动的生存期间的躯体联系在一起(假说 B)，就像我能以同样的方式证明关于我自己的相似的假说。(p. 137)

这是怎样在人们自己身上实现的呢？下面就是其具体实现过程：

在任何特定的时候，我都有许多关于我过去似乎拥有的经验的记忆。这些记忆并未把这些经验与别的躯体联系起来，即

使某些记忆可能与任何躯体没有联系,但我仍有这样的一些记忆,它们可以使我合理地假设:我的所有经验在因果上依赖于我的身体,在这方面,它们对任何别人的躯体没有依赖性。我所具有的经验自然比我能记住的要多得多,但没有理由假设:我不能记住的经验有别于我能记住的经验,就像它不能证明我的这样的想法,即前者与某种别的躯体有联系,而与我自己的则没有联系。基于这些理由,我可以合理地推论说,我的一切经验都与同一个躯体相联系,因此至少有一个人在躯体的整个生存过程中一直驱动着我的躯体。(p. 137)

从根本上来说,就人们自身而言,出路就在于主张:人的经验就是如此,因此假说 B 是最好的假说。而就别人的躯体而言,断言则是,与这些躯体相联系的经验是这样的,以至于假说 B 也是最好的假说:

当我转向他人的躯体时,为了辩护关于他人躯体的、相似于我关于我自己所形成的假说的假说,我需要的不过是这样的假设(正如我要论证的,即借助以事例的多样性为基础的类比论证而获得的假设):(I)这些躯体在它们的整个生命中都与心理事件联系在一起,(II)其中的某些心理事件是相似于我自己的、可公开的记忆中的东西。(p. 137)

但是就一个自己而言,至关重要的是他并没有与他人的身体相联系的经验。这一点得到了下述事实的证实,即不存在把经验与别人身体联系起来的记忆。因此上面所述的两个假定事实上不是必需的东西。某些记忆符合于所需要的模式,这与别的人做不

918 第七部分

到这一点是一致的。

关于别的人的关键证据应如何得到呢(不求助于不合法的、单一例证的类比推论)?唯一的办法就是根据没有有关的记忆报告去推论:有关的经验没有出现。但是在缺乏这类经验时,不可能有直接的证据证明:这类经验会被报告。以为所有记忆都会被报告显然是错误的(大多数不能被报告)。因此出路就是不得不求助于这样一种极其模糊的主张:对于任何经验而言,至少作出了某种记忆报告。即使把它解释为:某些身体至少在某些情形下发布了有关的报告,但也是极为可疑的。如果断言它适用于我们每一个人,那么就肯定是错误的。既然那是在这种情况下唯一可得到的观察材料,因此它必定适合于我们每一个人。在没有这种经验时,仅仅主张这种经验一定会被报告,就难以令人置信了。我的结论是:假说B不适用于别人的身体。

最后,在转到我先前关于多例类比推论的批评这一论题上来时,有必要注意的是:记忆在西科拉的两阶段假说推论中起着至关重要的作用。记忆报告与记忆之间的联系不可能是没有例外的。而且在为他的多例类比推理的第一阶段进行辩护时,所求助的根据对于保证这种特殊的联系似乎是无济于事的。因此让人捉摸不透的是,西科拉为什么会认为:在第一阶段他能说明,对于每个人的身体来说都有相关的记忆。^⑥

因此似乎很清楚,对他心的类比推论不可避免地是基于单一的事例所做的概括。我认为:这就是我们每一个人从直观上都知的东西(而又是哲学家们所惧怕的东西)。如果类比推论应予重视,那么那个概括也应如此。这正好就是下面要予考察的主要的

责难。不过我首先要考察的是一个有时容易与之相混淆的责难。这就是：被考查的例示与未被考查的例示之间的比率不是太高。唐·洛克抱怨说，对他心的类比推理：

是从最小的可能基础到最大的可能结论的推理；或许，它是把一个单一的例示推论到无限的事例上。^⑦

这个责难与对他心的类比推理的本质的误解有关，因为它主要是把那种推理当作一种统计论证。事实上，正如下面的论证要表明的那样，关于他心的类比推理的合理性是独立于被考查的事例与未被考查的事例之间的比率的。

在试图确认另一个人有心灵时，假设我用了他心的类比推理，那么对被考查的事例和未被考查的事例之间的比率是不可能有什么抱怨的，因为这个比率大大高于大多数归纳概括的事例中的比率。但是很显然，这个他人可能是任何他人。因此如果用他心的类比推论证明一个他人有心灵，这样做是合理的，那么对于任何他人来说，不管那个人是谁，用那个推理证明他们有心灵也是合理的；当然得假定，他们在行动上而不是别的方面充分地相似于我自己。被考查的事例与未被考查的事例之比为什么与类比论证的合理性无关，其理由在于：类比论证的合理性依赖于被考查的事例与未被考查的事例之间的相关的类似性。如果被考查的事例与某个特定的、未被考查的事例之间的类似性足以支持那个推理，那么如此类推，在类比所涉及的任何未被考查的事例中，不管有多少这样的事例，我们均有权作出所说的那个推论。

针对类比推理的责难首先是由 N. 马尔科姆所提出的责难。

920 第七部分

人们几乎一致认为:这个责难是成功的,致命性的。它的确有望成为一个要求得到令人满意解决的问题。马尔科姆说:

……这是非常弱的归纳推理,因为它是基于对单一的例示的观察。^⑧

这个责难矛头所指的不是被考查的事例与未被考查的事例之间的低比率,而是这样的事实,即只有一种被考查的事例。乍一看,这本身是一个有力的驳难,因为无可争辩的是,我们常常要求根据类比的论证建立在一个以上的例示的基础之上。

然而这个责难不应与下述责难相混淆:

感觉到生气时我在皱眉,我怎么知道这不是我个人的独特之处呢?^⑨

人们可以追问归纳的基础:我怎么知道取样事例所处的状态就不是它们的独特之处呢?既然它可能是针对任何类比论证而提出的,那么它就不是对他心的类比推论本身的责难。这也适用于这种责难的改进了变种,因为它指出,他心的类比推论试图表明的事实上:我不是唯一的。但是据声称,人们有必要知道这是可以排除下述事实的,即我与他人之间的差别是无关紧要的,因此他心的类比推论是以未经证实的假定为基础的。再者,人们可以认为任何类比推论就是想表明:取样事例不是独一无二的。因此改进了的观点也以失败而告终。^⑩

现在是考察下述责难的时候了,即他心的类比推论仅以一个事例为基础。即使我们常常要求:类比推理应以一个以上的例示为基础,然而却有这样的类比推理,我们常常以之去建立具有很高

概率的结论,而这些结论仅仅是从一个被考查的事例中推论出来的。我要论证的是:他心的类比推理就是这样一种推理。我的论证将通过对那个照现在的情况看需要根据一个以上的例示的类比论证的讨论而展开。

假如我把一个蛋掉在地板上了,摔破了。我有权推论说:再掉一个蛋还会摔破吗?如果它以掉下去、摔破了的大量的蛋为基础,那么我们都会为有这样的推论而更加欣喜不已。为什么我们应更加欣喜呢?是因为我们应该具有统计上的、更重要的样本吗?肯定不是,因为蛋的总数如此之大,以致更多的事例在被检验的蛋的百分比中并不会导致显著的增加。答案肯定是:如果我们作出更多的检验,我们就会排除这样的可能性,即在那些被检验的(被摔破的)蛋中,由于被摔破之外的某种原因引起的破损的可能性。由此似乎可以得出结论说:如果不知由于什么原因,我居然知道:第一个蛋破了,只是因为它被摔破了,那么再进一步检验就没有必要性。我将用更一般的形式表述这一点。假设我根据事实: $X \cdots X_n$ 是 F 和 G, Y 是 F, 推论出了这样的结论,即 Y 是 G。在什么时候,我需要 n 是大数呢?我提出的设想是:只有当(I)大数的样本会为我们提供关于 X 是 F 和 X 是 G 之间的(因果)关系的信息,(II)这个信息对于推论 Y 即 G 是必要的,(III)我们尚无这种信息时,我们才需要 n 是大数。这里的关键是(III)。如果我们已有所需的信息,那么那个样本就用不着是大数。

现在我可以这个设想来说明:他心的类比推论用不着根据一个以上的事例进行推理。在第 3 章中,我略述了到达他心所涉及到的步骤。关于他心在我们行为中所具有(或不具有)的因果作

用,我们没有直接的知识。这涉及到以我们能获得的根据为基础的推论。假如有那个证据,我们的当务之急就是要确定我们的心理状态的因果作用。然而这样的裁决与他心问题无关。它还不是使我们得以到达他心的东西。裁决他心有因果作用是无济于事的。有关的原则是不确定的,尽管它是从什么引起了我们的行为到什么引起了他人的行为的推论,因此依赖的是相同的结果有相同的原因这样的原则。幸运的是,那不是我们到达他心的步骤。我们是这样到达他心的,即先就我们自己进行推论:我们的心理状态是由我们的(某些)物理(大脑)状态所产生的,然后推论别的人也是如此。同样的物理状态会有同样的结果。相关的原则是有用处的;不是同样的结果有同样的原因,而是同样的原因有同样的结果这样的原则。事实上,这是对自然齐一性的求救。

在这些接近他心的步骤中,关键的环节是:我们的心理状态是由我们的(某些)物理状态所引起的。这就是有关的因果链。就我们自己来说,这是已知的。把样本加以扩展并不会增加那个推理的力量,使其到达存在于物理状态和心理状态之间的那个链条。心理状态由物理状态所引起毕竟不是观察的问题。它是推理的问题。居于他心的类比推理中心位置的东西是某些物理状态与心理状态之间的因果链。那才是我们所需要的东西。它能够在我们自身中获得。那个事例提供了大量有关的相互关联。

已述的关于他心的探讨(以及对单例责难的反驳)最初是在我的“作为理论实在的他心”一文中阐述的。^⑪它表达的观点与此前所述的 A. 希斯洛普和 F.C. 杰克逊的^⑫“他心的类比推论”一文所提的方案有严重的分歧。其侧重点从心理状态的因果作用转到

了心理状态的因果源头。C.S. 希尔^⑩对他心的类比推理作了强有力的辩护,对指责其归纳基础太弱作出了反击。这个辩护与上述的方案以及希斯洛普、杰克逊的(替换)方案是部分一致的。希尔强调了生物学属类的重要性,而我则认为它与我诉诸于自然齐一性是不谋而合的。不过,希尔对类比论证的阐述以及随后的讨论似乎授人以柄,易受这样的批评,即被辩护的推理是从相同的结果到相同的原因的推理。希斯洛普和杰克逊肯定也是这种指责的对象。希尔所强调的似乎是心理状态的因果作用。

注 释

① 这一点在本章后面再予以阐述。

② 我自然也相信:对于每一活着的人体来说,确实存在着一个人;但这不是我们指望他心的类比推理应认定的东西。

③ R.I. 西科拉:“类比论证不适合于对他心的论证”,《美国哲学季刊》,14(1977),137-141。

④ “维特根斯坦的哲学研究”,载 V.C. 查普尔编:《心灵哲学》,普伦蒂斯-霍尔,新泽西,1962。

⑤ 参阅 D. 戴维森:“作为哲学的心理学”,载 J. 格罗弗编:《心灵哲学》(OUP,1976),还可看他的“心理状态”,载 L. 福斯特和 J.W. 斯旺森编:《经验与理论》(达克沃思,1970);保罗·齐夫,“他心的简单性”,《哲学杂志》,62(1965),575-589;J. 金:“论心物同一理论”,《美国哲学季刊》,3(1966),227-235;D. 刘易斯:“同一理论之论证”,《哲学杂志》,63(1966),17-25。

⑥ 我对西科拉的答复是我的“多例类比论证与他心”一文的改进了的观点,该文见《澳大利亚哲学杂志》,57(1979),330-336。

⑦ 唐·洛克:《自我与他人》(OUP,1968),p. 49。

⑧ N. 马尔科姆:“关于他心的知识”,载 V.C. 查普尔编:《心灵哲学》,pp. 151-159(参看 p. 152)。

⑨ J. 戴:《归纳概率》(劳特利奇,1962),p. 64。

924 第七部分

⑩ R.J. 帕吉特：“他心的科学推理”，《澳大利亚哲学杂志》，62(1982)，160。

⑪ 《澳大利亚哲学杂志》，54(1976)，158－161。

⑫ 同上。

⑬ “对类型唯物主义的辩护”，《综合》，59(1984)，295－320，引自 pp. 310－314。

人^{*}

P.F. 斯特劳森

I

在《逻辑哲学论》(5.631 – 5.641)中,维特根斯坦谈到了哲学中所出现的“我”,谈到了经验主体的哲学观念。他首先说:“思维的、描述的主体——这种东西是没有的。”在稍后,又说:“在一重要的意义上,主体是不存在的。”紧接着的话是:“主体不属于世界,而是世界的一个界限。”再后就跟着这样的段落:“(因此)确实有一种意义,据此我们可以在哲学中以非心理学的方式谈论我。这个我是由于‘世界是我的世界’这一事实而出现在哲学中的。这个哲学上的我不是人,不是人的身体或心理学所研究的人的灵魂,而是形而上学的主体,是世界的界限而不是世界的一个部分。”这些评论令人震惊,但也令人迷惑难解。在读到它们的时候,人们可能会想:好的,我们姑且赞成说人的身体,说心理学所讨论的、作为世界的一部分的人的灵魂,而且也不理会形而上学的主体。但是我们可能又会觉得:不行,当我谈到我自己的时候,我毕竟是在谈论那

* 译自 H. 费格尔等编:《明尼苏达科学哲学研究》,卷 2,明尼阿波利斯市:明尼苏达大学出版社,1958,pp. 330 – 53。

926 第七部分

个具有我的全部经验的东西,我确实是在谈论我的经验的主体——而且也是在谈论作为世界的一部分的某种东西,因为当我死的时候,是这个东西,而不是这个世界走到了尽头。我的世界的界限并不是——而且我认为不是——这个世界的界限。要解释这个关于既是经验的主体又是世界的部分的东西的观念,可能是很困难的。但它是我们所具有的一个观念:它应该是我们能够解释的一个观念。

我们不妨想一想我们通常用来谈论我们自己的某些方式,想一想我们通常将其归属于我们自己的某些事物。它们有多种样式。我们归之于我们自己的东西是:行动和意图(我正在做、过去做过、将要做这件事);感觉(我暖和、我疼);思想和感情(我思考、好奇、想要这个,我生气、失望或者满意);知觉和记忆(我看到这,听到别的,记得那)。我们在两种意义上将方位归属于自身:地点(我在沙发上)和姿势(我正躺着)。当然我们不仅把暂时的环境、状态、处境如此之类的大多数东西归属于自己,而且也把持久的特征,包括像身高、肤色、体形和重量这样的物理特征归属于我们自己。这就是说,在我们归属于我们自己的事物中,有一类事物我们也归属给了物质的身体,但我们却没有想过把我们归属于自己的事物中的其他事物归属于物质的身体。于是似乎无须解释这样的事实,即这些归属于我们自己的特定的身高、肤色和物理位置等也应该归属于某一事物或其他的事物;因为人们称之为自己身体的东西至少是一个身体,一个物质的东西。它可以从其他东西中区分出来,可以用常规的物理标准来鉴别并且用常规的物理学术语来描述。但是把人们的意识状态、人们的思想和感觉也归属于这

些物理特征、这种物理条件所归属的、完全一样的东西,这可能就需要而且已经表明需要解释了。为什么把人们的意识状态像某一肉体特征、某一物理条件等一样归属于一个完全一样的东西呢?这个问题一经提出,另一问题便会接踵而至,即究竟为什么要把人们的意识状态归属于(据说是或者属于)某一东西呢?不应设想,对这些问题的回答是互不相关的。

确实可以设想的是,上述两个问题的答案可能到每一个人的身体在他的经验、特别是他的知觉经验中所起的独一无二的作用中寻找。所有关心这些问题的哲学家都已论述过这种作用的独一无二性。(笛卡尔非常清楚地意识到它的独一无二性:“我不是像一艘船里的驾驶员那样居住在我的身体中”)这种独一无二性基于什么呢?当然是基于大量的事实。我们可以对这些事实中的一些作这样的概述,即对于每一个人而言,都有一个身体,它处在同那个人的知觉经验有关的因果关系的位置,而这个因果关系的位置对于这个人所具有的各种各样的知觉经验中的每一个来说,在各方面都是独一无二的。而且——作为进一步的结论——这个身体作为他所具有的各种知觉经验中的一个对象,对于他来说也是独一无二的;不仅如此,那个单一的身体的这种复杂的独一无二性似乎是一个偶然的事件;或者更确切地说是一组偶然的事件;我们可以,或者似乎可以设想:我们的知觉经验对于一个以上的身体的物理状态或状况的依赖性与独立性有许多特殊的组合。

现在我必须直截了当地说,关于每个人的身体在他的经验中所起的独一无二的作用的这组明显的偶然事实,在我看来,本身并没有为我们的问题提供答案。当然这些事实也有某种解释作用,

928 第七部分

它们提供了非常充分的理由,可以说明:经验主体特别关注的为什么只是一个身体,他为什么认为它是独一无二的而且或许比任何其他身体更重要。它们解释了——如果我可以如此表述的话——我为什么感到特别地从属于那个我事实上称为自己身体的东西;甚至还可以说,它们解释了:假如我准备把某个身体说成是我的,那么我为什么要把这个身体(我的确称为自己的身体的那个身体)说成是我的。但是它们却根本没有解释:我为什么有关于我自己的这个概念,我为什么要把我的思想和经验归属于某一东西。更进一步地说,即使我们对某种别的解释如一个人的意识状态(思想、感情和知觉)为什么被归属于某一东西感到心满意足,即使我们确信,这里所说的事实足以解释:一个特定的身体的“占有”为什么要被归属于这同一个东西(即足以解释:一个特定的身体为什么要被说成是处在同那个东西的某种特别的关系即所谓的“被占有”关系之中),但是,上述事实依然没有解释:为什么我们不是像我们通常所做的那样把某些身体特征仅仅归属于与“我们把思想、感情等等都归属于其上的那个东西”有着这种特别关系的身体,而是把这些身体特征归属给了我们把思想、感情都归属于其上的那个东西本身。(因为我们既说“我秃了”,也说“我寒冷”,既说“我正躺在壁炉前的地毯上”,也说“我看见天花板上有个蜘蛛”。)简言之,上述事实解释了:一个经验主体为什么要把一个身体与其他身体区分开来,也许还要给它一个尊贵的名称,并且把它所具有的一切特征都归属于它;但是它们却完全无力解释经验为什么要被归属于主体;而且即使经验被归属给了某一东西,它们也无法解释:它们和那些可能真的被归属于特定的身体的身体特征一起,为什么被

归属于同一的东西。所以上述事实并没有解释我们对“我”这个词的运用,或者说一个词怎样才能具有该词所具有的用法。它们没有解释我们所具有的、关于人的概念。

II

对这个观点的一个可能的反应就是说,我们所具有的这个概念是错误的或者是含混不清的,或者说,假如我们不习惯于将我们所具有的这个概念说成是含混不清的话,那么也可以说,我们藉以将如此不同种类的谓词归属于或好像归属于同一个东西的用法是令人困惑的,它隐藏了所涉及的概念或这种类型的某东西的真实本质。这种反应可能存在于关于这些问题的两种非常重要的意见之中。其一是笛卡尔主义的,即笛卡尔及其追随者的观点。关于第二类意见的归属我还拿不定主意;但是有证据表明,维特根斯坦可能还有石里克在某一时期曾持此观点。按照这两类意见,我们正在思考的问题中的一个,即“我们为什么要把我们自己的意识状态像某些身体特征等那样归属于一个完全相同的东西”,是一个不值得讨论的假问题;因为根据这两类意见,严格地把两类谓词归属于一个相同的东西,以为两类谓词有一个共同的所有者或主体,这不过是语言学上的幻象而已。而按照第二类意见,我们正在思考的另一个问题,即“我们究竟为什么要把自己的意识状态归属于某一东西”,也是一个不值得讨论的假问题;因为按照这种意见,一个人将自己的意识状态完全归属于某一东西,以为这种显而易见的归属有一个严格的主体,意识的状态属于那事物或者就是它的状

态等,这也不过是语言学上的幻象而已。

笛卡尔坚持第一类意见,这是众所周知的。当我们谈到一个人的时候,我们事实上指称的是两种各不相同的实体(两种不同类型的实体)中的一个或者两个,其中的每一个都有适合于自己的状态和属性类型;而且任何一个的状态和属性中没有一个可以成为另一个的状态和属性。意识状态属于这两个实体中的一个实体,而不属于另一个。关于笛卡尔主义的意见,在这里我就不多说了——我要说的以后再说——唯一要再次提到的是,尽管它避开了我们的一个问题,但是并没有避开、而且事实上招致了另一个问题,即“一个人的意识状态究竟为什么要被归属于某一主体?”

我将把这些意见中的第二种称之为关于自我的“无所有权”或“无主体”学说。概括地说,不管是否有人旗帜鲜明地坚持这种意见,但值得重构或思考。^①因为这种意见所犯的错误具有启发意义。可以推断,“无所有权”论者是以说明了一种物质身体在一个人的经验中所处的独一无二的因果地位的那类事实作为他的解释的出发点的。他主张,身体的这种独一无二性足以致使这样的观点,即一个人的经验可以归属于某一特定的个别的东西,可以说这个东西拥有或占有那些经验。他认为,该观点尽管是用所有权的术语不恰当地和错误地予以表述的,但是只要我们把这个个别的东西、这个经验的所有者理解为身体本身,这一观点就有某种合理性,就有某种意义。只要我们用这种方式思考,那么把意识的一个特定状态归属于这个身体、这个个别的东西,至少可以说是值得怀疑的,可能有或早就可能有某种错误。它可能是一个错误的归因;因为这里所说的经验可能或者本来有可能从因果上依赖于某种别

的身体的状态;就“属于”的现在这个尽管不恰当、但是值得考虑的意义而言,它本来有可能属于另一个个别的东西。然而就在这里,“无所有权”论者则表达了一种已变得混乱不堪的意义:离开了原来那个尽管不恰当、但是可以考虑的意义,而滑向了这些表达式的、不值得考虑的、空洞的意义,在原先的意义上,可以说某个人的经验属于某特定事物或者为其所有;在这个新的、不值得考虑的意义,这个被设想拥有经验的特定的事物不再被当成是身体,而被当成了另外的某事物,比如说自我。

假设我们将第一类所有关系即确实是一种因果依赖关系称之为“拥有₁”,而把第二类所有关系称之为“拥有₂”;把第一类个体称之为“B”,而把第二类假设的个体称之为“E”。于是差异就出来了,即我的全部经验由 B 所拥有₁,这确实是一种偶然的事情,而我的全部经验由 E 所拥有₂则似乎是一个必然的真理。但是对 E 和拥有₂ 的信念只是一个幻觉。只有其所有权在逻辑上是可转移的那些事物才真的能被占有。因此经验是不能被任何事物所占有的,除非在因果上依赖于一个特定的身体的状态这一含糊的意义上予以理解。这至少是与一个事物的一种真正的关系,因为经验可以处在同另一个事物的关系之中。既然 E 的全部功能,从“占有”的逻辑上不可转移的意义来说,就是占有经验,既然经验在此意义上不为任何事物所“占有”,因为“占有”没有这样的意义,因此必须把 E 从这种图景中彻底地清除出去。它的出现完全是源于混乱。

我认为,这种关于此问题的说明尽管也包含着某些事实,但显然是前后矛盾的。之所以如此,主要是因为持有这种观点的人在

陈述他的否定的理由时,被迫使用了他否定其存在的那种所有关系的意义。当他试图去陈述他认为导致了这种关于“自我”的幻象的偶然事实时,他必须用下述方式陈述那个事实,如“我的全部经验是为身体 B 所拥有₁的(即一无例外地依赖于身体 B 的状态)。”因为任何要消除“我的”或具有相同的拥有力量的其他表达式的企图都需要提供这样的某事物,它完全不是一个偶然的事实。例如,全部经验在因果上依赖于单个的身体 B 的状态这样的命题正好是错误的。持此论的人想说的是,某个人所具有的全部经验偶然有这样的依赖性。而且他无法前后一致地论证说,“某个人 P 的全部经验”与“偶然地依赖于某个身体 B 的全部经验”的意思是完全一样的;因为那样的话,他的命题就不会像他的理论所要求的那样是偶然的,而是分析的。他必须想到说出下述成员的某种经验,即这些经验都依赖于身体 B 事实上偶然适用于这些成员。这类经验的限定性特征事实上就是,它们要么是“我的经验”,要么是“某人的经验”,在这里“所有关系”的意思正好是他表示怀疑的那种意思。

当这种内在的不一致性否定了下面显而易见的事实时,如某人的确将自己的意识状态归属于某事物即他自己,而且这种归属的确像上述理论的持有者所发现的那样不能令人满意,即它似乎无法解释这一主张,如事实上是某人自己的同一种疼痛可能也是另一个人的疼痛,这种不一致性就成了一个严重的问题。为了理解这种逻辑上不可转移的所有权在我们的一般思维图式中的地位,我们没有必要深究下去。因为如果我们考虑到,在言语时,有对特殊的意识状态或私人经验的指称进行分辨的必要,那么我们

就会明白,这些特殊项目除了作为某个被分辨的个人的状态或者经验之外,是不可能在分辨中被涉及到的。人们可能会说,状态或经验作为特殊项目的同一性完全取决于人的同一性,前者正好是后者的状态或经验。由此可以直接得出下面的结论,即如果它们可以认定为特定的状态或经验,那么它们一定这样被拥有或归属,即以无所有权论者所嘲笑的方式,亦即下述逻辑上不可能的方式:某人事实上所具有的一个特殊的状态或经验又为其他任何人所具有。这种同一性的要求排除了所有权在逻辑上的可转移性。因此持此论的人要坚持他的观点,就只有完全否认我们能够提及特殊的状态或经验。而这种观点当然是荒谬可笑的。

即使是现在,我们仍然可以看到无所有权理论与笛卡尔主义的观点之间的一种可能的联系。开门见山地说,后者是关于两个主体(两种类型的主体)的一种二元论。而前者也可以有点悖理地被称作二元论:一种关于一个主体(身体)和一个无主体的二元论。我们可以推测,如此悖理地被称呼的第二种二元论是起源于如此非悖理地被称呼的第一种二元论;换句话说,假如我们试图把某人的意识状态所归属的那种东西看作是同某些身体特征所归属的东西完全不同的东西,那么要弄清楚意识状态为什么要被归属于或被认为属于任何东西就难上加难了。而且当我们想到这种可能性时,我们也可以想到另一种可能性,即笛卡尔主义者和无所有权论者两者都犯了一个深刻的错误,因为两者必定会认为,“我”有两种用法,在其中一种用法中,它意指的是它在另一种用法中没有意指的某种东西。

III

无所有权论者没有说明全部的事实,他只说明了部分事实。他正确地暗示,单个的身体在人的经验中的独一无二的地位或作用不能充分地解释这样的事实,即某人的经验或意识状态以这里所说的那种独特的不可转移的所有关系被归属给具有它们的某个事物。它可能是那种解释的必要的部分,但是它本身并不是一个充分的解释。就我们所知,那种论者会进一步主张,它或许是对另外的某东西的充分的解释,即它是对我们将意识状态当作是以这种特别的方式归属于某事物的那种混乱的和错误的思想的一个充分的解释。正如我们所看到的,这种主张是缺乏连贯性的:因为它包含着这种否定,即某人的意识状态是任何人的意识状态。然而如果我们赞成说,单个的身体在某人的经验中的特殊作用不足以解释:那种经验为什么应该被归属于某一身体,那么我们就避免了这种否定的不连贯性。存在着这种特殊作用,这一事实本身并不能充分地说明:我们当作经验主体的东西对于他自己作为这样一个主体的概念为什么是有用的。

当我说无所有权论者的说明由于没有考虑到全部事实而失败的时候,我想到的是一个非常简单、但是在这个问题中非常重要的思想:即一个人也把意识的状态、经验归属于(或准备归属于)不是他自己的其他人,这是他像人们通常那样把它们归属于自己的一个必要条件。^②这样说含义是很深广的,例如它意味着,归属这一短语在主体是另一个人时同在主体是自己时一样,应在相同的意

义上加以使用。这也是这样的看法,例如,不管一个人说“我疼痛”还是说“他疼痛”,“疼痛”的意思都没有什么两样,这种看法对于不是哲学家的人当然不会造成什么麻烦。对于描述意识状态的每一表达式,辞典不会给出两组意思:第一人称的意思和第二及第三人称的意思。但是对于哲学家来说,这种想法会造成困惑。事实上它已确实造成了困惑。当证实方法在两种情况下如此不同的时候——或更准确地说,在一种情形(即他人的情形)之下,存在一种证实的方法,而在另一种情形(即自己的情形)之下,严格地讲却没有这种方法——其意思怎么可能相同呢?或者还可以说,在自己的情形之下,谈论归属怎么可能是正确的呢?因为只有当存在着或可能存在着对要被归属的东西进行分辨这样的问题时,才能肯定有一个关于归属的问题吗?而且当那个处在疼痛中的人是另一个人的时候,虽然存在着对其进行鉴别的问题,可是当这个人是他自己的时候,怎么可能还会有这样的问题呢?但是一旦我们想起我们本是为了得到关于他人的信息而同他人说话时,上面最后一个问题便迎刃而解了。事实上,在某种意义上,当我就是那疼痛者时,并不存在我必须对“处在疼痛中的人是谁”作出判断这样的问题。而在另一意义上,我可能必须说出那是谁,即让他人知道那是谁。

我刚才所讲的或许已经解释了:假如某人把意识状态归属于其他人,在严格的意义上,他是怎样把它们归属于他自己的。但是他是怎样将它们归属于他人的呢?对此,有一点是可以肯定的,即某人在把意识状态归属给他人时,如果所归属的那些事物被看作一组笛卡尔主义的、只有私人经验才能在正确的逻辑语法中被归

936 第七部分

属于其上的自我,那么这个问题就是无法回答的,而且这个难题就不可能得到解决。在对意识状态应被归属的那些事物作出甄别时,如果私人经验就是某人所必须经历的全部,那么从某人自己的观点来看,根本就不存在判断私人经验就是他自己的经验这样的问题,基于同样的理由,也不存在判断私人经验就是另一个人的经验这样的问题。所有的私人经验、所有的意识状态都是我的,而不是某人的。简言之,只有当某人能够把意识状态归属于他人时,他才能将其归属于自己;只有当他能分辨的经验主体时,他才能够将意识状态归属于他人;如果他只能把他人甄别为经验主体或意识状态的所有者,那么他就不能识别他人。

可能有人会责难说,这样看待笛卡尔主义未免太肤浅了。毕竟,把身体相互区分开来不会有什么困难,对其加以鉴别也没有什么困难。而且如果保留笛卡尔主义的模式,这不是为我们提供了一种鉴别经验主体的间接的方法吗?例如,难道我们不能把这种主体看作是这样的主体吗,它“与那个身体所处的特殊关系就像我与这个身体所处的关系一样”;或者换句话说,看作这些经验的主体,它“与身体 N 所处的独一无二的因果关系同我的经验与身体 M 所处的特殊的关系是一样的”?然而这个建议是无用的。它要求我注意的是:当对我的所谓的经验的全部的谈论正好是正确的时候,我的经验就与身体 M 处在特殊的关系之中。(它要求我注意我的经验与身体 M 所处的特殊的关系;但是它还要求我把这当作一个可以对其他经验主体进行鉴别的条件,即当作拥有关于我自己作为经验主体的观念的一个条件,当作将任何经验都认作是我的经验的一个条件。)在这种解释模式中,只要我们一方面坚持

谈论经验,另一方面坚持谈论身体,那么我们可以注意到的最多的东西就是:经验,所有的经验都与身体 M 处在一种特殊的关系之中,而身体 M 正是通过这种途径才成为独一无二的,这就是使身体 M 从众多身体中独立出来的东西。(由于“经验”一词的出现,这里所说的“最多”也许太多了。)以上所作的解释告诉我们:“另一个经验主体可以被区分和确认为下述那些经验的主体,它们与身体 N 所处的独一无二的因果关系就像我的经验与身体 M 所处的某种关系。”相反的意见则是:“但‘我的’一词在上述解释中究竟起到了什么作用呢?(离开了该词,上述解释就不能成立了。)”

我们自己要想摆脱这些困难,就必须承认人的概念的原始意义。我所说的人的概念指的是关于这样一类实在的概念,即对意识状态进行归属的谓词以及对肉体的特征、物理状况等进行归属的谓词这两者都可以同等地适用的那个单一类型中的单个个体。而我说这一概念是原始的,其意思可从多方面加以表述。其一是重新回到我先前提出的那两个问题上来:即(1)意识状态为什么要归属于某一东西,(2)它们为什么像某些肉体特征、某种物理状态等一样被归属于相同的事物?一开始我就评论说,不应设想对这两个问题的解答是彼此孤立的。现在我要说的是,对它们的回答是这样相互联系着的:即意识状态被完全归属的一个必要条件就是它们应该像某些肉体特征、某种物理状态等一样被归属于相同的东西。也就是说,除非把意识状态归属于下述意义的人,即我赋予该词的那种意义,否则意识状态就根本不可能得到归属。我们常常被诱使去将人设想为由两类主体——一方面是经验主体(纯意识、自我)、另一方面是具有肉体属性的主体——所组成的复合

物。

当我们这样设想时,许多问题便接踵而至。尤其是当我们问自己:应该如何去构想、去利用这一关于两个主体的复合物的概念的时候,这种图景——假如我们是诚实谨慎的话——就会立即从两个主体的图景转变成一个主体加一个非主体的图景。因为,假如这种关于不同的、可以分辨的和可以鉴别的经验(即不同的意识)的主体的观念可以看作是关于人的由两个主体构成的复合观念中的逻辑上原始的东西,看作是逻辑上的构成成分,那么就不可能弄清楚,我们是如何得到这个观念的。因为永远不可能有把经验本身归属于除某人自己之外的任何其他主体的问题;从而也永远不可能有把经验归属于自己这样的问题;也不可能有将它归属于全部主体的问题。所以关于纯个体的意识(即纯粹的自我)的概念是一个不可能存在的概念;或者说至少是不可能作为一个初始概念而存在的概念,因为根据初始的概念,人的概念可以得到解释或分析。如果真的有此概念,那么它只能作为从属的、非原始的概念而存在,这样的概念本身是可以根据人的概念予以解释和分析的。它是这样的存在,即与关于纯粹意识、自我—实体的虚幻的基本概念相一致的东西。休谟在观察他自己时,在抱怨:他没有知觉就不可能发现他自己、而除了知觉他什么也发现不了的时候,他所要寻求的、或者讽刺性地假装要去寻求的正是这个概念。更为严峻的是——这里没有讽刺,对休谟来说,只有混淆,只有混淆的复仇女神——它是这样的存在,休谟曾经徒劳地寻找过关于它的统一性的原理,承认自己在这里陷入了困惑,并惨遭失败。休谟的探寻之所以是徒劳的,其原因在于,在没有差异性的原理的地方是找

不到这种关于统一性的原理的。它还是这样的东西,康德在这里的作法比休谟的巧妙得多,他为其提供了一个纯形式的(“分析性的”)统一性——即伴随着我的知觉并且因此等于什么也没伴随的那个“我思”的统一性。最后,它也许是这样的东西,即维特根斯坦在论及主体时所讲到的东西,第一,没有这种东西;第二,它不是世界的一部分,而是它的界限。

因此,“我”这个词所指称的从来就不是这个纯粹的主体。但是这并不是说,“我”这个词像无所有权论者所必须认为的或者维特根斯坦至少在某一时期所认为的那样,在某些情况下完全不指称任何东西。由于“我”是他人中的一个人,所以它有指称。实际上不可能属于纯粹主体——如果能被指称的话——的那些谓词,严格地说,只属于“我”所指的人。

人的概念在逻辑上先于个体意识的概念。不应把人的概念分析为一个被赋予了灵魂的身体的概念或者说一个肉体化了的灵魂的概念。但这并不是说,纯粹的个体意识的概念不可能有逻辑上从属的存在,如果人们认为或觉得有必要的话。我们常常说起一个死人——一个身体——而且在同样是从属的意义上,我们至少还可以设想一种灵魂离体的人,同时又不让个体性这种逻辑上的优点从已有的人身上失去。^③

IV

意识到人们在确认人的概念的逻辑上的原始性时所能作出的最大限度的确认,是非常重要的。下面我要简要地说明一下这种

940 第七部分

论证的步骤。除非一个人把意识状态或者经验归属于其他的个别实在,这些实在与他把自己的意识状态所归属的那个事物有着相同的逻辑类型,否则就不可能有关于把他自己的意识状态或者经验归属给某一东西的问题。把自己看成是这些谓词的主体的条件就是,他也把他人看成是这些谓词的主体。进而,使之成为可能的必要条件则是,他可以把这些谓词的不同主体即有关类型的不同个体相互区别开来(加以分辨确认)。而且,使这个条件成为可能的条件,又是有关的个体包括他自己在内,应该属于某种唯一的类型,即是说,属于一种类型,以至于对于那种类型中的每一个体来说,必须有被归属或可被归属的意识状态和肉体特征这两种东西。可是对这种类型的描述依然是很模糊的,难以清楚地说明它所包含的东西。为了说明这一点,我必须将那些严格适用于该类型的个体的谓词粗略地分为两大类。其中,第一类是由严格适用于物质的身体的那些谓词所组成的,而我们当然不想把归属意识状态的那些谓词用到这些物质的身体之上。我将把这第一类谓词称作M-谓词,它们包括像“重10英石”、“在起居室”等谓词。第二类则是由我们用之于人的其他所有谓词组成的。我将把它们称作P-谓词。P-谓词无疑是多种多样的,它们包括“在笑”、“在散步”和“在努力思考”、“信仰上帝”等。

迄今为止,我所论述的是,人的概念应理解为关于一类实在的概念,因此归属意识状态的谓词和归属肉体特征或物理状况等的谓词都同样可以应用于那种类型中的每一个别的实在。而且关于人的概念是原始概念这一说法的含义,我所说的不过是,不应该用一种方法或几种方法对它进行分析。例如,我们不应把它看成是

与两个基本类型即特定的意识和特定的人类身体有关的从属性的实在。我还暗示说,笛卡尔主义的错误仅仅是更普遍的错误中的一个特例,它以各种不同的形式出现在无所有权理论、以及因为各种被归属于这种被命名的实在的谓词而认为对人的命名或者表面上的命名没有明确地指称相同的事物或实在的各种理论之中。就是说,如果我们想避免这种普遍性的错误,就一定不能认为“我”或者“史密斯”有类型-歧义性的错误(如果我们要界定类型-歧义性,最好还是用这样的谓词如“在起居室”、“被一个石头打了”等等界定它,并且指出,当这些谓词用于物质对象时是一个意思,而用于人时则是另一个意思)。

这就是迄今为止我对人的概念是原始概念这一说法的意义所作的全部说明或提示。有必要进一步说明的是,这种说法对于我们用来归属意识状态的谓词的逻辑特征究竟有什么意义。为着这一目的,我们最好是对 P-谓词作一般性的考察。因为尽管并非所有 P-谓词都是我们所说的“归属意识状态的谓词”(如“去散步”就不是),但是我们可以认为它们共同具有下面这一性质,即就它们被归属的东西而言,它们表明了对意识状态的占有。

那么这种观点对于 P-谓词的特征究竟有什么影响呢?我认为它们是这样的,很清楚,对于一种特殊类型中的任何个体、任何 P-谓词来说,不管那个体是否具有那个 P-谓词,除非在原则上有所谓某种判断方法,否则谈论这个特殊类型即一种类型中的可辨认的个体,以至于主张它们同时具有 M-谓词和 P-谓词,这是毫无意义的。至少是在某些 P-谓词的情况下,判断的方法在某种意义上形成了关于 P-谓词归属的逻辑上充分的标准类别。因为绝

942 第七部分

不能假设这些判断方法构成了逻辑上充分的标准类别。因此我必须按照下面的思路来思考这些判断方法与 P-谓词所作的归属(或者所作的一部分归属)之间的关系:我们必须把这些判断方法视为这个不同的东西(意识状态)在有关个体中存在的标志。但是这一来,通过观察两者之间的相互联系,我们所能发现的仅仅是:判断的方法是由 P-谓词所归属的不同的东西之存在的一个标志。但是我们每一个人只能在一种情况下,即在我们自身的情况下作出这样的观察。现在我们又退回到那种认为我们对此问题的方法有缺陷的笛卡尔主义的辩护者的观点中来了。因为在这里,“我们自己的情况”意味着什么呢?除非进行归属的人已经知道如何把至少某些意识状态归属给其他人,否则主张把意识状态归属给自己或归属给所有事物便是没有意义的。因此他不能(或者一般不能)“从他自己的情况”出发推论出关于如何这么做的结论;因为除非他已经知道如何这么做,否则他就没有关于他自己的情况或者任何情况(即任何经验主体)的概念。相反,只有当某个身体以某些方式受到影响时,他才有证据说,疼痛等是可以预期的,而当其他身体受到影响时,他则没有这种证据。

当然,这个结论并非首创。我所要说的是,人们基于对他人的行为的观察而把 P-谓词归属于他们;人们所遵循的行为标准并非 P-谓词所意指的东西存在的标志,而是一种关于 P-谓词归属的逻辑上充分的类别标准。然而,为着这个结论,我敢断言,它来自于对把意识状态归属给任何东西所必须的条件思考。这种观点不是说,我们一定要接受这个结论以避免怀疑论,而是说为了解释这个怀疑论问题据以陈述的概念图式之存在,我们必须接受

这个结论。但一旦接受了,这种怀疑论问题也就不会产生了。(而且怀疑论问题的一般原则也是如此:它们的陈述包含着对一种概念图式的虚假的接受,同时又悄悄否认它得以存在的一个条件。这就是它们用它们据以得到陈述的术语为什么不可能被解决的原因。)不过这还仅仅是关于P-谓词的描述的一半。

下面我们来看另一半。当一个人将P-谓词归属于自己时,他并不依靠对他据以将这些P-谓词归属于他人的那些行为标准的观察,毫无疑问,这样说至少适用于P-谓词的某些重要类别,而不适用于所有的P-谓词。一般来说,它不适用于带有特征和能力评价的那些P-谓词:在进行自我归属时,一般来说,这些P-谓词被归属所依据的与它们归属给他人所依据的是相同类型的根据。人们一般不会根据下述行为标准即他据此把P-谓词归属于他人的标准,而把它们归属给自己,在它所适用的那些P-谓词中,有许多也符合下述说法,即它们的归属可能受到自我归属者以此为基础所作的矫正。但是在许多情况下,一个人有完全充分的根据把一个P-谓词归属给自己,可是在这些情况下,这个根据与他将这个P-谓词归属给他人所依据的根据是完全不同的。(因此,人们在报告当下的思想或感觉状态时,就会说:“我感到累,我失望,我疼痛。”)这个事实怎么可能与下述认为一个人将P-谓词归属给他人时所依据的标准就是这种归属的逻辑上充分的标准的学说相调和呢?

实现这种调和的明显困难可能把我们引向许多方向。例如,它可能诱使我们否认这些自我归属是真正的归属;还可能诱使我们把对意识状态的第一人称归属比作那些别的形式的行为,它们

944 第七部分

构成了人们据以将 P-谓词归属给他人的标准。这种策略似乎避免了那一困难;在任何情况下,它并非完全不适用。但是它掩盖了事实,而且是没有必要的。它不过是认识 P-谓词(或者至少是 P-谓词的一种至关重要的类别)的独特特征的过程中的一种不落俗套的失败形式。因为正像(一般)没有学习或教给自己这类谓词的内在的、隐私的意义的优先的过程一样,因此也不存在这样的另一种过程,即根据在自己身上所观察到的、与某些形式的行为的相互联系,学习把这些谓词用之于他人。同样地,(一般)不存在根据行为标准把这些谓词用之于他人的优先的学习过程,因此也不存在另一种过程,即获得表现一种新的行为形式,即第一人称的 P-表露的从属性技术的过程。以上两种描述都拒绝承认有关谓词具有独一无二的逻辑特征。

假设我们把“ Px ”看作是这样一种谓词的命题函项的一般形式,那么按照上面第一种描述,在这个形式中首先取代“ x ”的表达式的就是“我”这个第一人称单数代词;它与其它替换形式一起使用就是从属的、派生的和不确定的。另一方面,根据上述第二种描述,以这种形式首先取代“ x ”的是“他”、“那人”等等,而他与“我”一起使用也是从属的、特殊的,而不是一种真正的归属性的用法。但对于这些谓词的特征来说至关重要,它们同时具有第一和第三人称的从属性用法,而且它们除了可根据对它们主体的行为的观察进行自我归属以外,还可根据行为标准进行他人归属。要了解它们的用法,就要了解它们的用法的这两个方面。一个人要得到这类概念,就必须既是这些谓词的自我归属者又是他人归属者,而且必须把每一个他人都看成是自我归属者。要理解这类概念,

一个人还必须承认,有这样一种谓词,无论是根据对这种谓词的主体的观察,还是不以此为根据(独立于对主体的观察),它都是意义明确的和可以适当地归属的:这里的第二种情况是这样的情况,在那里,归属者本人也就是主体。如果没有与我刚才所作的描述相一致的概念,那么我们就肯定不会有关于灵魂的哲学问题;但是同样地,我们也就不会有我们的关于人的概念。

为了用这种类型的一种特殊的概念,比如说抑郁的概念,表述这一带有某种不可避免的不成熟性的观点,我们就得述及:用抑郁的方式行动(抑郁的行为)和被压抑的感情(感情的抑郁)。人们往往论证说,感情可以被感觉到,但不能被观察到,而行为则可以被观察到,却不能被感觉到,因此在这里一定有导致逻辑纷争的土壤。但是抑郁的概念包含着人们想引起这种纷争的条件。我们可能会说,一个概念要成为一个关于 X 的抑郁即 X 具有那种抑郁的概念,这个概念就必须既包括被 X 感觉到、却不能观察到的东西,又包括 X 之外的他人观察到、却不能感觉到的东西(对于关于 X 的全部评价来说)。然而这样说或许更好: X 的抑郁是某种同一的东西,它被 X 感觉到却观察不到,被 X 之外的他人观察到而又感觉不到(而且可以被观察到的东西当然也是可以伪装和伪造的)。拒绝接受这种说法也就是拒绝接受我们用来谈论抑郁的这种语言结构。在某种意义上,这一点也不错。人们可能会放弃言语,或者设计一种用来自言自语的不同的语言结构。而错就错在假装去接受那种结构,而同时又拒绝接受它;即用具有那种结构的语言表达某人的拒绝。

从这个角度看,我们肯定可以在关于心灵的话题中发现其他

946 第七部分

一些熟悉的哲学难题。因为其中的某些难题之所以产生,正是由于未能承认或充分肯定我一直主张至少为某些 P-谓词所具有的那个特征。人们没有注意到,这些谓词的两方面的用法(自我归属性和非自我归属性的),如果没有一方面的用法也就不可能有另一方面的用法。相反,如果认为它们的用法的一方面是自足的(其实这是不可能的),那么另一个方面就似乎是悬而未决的。正因为如此,我们才徘徊于怀疑论和行为主义之间。如果我们把某一 P-谓词(比如说“抑郁的”)自我归属方面的用法当作是原始的,那么在我们据以判断另一个人抑郁的标准和抑郁的真实状态之间似乎会产生一个逻辑上的鸿沟。我们没有意识到的东西是,这个逻辑上的鸿沟一旦被允许出现,它不仅可以吞掉那个人的抑郁,而且还会吞掉我们的抑郁。因为如果存在这个逻辑上的鸿沟,那么抑郁的行为,无论其多么强烈,也只不过是抑郁的一个标志。而且仅仅因为它与抑郁之间的、被观察到的相互关系,它才得以成为抑郁的一个标志。但它是谁的抑郁呢?人们可能被引诱说,只能是我的抑郁。但是如果仅仅是我的,那么就根本不是我的。这种怀疑论的观点常常描述说,鸿沟上的这种通道充其量是一种不可靠的推论。但是关键在于,如果这个鸿沟存在着,那么那个推论的前提的句法甚至就不存在。

另一方面,假如我们把这些谓词中他人归属的用法看作是自足的,那么我们最终可以说,这些谓词作为谓词所具有的全部意义就成了我们医藉以将这些谓词归属给他人的标准。从上述对那个逻辑上的鸿沟的否定难道不能得出这个结论吗?不能。如果以为如此,那就是忘记了这些谓词的自我归属的用法,就是忘记了我们

必须使一类谓词与它的下述必不可少的意义关联起来,这意义是:当自我归属不是在他人归属所依据的观察的基础上,而是在另一个基础上作出的时候,这些谓词应该既能自我归属于、又能由他人归属于同一个个体。这并不是说这些谓词有两种意义。而是说,对于这些谓词确实具有的单一的意义来说,至关重要的是;对它们作出归属的两种方法都完全是适当的。

如果一个人玩扑克牌游戏,那么某张牌上特有的标记就构成了我们称呼此张牌比如说“小王”等的一个逻辑上充分的标准。但是按照这个游戏的规则,当一个人将该牌称作小王时,也同时将这些标记所具有的属性之外的属性归属于它了。这个谓词是从这个游戏的整个结构中获得它的意义的。所以它与对 P -谓词进行归属的语言密切相关。我们藉以将 P -谓词归属给他人的那些标准对于这种归属来说是逻辑上充分的标准,这样说并不意味着对于这些谓词的归属意义而言,除了这些标准以外再没有什么东西了。如果这样说,那就是忘记了它们是 P -谓词,忘记了它们所属的其他语言结构。

V

我们现在的困惑可能表现为另一种形式,即这种形式的问题:“但是一个人怎么可能不依据观察而把其他人在观察的基础上、根据逻辑上充分的理由归属给他的同一个东西也归属给自己呢?”这个问题还可以合并到一个更宽泛的问题之中,这个更宽泛的问题就是:“ P -谓词是如何可能的?”或者说“人的概念是如何可能

948 第七部分

的?”这就是我们用来取代那两个较早的问题的问题,先前的问题是:“意识状态为什么要被归属,为什么要归属给某个东西?”以及“它们为什么像某些肉体特征一样被归属于同一个东西?”我们只有承认人的概念的原始性以及由此产生的 P-谓词的独一无二的特性,才能找到关于这两个原始问题的答案。所以剩下的难题不得不用这种新的方法来进行构想。因为当我们承认人的概念的原始性并且因此承认 P-谓词的独一无二的特性时,我们可能还是想追问:在自然的事实中,是什么使我们拥有这一概念成为合理的,并且要求它有望成为一个有价值的答案。^④我不敢妄言,对这种要求已完全心满意足了。但是我可以提出两个不同的方面,它们可以算作是一个答案的出发点或者组成部分。

首先,我认为可以通过把某类 P-谓词放到这个构想的中心以形成一个出发点。粗略地说,它们是一些涉及到“在干什么”的谓词,它们清楚地说明了意图或一个心理状态或至少一般性的意识,而且它们表明了身体运动特有的模式或模式的范围,而完全没有明确地表明任何确定的感觉和经验。我这样说指的是“散步”、“卷绳子”、“玩球”、“写信”这类事情。这些谓词具有下述许多 P-谓词的有趣的特点,即人们一般不会根据观察把它们归属给自己,而却根据观察把它们归属给其他人。但是就这些谓词而言,人们起码会觉得,承认以这两种不同的方式被归属的东西是同一个东西,是出于无奈。这是因为在他们所归属的东西中,身体运动的相当确定的模式占有明显的优势,而任何特殊的经验则明显不存在。它们使我们摆脱了下述观念,即我们不靠观察或不靠推理或两者都不靠所知道的唯一东西就是私人的经验。即使不借助这些手段

中的任何一种去判断,我们也能知道身体现在的和未来的运动。当然,身体的运动肯定也是我们借助观察和推理能够知道的东西。

与我们不借助观察也对之有知识的运动相似的身体运动是我们能观察到的事物中的东西,而对立于我们不借观察就知道的东西。重要的是,我们要理解这些被观察到的运动。它们影响着并且作为条件规定着我们自己。事实上,我们要理解它们、解释它们,唯一的办法就是把它们看作这样的行动计划或图式中的要素,即我们无须观察现在的相关活动也可以了解这些行动的现在的过程和未来的发展。但是这也就等于说,我们要把这种运动(被观察到的他人的运动)看作是行动,我们要用意图来解释它们,我们要把它们看作是这样的类型中的个体的运动,即我们无须通过观察就可以知道其现在和未来运动的那种个体所属的类型;我们不是根据观察,不是根据我们以观察为基础归属给他人的东西,而把他人看成是自我归属者。

当然,这些议论也无意于表明:如何才能解决“他心问题”,或者指出怎样对我们关于他人的信念作出一般的哲学“证明”。我已经论证过,这样一种“解决”或者“证明”是不可能的,对它的要求不可能得到一以贯之的陈述。这些议论也无意于成为一种先天的发生心理学。它们追求的不过是,在关于此课题的哲学历史发展的这个阶段,尽力去说明:我们所掌握的这个概念图式似乎是我们能够理解的。我这样说的意思是,如果我们首先想到这样的事实,即我们能够行动,我们相互作用,我们按照人类的共同本性而行动,那么就更好理解:我们怎么可能相互把对方(以及我们自己)当作人看待。“相互当作人看待”涉及很多东西;而不是孤立的无关的

950 第七部分

东西的堆积。我们置于这幅图景的中心的这类P—谓词不是杂乱无章地存在于那里的,与别的不相干的东西完全隔离,相反,它们与别的东西不可避免地紧密联系在一起,相互交织在一起。因此心灵问题不能划分为互不相关的课题。

我刚才提到了人类的共同本性。但是还有另一种意义,据我们所具有的概念图式得以存在的一个条件就是人类本性不应是共同的,即是不应是一种共同体的本性。哲学家们经常讨论是否存在或可能存在像“群体心理”之类的东西。对于某些人来说,这种观念富有特殊的吸引力,然而对于另一些人来说,这种观念似乎荒谬绝伦、毫无意义,同时是极其有害的。不难明白,人们为什么觉得这些是有害的:他们在下述看法中发现了令人恐惧的东西,这种看法认为,人们应该放弃他们过去对个人所采取的那种态度,而采取在某一方面类似于对群体的态度的那种态度;他们可能不再为自己去决定个人的行动过程,而只是参与团体的活动。但是他们发现这一想法有害,就足以表明:他们理解了被他们称为荒谬怪诞的那个观念。我们发现,将能够个体化为有机的身体的东西所组成的某种类别的成员个体化为人,这是自然而然的,这一事实并不意味着,这样的一个概念图式对于与我们自己完全不同的任何种类的存在而言都是不可避免的。

我们难道不能构想关于特定类型的社会世界的观念吗,在那里,不使用关于个别的人的概念,而只使用关于群体的类似的概念。首先,我们想想关于现实的人的存在的某些方面。例如,想想从事某种竞争性的但又是共同性的活动如战争中的两群人,他们都已经为此受过良好的训练。我们甚至还可以设想,命令是多余

的,即使也要传递信息。不难想象,这两个群体的成员在致力于这种活动时,完全不顾及个别的人,完全用不上个人的名字或代词,然而他们却必须提及群体,并且把类似于我们通常用于个人的、归属有目的的活动的那些谓词的谓词用到它们之上。事实上,在这种情况下,他们可以使用“我们”和“他们”的复数形式;但它们并非真正的复数,它们是没有单数的复数,就像我们在“我们已经占领了的城堡”,“我们已经输掉了的比赛”等句子中使用的复数一样。他们也可以谈及群体的组成部分,提到群体的成员,但是他们只是用这样的术语,即从在共同活动的那些成员所起的作用中获得它们的意义的术语。(因此我们有时用“敲击”或者“抓获”来谈论事实上是人的东西。)

当我们考虑到这些情况时,我们便会明白,在我们社会生活的一个方面——我要谨慎地说,不是一个很大的方面——我们自己的确操纵着一个概念图式,在该图式里,个人的观念是没有地位的,它在那里的地位被取代了,即是为群体的观念取代了。但是我们难道不能这样来设想共同体或群体吗,即认为其成员生活的这一部分就是其主导部分或者其全体?人类群体有时有这样的情况,即其成员会像我们所说的那样,“像一个人一样”思考、感觉和行动。我想阐述的观点是,关于个体的人的概念得以存在及运用的条件就是,这样的情况只在某些时候才发生。

就此而言,下述说法是绝对没有用的:纵使这样的情况总是发生,但是群体的每个成员同样有一个个别的意识,同样是一个个别的经验主体。有必要再一次强调的是,这样谈论个体的意识和个体的经验主体是没有意义的,因为我们无法对这种纯粹实在加以

952 第七部分

鉴别。^⑤当然,当我提到这个幻象的时候,我事实上已经把我们关于个体的人的概念当成了一种出发点。正是这个事实使上述无用的反对变成了一种自然而然的反对。但反过来,假设我提出下面这样的主张:即人类身体的每一部分、每一器官、每一成员都有一种个体的意识,都是一个独立的经验中心,那么这个主张和上述反对意见一样,只不过更明显一点,都是一种无用的主张。接下来再来想象所有中间的情况,例如这些情况:存在着一类运动着的自然物体,它们被划分成组,每一组都表现出同样特点的活动模式。在每一组内部,有某些伴随着功能差异的外表上的差异,尤其是每一组都有一个具有独特外表的成员。难道人们不能设想不同的观察组,它们在某种情况下让我们把某个特定的成员看成是这一组的代言人或喉舌,而在另一种情况下它们让我们把它们看成是它的嘴,把那个组当作是一个单分散的身体吗?这意思是说,一旦我们接受了后一种思维方式,那么我们就想抛弃前一种;我们将不再受到以第一种形式出现的人类类比的影响,只会受第二种形式的影响,我们再也不想说:“那些成员们也许有意识。”要理解我们这里的思想运动,我们只需记住:“身体及其成员”这一短语的那种令人吃惊的歧义性就够了。

VI

我的这种解释的尝试就到此为止。我论证的主要是,我们应该承认人的概念的逻辑上的原始性,并且因此承认某些谓词的独一无二逻辑特性。一旦承认了这一点,某些传统的哲学问题就

再也不会被当作是问题了。尤其是,那个似乎使休谟迷惑不解的问题^⑥也不复存在了——这个问题就是被视之为首要的特例的特定意识即“知觉”(经验)的特定主体的统一性原则、同一性原则的问题。根本就没有这样的问题,根本就没有这样的原则。如果有这样的原则的话,那么我们每个人为了确定他的现在的经验是自己的还是其他的某人的,就不得不用到它;而且这种主张是毫无意义的。(当然,这并非要否认,一个人在某一方面也许无法肯定他自己的同一性,例如,他也许无法肯定某一行动或者一系列行动是否是由他自己完成的。于是他就用与其他任何人在解决对他自己的怀疑时所用的方法相同的方法(在原则上相同的方法)来解决这个对他自己的怀疑。这些方法仅仅与人格同一性的通常标准之使用有关。而这些标准实际上是什么,它们的相对的重要性是什么,如此之之类的问题依然故我。但是一旦从这些虚假的问题中解脱出来,这种问题也就成了一种更简单的哲学问题了。)

在休谟犯了错误或者似乎犯了错误的地方,康德和维特根斯坦表现出了更好的洞察力。这一点也许从来没有用最愉快的方式表现出来。因为康德的学说,即“意识的分析统一性”既不要求也不能衍推关于统一性的任何原则,并不像人们所希望的那样清楚。而维特根斯坦(在某一时期)的评论对我们实际所用的概念图式似乎进行了不必要的嘲弄,其大意是说,意识的材料是不能被占有的,琼斯在谈到他自己的感觉等时所用的“我”指的并不是另一个人用“琼斯”一词所指的东西。当史密斯说“琼斯有点疼痛”而琼斯说“我有点疼痛”时,否认或者看起来否认他们谈论的是同一个实在,说的是关于它的同一件事,这是没有必要的、是悖理的,而否认

954 第七部分

琼斯对他自己疼痛的肯定也是如此。反过来,如果不否认那些自我归属的意识状态的确被归属了,同时又主张:对于语言的每个使用者而言,正好有一个人,在把意识状态归属于他时,他没有必要运用关于那个人的被观察到的行为的标准(尽管他没有必要不这样做),那个人就是他自己,那么这样说就与我们的实际的话语方式更加一致了。上述说法至少遵守了我们所使用的概念图式的结构,同时又不妨碍对它的进一步考察。

注 释

① 关于维特根斯坦曾经坚持这样一种观点的证据请参阅:摩尔在《心灵》上发表的第三篇论文:“维特根斯坦 1930 - 1933 年讲演录”(《心灵》, 1955, 专辑, pp. 13 - 14)。据说,维特根斯坦认为,在“我牙疼”或者“我看见一个小红斑”的情况下,“我”的用法和在“我有一颗坏牙”或者“我有一个火柴盒”的情况下“我”的用法是完全不同的。他认为,“我”有两种用法,在其中的一种用法中,“我”可以被“这个身体”所取代。以上的观点可能是笛卡尔式的。但是他还说,在另一种用法中(即在与“我有一颗坏牙”相对立的、由“我牙疼”所例示的用法中),“我”指的不是一个占有者,而且在思考牙疼或者经历牙疼时,自我也未被涉及到;他显然同意利希腾伯格的格言,认为我们(或者笛卡尔)不该说“我思”,而应说“存在着思”(“Es denkt”)。

认为石里克也有这种观点,主要根据是他的“意义与证实”一文的第五节(见《哲学分析文选》,H. 费格尔和 W. 塞拉斯编)。像维特根斯坦一样,他也引用了利希腾伯格的话,然后进一步指出:“因此我们知道,除非我们决定把自己的身体称为这些材料(直接的经验事实)的拥有者或载荷者——这似乎是一种相当错误的表达式——否则我们就必须说,这些材料没有拥有者或载荷者。”然而,他的文章的整体意思对我而言是难以捉摸的,而且完全有这样的可能,即上述被引用的那一个句子给人以一种错觉。我要说的不过是,我在构想我假设的“无主体”论者的案例时,引用了他的文章;但是我不敢妄言集中体现了他的观点。

利希腾伯格的反笛卡尔主义的格言,像后面的论述所要说明的那样,如果适当地加以运用,就是我所赞成的格言。但是笛卡尔主义的许多批判者是生吞活剥地加以复述的。

关于维特根斯坦和石里克曾经持有“无主体”观的证据似乎不是判决性的,因为有关评论的初衷很有可能是想批评笛卡尔主义的观点,而不是要陈述真实的观点。

② 对这种陈述的不合格的形式,我可以设想一种驳斥,它可以表述如下:关于一个可以唯一地加以使用的谓词(即事实上仅属于一个个体的谓词)的观点肯定不是荒谬的。而且假如它不荒谬,那么能够予以断言的就肯定是,人们得以把某种类型的谓词归属于一个个体(他自己)的一个必要条件就是他准备或乐意在适当的情况下把这些谓词归属于其他的个体,因此这个人就应该有一个关于归属它们的适当的情况是什么的概念;但他不一定在任何情况下真的这么做。

赞成这种反对意见的最简捷的方法就是承认它,或者至少不去反驳它;因为尽管根据更重要的主张去实施这种反对意见略微简单一些,但次要的主张就是这一论证所严格要求的。然而应进一步指出的是,我们说的不是单个的谓词,或者仅仅某一组或其他组的谓词,而是一大类谓词的全部,因此这些谓词或者它们的否定形式的使用决定了个体的主要逻辑类型或者范畴。从这个角度看,坚持次要主张与重要主张之间的区分,就是要把这种区分从一个它在其中明显是正确的层面带到另一个它在其中似乎没有用处或者没有意义的层面上去。

这里的主要观点是纯逻辑的观点:关于谓词的观念与关于下述可分辨的个体的范围的观念是联系在一起的,即关于这些个体的谓词是可以有意义地予以确定的,尽管不必然为真。

③ 略加思考就会发现,这种让步是多么有限,不过我现在暂不讨论这个问题。

④ 一个中项,它希望得到的答案不仅仅是说:“噢,在世界上有人。”

⑤ 更精确地说:对它们的鉴别必然从属于对人的鉴别。

⑥ 试比较:休谟《人性论》附录。

第八部分

行 动 哲 学

行动、理由与原因*

唐纳德·戴维森

当理性通过为行动者的所作所为提供理由来解释那种行动时,理由与行动之间究竟是什么关系呢?我们可以将这样的解释称为合理化解释,进而可以说理由合理地解释了那种行动。

在本文中,我想为下述古老的——而又习以为常的——观点作出辩护,它认为合理化解释是因果解释的一种形式。这种辩护,正如新近的许多论者^①所阐述的那样,无疑需要重新加以调整,但似无必要放弃这一观点。

I

理由要对行动作出合理化解释,唯一的条件就是,理由能使我们看到或想到行动者在其行动中所看到的某事——行动的某种特征、结果或方面,它是行动者需要、渴望、赞赏、珍视的东西,并认为对之负有责任、义务、能受益、能接受的东西。如果只是简单地那种特定的行动吸引了某人,那么我们就不能解释他为什么会有

* 原为提交给美国哲学协会 1963 年年会关于“行动”的专题讨论会的一篇论文,发表于《哲学杂志》60(1963)。后收入作者自己汇编的《关于行动与事件的论文集》(克拉伦敦出版公司,牛津,1980 年第 1 版,2001 年第 2 版)中。现根据论文集译出。

960 第八部分

如此的所作所为;我们必须指出,与吸引他的那个行动有关的是什
么。因此,一旦某个人由于一个理由而做某事时,就能对他作出如
下的描述:(a)对于某种行动有某种前态度;(b)相信(或知道、觉
得、注意到、记得)他的行动属于那一类别。在(a)之下,应包括有
愿望、需要、冲动、激励和形形色色的道德观、审美原则、经济上的
成见、社会习俗以及公众和私人的目的与价值,因为所有这些都能
被解释为行动者导向某种类型的行动的态度。“态度”一词在这里
是极为有用的,因为它不仅必须涵摄表现于行为过程中的稳定持
久的性格特征,如热爱孩子或热衷于社交等,而且还必须包括那种
触发奇异行动的转瞬即逝的怪念头,如想抚摸一下某个女人的手
肘的突发的愿望。一般来说,前态度一定不能被误认为是这样的
确信,即确信一类行动中的每一个应能得到履行,也值得履行,或
概而言之说是值得想望的,不管这些确信多么短暂。相反,一个
人在他的一生中可能有一种心瘾,如想饮下一罐颜料,而从不相信
值得这样去做,甚至在冒出那想法的一刹那也不相信。

说出行动者为什么做某事的理由常常就是列举前态度(a)或
相关的信念(b)或二者并举;因此我们不妨将这二者称之为行动者
为什么做出那种行动的基本理由。现在,通过陈述下列关于基本
理由的两个命题,就有可能一方面重新阐释合理解释就是因果
解释这一论断,另一方面赋予该论证以条理性:

1. 为了解任何一种理由如何合理解释一种行动,必要而充
分的条件是,我们至少应大致地搞清楚,怎样建构一种基本理由。

2. 行动的基本理由即是它的原因。

我将依次阐述这些论点。

II

我按一下开关,打开了灯,照亮了房间。我意想不到的是,我还向小偷展示了一个事实,即我正在家中。这里,我无须做四件事情,而只是一件事情,对此却有四种描述。^②我按开关,因为我想开灯,当我说我想开灯时,我便解释了“按”这种行动(给出了我的理由,作出了合理化的解释)。但是在交待理由时,我并没有对警示小偷或照亮房间作出合理化解释。既然当某个人的所作所为以某一种方式被描述时,理由可以对之作出合理化解释,而当以另外的方式进行时便不能如此,那么我们就不能简单地把所作的事情当作这样的句子即“我按开关的理由就是我想开灯”中的一个词语;否则,从“按开关”同一于“警告小偷”这一事实,我们就不得不得出结论说,我警告小偷的理由也正好是我想开灯的理由。我们不妨通过更确切地陈述基本理由的必要条件,来表明合理化解释过程中的行动描述的这一准内涵的^③特征:

C1. 只有当 R 包含着行动者对于具有某种属性的行动的前态度和行动者相信 in d 描述下 A 具有那种属性的信念时, R 才是解释行动者为什么在 d 描述下作出了行动 A 的基本理由。

既然我想开灯似乎缺乏必需的普遍性因素,因此它怎么可能成为一个基本理由(的部分)呢? 我们也许会为“我开灯”与“我想开灯”之间的词语上的近似性所欺骗。前者指的显然是一个特殊的事件,因而我们断言,后者具有同样的事件作为其对象。当然毋

962 第八部分

庸置疑,我开灯这一事件并不能由上述两种语句以同样的方式来指称,因为那个事件的存在是“我开灯”的真值所必需的,而不是“我想开灯”的真值所必需的。如果在两种情形中指称是相同的,则第二个句子便可衍推出第一个句子;然而事实上这一句子在逻辑上是独立的。尚不明朗的是——至少直到我们注意到它时是这样——那种其发生使得“我开灯”这一陈述为真的事件,无论如何有意向,绝不能被称之为“我想开灯”的对象。如果我已打开灯,那么我一定在一个确定的时刻、以特定的方式完成了这件事情——其中每个细节都是固定的。但是让我的需要指向在什么时候已完成的一个行动或以某种独特的方式完成的行动,那则是毫无意义的。无数行动中的任何一个行动都将满足这种需要,并同样适合于被视之为它的对象。需要和愿望常常指向物理的对象。尽管如此,“我想要橱窗里的金表”并不是基本理由,而仅仅是因为它使人想到一个基本理由——例如,我想买那块手表,才能解释我为什么走进商店。

因为“我想开灯”与“我开灯”在逻辑上是彼此独立的,第一个陈述可以用来作为说明第二个陈述为什么为真的理由。这样一种理由提供了最低限度的信息:它意味着这一行动是有意向的,而“需要”总倾向于排斥某些另外的前态度,诸如职责或义务感等。但是这种排斥在很大程度上取决于行动和解释的语境。“需要”除了欲望以外,似乎是苍白无力的,但是否认某个贪恋女人或一杯咖啡的人需要得到该女人或咖啡,则是不可思议的。事实上,将“需要”当作一个包括作为种的所有前态度的属是合情合理的。当我们这样做时,当我们知道某种行动是有意向的时,就不难回答这样

的问题：“你为什么干那事？”答曰：“没有任何理由”，这意思不是说没有理由，而是说没有更进一步的理由，从那个行动有意向地完成这一事实不能推出没有理由；换言之，所谓“没有理由”是不包括想要做那件事的。这最后一点对于现在的论证并不是必不可少的，但却是有好处的，因为它为将一种有意向的行动定义为出于某个理由而完成的行动这一可能性提供了辩护。

所谓基本理由包括信念与态度，但同时述及这两方面往往徒劳无益。如果你告诉我，你正在放松船首三角帆，因为你认为那将能阻止主帆翻转，那么我无须再被告知，你想要阻止主帆翻转；而如果你说，你用手指指向我，是因为你想侮辱我，那么再补充说，你认为，通过将手指指向我，便能达到侮辱我的目的，就是没有意义的。同样，要完成这一陈述，许多按照并非基本的理由而对行动进行的解释并不需要述及基本理由。如果我说我拔除杂草是因为我想要让草坪漂亮起来，那么再画蛇添足地加上这样的说明：“所以我明白，在任何的确会使或极有可能使这草坪漂亮起来的行动中，有某种东西是有望得到的”，就是愚不可及的。为什么要坚持认为，在愿望从一种目的——它不是行动——到某人设想为手段的行动的转换过程中，存在着任何逻辑上的或心理学上的环节？同样可作为论证使用的是，只有当行动者相信其作为手段的东西被想望时，所想望的目的才能解释那个行动。

幸运的是，没有必要对各种情绪、情感、心境、动机、激情和渴求进行归类和分析，对它们的述及就可回答这一问题：“你为什么做那事？”为了弄清这种述及怎样、何时合理地说明某一行动，就得涉及到一个基本理由。当某个人离开鸡尾酒会时，可以用幽闭恐

964 第八部分

怖症作为理由加以解释,因为我们知道,人们需要在其自身与其所恐惧的事情之间保持一定距离,想予以回避、逃避进而得到安全。在一个投毒行动中,嫉妒就是动机,因为投毒者格外相信,他的行动将打击他的敌人,消除他极度痛苦的根由,洗清不白之冤,一个嫉妒的人想要做的便是这些。当我们听说一个人出于贪婪而欺诈其儿子时,我们再无须知道其基本理由是什么,但我们知道肯定有一个理由以及它的一般本质。赖尔从“他由于虚荣而自夸”分析出:“他遇到陌生人便自夸起来,他这样做合乎一种似规律的命题,即一旦他发现有机会获得赞美和别人的忌妒,他便会做出他以为会获得这种赞美和忌妒的事情”(p. 89)。这样的分析常常受到可能是公正的批评,其根据是,某个人可能只是偶尔地出于虚荣而自夸。但是如果赖尔所说的自夸者确实出于虚荣而如此作为的话,那么赖尔的分析所衍推出的某种内容便是真的:那个自夸者想要获得别人的赞美和忌妒,并且相信他的行动将会引出这种赞美和忌妒;不管是真还是假,赖尔的分析并没有省去基本理由,而恰恰是依赖于基本理由。

弄清某个人为什么如此行动的基本理由,即是弄清该行动据以做出的意图。如果我在岔道口向左转,那是因为我想去加德满都,我向左转的意图即是去加德满都。但是,知道了意图不一定就详尽无遗地知道了基本理由。如果詹姆斯怀着取悦于他母亲的意图去教堂,那么他对取悦于他母亲肯定有某种前态度,但是要判断他的理由究竟是想取悦于他的母亲,还是他认为这是正当的,是他的责任或是他的义务,则需要更多的信息。“詹姆斯去教堂所抱的意图”这种表达式具有一种描述的外在形式,但是事实上它是非自

足的语词,而且不能认为它指称了一个实在、一种状态、一种倾向或一个事件。其在语境中的功能乃是根据行动的理由来形成关于行动的新的描述;因此“詹姆斯怀着取悦于其母亲的意图去教堂”这一表达式产生了一种关于由“詹姆斯去教堂”所描述的行动的新的、更充分圆满的描述。同样的过程在我回答下述问题时也必然会发生:“你为什么那样不停地快速移动?”我回答说:“我在抓、缚、锻炼、摇、亲近、训练跳蚤。”

对一个蓄意的结果的直接描述,常常比指出该结果是蓄意的或所期望的能更好地解释一个行动。“它将使你平静下来”解释了我为什么为你倒了一杯酒,这种解释同“我想做点什么以便让你平静下来”一样有效。因为在解释的语境中,前者暗含着后者;但是相比之下,前者更好一些,原因在于,如果它是真的,该事实将证明我对行动的选择是正当的。因为证明一个行动为正当和解释一个行动常常是形影相随的,以至于我们常常这样来表明行动的基本理由,即作出断言说,如果它是真的,那么将证实、维护或支持那个行动者的有关的信念或前态度。“我知道我应该归还它”,“这张报纸说天将要下雪”,“你踩了我的脚趾”,所有这些,在给出恰当理由的语境中,执行的都是这种常见的、二元的功能。

根据这种阐释,理由的这种辩护的作用,依赖于解释的作用,但是反过来则不成立。你触犯了我这一点既不能解释我触犯了你,也不能辩护我触犯了你是正当的,除非我确信你真的触犯了我,但是这一信念无论是真是假,只能解释我的行动。

III

从基本理由的角度看,行动总是以与行动者的某种或长或短的、或独有或非独有的特征一致的形式显现出来,而那个行动者则以有理性动物的角色出现。与这种关于行动的基本理由的信念和态度相一致,我们能一无例外地(以一定的创造性)建立三段论的前提,根据这些前提,可以推论说行动具有某种(正如安斯康帕所谓的那样)“值得想望的特性。”^④因此就存在一种不可还原的——尽管有点贫乏的——意义,在此意义下,每一合理化说明都能证明:从行动者的观点来看,当他行动时,对于该行动来说,肯定有要说明的某东西。

某些哲学家由于注意到非目的论的因果解释并未揭示由理由所提供的辩护的要素,因此便断定,适用于别处的原因概念不能应用到理由与行动之间的关系上,而且这种辩护模式,就理来说,提供的是所需要的解释。但是,假如我们姑且承认在解释行动的过程中理由能够单独地证明这些行动的正当性,也不能由此得出结论说,这种解释仍不是——而且必然不是——因果解释。毋庸讳言,我们关于基本理由的第一个条件(C1)就是想促进合理化解释与其他各种解释的分离。如果合理化解释像我想要证明的那样,是一种因果解释,那么在 C1 所给予的那种意义上,辩护至少是那种对属性作出鉴别的辩护。另一主张又如何呢?它认为辩护就是一种解释方法,因此无须引进通常的原因概念。这里必须裁决的是,辩护中究竟包含着什么。它能够涵盖的可能只是 C1 所

要求的東西：即行動者有信念和態度，正是基於此，該行動才是合理的。但是這樣一來又肯定忽視了某些必要的東西，因為一個人可能為一個行動以及這一行動的實施找一個理由，然而這個理由並不是他為什麼採取該行動的那個理由。對於理由與其所解釋的行動之間的關係至關重要的是這樣的觀點，即行動者之所以實施那一行動，是因為他有那個理由。當然，我們也能把這種觀點包括在辯護之中。但如此一來，在我們能夠說明“因為”的實質之前，關於辯護的這一概念將像理由概念一樣隱晦難解。

當我們問某個人為什麼如此作為時，我們想得到的是一種解釋。他的行為看起來怪異，難以容忍、荒誕、無聊、與其個性不符，且顯得突兀；或許在它裡面，我們甚至看不出是一種行動。當我們了解到他的理由時，我們便有了一種解釋，一種關於他所作所為的新的描述，這種描述把它放进日常的圖景之中。這一圖景包括有行動者的信念與態度，也許還有意圖、目的、天賦、普通的個性特征、德行或惡習等等。除了這之外，由一個理由所給予的關於行動的再描述，可能將該行動置於一個更廣闊的社會的、經濟的、語言的或價值的語境之下。通過了解這一理由，去了解：那個行動者將其行動設想為一個謊言，一筆債務的償還，一種侮辱，一項慈愛的義務或一種騎士風度的實現，就是要在其所處的規則、習俗、慣例和期望的背景中把握該行動的意義。

由後期維特根斯坦所引發的這樣一些言論，已得到了許多哲學家的敏銳而深入的闡述。下面一點毫無疑問是真實的：當我們通過給予理由來解釋一個行動時，我們的确是重新描述了那個行動；而重新描述該行動就是將該行動置於一種模式之中，於是該行

968 第八部分

动便以这种方式获得了解释。在这里,人们往往被诱使冒险作出了两个不可能推论出的结论。其一,从给出理由只是重新描述该行动,以及原因与结果相分离这样的事实出发,我们并不能推论出,理由不是原因。理由,作为信念与态度,肯定不能同一于行动;但更重要的是,事件常常按照它们的原因来重新描述。(假如某个人受伤了。我们可能“按照原因”、通过说明他被烧伤了来重新描述这一事件。)其二,它错误地以为,因为通过将那个行动置于一个更大的模式中而予以解释,因此我们现在便把握了所涉及的那类解释。关于模式和语境的话语并没有回答理由如何解释行动这一问题,因为有关的模式或语境既包含了理由,也包含了行动。我们得以解释事件的一种方式就是将它置于其原因的语境之中;在我们以及其他大多数人所理解的“解释”的意义上,原因和结果共同构成了能解释该结果的那种模式。如果理由和行动阐明了一种不同的解释模式,那么就必须认为那种模式是同一的。

下面我将联系梅尔登的一个例子来进一步阐明这一点。有一个开车的人为了发信号,抬起了他的手臂。通过将他抬手臂这一行动再描述为发信号,这样他的意图即发出信号,便解释了他抬手臂的行动。解释这一行动的模式究竟是什么?它就是那种常见的由于一种理由而采取行动的 mode 吗?假使那样的话,它的确解释了这一行动,但这仅仅只是因为它假设了我们想要分析的理由与行动之间的关系。更确切地说,这种解释模式是:这个人开着车,正驶近拐弯处;他知道他应当发信号;他也知道怎样通过抬起手臂发信号。在这种语境之下,现在他抬起了他的手臂。也许正如梅尔登所设想的那样,如果所有这些都发生了,他的确会发信

号。假使那样的话,这种解释就会是这个样子;如果在这些条件下,一个人抬起他的手臂,那他就是在发信号。当然,其困难在于,这种解释并没有触及到他为什么抬起他的手臂这一问题。他有理由抬起他的手臂,但这并未被说明就是他为什么做这件事的那个理由。如果这种关于“发出信号”的描述,通过交待他的理由而解释了其行动的话,那么发出信号就一定是有意向的;但是,根据上面的阐述,它可能并不是有意向的。

根据梅尔登的主张,如果因果解释“与我们所要寻求的关于人类行动的理解风马牛不相及”(p. 184),那么我们在“他如此作为是因为……”这种描述中,就没有关于那个“因为”的分析,因为在这里,我们要进一步列举理由。关于理由与行动之间的关系,汉普希尔指出:“在哲学中,人们确实应当去发现这种……联系,尽管它很神秘”(p. 166)。汉普希尔反对亚里士多德通过引入需要概念作为原因要素来解决这种神秘性的企图,其根据是,那样做结果就会使理论太绝对、太确定了,以至于不能适用于所有的情况,而且“依然没有令人非相信不可的根据来坚持认为,‘需要’一词必须进入每一关于行动理由的完全陈述之中”(p. 168)。我同意这种认为需要概念过于狭隘的观点,但是我也论证过,至少在大量典型的事例中,如果关于行动者之行动的理由的陈述是可理解的,那么就必须假定某种前态度的存在。汉普希尔并不明白亚里士多德的图式怎样被评判为真或假,“因为没搞清楚的是,什么能成为评价的基础,或者什么样的论据才可能是判决性的”(p. 167)。但是我要极力主张的是,在尚无令人满意的抉择时,对亚里士多德那样的图式的最好论证就是,唯有它才有望说明理由与行动之间的“神秘的联系”。

IV

在“他锻炼身体和他想减肥,并认为锻炼身体将达此目的”这一陈述中,要想把“和”变成“因为”,首要的一环^⑤就是我们必须有下述论证条件 C2:

C2.行动的基本理由就是它的原因。

赞成 C2 的根据,我想至此已很明显了。在本文的余下篇幅里,我愿针对形形色色的攻击,对 C2 作出辩护,而且在这一过程中,阐明所涉及到的因果解释的概念。

A.第一类攻击是这样的。基本理由由态度和信念所构成,而态度和信念又是状态或倾向,而不是事件;因此它们不可能成为原因。

对此,不难作出答复说:状态、倾向和条件通常被称之为事件的原因:桥梁坍塌是因为结构上的缺陷;飞机起飞时坠毁是因为气温异常之高;盘子破碎是因为它原本有个裂缝。不过,这种答复并没有触及到直接有关的要害。述及一个事件的因果条件,仅仅是根据还有一个在先的事件这一假定,交待一个原因。但是,那个引起一个行动的在先的事件究竟是什么呢?

在很多情况下,可以不费吹灰之力地找到与基本理由紧密地联系在一起的事件。状态和倾向并不是事件,但受到状态或倾向的冲击则是事件。当你激怒我时,我便可能产生一种想要伤害你的感情的愿望;当我看见一个甜瓜时,我可能想吃甜瓜;当我们注

意、发现、了解或记住某事时,可能马上产生某些信念。那些主张心理事件没有资格作为行动的原因的人常常忽视了这种显而易见的事实,因为他们坚持认为,心理事件是被观察到的或被注意到的(而不是一种观察或一种注意),或者说它就像一阵刺痛、一种疑虑、一种悔恨或一阵颤抖、良心的一种神秘的激发,或随意的行动。梅尔登在讨论那位通过抬起手臂来发拐弯信号的驾车人时,就对那些人提出了挑战,他们试图从因果上把行动解释为同一于“所有这些事例中常见且独有的事件”(p. 87),也许这事件就是一种动机或一种意图,至少是“某种特殊的情感或经验”(p. 95)。但是心理事件自然是存在的;在某一时刻,该驾车人注意到(或认为他注意到)他将要拐弯了,这正是他发信号的时刻。在任何连续的活动如驾车中,或在小心翼翼的实施过程如在畅游达达尼尔海峡的过程中,或多或少有决定整个活动的方向和形式的混合的目的、规范、愿望和习性,而且还有关于我们正在做什么、关于环境变化的持续不断的信息输入,正是根据它们,我们才能控制和调节我们的行动。拔高驾车人对他已经来到拐弯处的意识,把它称之为一种经验甚或一种情感,这无疑是言过其实的夸大,不管它是不是应享有那个称号,最好把它当作是解释驾车人为什么抬起他的手臂的理由。倘若如此,且在一般情况下,就不可能有我们可称之为动机的东西,但是如果提到了诸如想稳妥地达到某个人的目标这样一种一般的目的,那么该动机显然就不是事件。那个驾车人据以抬起他的手臂的意图也不是一个事件,因为它什么也不是,既不是事件、态度、倾向,也不是对象。最后,梅尔登请因果论者去找这样一个事件,它是一个人在其中有意向地抬起手臂的各种情况下、为

这些情况所常见且独有的。必须承认,这是不可能得如所愿的。因而桥梁坍塌、飞机坠毁或盘子破碎的常见和唯一的原因也是不可能出现的。

那位发信号的驾车人能够回答这一问题:“当你抬起手臂时你为什么还要这样做?”从他的回答中,我们能认识到引发那行动的事件。但是,一个行动者能一无例外地回答这样一种问题吗?有时候这种回答将提及一个并未给出理由的心理事件:“最后我终于打定了主意。”不过,似乎还有这样一些有意向的行动的事例,在那里我们完全无法解释:我们在行动时,我们为什么要这样行动。如果是这样,根据基本理由作出的解释相似于根据结构缺陷而对桥梁坍塌所进行的解释:我们并不知道这一事件或导致(引发)坍塌的事件的后果,但我们确信一定有这样一个事件或事件的后果。

B. 根据梅尔登的观点,原因一定“在逻辑上有别于那个所谓的结果”(p. 52);但是行动的理由在逻辑上并不有别于那行动;因此,理由并不是行动的原因。^⑥

这种论证的一种可能的形式已经产生了。既然理由借助对行动的重新描述而使行动成为可理解的,那么在不同的描述之下我们就没有两个事件,而只有一个事件。然而因果关系又要求有不同的事件。

有人可能会误入这样的歧途,即错误地以为,“我按动开关”使我把灯打开了(事实上它只是导致灯开始发光)。但是不能由此说,下述看法是错误的,这看法是:“我按动开关的理由就是我想开那只灯”,部分衍涵着“我按动了开关,而且可以进一步描述说,这种行动已为开灯的需要引起了”。根据事件的原因描述一个事件,

并不等于把事件与其原因混为一谈,通过重新描述而作出的解释也不排斥因果解释。

这一例证还可以用来批驳下述主张:如果不运用把行动与所谓的原因联系起来的词语,我们便不能描述该行动。在这里,该行动应在“我按动开关”这一描述之下予以解释,而且所谓的原因就是“我开灯的需要。”在这些表述之间,应假设存在什么样的相关的逻辑关系呢?更为合理的似乎就是强调:“我开灯”与“我需要开灯”之间具有一种逻辑联系,但是即使在这里,这种联系一旦加以审视,实际上是语法上的,而不是逻辑上的联系。

在那种认为因果关系是经验的而不是逻辑的关系的观点中,无论如何存在着某种非常奇异的东西。这意思是什么呢?诚然,并非每一种真实的因果陈述都是经验的。因为假设“A引起B”是真的,那么B的原因便是A;如此进行替换,我们便有了“B的原因引起B”这样一种分析的陈述。一种原因陈述的真理性取决于把事件描述成什么;其身份究竟是分析的还是综合的则依赖于该事件如何被描述。尽管如此,有一点还是可以坚持的,那就是:只有当这种描述恰当地确定下来了,而且这种恰当的描述在逻辑上并非独立的时候,理由才能合理地解释行动。

假设,说一个人想要开灯即意味着他将实施他相信会达到其目的的任何行动,那么关于他按动开关的基本理由的陈述就可衍推出:他已按动了开关——即像亚里士多德所说的那样:“他直接在行动”。如果是这样,理由与行动之间就肯定存在着逻辑上的联系,这种联系正如同“它可溶于水并被放入水中”与“它已被溶解了”之间的联系一样。既然这一含义从原因描述延伸到了结果描

974 第八部分

述,而不能倒过来,那么列举这一原因还是提供了信息。而且,尽管这一点常常被忽视,但是“将它放入水中便使它溶解了”并不能衍推出“它可溶于水”;这样一来,后者便有了额外的解释力。然而,如果在用解释与被解释的事件之间的明显确定的联系来代替这种可溶性的时候,我们能参照某种属性,比如说一种特殊的水晶结构,它与在水中溶解之间的联系只能通过实验来认识,那么这种解释将更有意义。因此像愿望和需要这样的基本理由为什么不能以如同可溶性解释溶解那样的相对通行的方式来解释行动的问题,便变得清楚明白了。所谓可溶性,我们假设是一种纯倾向性的属性:它是根据单个的准则而得到定义的。但是愿望则不能按照它们可以合理化地解释的行动来界定,即使愿望与行动之间的关系不完全是经验性的。关于愿望,例如它们在情感中和在它们不能合理化解解释的行动中的表现,还有另外的、同样必要的标准。某个有愿望(或需要或信念)的人,在正常情况下完全不需要什么标准——即使没有任何可为别人利用的线索,他一般知道,他想要的、欲求的和相信的是什么。基本理由的这些逻辑特征表明,正好不乏这样的聪明才智,它阻止我们将基本理由界定为由于这些理由而行动的倾向。

C. 根据休谟的观点,“我们可以把原因界定为对象,它为另一个对象所伴随,这一来,所有与第一个对象相类似的对象都为与第二个对象相类似的对象所伴随。”但是哈特与奥诺雷却断言,“那种关于一个人因为例如另一个人威胁他而做某事的陈述,并不蕴涵着这样的含义或秘密的主张,即如果那些条件重复出现,便会伴随着同样的行动”(p. 52)。哈特和奥诺雷认为,当休谟说普通的单称

因果陈述蕴涵着普遍原则时,他是正确的;但当他因为这个理由而假定动机和愿望即是行动的常见的原因时,则是错误的。简言之,规律必然要卷入日常的因果解释之中,但不会出现在合理化解释之中。

回应这一论证的常见的作法就是假定,我们确实具有连接理由与行动的初步的规律,而且在理论中,它们还可以得到完善。诚然,受到威胁的人不一定都以同样的方式来作出反应;但是我们可以对威胁本身进行区分,而且同时还可以根据行动者的信念和态度对行动者作出区分。

不过,这种意见确然是虚妄的,因为连接理由与行动的普遍原则并不是——也不能被纳入——准确预言能据以可靠地作出的那种规律。如果我们反思一下理由用来确定抉择、决策和行为的方式,那么便不难明白,事情为什么会是这样。对于当时正在行动的行动者来说,在事后的解释和辩护的环境下作为理由经常出现的只是众多理由中的一种考虑,即一种理由。任何依据理由预见行动的严肃理论都必须找到一种方法,来评估各种愿望和信念在决策孕育过程中的有关的效力;任何严肃的理论不能把有望来自于单个愿望的改铸过的形式当作评估的出发点。关于实际活动的三段论在让一个行动从属于一个理由时是没有用武之地的;因而它不能融入关于实际活动的推理的重构之中,因为这种重构牵涉到对诸相反理由的估价。这种关于实际活动的三段论提供的一种模式,既不适用于关于行动的预言性科学,也不适用于关于评价性推理的规范性说明。

对于有足够预言力的规律的无知并不妨碍有效的因果解释,

976 第八部分

不然的话,就几乎不可能作出因果解释。我敢肯定,那个窗户的破碎是由于被一块石头击中了——我亲眼看见这一切是如何发生的;但是我并没有掌握这样的规律(也许任何人都是这样?),即我能据以预言什么撞击将打破哪些窗户的规律。一个普遍原则,如“窗户是易碎的,当撞击足够强大时,尽管其他方面都很正常,这种易碎的东西也将趋于破碎”,大致说来,并不是一种预言性的规律——所谓预言性规律,如果我们有的话,它将是定量的,并使用了完全不同的概念。这种普遍原则,如同我们关于行为的普遍原则一样,具有一种不同的功能,即它为证明涵盖眼下这种事例的因果规律之存在提供了根据。^⑦

通常,我们对于单一的因果联系比对支配这种事例的任何因果规律要有把握得多;这是否表明休谟关于单一因果陈述可衍推出规律的论断是错误的呢?未必如此,因为休谟的主张,正如上文所引证的那样,乃是模棱两可的。它的意思可以是,“A引起B”可衍推出可与“A”“B”两个描述所用的谓词有关的特殊规律,或者它也可以意味着,“A引起B”可衍推出:存在着由关于“A”与“B”的某些真实的描述所例示的因果规律。^⑧很显然,休谟理论的这两种阐释使那种认为单一的因果陈述可衍推出规律的主张成为有意义的,而且两者都为那种认为因果解释“与规律有关”的观点提供了支持。但是第二种阐释太弱了,因为并没有什么特殊的规律能从一个单一的因果断言中衍推出来,而且一个单一的因果断言如果需要辩护的话,也能在不为任何规律提供辩护的情况下得到辩护。只有休谟理论的第二种阐释才能使之符合多数因果解释;它同样也适用于合理解释。

对事件的最初的解释为它提供原因;而更深入细致的解释则可以交待更多的内情,或者说通过建立有关的规律,或为相信其存在而提供理由,就可对这种单一的因果论断作出辩护。但是以为只有等建立了规律时才能给出解释这一想法显然是错误的。与这些错误相联系的正是下述观念,即认为单一的因果陈述必然通过它们所用的概念,来标示将在被衍推的规律中出现的那些概念。假设一场飓风,它是在星期二的《泰晤士报》第5版上报道的,结果引起了星期三的《论坛报》第13版上所报道的一场灾难。这样一来,星期二《泰晤士报》第5版所报道的事件引发了星期三《论坛报》第13版上所报道的事件。我们应该寻找一种把这两类事件关联起来的规律吗?寻找把飓风与灾难关联起来的规律几乎没有什么荒谬之处。而准确预见那场灾难所需的规律对于像飓风和灾难之类的概念来说理当然是没有用处的。预报天气所面临的麻烦在于,事件在其之下受到我们关注的描述——“寒冷、多云、午后有雨”——与更确切知道的规律所用的概念之间只具有微乎其微的联系。

我们可以确信,这些规律——如果理由是行动的原因,那么这些规律的存在才是必要的——并不委身于普遍原则必须委身的概念。如果一类事件(行动)的原因属于某一种类(理由),并有一种规律来支持每一个单一的因果陈述,那么不能因此说,存在着把被分类为理由的事件与被分类为行动的事件联系起来的规律——这种分类法甚至可以是神经病学的、化学的或物理学的。

D. 据说,人们所具有的、关于自身行动的理由的那种知识有悖于理由与行动之间的因果关系之存在:一个人无须观察或归纳

978 第八部分

就能确凿可靠地知道他在行动中的意图,而且任何常见的因果关系都不能以这种方式来认识。毫无疑问,我们关于自身在行动中的意图的知识将表明:关于某人自身的疼痛、信念、愿望等等的第三人称知识具有许多奇特之处。唯一的问题是,这些奇特之处是否能证明:至少在任何日常的意义,理由并不引起它们合理化解释的行动。

在关于“我正在毒害查尔斯,因为我想帮他摆脱痛苦”这种形式的陈述的真实性问题上,你很容易犯错误,其原因在于,你在关于你是否正在毒害查尔斯这一点上你可能犯错误——你可能自己误饮了有毒的那一杯水。但是似乎还有这种情况,你可能把你的理由弄错了,特别是当你对于一个行动具有两种理由时更是如此,其中一个正合你的意思,而另一个则相反。例如,你的确想帮查尔斯摆脱痛苦;你也想让他死去。究竟是哪一个动机使你做出那件事,你也许会弄错。

你可能犯错误这一事实并不表明,问你怎样知道你的理由曾是什么或要求你拿出你的证据,这样做一般是有意义的。虽然你也许在罕见的情况下接受公开的或私下的证据,由此表明你在你的理由问题上是错误的,但你通常并没有证据,也无话可说。这样一来,你关于你自己的行动的理由的知识一般不是归纳的,因为在有归纳的地方,便有证据。这难道就表明这种知识不是因果性的知识吗?我看不出它是这样。

因果规律不同于真实而又非规律似的普遍原则,因为它们示例证实了它们;因此归纳确实是了解规律的真实性的良好途径。但不能因此说,这是了解规律的真实性的唯一途径。在任何情况

下,要知道一个单一的因果陈述是真实的,没有必要知道规律的真实性;唯一有必要知道的是,某些涵摄眼前的事件的规律存在着。尚不明显的是:归纳,而且只是归纳,能提供那种关于满足了某些条件的因果规律存在的知识。或者换言之,正如休谟所承认的那样,一个事例就常常足以使我们相信存在着规律,这等于是说,即使没有直接的归纳根据,也能够使我们相信存在着因果关系。

E. 最后,我愿意就一些哲学家在讨论行动的原因时所感觉到某种担忧谈一点自己的看法。例如,梅尔登提出,行动常常同一于躯体的运动,而躯体的运动总是有其原因的;不过他否认这些原因就是那些行动的原因。我认为这是自相矛盾的。他是基于下述这样一种考虑而得出这一结论的:“试图通过愿望的因果有效性解释行为是徒劳无益的——它能解释的不过是进一步发生的事件,而不是行动者所实施的行动。行动者在碰到这些事件出现于其中的因果关系时,便成了发生在他身上、并为他所想到的一切事件的无用的牺牲品”(p. 128、129)。除非我弄错了,否则这个论证,如果能成立的话,就能表明行动不可能有任何原因。我无意于指出在将行动完全从因果关系领域中驱逐出去时所面临的明显的困难。但是尝试揭示这种麻烦的源泉也许是值得的。一个原因究竟为什么会使一个行动变成一个单纯的事件,使一个人变成一个无用的牺牲品? 这是否是由于我们总倾向于假定,至少在行动的活动场所,一个原因要求有一个引起者,一种能动作用要求有一个动原? 因此我们不妨提出这样的问题:如果我的行动被引起了,那是什么引起它的? 如果是我引起的,那么便会有无穷倒退的荒谬性;如果不是我引起的,那么我便是一个牺牲品。当然,可供选择的项目并

980 第八部分

未穷尽。某些原因就没有动原。人身上的状态及其变化就属于这些无动原的原因,因为它们既是理由也是原因,从而它们就使某些事件成了自由的和有意向的行动。

注 释

① 具体例证有:吉尔伯特·赖尔:《心的概念》;安斯康帕:《论意图》;S. 汉普希尔:《思想与行动》;H. 哈特与 A. 奥诺雷:《规律中的因果关系》;W. 德雷:《历史中的规律与解释》以及 R. 霍兰编辑的丛书《哲学心理学研究》中的大多数著作,包括 A. 肯尼:《行动、情感与意志》,梅尔登:《自由行动》等,文中括号中的页码均参见这些著作。

② 我们可能不把我对小偷的无意向的警示称之为行动,但是不应由此推论说,对小偷的警示因此便是某种与按开关不同的事情,它恰恰是它的结果。行动、行为和不涉及意图的事件都是一样的,因为它们常常部分地以某种终极的阶段、结果或后果的形式而被述及或定义。

在日常话语中,“行动”一词并不经常出现,当它出现时,它通常只适用于相当怪异的场合。在把行动者有意地做的事情包括有意的疏忽称之为行动时,我将遵循一种实用的哲学惯例。要在下述鸿沟上建起连接的桥梁,真正需要的是某种恰到好处的全称术语,这鸿沟是,假设“A”是关于一个行动的描述,“B”是关于自愿做某事的描述,尽管不是有意向地做某事,而“C”则是关于某种非自愿地且无意地做某事的描述;最后假设 $A=B=C$ 。那么,A、B与C就是相同的——不是吗?“行动”、“事件”和“所做的事情”,当它们与某种错误的陈述结合在一起时,至少在某些语境中,每一个都有一种奇异的循环。只有这样的问题,即“为什么你(他)作出了A?”才具有所需要的真正的普遍性。显然,如果我们像梅尔登那样假定,一种行动(“举起某人的手臂”)能等同于身体的运动(“某人的手臂举起来”),则该问题将变得更为严峻。

③ 之所以是“准意向的”,是因为除了其有意向的一面之外,在合理化解释中对行动的描述一定还会提到,否则就有这样的情况,即一行动基于某一理由做出来了,然而那行动并未完成。试比较“乔治四世知道《韦弗利》的作者写了《韦弗利》”这一描述中的“《韦弗利》的作者”。行动描述的这一语义学

特征在本书第3和第6篇论文中有进一步的讨论。

④ 安斯康帕否认关于日常活动的三段论是演绎性的。她之所以如此,部分是因为,她像亚里士多德一样认为,这种关于日常活动的三段论与关于日常活动的推论是一致的(而在我看来,它只是关于某个人据以采取行动的理由概念的分析的一部分),因此她还是坚持效法亚里士多德,认为一个关于日常活动的三段论的结论与一个判断是一致的,不仅行动有一种值得想望的特征,而且行动也是值得想望的(合理的、值得去做的,等等)。关于日常活动的推论问题,在论文2再予以讨论。

⑤ 我说“首要的一环”,旨在废除那种认为C1和C2结合起来足以界定理由对于它们所解释的的行动的关系的主张。关于这一问题的详尽讨论,可参阅引论和论文4。

⑥ 这种论证不止一种变种,如在肯尼、汉普希尔、梅尔登的论著中有之,还出现在P. 温奇的《社会科学的观念》以及R.S. 彼德斯的《动机的概念》等中。作为其形式之一,这里的论证自然是由赖尔在其《心的概念》一书中对动机的讨论所促发的。

⑦ 论文11、12和13将专门讨论这一段及前一段中的问题。

⑧ 我们不妨粗略地将这里所暗示的、对单一因果陈述的分析刻画如下:“A引起B”为真,当且仅当有关于A与B的描述,以至于通过这些关于A与B的描述放入“A引起B”之中而获得的命题是来自于一个真实的因果规律。由于并非所有真实的普遍原则都是因果规律,才使得这种分析摆脱了繁琐性;因果规律之所以被相互区别开来(当然尽管这不是分析),一方面是由于它们从归纳上为它们的例示所确证;另一方面是由于它们支持反事实和虚拟的单一因果陈述。关于因果关系问题,论文7有更进一步的论述。

参考文献

G. 赖尔(1949):《心的概念》,巴恩斯与诺布尔,纽约。

G.E. 安斯康帕(1959):《论意图》,布莱克韦尔,牛津。

S. 汉普希尔(1959):《思想与行动》,查托与温达斯,伦敦。

H. 哈特与A. 奥诺雷(1959):《规律中的因果关系》,克拉伦敦出版公司,牛津。

982 第八部分

- W. 德雷(1957):《历史中的规律与解释》,牛津大学出版社,伦敦。
- A. 肯尼(1963):《行动、情绪与意志》,劳特利奇与基根出版公司,伦敦。
- A. 梅尔登(1961):《自由行动》,劳特利奇与基根出版公司,伦敦。
- P. 温奇(1958):《社会科学的观念》,劳特利奇与基根出版公司,伦敦。
- R. 彼德斯:《动机的概念》,劳特利奇与基根出版公司,伦敦。
- D. 休谟:《人性论》,塞尔比一比格编,克拉伦敦出版公司,牛津,1951。

行动的心理原因与解释*

辛西娅·麦克唐纳

近来,有人指责说,非还原的一元论要么陷入了自相矛盾,要么陷入了副现象论。^①这种二难困境被认为是由该观点试图将下述三种显然自相矛盾的原则调和起来这一事实所使然。其一是主张,心理与物理事件之间存在着因果上的相互作用,在那里这种因果关系据说存在于一切个别事件之间。其二是坚持因果关系的法则学特征,即这样的论点:哪里有因果关系存在,哪里便有把如此相关的事件(但只是在某种特定的描述之下)连接起来的(普遍)规律。最后是这样的论点,即并不存在连接心理事件与物理事件的因果规律。(这一特征有时被描述为“关于心理的异态一元论”,它在逻辑上意味着心理学理论对于物理学理论的不可还原性。心理异态一元论包含的不仅仅是对于心物因果规律的否定;它还否定任何种类的心物规律、以及纯心理学规律的存在。在这里,我们不想就异态一元论的后两个方面发表意见,因此我们采用这一术语旨在把握与本文关系最为密切的那一个方面——即那种认为没有心物因果规律的主张。)这些原则是可以调和的,只要再采纳一个

* 本文原载 L. 斯蒂文森等编的关于行动哲学的论文集:《心灵、因果律与行动》,巴兹尔·布莱克韦尔,1986。现据此译出。

论点就行了,这论点就是主张,每一标记心理事件也即是一个物理事件。既然它可以这样来理解,即成为一个心理(或物理)事件就是以某种方式予以描述(例如,以命题态度的词汇来描述),那么这种更进一步的论点便等于断言,每一心理事件都有一种物理描述,因此为物理规律“所涵盖”。总而言之,这种观点就是主张,尽管心理属性事实上不可还原地有别于物理属性,但每一心理事件就是一个物理事件。

断言这一观点陷入了副现象论,其主要渊源有两方面:一个与因果关系的本质联系在一起,另一个则和因果效力与解释价值之间的假定的纽带有关系。化解这两种责难是一件相当棘手的事情,因为它们追随者常常把它们弄得混乱不堪。然而,因果关系的问题必须与解释的问题区分开来,所以在本文中我们将分别讨论上述两种反对意见。

I

支持非还原的一元论的论证之所以有用,是根源于因果关系的外延性和推理法则的内涵性。如果 A 与 B 是因果相关的事件,那么不管怎样,它们就是如此被描述的。不过, A 与 B 从属于其下的那种因果规律乃是由于(部分地)包涵于其中的描述而成为一条规律的。反对的人会附和说,这也许是真实的,但那只是事情的一部分。将那些在因果上相关的事件加以分解将会表明,那些事件中的某些属性并不存在因果上的关联。下面出自洪都里奇的一大段引文阐述了这一点:

抚平餐巾的,不是茶壶的年限或(原文如此)光泽,而是它的重量。使门能够反射光的東西,不是门的重量而是它的色彩。因此,对于究竟是什么引起某物这一问题的最自然的回答便是普通事物的一种属性。不过,有必要立即予以拒斥的是,这里所说的东西是一种一般性的属性,一种普遍的属性。抚平餐巾的东西并不是重达1磅这样的一般属性,因为这种一般属性是重为1磅的茶壶之外的东西。如果该茶壶的重量发生了变化而该餐巾又没有被抚平,那么就有一般属性……这一来,我们便进到了这样的观念,即抚平该餐巾的是该茶壶的重量,亦即该茶壶的个别的属性……它不是那个茶壶的全部性质,或者说它重为1磅之外的任何个别的属性,而是重为1磅这一一般属性的一个例示……严格说来,原因就是个别的属性……^②

心理的因果关系的真谛在于:如果心理事件与物理事件真的能相互影响,那么它们一定是由于它们的某些(而不是全部)属性而如此的。而且这些属性,即“因果相关”的属性,是在连结这样一些事件的规律中可能出现的属性的主要候选者。事实上,洪都里奇认为:^③“关于因果相关属性的法则学特征的原则”是关于因果关系的任何合理说明所必需的。这样一来,这种对因果性与规律之间的关系的理解便为异态一元论带来了麻烦。因为随之而来的问题是:究竟是由于哪一些属性,一个心理事件才与物理事件发生因果上的相互影响?如果回答是“心理的属性”,那么接踵而至的似乎是对异态论的否定,因为它是关于出现在连结因果相关的事件的规律中的那些属性的表达式。另一方面,如果回答是“物理的

属性”，那么任何已知心理事件的“心性”似乎不再相关于它与别的事件的相互影响，进而副现象论似乎便粉墨登场了。^④

我们不妨假设，因果性与规律之间的关系是已述的那种关系。这一来，心物之间的无规律会导致心理的无用吗？对此不乏更深入的论证，而我们认为这种更深入的论证依赖于二元论的偏见。长期被忽略的是属性与其例示之间的至关重要的区分。洪都里奇正确地强调，正是那个得到例示的属性（“个别的属性”）才具有因果上的有效性。不过，当他转向对非还原的一元论的抨击时，似乎又忘记了这一点。如果我们真的坚持认为，因果关系是标记事件之间的关系，而且正是与事件类型相联系的属性的例示具有因果上的有效性，那么那种“关于因果相关属性的法则学特征的原则”便应予修正，以便能完成“……因果相关属性的例示”。这种修正将使那种注定会导致副现象论的论证失去效力。一些实例也许有助于我们弄清事情为什么会是这样。

与任何事件类型——诸如射击——相关联的便是属性，例如正在完成一项射击的属性。这种属性将由作为那种类型的标记的任何事件所例示。而且，任何标记事件都“有”这样一种属性，这种涵义就在于它是那个属性的例示这一事实。例如，1984年3月4日星期日午后12:00约翰射中乔，便是这种类型即“射击”的一个标记，而且那一事件由于例示了它而具有作为一项射击活动的属性。^⑤

假设有一个心理事件，如苏珊想要喝酒，引起了苏珊的身体上的物理事件，比如说她的手臂运动。那么有一点可以肯定，作为属性的例示的一个事件，即想要喝酒（也许是其他饮料），就在因果上

相关于作为属性的例示的一个事件,即苏珊的手臂运动。但是确定无疑的是,任何事件都可以是一种以上属性的例示。即是说,不同的属性可以在同一个“个别属性”(用洪都里奇的术语说)中得到例示。例如,苏珊想要饮酒的愿望不仅可以是作为饮酒愿望这一属性的一个例示,它还可以是不同的属性如作为饮水的愿望的例示。约翰向乔射击是作为一种射击活动这一属性的一个例示;但它也完全可以作为手指的运动这一属性的例示,以及作为扣动扳机活动的属性的例示。

关于因果关系的法则学特征的原则要求的是,相互之间具有因果关联的事件必须是出现在因果规律中的表达式所述及的属性的例示。它并不要求,为一个事件所例示的每一个属性都必须如此:更确切地说,洪都里奇关于“因果相关”的属性的描述在一定程度上就是想阐明这一观点。而关于因果上相互作用的原则要求的则是,因果上相关的事件由于它们的某些属性而必须如此;因此,如果作为属性的例示的任何事件,如有一种想要饮酒的愿望,引起了一个也是作为属性之例示的事件,如手臂的运动,那么由于它的某些属性,它就一定是如此。但是,这里的“属性”一词在属性与其例示之间模棱两可。既然任何事件都能够是两种或更多的不同属性的例示,那么情况也完全可能是:同一个事件既可以是作为想饮酒的愿望这一属性的例示,也可以是另一属性,比如说作为一种大脑事件的物理属性的例示,在这里,作为前一种属性的一种例示正好也是作为后一种属性的一种例示。这就是说,不同的属性是可以共例示的;就这个特定的事件而言,它们“分享”着同样的例示。它们在属性层面上的差异是由两种属性所决定的,而不总是共例

示的。这样,想饮酒(一种心理属性)这种属性的例示将符合关于“属性的因果相关例示的法则学特征原则”的要求;由于该例示为规律所涵盖,因而它是因果相关的。

这样的心理事件的“心性”究竟在何种意义上不再是因果上有效的呢?是在这样的意义上吗,即心理属性如想要饮酒的属性不能包括在连接心理的东西与物理的东西的规律中?这完全忽视了下述观点,即因果关系乃是标记事件之间的关系,因此也许是某些属性的例示将一些事件与其他事件因果地连接起来,也许不是。心理属性的例示完全可以具有因果上的效力。而且,如果一种心理属性的例示正是一种物理属性的例示,那么尽管这些属性本身有差别,但异态一元论者便能理所当然地坚持说,前者既可以具有因果上的有效性,又可以是能够涵摄在因果规律中的属性的一个例示。在我们看来,坚持别的观点就等于主张,作为心理属性之例示的事件绝不可能是(同一于)一种物理属性的例示。而这究竟是否更甚于二元论的偏见一时还无法明晰。

这种回答也许是不能令人信服的。其理由就在于存在着作出下述假定的倾向,即认为,为了使一种心理属性在因果上有效,其任何例示都必须有别于一种物理属性的任何例示。因此,如果每一心理事件都是一种物理事件,那么任何作为心理的东西的事件的因果有效性将似乎要求:它必须同时既是一种心理属性的例示,又是一种物理属性的(不同的)例示。

但是,我们为什么要假定事情一定是这样呢?假设心理属性随附于物理属性,其意思大致是说,如果任何两个对象(如有机体)相对于后者来说是难以辨别的,那么它们在前者那里就是不会有

歧异的。^⑥进而便有理由假定：虽然心理属性与任何单个的物理属性未能整体地相互关联起来（更不用说以任何类似规律的方式），但它们可以与这样一些属性的一种析取（也许是一种无限的析取）^⑦相互关联。这促发了下述思想。许多物理属性，如作为有颜色的或作为动物的属性，显然有别于它们可以表现出的许多特定形式（诸如是红色的或绿色的，或是一只老虎），因为由一个对象对一种更确定的属性（例如红色的属性）的占有可衍推出对更易于确定的属性（例如有色彩的）的占有，而反过来则不是如此。不过，没有人愿意假定，一个对象的彩色范例比如说红的范例首先要求，它必须是作为红色的属性的一种例示；其次要求，它必须是第二种有关的属性即作为彩色的属性的一种（不同的）例示。如果是这样，作为前者的一种范例，恰恰也是作为后者的一种范例，尽管这些属性本身有差异。这是否意味着彩色在因果上是无效的呢？当然不是这样；因为因果上有效的任何事例——在其中，那种属性的一种更确定的形式得到了例证——也就是这样的事例，即就因果关系的外延来说，彩色本身的范例在其中是有效的（试想具有重量的属性与重达 2.3 公斤的属性）。

同样，还可以论证说，心理属性以一—多的方式与物理属性相互关联着（尽管不是以整体的方式），亦即与下述推断相互关联着，这推断是，前者的任何例示同时也是一种或另一种更确定的物理属性的一个例示。正是由于红色也就是彩色，因此人们可能会说，成为那种属性的例示，即成为一大脑事件 B，可以成为那个心理属性的例示，如有疼痛。而且，如果前者的例示在因果上是有效的，那么后者也将同样如此。人们可能被诱使作出这样的责难，即这

种红颜色的情况完全不同于上述所设想的大脑事件 B 或疼痛的实例,因为正是在前者那里而不是在后者那里,一个对象(有机体)对被随附的(“基础的”)属性(红)的占有在逻辑上可衍推出对随附属性(彩色)的占有。但是这一点,即便是真实的,也与现在的主题没有关系。这里所描述的随附性,指的是属性之间的这样的关系(或者说是一种属性与一组(不同的)属性之间的关系),以至于如果任何两个对象(有机体)对被随附的属性来说是没有区别的话,那么它们在对随附的属性的占有方面便是不可能有所歧异的。不过,关于“不能”这一模态的理由或根据,会因情况变化而相互有别,关于这一点,我们将在第三部分再予以阐述。

我们早就说过,一旦副现象论者的论证以必要的方式得到补充,它仅仅通过二元论的偏见就能达到所企求的结论。坚持心理属性的例示一定有别于物理属性的例示,其实就是展示这样一种偏见。可怕的是异态一元论的一贯性。我们相信,即使我们赞成“关于属性的因果相关的例示的法则学特征的原则”,但我们已经说明了,它是一贯的。^⑧不过,通向副现象论的结论有一种不同的路径,即经过对行动的解释的讨论而迂回进行的途径。我们下文要讨论的正是这一点。

II

有人断言,对副现象论的否定必须依赖于对“对象据以被假定有相互作用的那个特征的内涵解释的恰当性”,^⑨或者说依赖于这样的“确信,即相信作为心理的事件乃是任何关于行动的充分解释

的不可排除的一部分”。^⑩这些表述意味着,这种反对意见有两个组成部分。其一是心理主义的因果解释是不可能的;其二是即使这是可能的,也将是多余的,因而是可以排除的。我们将依次讨论这些反对意见。

也许可以认为,有这样一种非常简捷的方法,心理解释的有效性可以以此建立起来,这种方法走的就是那种常见的“理由即原因,因此理由—类型的解释即因果解释”的道路。^⑪我们的主张是,只有某种关于单一因果解释的说明被接受时,这一点才能被认可。理由就是行动的物理原因这一假设,显然不足以证明那种认为理由—类型的解释即是因果解释的断言。人们不得不说明下述两点以补充这一假定,(a)理由—类型的解释即是解释,(b)它们即是因果解释。^⑫就我们的目的而言,只须将讨论的中心集中于围绕因果解释所产生的问题,特别是心理的因果解释是否可能,以及如果可能,它是否基本上是多余的这样的问题,就足够了。

提供因果解释有两种不同的途径;其一是通过因果规律的途径(法则学上的因果解释),其二是通过单一因果陈述的途径(即单一因果解释)。副现象论者的反对意见很显然集中在前者上。根据这种观点,两个在因果上相关的事件,A与B,只有在某些描述如A'与B'之下才能例示因果规律。这些描述构成了详细说明这些事件所从属的规律的条件句的前件与后件。粗略地说,当一个人拥有这种类型的解释时,他将有理由预期,每当事件A的、与原因有关的属性出现时,事件B的、“与结果有关的”属性将随之出现。正是这些关于属性的标准的描述,占据着法则学上的条件句的前件与后件的位置。

如果这就是对因果解释来说所能有的一切,那么很显然,解释的副现象论将是根源于赞成对心-物因果规律的否定。不过,也许有人责难说,上述关于法则学的因果解释的说明太苛刻了;由于不是 A' 和 B' 构成似规律陈述的前件和后件,因此这种反对意见进一步指出,人们能够得到对 A 和 B 的描述,这些描述接近于那个规律中所例示的那些描述。例如,“砖头的抛出”能解释“窗户破碎”,即使“砖头的抛出”和“窗户破碎”并未包含在似规律的陈述中,如果从该解释中所用的词汇到规律的严格陈述中所用的词汇,存在着一种相当平和的转换的话。我们乐于接受这种修正,但它未必能帮助异态一元论者,因为正是心理词汇的这种特殊性常用来论证还原的不可能性。不管“平和的转换”可能意味着什么,它也未必能够消融内涵的词汇与外延的词汇之间的尖锐冲突。由此人们不得不推断说,我们的一元论者必须放弃那种认为人们能够拥有这种类型的心理的和因果的解释的主张。

这一结论并不令人诧异,因为它正是这种认为心-物因果规律并无用处的一元论原理的一个组成部分。^③不过,因果解释有可能从这种结论中幸免于难,因为在单一的因果解释中,我们可以发现足够的外延来保证下述结论的成立:即从心理上得到描述的事件 A,可解释从物理上描述的事件 B 为什么发生。在某种程度上,一项决定总是以这样或那样的方式依赖于人们所理解的“解释”这一术语所起的作用。如果某人希望所有的解释都有一种预见功能,那么他就不大可能将这种类型的单一因果解释视为纯粹的解释。其理由是不难明白的:该外延(更严格地说,指称的透明性)所利用的同一性只是一种标记-标记同一性。即是说,心理的和物

理的属性并不同一,而只是它们的例示相同一。如果表示心理属性的表达式不属于涵盖了类似因果规律陈述的那个类型,那么对(从心理上描述的)此事件解释(从物理上描述的)彼事件为什么发生的详细说明,将不会使解释者有条件预见:一旦此心理属性再发生时,彼物理属性也将再次发生。不过,如果人们撇开预见,只考虑解释,那就没有理由不将这种情况视为纯粹的解释。毕竟,它起到了一种因果解释必须起的作用;它详细说明了在待解释的事物中何种事件引起了所描述的后果。更重要的是,它还认可了这样的反事实,即如果在该解释中被提到的事件没有发生,那么情况就会完全不同。它之所以如此,是因为被提到的属性的例示(如想要喝杯饮料)在因果上是有效的。倘若这个例示不发生,那么(除了由多种因素决定的事例),被解释的事件(她的手臂伸向水)也将不会发生(或至少只有很小的可能性)。一般意义上的科学解释也具有这种单一因果解释的特征。简言之,“我们可以说,科学解释,如同单一因果解释一样……总是关于实际的结果而非某些其他可能的结果为什么发生的解释;这样的解释总是这样进行的,即说明:基于实际上已知的条件,实际的结果比它在其他情况下更偏向于这些选择。”^④

诱使人们以为这种解释缺乏基本构成成分的东西常常是源于这样的一种感觉,即除非一种解释被证明是正当的,否则它便不是真正的解释。而要作出证明又有赖于预见性和规律性。在没有这种证明的地方,人们就会被诱使着认为,我们的确没有解释。不过,在这种情况下,需要证明为正当的并不是心理和物质属性之间的恒常连结,而是那种认为此心理事件引起彼物理事件的主张。

后一种主张,假如是一元论,显然是可检验的,不管心理的和物理的属性之间的恒常连结是不是出现。因此,单一因果解释是可能的,即使在该解释中所采用的描述指的并不是法则学上相关的属性。

完全可以感觉到,甚至在这种解释中,在一种既定的情形之下,并非任何关于原因的描述都是解释性的。这无疑是正确的,但其本身又不妨碍借助于有关的心理描述而遴选原因的真正解释的可能性。一种无关的、而且完全可能失了解释意图的描述,就是这样的描述,它把原因说成是这样的事件,比如说“我的祖母今天早晨嚼烤面包时想到的那个事件。”这是一个基本上难以理解的遴选原因的方式,人们几乎不会被诱使说,那个描述能让我们洞察被指称的事件的因果相关特征的本质。它可能曾经是解释性的,那可能仅仅根源于解释的实用本质。如果某个需要这种解释的人对于所指称的事件的本质具有独立的知识,进而能在所给予的描述下将它分辨出来,那么我们就可以那样看待它。不过,即使我们足够宽容地承认,在这种情况下,如果一种解释被给予了,那么也仅仅是由于提问者所能获得的独立的知识的许可才奏效的。如果心理主义的因果解释遵循这种模式,那么它确实是值得怀疑的。但是幸运的是,并没有理由怀疑它是这样。常常涉及到这样一些解释的描述提供了接近那些属性的直接通道,这些属性的例示在因果上是有效的。即使不求助于关于该原因的本质的独立的知识,那些探寻解释的人也能如愿以偿。(他们甚至不必知道,因果上相关的例示同时也是一种物理属性的例示。)

如果副现象论者的反对意见归结起来就是这种认为任何心理

因果解释都不可能的论断,那么单一因果解释的有效性将足以平息这种反对意见。不过,还有一种不同形式的反对意见,它考虑到了这类解释的可能性,但坚持认为它们基本上是多餘的。^⑮据说,这种多餘性是源自物理理论的无所不包的本质。对于未来的任何物理状态,预言这种状态的最理想的理论将是物理理论,即物理学。根据一元论,物理学仿佛就是我们所需要的全部,而所有其他关于将来的状态为什么会发生的解释好像基本上都是可以排除的。

这种反对意见好像能被轻易地驳倒,因为它似乎采取了下述形式:(a)心理解释对解释行动似乎是必不可少的;(b)物理学的解释可以解释行为;(c)任何行动都同一于某些行为;(d)因此物理学的解释能够解释行动;(e)这样一来,心理解释便是多餘的了。鉴于解释的假定的内涵性,因而从(c)到(d)的演进是可疑的。

不过,这种观点更为复杂。甚至在经典的演绎法则学解释中,也容许某些替换形式。尤其是,已有论证表明,替换形式由于(I)保持了属性同一性,(II)包含了法则学上相关的属性,因而也许是合法的。^⑯很显然,(I)并没有为那个替换形式提供必需的说明根据,而且非还原的一元论者好像能避免(II)所特许的替换形式,其方式是坚持主张,随附性能与还原清楚地区别开来,因为它的目的就是要避免基础属性与随附属性之间的这样一些法则学上的联系。这一目标是否能实现,以及这是否确实是它的目标,仍是聚讼纷纭的问题,因此,我们现在必须转向对随附性问题的讨论。

III

随附性已有多种多样的定义,详尽地评述所有这些定义超出了本文的范围。^⑦我们现在的兴趣主要限于下述问题,即在对随附性的任何合理的阐述中,法则学上的联系是否必要?如果是必要的,那么这与非还原一元论的一个重要特征是否相矛盾?如果法则学的联系被证明是必要的,那么这第二个问题似乎是可以解决的(与非还原的一元论相反);这似乎正是黑尔和布莱克本两人的立场。^⑧但是,戴维森的异态一元论的基本前提是这样一些前提,它们要求对因果相互作用提供似规律的支持,并在否定任何心理主义-物理主义的规律的有效性的同时,断言心-物因果相互作用能够发生。这种被认为是无效的规律类型显然具有因果关系的性质,而且应该明确的是,如果在基础(物理的)属性与随附于它的(心理的)属性之间有一种法则学上的联系,那么它就不是一种因果上的法则学联系。我们关于随附性的种种定义要求的是,心理的东西是与物理的基础同时得到实现的;它们并不认为,随附属性之形成是因为物理属性的一种更早的构形。

现在情况将很可能是这样:如果心理的东西与物理的东西具有法则学上的联系,那么一种派生的“暂时的规律”便会应运而生。如果我们有下述这些法则学的条件句:

$$(1) P1 \ t^1 \rightarrow M1 \ t^1$$

$$(2) P2 \ t^2 \rightarrow M2 \ t^2$$

$$(3) P1 \ t^1 \rightarrow P2 \ t^2$$

那么将可能推出:

$$(4) P1 \ t^1 \rightarrow M2 \ t^2$$

(其中 P 与 M 分别表示的是物理的和心理的状态,且 $t^1 < t^2$).

这样推导出的规律的存在不一定威胁那些前提,因为“无因果规律”的要求完全可以改写成“没有非派生的因果规律”,进而该论证便能成立(当然,对因果规律要求的修改表明,非派生因果规律对于支持因果相互作用来说是必不可少的)。但是,还应该指出的是,在这种对非还原的一元论的辩护中存在着一种隐伏的困难。戴维森对心-物因果规律存在的否定似乎依赖于那些一般不利于心-物法则学联系的主张(例如,关于心理的整体论,以及心理的和物理的术语对应于不同的示证规则)。如果随附性需要这样一种联系,那么要否定连接两者的(非派生的)因果规律,就不得不寻找另外的根据。然而,随附性需要法则学的关系吗?

我们的讨论将依照 J. 金近来尝试的关于随附性的两种不同概念的定义而展开。“弱随附性”可以这样予以描述,即对于两个非空属性族 A 与 B 来说:

“A 弱随附于 B,当且仅当,对于任何 x 和 y 都必然是,如果 x 和 y 都具有 B 中的所有属性,则 x 和 y 也都具有 A 中的所有属性——即是说,关于 B 的不可辨别性可衍推出关于 A 的不可辨别性。”^⑨

对我们来说有关的一点在于,弱随附性实在太弱,因而不能为我们相信物理的东西决定心理的东西提供任何理由。正如定义所述,只是在任何可能世界内部,它妨碍了两事物在基础属性中的一致的可能性,而同时又妨碍了它们在随附属性中的相异的可能性。

换言之,跨世界 B 的不可辨别性容许跨世界 A 的可辨别性。这与我们的下述直觉是背道而驰的,这直觉是,如果 A 随附于 B,则若任何对象变得与 B 类似,那么它也将变得与 A 类似。即是说,任何具有有关的 B 属性的世界中的任何对象,相对于 A 属性来说,与任何具有那些 B 属性的世界中的所有那些其他的对象应该是不可分辨的。如果事实不是这样,那么我们便面临着一个难题:那些世界为什么被如此分隔开了?用布莱克本的话说,为什么没有“混合的”世界?^②对于这一问题似乎没有合理的回答,这表明关于随附性的界定有必要像下面这样再加强一些:

“A 强随附于 B,仅仅在这样的条件下,且必然地,对于每一个 x 和 A 中的每一个属性 F 来说,如果 x 有 F,那么 B 中便有一种属性 G,以致 x 有 G,而且必然地,如果任何 y 有 G,那么它便也有 F。”^②

强随附性与弱随附性之间的根本区别在于这第二个“必然地”,很显然,正是这一点导致人们主张,随附性需要法则学上的关联。因而它还引出了这样的建议,即(物理主义者)关于行为的解释便足够了,而关于行动的心理主义解释则可能被取而代之。如果在上述的界定中,人们将“G”当作是一种物理属性(或物理属性的一种合取),将“F”当作是一种行动-属性,那么关于某物为什么是 G 的一种解释便应该能解释它为什么是 F。或许,这就是多余论者的主张。

当然,事情决非如此简单。还存在着这样一些关于被随附的基础之中的属性的无穷集合的问题,如人们能否完成关于心理属性的布尔运算的问题;从这些布尔运算中能获得什么结果的问题

(例如,物理属性本身的补集还是一种物理属性吗?物理属性的合取能形成一种自然类别这样的东西吗?)。我们将假设,这些疑问能够得到令人满意的解决,以便全神贯注于两个相关的争论,即被设定的认识论问题与所涉及的必然性的本质问题。我们之所以说这些问题是相关的,是因为如果这种必然性是一种科学规律的必然性,那么人们从关于基础属性的知识中“迅速读出”随附的属性,这似乎是不成问题的。如果人们以为随附的唯物主义像还原的唯物主义那样发生作用,那么这种观点就得到了加强。在后者那里,还原的发现旨在为唯物主义提供辩护;因此可以认为,对前一种唯物主义的辩护有赖于对这些随附的依赖性的发现。

我们的看法是,这一点是错误的,因为随附性之被引入是为了使基于不同的理由而接受的唯物主义保持一贯性。为了证明唯物主义的正确,没有必要发现这些随附的依赖性。它们更有可能是基于推论的根据而约定的,这些根据是相当普遍的,从而使随附的唯物主义成为一种合理的形而上学命题。我们无需深入讨论这种形而上学命题的理由,但的确有必要注意到,如果上面的论述乃是关于随附的唯物主义的一般定位的正确描绘,那么所涉及到的那种类型的必然性就将是形而上学的必然性。

这与解释的多余性问题有什么关系呢?这种认为多余的反对意见的要点在于,随附性必然导致对心理主义解释的排除。这种主张由于下述假设而貌似有理,这假设是,假如我们能从物理上解释所有行为,则我们便能解释行为在法则学上与之关联的东西。我们已经说过,只有当这些依赖性是被认识的时候,这种替换的假设才能得到辩护,同样,只有当这些依赖性类似于科学上的规律

1000 第八部分

时,它们才能被认识。如果它们不是这样,而事实上是形而上学的规律,那么那个认识论的观点就是莫明其妙的。其实,心理的属性无需从物理的基础中“迅速地读出来”。如果它们不是这样,那么这将使这种替代性的论证变得苍白无力。心理解释在某一天是否将被证明是多余的这一问题由此也就成了一个悬而未决的问题。

IV

我们关于这种主张多余的反对意见的讨论尚不是定论。在文章的结尾部分,我们最好是来看看这种反对意见的效力。人们会被诱使认为,如果物理理论比其心理主义的对立面更有力量,那么对物理主义的解释来说,心理主义的解释便只是一个填空补缺的东西,由于目前的无知,它将出现,但暂时又无法得到。^②目前,隐藏在这种工具主义的反驳之下的是那种实证主义原则的残余,根据此原则,对于专门科学唯一能采取的态度就是,要么认为它们专门探讨一个不同于物理理论的论题(二元论),要么认为它们(在原则上)可还原为物理理论,因而是可以排除的。但是这似乎被证明太过分了:它不仅不利于预期的目标即心理主义解释,而且有碍于所有那些主张存在着某种不同于物理理论的独立性的科学。简言之,它将一种工具主义的阐释强加到所有那些不合法则学上的因果模式的解释之上,例如生物学理论中的功能主义解释。^③如果——似乎也可能如此——任何声称探讨随附于物理属性的属性(但不能还原于它)的科学,将被禁止采用法则学上的因果解释,那么任何关于随附性的科学也将被认为提供了基本上多

余的解释。

尽管工具主义的一般问题在这里不可能得到裁决,但我们所述的足以表明这种反对意见的力量(反对所有“关于随附的科学”)与源泉(二元论或还原)。当然,更讲究实际的人也可能乐于主张,只有物理学能为我们提供真正的解释,而其他的科学在原则上一定能还原为物理学。由于这种主张为中间的立场为什么是稳固的、随附性为什么能够是真实的、不可还原的、且不依赖于二元论的本体论等问题提供了一种论证,便将它嘲笑为一种实证主义的偏见,其实是无济于事的。人们不得不加以辩护的论断是:一组属性不是可还原的但却是真实的;否则,例如通过解释随附性怎么可能独立于随附的属性的假设的真实性而存在,就能驳倒这一观点。^①假如这种辩护的本质将依有争论的随附属性的种类的不同而不同,那么一种一般的辩护便似乎是不可能的。在这样一些辩护中,尽管我们认为最为重要的东西是被假设为真实的随附属性的因果地位,但在本文的第一部分中,我们所提供的正是这样一种说明,即说明心理的东西的因果地位如何在反击副现象论的那种进攻中得到保护。在其后的几个部分里,我们已经表明,副现象论者在否定心理主义解释具有真正的解释力的过程中,正好也否定了在二元论和还原论之间存在着一种稳固的中间立场的可能性。这样一种折中方案显然是可能的,但是其现实性不得不依据充分的理由逐一予以论证。本文所论证的是,非还原的一元论的根据并不像副现象论者的反对意见试图使我们相信的那样脆弱。

注 释

① 例如,参见泰德·洪都里奇:“论异态一元论”,《分析》,42(1982),pp.

1002 第八部分

59-64;以及“心理的类规律联系及其问题”,《探索》,24(1981),pp.277-304;H. 鲁滨逊:《物质与感觉》(剑桥,1982),特别是其中 pp.8-13;还参见 J. 金:“心-身问题中的因果性,同一性与随附性”,载《中西部哲学研究》,4(明尼波利斯,1979),pp.31-50(特别是其中 p.47)。关于异态一元论的经典表述,参见 D. 戴维森:“心理事件”,载福斯特与斯旺森编:《经验与理论》(阿姆斯特,1970),pp.79-101,重印于戴维森本人的《关于行动与事件的论文集》(牛津,1980)。

② T. 洪都里奇:“原因与假设 P,即使 x,依然为 q”,《哲学》,57(1982),p.292。

③ 参见 T. 洪都里奇:“论异态一元论”,同前,p.62。

④ H. 鲁滨逊表达了一种几乎完全相同的批评:“……相互作用并不取决于某种外延的因果关系,而取决于关于下述特征的内涵解释性关联性,正是由于这一特征,那些对象才发生假定的相互作用。”同前,p.12。还可参阅 J. 金:“疼痛只是呻吟的随附的原因,它的因果作用依赖于它的中枢关联物的因果作用,后一作用对呻吟有直接的因果作用……从某种重要的意义上说,这种疼痛并没有独立的因果力量……。”不过,金的结论比洪都里奇或鲁滨逊的结论更肯定一些。“如果这就是副现象论”,他说,“那就让我们充分利用它吧!”金,同前,pp.46-47。

⑤ 一方面的事件类型、标记与另一方面的与这些类型相关的属性及其例示之间的类似性,不管是赞同戴维森式的事件类型观还是赞同金的事件类型观,都是成立的。二者都将事件解释为非重复的、有时间性的特殊事例,因而是能占有属性的个体。

⑥ 这便是金在“论心理物理随附性”一文中关于它的描述,载《哲学研究》41(1982),pp.51-70,也可参见《美国哲学季刊》15(1978),pp.149-56。有许多困难与随附性概念紧密相联(例如,参见 H. 鲁滨逊在《物质与感觉》一书所作的讨论,同前),但是这些困难对这里的讨论无足轻重,因此,我们在这里将不着力于这些问题的讨论。

⑦ 例如,参见金:“随附性与法则学上的不可通约性”,同前。

⑧ 在我们看来,正是这种在属性与其例示之间的区分中的失败,构成了诸如洪都里奇和史密斯这样一些人争论的核心。参见洪德里奇:“异态一元

论:对史密斯的答复”,《分析》43(1983),pp.147-149;以及“史密斯与莫夫的成功”,《分析》44(1984),pp.86-9;彼特·史密斯:“对异态一元论是坏事吗?”《分析》42(1982),pp.220-4;以及“异态一元论与副现象论:对洪都里奇的答复”,《分析》44(1984),pp.83-6。

⑨ 参见鲁滨逊,同前,p.12。

⑩ 洪都里奇:“论异态一元论”,p.64。

⑪ 试比较 S.P. 斯蒂克的论述,他宣称,“关于行为的民众心理理解释应当理解为日常的因果解释”,这一断言现在看来乃是“一种非常正统的观点”,载 A. 伍德菲尔德编:《思想与对象》(牛津,1982),p.156,我们反对这种“正统观点”。

⑫ 关于认为说明(a)便足够了的人,可参阅 C. 麦金:“行动及其解释”,载 N. 博尔顿编:《心理学中的哲学问题》(伦敦,1979)。

⑬ 这通常被陈述为对心-物规律的否定。不过,并非所有的规律都具有因果关系的性质,因此我们的更有限的否定本身并不导向彻底的异态论。人们可能想要坚持的是,存在着心-物属性例示的规律。这并不影响我们的论证。

⑭ 参见 J. 伍德沃德:“论单一因果解释”,《认识》21(1984),p.240。我们同意伍德沃德关于单一因果解释这一特征的阐述,但拒绝他关于其外延的论述,他认为其外延依赖于那种关于保留了“对比中心”的句子的概念。

⑮ 洪都里奇(在其“论异态一元论”中)似乎就采取这一路线,认为心理主义者的因果解释是多余的,同样的观点也见于金的“心-身问题中的因果关系、同一性与随附性”一文。

⑯ 试比较 J. 伍德沃德:《解释的不对称性》,载《科学哲学》51(1984),pp.421-442。伍德沃德关心的是量值的同等性,因此我们的术语表达将与他的有所不同。

⑰ 关于这一问题的概览和文献材料,最好参见:《南方哲学杂志》,增刊, T. 霍根编(1983年斯平都(Spindel)会议:随附性)以及 J. 金:《随附性的概念》,载《哲学与现象学研究》45(1984),pp.153-177。

⑱ 参见 R.H. 黑尔:“论随附性”,载《亚里士多德学会增刊》,58(1984),特别是其中 pp.15-16;以及 S. 布莱克本:《随附性问题再思考》,载 I. 哈金

1004 第八部分

编:《分析的思潮》(剑桥,1985),尤其是其中的 pp.58-61。

① 参见金:“随附性的概念”,p.158。

② 参见 S. 布莱克本,同前。

③ 参见金:“随附性的概念”,p.165。事实上,我们并不相信这种定义是充分的,但这种不充分性与我们现在所讨论的问题并无干系。

④ 这似乎是丹尼特的观点,参见其《脑猝病》(剑桥,麻省理工学院出版社,1978)。

⑤ 关于某些生物学理论中有说服力的非还原论观点,参见 J. 杜普雷:“科学的不统一性”,载《心灵》92,pp.321-46。

⑥ 参见 S. 布莱克本关于道德属性的随附性的说明,载《词语的传播》(牛津,1984,第6章)。

第九部分

心理解释与民众 心理学的地位

民众心理学及其职责*

威廉·利康

我被分派的任务几乎是一项痛苦的、低级的工作：为围绕“民众心理学”与科学心理学/生物学之间的关系而展开的几个争论勾勒一幅全景图。

第一个争论,更确切地说一个前-争论,就是:作为人文科学术语的“民众心理学”实际上包含哪些意思?^①这个问题几乎没有受到特别的关注。我将按力度或倾向性上升的顺序罗列一些各具特色但又很实在的描述。

什么是“民众心理学”?

直到最近,民众心理学一直建设性地被理解为一种学问或学说的集合,不管其内容概念中存在着多少未被承认的变数。不过,B. 冯·埃克德(1994)极力论证过一种几乎未获认可的用法,因此,我将从她的描述开始。

(a)“在最低限度上(……),它指的是一组归因性的、解释性的和预言性的实践(针对人自己和他人的心理状态与外显行为),以

* 选自 M. 卡里尔与皮特·K. 马查默编:《心灵图景:哲学·科学与心灵》,匹兹堡大学出版社,1997。

1008 第九部分

及在那些实践中所用的一组观念或概念”(1994, p. 300)。

(b)关于心灵的“老生常识”或“常识性”普遍原则之集合,至少为大多数人心照不宣地所接受,以及有望被接受(刘易斯,1972)。

(c)关于心灵的极不严格的因果上的(原因的)老生常谈或常识性普遍原则之集合,而不武断地裁决“理由是不是原因”这样的问题(丹尼特,1978;1987;1991)。

(d)关于心灵的更严格的、因果的、常识性的普遍原则之集合,这是按照对莫顿(1980)最初所说的“‘理论’理论”^{*}的某种描述所作的理解(塞拉斯,1956;1963;^②阿姆斯特朗,1968;刘易斯 1994; P.M. 丘奇兰德,1981;1991;1994;P.S. 丘奇兰德,1986)。丹尼特(1991b)持折中态度。

在这一点上,不同的哲学观点已开始集中在“民众”一词上。^③

(d')或者说,一组更严格的因果普遍原则,可由关注人的本质的普通的观察者得到,并可予以陈述,而无须利用任何有自我意识的科学方法(P.M. 丘奇兰德,1994)。

正如福多(1981;1987)所补充的,民众心理学在此意义上是适宜于溶入科学的心理学之中的。

(e)一些更带倾向性的哲学观点,如表征主义认为,(d)或(d')有资格成为关于表征的“思想语言”说明,成为关于因果性的超强理解,成为这样的论点,即命题态度主要是根据它们的内容而行使其因果作用的(福多,1987),还可成为关于功能离散、模块性甚或

* 原文为“‘Theory’ Theory”或“‘Theory’-Theory”,指的是对民众心理学的阐释、重构。如果民众心理学是一种理论,那么对它的阐释、说明自然可视之为“理论”的理论。

形态学特征的论断(拉姆齐,斯蒂克和加龙,1990——对此我不打算作什么补充)。

我想,下一步就是相信民众忠诚地拥戴这样的主张,即J. 福多关于心灵所说过的一切都是对的。

但是我曾提到“‘理论’理论”。这里还要谈及,直到最近,哲学家们仍然假定,应当有一种相当明确的用法;哎呀,并非如此。

什么是“‘理论’理论”

还是按追求的目标的上升次序加以叙述:

(i)主张:民众心理学(无论在什么意义上)是一组经验信息,存在于我们关于可观察的物理对象及其可观察的属性的经验信息之外。(除了分析行为主义者,大概还有丹尼特(1978;1987),此外再没有一个人否定这一点。)

(ii)主张:民众心理学(无论在什么意义上)是对人的行为的“因为”解释,如前述一样,它不武断地裁决理由是不是原因这样的问题。(丹尼特肯定持此论。)

(iii)主张:民众心理学(无论在什么意义上)是更严格地根据原因所作的解释(这里也有中间的立场:丹尼特 1991b;利康 1993,它辩护的是以醒目的首字母缩略词“GASP”命名的理论。)

(iv)主张:民众心理学(无论在什么意义上)是从原因上所作的解释,其作者好像假定了无法观察到的事物(塞拉斯,1956;1963;奎因 1960;1975)

(v)主张:民众心理学(无论在什么意义上)是借助(至少是心

1010 第九部分

照不宣地)假定的、无法观察到的事物,即人们为了解释他人的行为实际上假定的内在状态和事件,从原因上所作的解释(刘易斯,1972,阿姆斯特朗,1968,福多 1981,以及后塞拉斯主义者)。

“理论”——理论家承认:大约在(iii)与(iv)之间,相对于日常宏观物理观察而言,民众心理学具有经验上的可错性。也正是由于这一点,我更加明确地把一个与冯·埃克德(1994)所强调的另一个区分开来,所依据的是,人们是不是把内在理论当作是在某处被明确表达的东西,即使我知道,没有这样的哲学家,他假定任何民众理论都是如此被表达的。

(vi)是(v)与下述主张的相加,即心理术语是民众心理学的理论术语,并在语义上以理论术语一般起作用的方式起作用。(刘易斯 1972,阿姆斯特朗 1968,福多 1981。丘奇兰德 1981,斯蒂克 1983,利康 1988 等。)

(vii)是(vi)加上民众心理学与有自我意识的科学理论之间的某种更强的类比。(我知道,没有人会严肃地提出这样的主张,即使批评“理论”理论的某些人有时提出了只适用于对(vii)的责难。)一般认为,P.M. 丘奇兰德有时喜欢夸夸其谈:

民众心理学不仅仅是一种理论,它还显然是这样的理论,即直到 20 世纪后半叶哲学家们认识到它为止,它一直深入人心,为什么是这样,这一定是一个巨大的秘密。民众心理学的结构特征完全类似于数学物理学的结构特征;唯一的不同只在于它们说明的是抽象实在的有关方面——在物理学中是数,在心理学中是命题(1981,p. 71)。

由于对“民众心理学”有五种或六种不同的见解,其中除了两个以外都以“理论”理论为前提条件,而“理论”理论本身又有七种不同的阐释,它们无疑还有我这里和有限的发表空间没有提及的更重要的差别,因此对“民众心理学”和科学心理学之间的关系的考察一定得稍微限定其注意范围。我将根据我认为目前大多数文献所关注的问题去做这项工作。

关于“民众心理学”的争论问题之限定

第一,文献把注意力几乎完全限制在命题态度上——也就是说,它们几乎专门关注信念与相信。^④

第二,对于任何问题来说,它是否是真正的问题,取决于人们所理解的“民众心理学”的意义。一般来说,争论只是围绕着“理论”-理论所说的那种意义(上述(d),(d')或(e))的民众心理学而展开的。

但是第三,即便如此,特定的问题是不是真正的问题仍取决于“理论”需要什么样的级别。在正常情况下,有一个真正的争论,它是环绕着把态度至少描述为不言而喻地被假定的内在原因的“‘理论’-理论”(等级(iv)-(vii))而展开的。

因此,让我们把考察限定在“理论”-理论所说的那种意义的命题态度“民众心理学”上,在这里,“理论”-理论被认为作出了这样的断言,即民众心理学假定该态度是内在的原因。(此后用“因果‘理论’理论”和“因果理论民众心理学”。)

在这样的限定下,存在着下述两个主要争论或争论家族:因果理论民众心理学的诚实性或确切性,以及因果“理论”理论的真理

1012 第九部分

性。(当然,它们有多方面的相互联系。)

因果理论民众心理学的诚实性

与因果理论民众心理学相对抗的有三条主要的路线,其中的前两条显然是相互对立的。

第一个劲敌。第一类敌手的攻击是从弱的方面进行的,挑战的是这样的主张,即命题态度在任何强硬的意义上都是内在的原因,以及因果理论民众心理学实际上是由常识所认可的(行为主义者、工具主义者、维特根斯坦主义者、解释学家以及丹尼特—后“行为主义者之流”)。既然这条路线的种种描述受到了过分的关注,并争论了几十年,因此我这里不打算进一步加以陈述。

第二个劲敌。第二类反对意见假定,实际上是坚持说:因果理论民众心理学充当的是民众的代言人,但否定有任何这类内在的原因,因此也否定有命题态度。这自然是取消论唯物论的观点。即使没有人无保留地、完全彻底地接受取消论唯物论,但 P.M. 丘奇兰德、P.S. 丘奇兰德,以及有点喜怒无常的 S. 斯蒂克还有其他人把它认作是合理的选择或好的赌注。

于是便出现了一种无休止的三角关系,每一个都赞成其中一个而否定另一个。相互对峙的三个命题是:(1)因果理论民众心理学;(2)在某种弱的意义上,民众心理学有真理性或正确性;以及正如我们将看到的;(3)不能为民众心理学辩护的科学的(认知的)心理学。硬核命题态度实在论者坚持前两个,反对第三个,主张“辩护”;取消论唯物主义者坚持(1)(3),把民众心理学作为彻头彻尾

的谬论而加以拒斥;行为主义者以及其他坚持(2)(3),要么否认因果理论民众心理学自身是真实的,要么否认它是关于命题态度的民众概念的正确表达。(我们的、也是未予陈述的第三类反对意见,在这种三角关系中居何地位有待详细讨论。)取消论唯物主义者和实在论者一致认为:如果有的话,态度是人的真正的内在的原因状态,但恰恰在下述经验问题上他们分道扬镳:人的任何真实的神经生理状态是否能真的对应于因果理论民众心理学的常识心理范畴。(取消论唯物主义对实在论者直截了当、心甘情愿地拿起和使用他们的经验武器大加赞赏;两者都蔑视行为主义者以及其他人的卑劣的逃避。)不过取消论唯物主义者赞成行为主义者的这样的观点,即以因果理论民众心理学形式出现的实在论是一种白日梦,而因果理论民众心理学实在论者又赞成行为主义者说:心理归因常常是真实的,并如此显然易见。

有趣的是,联结主义——粗略地说,是哲学上这样的观点,即人类认知应根据人工智能中的并行分布加工方案而加以理解——为了捍卫两个相互对立的反实在论观点的每一个而一直在与实在论者展开搏斗。(例如丘奇兰德 1989;拉姆齐、斯蒂克和加龙 1990 为取消论唯物主义的辩护;克拉克 1989 和贝希特尔 1990 在一定程度上为行为主义者及其他人作了辩护)。^⑤

因果理论民众心理学的第三个劲敌处在前述两者之间:他/她承认,因果理论民众心理学得到了常识的庇护,人们拥有在引起行为时起着似状态作用的内在状态,但对态度内容的确定性提出了难题。即是说,我的公开的分类学显然没有导致对因果理论民众心理学的态度内容的承诺,这是很重要的。令人震惊的是,正如普

1014 第九部分

特南(1975)、斯蒂克(1978)、福多(1980)以及伯奇(1979)所论证的那样,因果理论民众心理学把命题态度说成是认知者的“广泛的”特征——即不能随时随附于它们的主体的分子构成的特征;态度内容部分是由外在于主体头脑的因果的、历史的和社会的因素所决定的。但是如果一态度的命题内容取决于外在于它的所有者的皮肤的物理疆界的因素,那么这些因素是什么呢?更为重要的是,我们还能假设:态度内容是其所有者或主体的真正的属性吗?尽管第三种反对意见承认内在原因,但是,取消之幽灵还是再度出现了;也许它并不真正地、客观地符合于相信这或意欲那的人。

环境因素实际上是什么,它们怎样决定态度内容,这正好是这样的问題,即什么是J. 福多所说的“心理语义学”。但是人的信念和愿望本身的本体论地位又是怎样的呢?有这样几种可能的看法:

(i)对有机体的内在物理或功能状态的外在语义解释完全是可以得到的:任何解释只要使任何人觉得方便有用就足够了,如果两个解释者的解释相互冲突,那么不顾另一个,两个便都是错误的。(奎因 1960;戴维森在某些情况下尤其是 1974;丹尼特 1978, 1987;希夫尔 1981。)如果我们赞成这样的观点,那么我们就几乎不能把主体相信一件事而不相信另一件事称之为一种可靠的事实。

(ii)信念状态的语义解释是不可企及的:它至少不严格地为不同的情境因素所决定,只有当别的解释是错误的时,某种解释才是正确的。但是这里所谓的情境因素是极其粗糙的、混乱的、社会性的、受利益左右的,完全不适合于溶入任何真正的科学之中。把信念P归之于某人是不科学的,没有解释性的或没有用的,就像不能说某物是好心的一样(斯蒂克 1983)。

(iii)信念的语义解释不严格地为各种情境因素所决定,某些解释只有在其他解释是错误的时候才是正确的,这就是主体相信一件事而不相信另一件事的原因之所在。主体环境的某种复杂的因果-历史的或目的论特征使主体相信这件事而不相信那件事这一点成为客观上真实的东西,不管那特征是不是科学上有利的或正确的(利康 1988 精神抖擞地为此观点作了辩护。)

(iv)关于这里所说的情景因素不存在混乱的、牵涉到利益的东西,即使它们扩展到了皮肤疆界之外。它们是简单的法则学的或目的论的事实。这一观点尤其得到了“指示词”语义学家的首肯(德雷特斯克 1981;1986;1988;斯托纳克 1984)。

(i)实质上是一种取消论观点,对因果理论民众心理学肯定持蔑视态度,同时否认作为命题态度的命题态度是世界上真正的原因要素。(ii)承认:态度归因可能是真的(不管怎样牵涉到利益),因此承认人们相信一件事而不相信另一件事,但(ii)仍拒绝了因果理论民众心理学的下述观点,即态度确实能引起行为。(iii)和(iv)完全可与因果理论民众心理学和谐共存。(i)与(ii)彼此有别,似乎只是程度上的,因此迄今的真正的的问题就是:(ii)或(iii)是否更加合理。但我这里不拟进一步考察第三个反对意见,因为取消论唯物论仍然是一更加引人注目的课题。

赞成取消唯物主义的论证

取消唯物主义者每一论证都以自然主义、唯物主义、物理主义或科学主义的某种一般观点为前提。

1016 第九部分

(E1)因果理论民众心理学预先假定命题态度对知觉或别的状态的随附因果关系,或行为对态度的随附因果关系。但是如果假定,态度依照其命题内容发挥了原因作用(在上述(e)之意义上,此即福多关于因果理论民众心理学的观点),那么就存在着“解释排除”的问题,同时导致对副现象论的驳难(洪都里奇 1982;金 1984;1988)。

(E2)如上所述,因果理论民众心理学把命题态度描述为认知者的“广泛的”特征,这就使态度成为令人讨厌的、非科学或反科学的而且还有可能是副现象性的东西。(斯蒂克 1983;还有关于“意向因果关系”的新的问题,就像莱波尔和洛伊尔 1987、福多 1989 所讨论的那样,见海尔和米尔 1993 编辑的论文集。)

(E3)因果理论民众心理学本身假定了这样一类事物,它们根本不可能存在于人脑之中或自然界的别的地方。(斯蒂克 1983;丘奇兰德 1989;典型的例证有,拉姆齐、斯蒂克、加龙用他们自己的具有(e)意义的特定观点批评了民众心理学。)

(E4)因果理论民众心理学是一种失败的或至少是一种糟糕的理论,将为更好的认知科学或神经科学所取代,因此应作为一种谬论而加以拒斥(丘奇兰德 1981;1989)。而且,丘奇兰德还补充道:我们完全可以学到新的、替换的理论,因此它将教给孩子们,并在父母的身边就可学到!

(E5)E3 与 E4 的混合物,因果理论民众心理学既不会得到认知科学、又不会得到神经科学的“辩护”,因此应作为谬误而加以拒斥(丘奇兰德 1989)(注意,这里至少有三种关于“辩护”的概念。我们可以说,如果因果理论民众心理学的“拉姆齐句子”的某种置

换一例示,由于在能用辩护理论的词汇描述的自然(原则上讲是在大脑)中发现的某种单元集合,而近似地为真,那么因果理论民众心理学便从弱的方面得到了辩护。此外,如果自然的项目单元为辩护理论的相当恰当的术语、也许是自然一类型术语所描述时,因果理论民众心理学便从强的方面得到了辩护。最后,只有当辩护理论由于自己的材料、议事日程和方法论的驱使而如此这般地假定了一种完善的术语,它们将还原因果理论民众心理学的“拉姆齐句子”,那么因果理论民众心理学将从非常强的意义上受到辩护。也正是在这里才真正涉及这样的问题,即究竟应向民众心理学中塞进多少东西。^⑥)

在有关文献中,这一些论证的每一个都从多方面得到了回应,尽管每一个仍聚讼纷纭。在不同场合,我曾说过:它们都是软弱无力的(反对这些论证、赞成对“辩护”课题的探讨,主要见于利康 1988,参阅霍根与伍德沃德 1985);但非常奇怪的是,我从未说服其最后一个名字叫丘奇兰德的人。

反取消唯物主义的论证

不足为奇,也有许多反取消唯物论的论证。

(F1)行为主义之流自然拒绝因果“理论”之理论。(这样一种观点本身当然不能算作是对取消唯物论的反击,而只是对上述论证(E3)、(E4)和(E5)的削弱。为了完成这项工作,行为主义之流不得不对论证(E2)发起进攻,方法是证明非科学性完全风马牛不相及——不仅因为因果“理论”理论是虚妄的——进而再指责其毫

1018 第九部分

无根据地赞成命题态度实在论。)

(F2)证明:假定的态度对预言、推理、深思、理解或对把握重要的宏观原则具有不可或缺性(基切尔 1984,霍根与伍德沃德 1985,伯奇 1986,丹尼特 1987,福多 1987,麦克兰洛克 1995)。一种不太健康的变种就是威尔克斯的主张(1984;1986;1991),即民众心理学不仅有利于前述的目的,而且有利于其他的、尤其是更实际的目的;与上述变种对立的,可看丘奇兰德(1995)。

(F3)一个摩尔式的论证。(奇怪的是,我记得,除我以外,我不曾听说过别的人作过这个论证,而且我也没有听到任何反驳。)大量的常识心理归因,如劳拉想要啤酒,相信在长靠椅下有一瓶,就个体(更不用说就群体)而言,比旨在使我们相信其反面的任何论证的纯哲学前程更加合理,而且将永远如此。(摩尔非常正确地意识到:纯哲学的假定只有很弱的认识根据。对取消唯物论的任何论证,都依赖于一个纯哲学的命题——这是没有错误的。)^⑦

(F4)“认知自杀”论证,大意是说取消唯物论实即无故的自我伤害或自我否定(贝克 1987,博格森 1990)。例如,如果被视作言语表达的取消唯物论是真的,那么取消唯物论就是有意义的。但是取消唯物论是有意义的(或有意义地被断定),当且仅当它表述了一定的信念,或它的言辞在某一方面是有意图的,如果是这样,某人在某处便拥有某种命题态度,如果是这样,取消唯物论就是虚妄的。因此只有当它虚妄时,它才是真的,如果前述的条件命题都是概念真理,那么这就说明取消唯物论是自相矛盾的。

所有这些反取消唯物论的论证自然也是有争议的。这里我只准备对(F4)作一番分析,它既是其中最引人注目的,又是迄今最

弱的一个。先看关于“认知自杀”论证的模棱两可的解释：第一，在宇宙和我们星球的历史上，直到最近，取消唯物论本身不可能有任何概念上的错误，取消唯物论显然是完全真实的；无论什么地方，不曾有过命题态度。第二，除非分析行为主义是真的，否则下述更强的建议就不可能有任何概念上的错误，这个建议是：尽管所有可观察的人类行为总是像现在那个样子，取消唯物论仍是真的。能断言的不过是：如果肯定的话语、交流等等在概念上离不开命题态度，那么取消唯物论就不能一以贯之地被断定和交流；但是这个条件从句为下述取消唯物论者心甘情愿地接受，他们要么拒绝它的前提，要么接受它的结论——或者是两者兼顾。

取消唯物论者有两个进一步的回答：(i)她/他有权成为一名还原论者。在反驳民众心理学时，她/他假定了意义以及它所蕴涵的、可作为临时假定的一切：假定民众心理学是真理，假定我们能陷入矛盾。对此，不存在丝毫的诧异或疑虑。(ii)存在着不同于命题态度的真理源泉：一种新型的原始语义学或生物语义学，也许是基于 P.M. 丘奇兰德与 P.S. 丘奇兰德(1983)所谓的“校准内容”(calibrational content)或 P.M. 丘奇兰德的“原型活动”(prototype activation)概念，即使从来没有哪一个人有过命题态度，它也能取消唯物论提供真理——即使新的意义上的“真理”除了必须适应于命题或句子之外，还必须适用于项目。

因果“理论”理论是真的吗？

有两条主要的反因果“理论”理论的路线。一条自然是行为主

1020 第九部分

义之流的路线,对此我已答应不予讨论。另一条是新近出现的反因果“理论”理论的自吹自擂的路线:R. 戈登(1986;1992a;1992b)所倡导并为 A. 戈德曼等人所接受的“模仿理论”。根据这种理论,因果-“理论”-理论家关于心理状态的图画,作为试图解释和预言他人行为的假设,是错误的。准确地说,为了解释和预言他人的行为,我们便设想我们处在他们的环境中,并设想我们自己继而作出了心理的和行为的反应;这等于在模仿,即我们自己的心理机制在脱机运作,因为我们的心理功能从根本上相似于我们的被试,因此这条路线是很管用的。

(实际上,在刚才的描述中,我把模仿理论的两变种揉在一起了,因为戈登和戈德曼的观点是不同的。戈德曼强调的是,用人们自己的心理机制所作的脱机加工;而戈登则把这称作“‘模型’模型”。即使戈登承认这种观点的有用性,但他一贯认为:模仿的方法论包括的主要不是:在如此这般的条件下我决定我自己做什么,而是:调节我的被试和我之间的有关心理功能差异的企图,以及关于那被试会做什么的设想。^⑧“调节有关心理功能差异”听起来像是冒险地用了某人所坚持的理论,但戈登强调并非如此,它指的是沉迷于一种想象的做作或佯装之中。)

几乎被忽视的问题是,因果“理论”理论是否为真,取决于“理论”理论被当作是一种关于什么的理论。这种依赖性几乎与行为主义之流的反对完全没有什么关系,因为行为主义之流在关注关于命题态度本质的本体论断时毫无异议地融合了他们的因果“理论”理论的目标。但是时下关于模仿理论的热烈讨论则表明:实际上存在着几种关于待解释词语的备选方案^⑨:

(a)心理术语的语义学。因果“理论”理论认为:心理术语具有科学—理论术语一般具有的那种类型的语义学。

(b)关于对应实在的形而上学。因果“理论”理论认为:心理状态是理论实在,即被观察到的事件的未被观察到的原因。

(c)心理状态与事件的基本的认识上的保障(如果你愿意的话)。因果“理论”理论主张:心理归因不仅就个体来说是难免错误的,就整体来说亦复如是:与行为主义之流的许多归因相反,它们可能都是错误的。

(d)关于心理实在的更详细的认识论。因果“理论”理论主张:它们是,或实际上似乎是为解释的目的而假设的,能证明它们正确的东西是(总是或首先是)去推论最好的解释。

(e)关于心理归因的实际的、成熟的心理学。因果“理论”理论自然可理解有这样的含义:我们经过构造解释性假说这样的过程完成我们的心理归因。

(f)关于心理归因的发展心理学。我知道,没有这样的“理论”理论家,他们对这一题目假定作出了严肃的、甚或真正的推测(塞拉斯关于天才的琼斯所讲的著名的发展故事显然是作为神话引入的);但某些模仿理论家和中立派人士一直把经验发展证据当作是对模仿理论与因果“理论”理论之间的一般争论有用的东西。

现在我们可以作一些清理工作。在“理论”理论家与模仿理论家之间不可能或不应当有一般性的的争论,因为至少存在着前述六个不同的探讨领域,它们是什么理论都可能关心的。更确切地说,人们必须逐一考察这样的领域,首先问:两个据说是敌对的理论是否述及它,然后再看哪一个胜利了。

1022 第九部分

这里,我可以对上述分析的结果作出快捷而武断的概述。第一,在(a)(b)(c)三个领域,因果“理论”理论与模仿理论并不是对抗的。因为模仿理论完全可与因果“理论”理论家关于语义学、形而上学和基本认识保障的观点和睦共处,甚至不提出敌对的备选方案。即是说,只有这些领域才是塞拉斯、费耶阿本德、奎因、福多和其他大多数因果“理论”理论家主张因果“理论”理论应致力于的三个领域。因此想在这些领域的任何一个或全部领域对因果理论发难的模仿理论家,就必须拿出一个特定的对立面:在(a)领域,模仿论者必须为心理术语设定一种特殊的、可供选择的语义学。在(b),他(她)必须提供一幅对抗性的形而上学图景。我知道,没有哪一个模仿论者做了或声称做了这样的事情。在(c)领域,模仿论者至少必须为某种不可错主义者或不可纠正主义者的学说提供辩护,我坚信,没有哪一个模仿论者情愿做这种事情。

现在来看(f)领域,即发展心理学。正像模仿主义者对(a) - (c)领域实在没有兴趣一样,因果“理论”理论家对发展心理学也没有正式的兴趣,因为我所熟悉的哲学家没有一个提出过实际的发展理论(不管有多少像福多这样的因果“理论”理论家欢迎来自于发展心理学的间接证据)。因此在(f)领域,因果“理论”理论家与模仿论者之间也没有根本的冲突。类似的说明尤其适用于心理学中的新生的“心灵理论”文献(例如可参阅:温默与佩尔内 1983;巴伦-科恩等,1985;阿斯廷顿等,1988;怀登 1991;阿斯廷顿 1993;卡鲁瑟斯等 1995;戴维斯等 1995a)。

(e)领域是不确定的。模仿论者在那里有重要承诺,并列了几种不同的经验材料予以支持。说因果“理论”理论对实际的心理

过程提出了一种主张,这是容易理解的,即使只有对因果“理论”理论文献的学究式梳理才能说明:其作者是否实际上提出了那个主张。一旦(e)和(d)之间的区分被确定了,那么这就成了因果“理论”理论的决定性观点。如果一特定的因果“理论”理论家发现自己准备接受与(e)有关的领域,那么她或他就一定会面对着模仿论者的明显是相互矛盾的经验材料。(但是再者,因果“理论”理论原来的哲学观点并不依赖于关于实际心理过程的经验主张。)

剩下的是(d),在这个领域,最终存在着真正的争论。纯因果“理论”理论坚持认为:心理归因永远是借推断最好的解释来证明的因果假说;模仿论者则坚持完全相反的观点,认为它们的大多数或许多都得靠脱机模仿技术来证明。但是,模仿论者在这里还有事情要做。因为尽管大家普遍承认:^⑩推论出最好的解释能够证明一个信念。这一点已得到了广泛承认,但我没法肯定,人们怎么可能让我信服:在“证明”的完全标准的、不同于询问者对最终怎样坚持那个信念的纯描述的意义,一种想象的假装或脱机模仿能够证明同一个信念。

一种可供选择的办法就是论证说:模仿告诉询问者:她或他想的是什么,在所设想的环境中又做了什么,接着询问者便从一前提过渡到这样的结论,即那被试在有关方面像她或他。但是这两步的每一步都有问题。第一步之所以有问题,是因为关于我们自己行动的反事实是冒险的;我在如此那般的条件下会做什么,对此我可能犯重大的错误。第二步之所以有问题,是因为被试可能完全不像那个询问者,除非涉及到像乐于避苦之类的一般特征,总之,人们会推测说,一个人要证明第二步就得有一种关于那被试的理论。

1024 第九部分

撇开标准的问题,还剩下什么呢?似乎很清楚:在(d)领域,关于某种重要的东西,每一方面都是正确的。模仿论者说,我们通过模仿有时也完成了心理归因,这是对的(为了论证的目的,我们承认这一过程证明了那些归因)。但是因果“理论”理论家说,我们借助对最好解释的推论完成了心理归因,这也是对的,尤其是当我们已知道:我们的被试在有关心理方面与我们不同、我们正设法领会他时更是如此。因此争论的焦点不是纯因果“理论”理论对模仿理论,而仅仅在于,哪种方法在一种或另一种意义上居于支配地位。这里的支配性,可能只是统计学上的,或者说任何派别都可能提出这样的主张:她或他的方法严格地说更根本,其他方法则寄生于它之中(记住:任何适用于(d)领域的支配性是规范的,而非心理学的)。

但我将不考虑这里的模仿争论,既然如此,我们的话题到此便完结了。

注 释

① 威尔克斯(1991)等人主张:“民众”这一修饰语让人有贬损之感,在其发明者看来也许是不可少的,也许是深思熟虑的。“常识心理学”肯定更恰当,但会议组织者安排给我的就是这个题目与术语。

② 正如 B. 布兰登提醒我的那样,塞拉斯的“理论”-理论化受制于一系列思想;他把“信念”仅仅看作是倾向性的,注意到这一点极为重要。

③ 由于这个理由,布莱克本(1991)不可能赞成(d),尽管他并没有强调要继续支持(c)。

④ 正如冯·埃克德所叹惜的那样(见本论文集)。

⑤ 关于心灵哲学中的联结主义的大量参考文献,可参阅前述著作。另见:贝希特尔和亚伯拉罕森 1991,利康 1991。其他论文汇集在:拉姆齐、斯蒂

克和拉梅尔哈特 1991。

⑥ 关于什么时候把实在还原为更基本的科学、什么时候取消它们之类的一般性问题,我认为,斯蒂克的论述是最好的(正在印刷中)。

⑦ 注意:这个论证不受常见的丘奇兰德-丘奇兰德反驳的影响,试比较炼金术、巫术和其他民众的但虚妄的理论。不管这些理论的支持者多么宠爱它们,但它们没有资格成为摩尔式的常识。信念作为常识性的东西,一定是这样一种信念,每个正常的人在其日常生活中都少不了它,诸如“在中饭前我吃过了早饭”或“这是一只手,这是另一只”或“太阳闪闪发光”。

⑧ 罗斯卡-哈蒂使我想到了那种区分,谢谢。在讨论中,C. 格里莫提出了第三种方案:人们有时通过使被试 A 相似于第三者 B,想象 B 在所设想的环境中做了什么,来预言 A 的行为。

⑨ 这里我得益于 B. 科恩,参阅其正在写作中的哲学博士学位论文,北卡罗莱拉大学。

⑩ B. 弗拉森就是著名的持异议者。

参考文献:

- D. 阿姆斯特朗(1993):《儿童的心灵发现》,哈佛大学出版社。
 J. 阿斯廷顿等(1988):《心灵的发展理论》,剑桥大学出版社。
 W. 贝希特尔等(1991):《联结主义与心灵》,牛津,巴兹尔·布莱克韦尔。
 T. 伯奇(1986):“个体主义与心理学”,《哲学评论》,95:3-45。
 P. 卡鲁瑟斯等(编)(1995):《关于心灵理论的理论》,剑桥大学出版社。
 P.M. 丘奇兰德(1981):“取消唯物论与命题态度”,《哲学杂志》,87:67-90。
 ——(1989):《神经计算展望》,麻省理工学院出版社。
 ——(1994):“民众心理学(2)”,载 S. 格滕普兰(编)(1994):《心灵哲学的伴侣》,牛津,布莱克韦尔。
 R.M. 丘奇兰德与 P.S. 丘奇兰德(1983):“原始认识机寻踪”,《心灵》,17:5-18。
 P.S. 丘奇兰德(1986):《神经生理学》,麻省理工学院出版社。
 A. 克拉克(1989):《微观认识》,麻省理工学院出版社。

1026 第九部分

- M. 戴维斯等编(1995a):《民众心理学:心灵理论争论》,牛津,布莱克韦尔。
——(1995b):《心理模仿:评价与应用》,牛津,布莱克韦尔。
- D. 丹尼特(1987):《意向态度》,麻省理工学院出版社。
- F. 德雷特斯克(1981):《知识与信息流》,麻省理工学院出版社。
——(1988):《对行为的解释》,麻省理工学院出版社。
- J. 福多(1987):《心理语义学》,麻省理工学院出版社。
——(1989):“让心灵更加物质”,《哲学论题》,17:59-79。
- A. 戈德曼(1992):“模仿理论辩”,《心灵与语言》,7:104-119。
- R. 戈登(1986):“作为模仿的民众心理学”,《心灵与语言》,1:158-171。
- J. 格林伍德(编)(1991):《民众心理学的未来》,剑桥大学出版社。
- J. 海尔与 A. 米尔(编)(1993):《心理原因》,牛津大学出版社。
- T. 洪都里奇(1982):“论异态一元论”,《分析》,42:59-64。
- T. 霍根等(1985):“民众心理学不会消失”,《哲学评论》,94:197-225。
- J. 金(1984):“副现象论与随附因果性”,载 P. 弗伦奇等编,《中西部哲学研究》,第 9 卷,明尼苏达大学出版社。
- D. 刘易斯(1994):“大卫·刘易斯:心灵的还原”,载 S. 格滕普兰编(1994)。
- W. 利康(1993):“思维表征理论的演绎论证”,《心灵与语言》,8:404-422。
- A. 莫顿(1980):“心灵的构架”,牛津大学出版社。
- W. 奎因(1960):《词语与对象》,麻省理工学院出版社。
- W. 拉姆齐等(1990):“联结主义,取消主义与民众心理学的未来”,载 J·汤伯林(编):《哲学展望》,4:“行动理论与心灵哲学”,阿塔斯卡德罗:里奇维出版公司。
- P. 拉姆齐等编(1991):《哲学与联结主义理论》,希尔斯德尔,新泽西:埃尔伯。
- S. 斯蒂克(1978):“自动心理学与信念愿望论”,《一元论》,61:573-591。
- B. 埃克德(1994):“民众心理学(1)”。载 S. 格滕普兰(编)(1994)。
- K. 威尔克斯(1991):“科学心理学与常识心理学之间的关系”,《综合》,89:15-39。

心灵哲学中的工具主义*

加里·富勒

常识与科学之间的关系是哲学关注的一个核心问题。毫无疑问,心灵哲学过去一直享有这种关注,今后还将继续如此。在这里,争论的焦点是,我们关于信念、愿望等的常识心理学概念(以及个性、道德动因和责任等有关概念)与蓬勃发展着的科学如传统的牛顿力学、几十年前的激进行为主义和现在的认知心理学、神经生理学之间的关系问题。一般性的问题恒常不变,它们依然是:科学与我们的民众心理学概念有严重冲突吗?或者说它能与它们相谐并存吗?当今,有两类人,一类是乐观主义的思想家,诸如 J. 福多,^①他们坚信:一种足够科学的心理学最终将十分相似于民众心理学,另一类是悲观主义者,诸如 P. 丘奇兰德,^②他们相信:发展着的心理学、脑科学与民众心理学将证明是完全不相容的,因而民众心理学将被抛弃。

哲学家保护民众心理学、使之免遭那些发展着的科学的可能威胁的一种方式就是对民众心理学作出工具主义的解释,据此,民众心理学术语诸如“相信”“愿望”等并不指称什么,而只应当被当

* 译自 R.M. 齐硕姆等编:《心灵哲学·心理学哲学》(第九届国际维特根斯坦专题讨论会论文集),维也纳,1985。

作是纯粹的预测策略。作为最有影响的心灵哲学家之一的 D. 丹尼特,在过去 15 年中以坚持这一立场而著称。这篇论文的目的就是要考察贯穿于丹尼特论著中的工具主义,并提出工具主义是否为民众心理学作出了辩护这一问题。我要做的事情有三件;(1)澄清工具主义概念;(2)对丹尼特的工具主义观点作出简要解释;(3)证明:至少是丹尼特的早期强工具主义未能对民众心理学作出辩护。

I

我们先来澄清工具主义概念。对于民众心理学来说,成为一个工具主义者意味着什么?

如果我们采取一种现今得到广泛认可的假定,即民众心理学是一种关于行为的理论,它隐含在我们根据信息、愿望等对行为所作出的解释和预言之中,那么对工具主义包含什么内容作出阐述就再容易不过了。根据这一假定,该理论的心理术语如“信念”、“愿望”等应看作是这样的理论术语,它们至少在表面上指称了状态类型和/或标记。

有了这样的假定,我们就能通过把工具主义与实在论加以对照而对之作出下述初步的描述:

(1)实在论:民众心理学的心理术语指称了状态类型和/或标记,它的句子有真值。

(2)工具主义:那个理论的心理术语没有指称,它的句子没有真值。毋宁说,这类术语和句子是旨在预言行为的预测策略。

请注意,根据我们迄今为止的描述,工具主义与心理类型/物理类型、心理标记/物理标记两种同一理论是不相容的。

然而,这种对工具主义的初步说明还可以加以修饰,如给(2)贴上“强工具主义”的标签,并把它与弱工具主义区别开来,后者可以表述如下:

(3)弱工具主义:民众心理学的心理术语没有指称;尽管如此,该理论的句子则有真值。

II

多年来,丹尼特一直致力于对民众心理学的工具主义方案的清晰阐述,并为之进行辩护。丹尼特的工具主义不是静止不变的;甚至可以举出很好的例子证明:他的工具主义经历了重大的变化。^③稍早一点,尤其是在《脑猝病》^④一书中,丹尼特坚持强工具主义立场,在其看来,民众心理学的句子没有真值。近来,尤其是在他的重要论文《三类意向心理学》^⑤中,他则采取了弱工具主义,据此,民众心理学的状态如信念,并不是真实的假定(那个),而仅仅是与实在(抽象)有关的预测,类似于引力中心。因此,根据新的观点,术语“信念”没有指称,但信念句子就像“关于”引力中心的句子一样却具有一种真值。

撇开细节不说,我们有必要提出一个一般性的问题,即丹尼特为什么赞成工具主义而不选择实在论立场。可以肯定,丹尼特的一个根本考虑就是:民众心理学(或者用丹尼特的话说,意向系统理论),如果从实在论角度加以解释,那么就可能被证明是错误的。

1030 第九部分

尤其是,民众心理学包含了合理的理想化成分,因而是不可能实现的,而且对民众心理学的实在论阐释似乎需要内在的“表征”,至少可以这样说,丹尼特对此常常持怀疑态度。

这种就其本身所作的考察显然还不够。如果从实在论角度理解的民众心理学确实是虚假的,那么为什么不将其完全抛弃呢?丹尼特想保护民众心理学是有许多理由的。首先,他相信:它在科学心理学中起着作用,例如它能在从上到下的战略中作为有用的出发点而发挥作用。其次,也许是更为重要的,民众心理学在某种意义上对于我们来说是必不可少的。因此丹尼特乐意接受工具主义的第二个重要的考虑就是:它使我们能够保护民众心理学。在一定的意义上,工具主义将为民众心理学提供辩护。

III

关于民众心理学的工具主义方案如果有错误的话,那么有什么错误呢?在这一部分,我将集中批评丹尼特的早期思想中的强工具主义,其实探寻的正好是对其早期观点的一种批判。我想问的问题是:强工具主义,不管它有其他什么优点与缺点,是否真的对民众心理学作出了辩护。

对民众心理学的强工具主义阐释为我们关于信念、愿望以及关于个性、责任等常识概念作出了辩护吗?

“辩护”在这里意味着什么?一种意思是,比如说一种主张或理论在怀疑论的打击面前能被认为得到了辩护,就是能说明它是完全真实的。但是这种含义自然不是我们要寻求的那种含义。

“为我们的民众心理学概念辩护”，其含义必须是：“说明我们有充足的理由继续使用这些概念”。因此以这种方式理解“辩护”，我们可能马上想到要论证：强工具主义为我们的常识概念作出了辩护：这些概念作为预测策略的有用性肯定为继续使用这些概念提供了充足的理由。

随之而来的麻烦就是：我们的民众心理学概念不仅仅是预测策略。更有甚者，像许多哲学家已经正确地论证过的那样，对于我们的信念、愿望和个性等的惯用语的强工具主义说明将产生不同于我们的事实概念的概念。我们的事实概念是实在论的：关于一个系统的内在状态（以及关于它们的相互因果关系、它们与输入、输出的关系）的事实与这些概念的应用密切相关，就像它们与相应的强工具主义概念的应用没有关系一样。例如，试与关于“人”（或“心灵”）的强工具主义概念进行比较，在丹尼特看来，根据该概念，只有当一个系统的行为借助意向战略的工具主义规则“能可靠地加以预言”时，^⑥该系统才能被看作是人。这两点是泾渭分明的，因为：

（1）我们能描述这样的系统，它们符合关于人的强工具主义概念，而不符合事实概念。S. 斯蒂克曾阐述过 N. 布洛克关于下棋计算机的例子，它的着法用意向战略一般是可以预言的，但它没有关于规则或目的的内在表征。

确切地说，这种计算机的记忆包含了关于每一可能的一局棋的丰富的、多种多样的树状分支表征，比如说长到 100 步……。如果我们偶然碰到（似乎如此）一个人，他的交谈、下棋和别的行为都是由大量的关于什么时候做什么事的预先编制

1032 第九部分

好的程序分支目录控制的,那么我认为我们的直觉就会抵制下述说法,即“那个人”相信:我准备攻击我的王后——或者说他相信别的任何东西!^⑦

(2)想出相反的例子也不难,例如有这样的系统,它符合关于人的事实概念而不符合强工具主义概念。我的同事保罗·俞提供了下面的例子。假如这是我们都不知道的事情,火星人是如此处理问题的,每当任何尝试作出来以预言琼斯的行为时,他的大脑状态就变得反复无常,即便在别的时候他都是完全正常的。因此,即使根据关于“人”的实际构想,琼斯被认为是人,但根据工具主义观点,他则不是。

因此,我们的民众心理学的事实概念不同于他们的强工具主义的对应概念。如果丹尼特只是因为民众心理学能够在科学心理学中,例如(如我们已看到的)在发展从上到下的战略中,发挥作用才对之有兴趣,那么缺乏这种同一性也许不会困扰他。但是如果他像他似乎所是的那样,想对我们的日常心理学概念和实践作出辩护,那么他就会陷入困境,因为引进工具主义的替换物,不管从科学上说如何合法,这怎么可能为我们关于愿望、信念和个性的事实的、实在论的概念作出辩护呢?

为了回答上面的问题,我们也可能尝试下述路线。我们常用我们的民众心理学概念作出某些区分,如在人与非人之间、进而在人能说明的行动和不能说明的行动之间等作出区分。像丹尼特所设想的那样,这些常识概念在某些方面也许是无效的,因而应为新的有效的概念所替换——在丹尼特看来,应代之以新的强工具主

义替代物。如果新的概念能像旧的概念一样作出大致相同的区分,那么我们就说新的概念为旧的区分作出了辩护,进而间接地为旧的概念作了辩护。如果采取这一思路,那么我们就说:工具主义的替换物就将为我们的事实的民众概念作出辩护。只要它们合理地、严密地保持它们的(可能的)外延的一致。

什么样的严密才是合理的严密?上述例(1)和例(2)已经表明:强工具主义的替代物并不能等同于我们的事实的民众心理学概念。在这里,可能有人依然认为:(1)和(2)所描述的情况是极不可能发生的,工具主义概念和事实概念至少在现实世界将被证明是共外延的。毫无疑问,这对辩护来说是足够的。

不幸的是(对于强工具主义而言),我们可以设想这样的事例,它们几乎是不可能发生的,当然其中也有歧异。试看下面的例子。

(3)琼斯把枪对准史密斯,有拉动扳机杀死他的意图。随后他真的扣了扳机,史密斯死了。尽管如此,促使琼斯扣扳机的东西并不是他的意图,或别的心理事件,而是某种与琼斯的心理状态完全无关的意外的大脑事件。

即使这个例子有必要更详细地加以描述,但有充分的理由断言:我们(用我们日常的民众概念)不想把史密斯之死的责任归咎于琼斯,而强工具主义者则会如此。

总之,我对丹尼特的早期强工具主义能否为信念、愿望和个性之类的日常民众概念辩护深表怀疑。诚然,我猜想丹尼特也有这些疑虑,而且它们是使他从早期观点转变为新的弱工具主义观点的重要驱动因素。新的观点是否也面临着同样的辩护问题,这则是有待进一步探讨的课题。

1034 第九部分

参考书目：

- ① 参见福多：《思想的语言》，纽约，1975。
- ② 见丘奇兰德：《取消论的唯物主义与命题态度》，载《哲学杂志》，78（1981），pp. 67—90。
- ③ 保罗·俞和我在《丹尼特批判》中论证过这个例子，见即将刊出的《综合》。
- ④ D. 丹尼特：《脑猝病》，剑桥，1978。
- ⑤ 载 R. 希利编：《还原、时间与实在》，剑桥（英国），1981，pp. 37—61。
- ⑥ “真正的相信者……”，载 A. 希思主编：《科学解释》，牛津，1981，pp. 55。
- ⑦ S. 斯蒂克：《从民众心理学到认知科学》，剑桥，1983，pp. 244—245。

联结主义、取消主义与 民众心理学的未来*

斯蒂芬·P. 斯蒂克 威廉·拉姆齐
和约瑟夫·加龙

I. 导言

在 T. 库恩发表《科学革命的结构》一书以后的岁月里,科学革命一词以与日俱增的频率应用于关于科学变革的讨论之中。在某人或别的人被诱使把一种革新称之为—场革命之前,这场革新所必须具有的重要性便奇怪地去无踪迹了。本章的主题是:如果联结主义的一系列假说被证明是正确的,那么它们便肯定可算是革命性的,即使根据严格的前库恩标准也是如此。毫无疑问,联结主义已经经历了许多重大的变化,就像许多认知科学家所设想的认知一样。不过,正如我们所看到的那样,使某些联结主义模型成为名副其实的革命模型的东西就是它们为一种彻底的取消论提供了支持,这种取消论要取消的是常识(或“民众”)心理学的核心假定。我们这里集中讨论的是信念或命题记忆,即使论证已直接

* 选自斯蒂芬·P. 斯蒂克:《解构心灵》,牛津大学出版社 1996,第 2 章,本章是斯蒂克与拉姆齐、加龙合写的。

1036 第九部分

泛化到所有其他命题态度上。如果我们对的,那么这类联结主义的结论将超出认知科学的疆界,因为这些模型如果是成功的,那么就有必要彻底改变方向,就像我们对自身的思考一样。

这里先对要论述的作一简要的概述。第2节扼要说明了取消论的主张,接着描述了取消论论证一般必需的两个前提。第3节阐述了我们关于民众心理学及其所假定的命题态度的看法。进而阐释了一种利用民众心理学假定并以之为基础的心理模型。第4节是专论联结主义的。被称之为联结主义的那些模型已形成了一种模糊的、异类合成的集合,其成员经常表现出的不过是粗略的家族相似性。不过,我们把联结主义与取消论结合起来的论证将只限于联结主义模型的有限范围,即以特定方式予以解释的范围;第4节的主要任务就是讨论:这个范围是什么,此范围内的模型应怎样被解释。在第5节中,我们将说明:与我们的限定相一致的联结主义信念模型可能是什么样子,然后我们再进行论证:如果这类模型是正确的,那么结果对常识心理学将极为不妙。第6节汇集了一些反对意见及答复。最后一节是简要的总结。

在具体论述之前,我们应强调的是:我们打算辩护的论点是一种条件论断:如果我们概述的那类联结主义假说最终是正确的,那么取消主义关于命题态度的观点也是如此。既然我们的目标只是表明,联结主义与取消主义怎样相互关联,因此我们用不着花力气去论证两个学说哪一个真,哪一个假。尤其是我们用不着为赞成我们的条件句之前件所必需的联结主义观点提供论证。确切地说,我们的观点是:满有把握地断言这类联结主义假说家族将怎样健康发展,还为时尚早——太早了。那些诚信联结主义的人自然

会诉诸于我们的条件句,将其作为铲除命题态度的更大的论证的一部分。^①但是正如 J. 豪格兰德曾说明过的那样,一个人的权利(ponens)就是另一个人的失去(tollens)。那些接受取消主义关于命题态度是荒谬的或不可思议的观点的人很可能把我们的论证当作反联结主义的更大的理由的一部分。因此如果对联结主义的强有力批评者,如福多和皮利辛,发现我们的条件句和论证是相当协调的,那么我们丝毫不会大惊小怪。^②

II. 取消主义和民众心理学

作为我们要用的术语的取消主义是为表述一简单论点而想象出来的名称。它主张,在对世界的常识或科学解释中所使用的、表示实在、过程或属性的某些范畴并不存在。如果这样来理解,那么我们都是关于许多类型的事物的取消论者。在民众理论的领域内,巫婆是标准的例子。曾经在某一时期,巫婆被广泛相信对各种地区性的灾难负有责任。但是人们渐渐相信,对与巫婆有联系的大多数事件可作出更好的解释。不存在只有巫婆才能完成的解释工作,因此明智的人断言:根本没有这回事。在科学的范围内,燃素、热的流质和传光的以太是取消主义津津乐道的例子。其中的每一个曾为实施复杂研究计划的严肃的科学家所援引。但在每种情况下,计划在主要方面均告搁浅,因此那些诉诸于这些实在的理论便为后续的、不给这些实在以地盘的理論所取代。科学共同体逐渐认识到:燃素之类根本就不存在。

正如这些例子所暗示的,取消主义论证中的关键步骤一般来

1038 第九部分

说就是证明:在其中诉诸于某些所谓的实在或过程的理论应予拒斥,而由更好的理论所取代。这便引出了下面的问题:我们将怎样说明一种理论比另一种更好呢?众所周知,提出此问题比给予回答要容易得多。不过,下述观点也得到了相当广泛的认可,即如果一种新的理论比旧的理论提供了更准确的预言和更好的解释,而且能适用于如此广泛的现象,如果新的理论能更好地与相邻领域完善地建立起来的理论相谐并进,那么就有理由认为:旧的理论是低劣的,新的理论是应予接受的。对于我们的目的来说,这尽管是足够的,但是几乎不能算作是对于一种理论优于另一种理论之条件的完满的说明。

不过,仅仅说明一类实在在其中起着作用的理论次于后续的理论,还不足以说明那些实在不存在。通常,更妥当的结论是:被拒绝的理论对在它的范围内的实在的某些属性或主宰这些实在的规律的看法是错误的,也许有严重的错误,而新的理论对那些实在为我们提供了更准确的说明。例如前哥白尼宇宙学对行星的本质、制约其运动的规律的看法就是非常错误的。但是说哥白尼和伽利略已说明了托勒密所说的行星不存在则又多少有点滑稽。^③

在其他事例中,得到正确断定的东西是:旧理论的假定可还原于新理论的假定。这里的标准例子有:温度可还原于平均分子动能,声可还原于介质的波动,基因可还原于一系列的多核苷酸碱基。^④基于这些流行的看法,就可从这些事例引出下述教训:即使命题态度由此发源的常识理论应被更好的理论所取代,但这还不足以表明:常识理论的假定不存在。

还需要什么呢?把一方面的如燃素、热质之类的事例与另一

方面的如基因或行星之类的事例区分开来的东西是什么？或者换一种方式表述这个问题：什么使燃素、热质之类成为被排除的候选者？为什么不能得出这样的结论，即燃素就是氧，热质就是动能，早期理论关于燃素和热质的某些属性的看法正好是极端错误的？

让我们引入一些术语。我们将把下述理论变革称之为本体论上保守的理论变革，在这种变革中，旧理论的实在和属性被保留下来了或被还原为新理论中的实在与属性。在本体论上不具有保守性的理论变革，我们将称之为本体论上激进的变革。基于这些术语，我们将提出的问题便是：怎样把本体论上保守的理论变革与激进的理论变革区别开来。

这还是一个容易提出而难以回答的问题。在科学哲学文献中，对于理论变革何时支持取消论结论，何时不支持，根本就找不到近乎合理与完全一般性的说明。在缺乏确定本体论上的排除何时为根本方法的情况下，我们所能做的充其量是，考察旧理论的假定——有被取消危险的假定——然后探讨在新的理论中是否有这样的东西，即可能被同一的或被还原的东西。如果新理论的假定给我们的印象是，在深层次上、在根本上不同于旧理论的假定，就像分子运动在深层次上、在根本上不同于热质理论所假定的“异常活跃的”流质，那么便可合理地断言：理论变革是激进的，取消主义的结论是稳妥的。既然对于两个假定究竟怎样“在深层次上、在根本上不同”没有简单明了的标准，因此我们所到达的结论就一定是一种判断预期。^⑤

为了证明某些联结主义模型支持关于命题态度的取消主义，我们就必须论证这些模型不是本体论上的保守模型。我们的战略是，

把这些联结主义模型即类似于第5节中要解释的那些模型与第3节后面要叙述的那样的本体论上保守的模型加以对比,以突出联结主义模型在本体论上如何激进。但是在这里,我们正在超越自我。在设法让你相信联结主义模型是本体论上激进的模型之前,我们必须考察一下联结主义模型扬言要取代的民众心理学理论。

Ⅲ. 命题态度与常识心理学

为了现在的目的,我们将假设:常识心理学可以合理地看作是一种理论,信念、愿望和其他的命题态度,可以合理地视之为该理论的假定。即使这不是一个没有争议的假设,但其他的人已精彩地阐述了关于它的理由。^⑥—当承认常识心理学真的是一种理论时,我们就可以预期,几乎每个人都会承认:该理论是被取代的一个可能的候选者。我们这样说,无意于毁谤民众心理学,或以假定为根据论证它所假定的实在的地位。我们的观点很简单,那就是,民众在心理学方面的智慧不可能把应知的一切都告诉我们。常识心理学像其他民众理论一样在许多方面注定是不完善的,在更多的方面很可能是不正确的。如果事实不是这样,那么细致的、定量的、实验的心理学科学就是多余的。除了一些顽固的维特根斯坦分子这样的可能例外,几乎每个人都乐意承认:在内在于常识的事实和原则之上,还有许多心理学的事实和原则。如果这是对的,那么我们就有取消主义针对信念、命题记忆和其他命题态度的论证所需要的第一个前提。假定这些态度的理论的确是要予以取代的第一个候选者。

即使常识心理学包含了关于信念、记忆、愿望、希望、恐惧和别的命题态度的经验知识财富,但在构造联结主义与取消论之间的联系纽带的过程中,重要的民众心理学信条是这样的主张,即命题态度是功能上具体的、语义上可解释的状态,它们对别的命题态度的产生、最后对行为的产生有因果作用。按照斯蒂克(1983)的这个建议,我们把这组主张称之为命题模块性(modularity)。^⑦(读者一定要当心,不要把命题模块性这个概念与福多(1983)所辩护的非常不同的模块性概念混淆起来了。)

可以援引大量的证据,以支持这样的论点,即:民众心理学承诺了命题模块性信条。常识心理学认为,信念和别的命题态度有语义属性,这一事实已得到了特别的强调。根据常识:

(1)当人们看到附近有一条狗时,他们一般会相信附近存在着一条狗。

(2)当人们相信:如果山上有雪,那么火车就会晚点,进而确实相信:山上有雪,那么他们一般会相信:火车要晚点。

(3)当讲英语的人说,“花园里有只猫”时,那么他们一般会相信:花园里有只猫。

如此等等,以至无穷。须知,这些常识心理学的普遍原则是依照态度的语义属性而予以表达的。正是由于有信念 p,特定的信念才有特定的结果或原因。因此常识心理学把表达了这些语义属性的谓项,如“相信火车晚点”这样的谓项,当作是可投射的(projectable)谓项——恰当地运用于法则学的或准法则的普遍原则中的那类谓项。

民众心理学赞成命题态度就是功能上具体的状态这一论点,

1042 第九部分

而阐明这一点的最好的办法就是指出：它常能完满地澄清下述主张的意义，即一个人已获得（或失去）了一个记忆或信念。例如，在特定的情况下，大概可以合理地断言：当亨利打盹后醒来时，他完全忘记了汽车钥匙藏在冷藏室中，即使他没忘记其他的事情。在说民众心理学把信念当作能为一个人同时得到或失去的那类东西时，我们无意于否定：拥有任何特定的信念可能以有关信念的基本网络为前提。相信汽车钥匙在冷藏室中这样的信念不是原始部落的人所能获得的那种信念，因为他们对汽车、钥匙或冷藏室一无所知。但是一当有关的背景出现了，正如我们设想它有利于我们和亨利那样，民众心理学似乎可以与下述可能性完全一致起来，即一个人可能得到（或失去）汽车钥匙在冷藏室这样的信念，而他的其他信念依然如故。命题模块性自然不否认：获得一个信念常常导致一组有关信念的获得。当亨利被告知：钥匙在冷藏室时，他可能会相信：钥匙没有丢在点火装置或他的夹克口袋里。但是下一次，他又不可能如此。甚至，根据民众心理学关于信念的概念，一个人完全可能有这样的持续甚久的信念：钥匙在冷藏室中，但却继续在卧室搜寻。^⑧

为了说明民众心理学以此把命题态度当作功能上具体、因果上能动的状态的方法，让我们叙述两个更复杂的例子。

在常识心理学中，行为常常诉诸于行动者的某些信念和愿望予以解释。因此，为了解释艾丽斯为什么要到她的办公室，我们可能会说她想发送一些电子邮件（而且她自然相信，她在办公室能做到这一点）。然而在某些情况下，行动者有几组信念和愿望，其中每一个都可能导致相同的行为。这样一来，我们就可以假定，艾丽

斯还想与她的研究助手谈谈,她相信他在办公室里。如果是这样,常识心理学就会假定:艾丽斯去她的办公室可能是由两个信念/愿望中的某一个引起的,或同时由两者所引起,确定这些选择中的哪一个付诸实施是一个经验问题。因此完全有这样的可能,即在这种情况下,艾丽斯发送电子邮件的愿望在产生她的行为时并未发挥作用。事实上,正是想与她的助手谈话这样的愿望促使她上办公室。不过,如果她不想与她的助手谈话,那么她也有可能去办公室,这是因为发送电子邮件的愿望,该愿望在实际作出决策时并未发挥原因的作用,但后来可能积极地卷入了。须知:在这种情况下,常识心理学愿意承认两种非常不同的、从语义学上被描述的状态,其中一个在因果上是能动的,而另一个则相反。

我们的第二个说明类似于第一个说明,但侧重于信念与推理,而不是愿望与行动。根据常识观点,有时候可能碰巧有这样的情况,即一个人有大量的信念串,其中任何一个都可能使他推论出某种别的信念。当他实际作出了推理时,民众心理学就假定:他根据什么把它推论出来,这是一个经验问题,而且这个问题一般有一个确定的答案。例如,假设检查官克洛斯相信:男管家说他在乡村旅店过了夜,他还说他是乘早上的火车回来的。假设克洛斯还相信:那个乡村旅店在那个季节关了门,早上的火车取消了。假如有这些信念以及某些普遍共有的背景信念,那么克洛斯很可能推论说,那个男管家在撒谎。如果他这样做了,那么民众心理学就会假定:那个推论要么是基于他关于旅店的信念,要么是基于他关于火车的信念,或者两者兼而有之。从常识心理学的观点看,完全有这样的可能:即使克洛斯早就知道哪个旅店在那个季节关门了,但在这

1044 第九部分

种特殊的情况下,这个信念在他的推理中并没有起什么作用。再者,我们知道,常识心理学诉诸于两种不同的命题态度,其中之一在特殊的情况下是能动的,而另一种是无因果作用的。

在心理学文献中,不乏关于人的信念或记忆的模型,它们步常识心理学的后尘,假定命题模块性是真实的。而且,先于联结主义而出现的、几乎所有的关于命题记忆的模型,除行为主义者所强调的那些以外,都非常和谐地见容于命题模块性。一般来说,这些模型把主体的信念或记忆储存当作是一堆相互联系的状态,这些状态在功能上是具体的、在语义上是可解释的,并以系统的方式相互作用。其中一些模型把个体的信念表征为似句子的结构——一串符号,通过把它们从长期记忆输送到中心加工单元的更有限的记忆,它们可以个别地被激活。另一些模型把信念表征为标记节与标记链的网络,激活的模式正是由此而扩散的。还有一些模型把信念当作产生性规则的集合。^⑨在所有三种模型中,一般有这样的情况,即对于任何特定的认知阶段来说,如完成特定的推理或解答一个问题,某些记忆状态就会主动地被卷入,而别的则停止起作用。

在图1中,我们将用柯林斯和奎利恩(1972)的方式展示记忆的“语义网络”表征的一个片断。在这个模型中,记忆中的每一不同命题都为一种椭圆形的节所表征,这个节又与它的关联于各种概念的标记链连在一起。通过增加下述假定,它们是关于问题或其他类型的记忆试探电极导致激活扩散到那个网络的方式的,该模型就使我们能对各种记忆实验研究的速度和准确性作出预测。就我们的目的而言,关于这一模型的三个事实都具有特殊的重要性。第一,既然每一命题都以功能上具体的方式被编码,那么直接

有关的就是从记忆中加上或减去一个命题,而让网络的其余部分保持不变。例如图 2 就描述了从图 1 的网络中减去一个命题的结果。第二,该模型把表达了信念或记忆的语义属性的谓词当作是

语义学网络

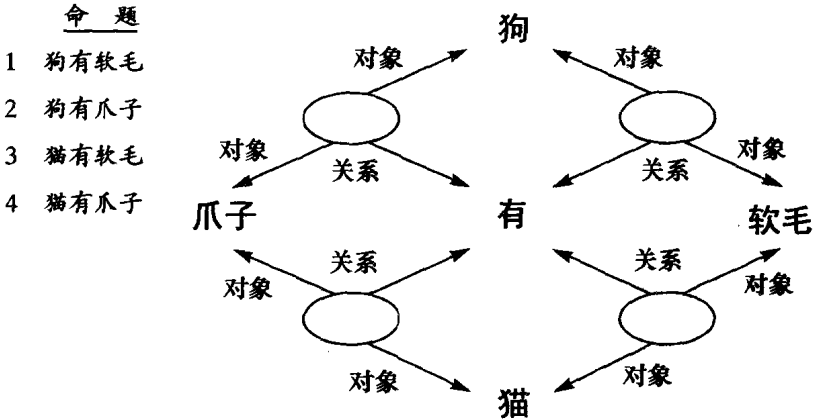


图 1

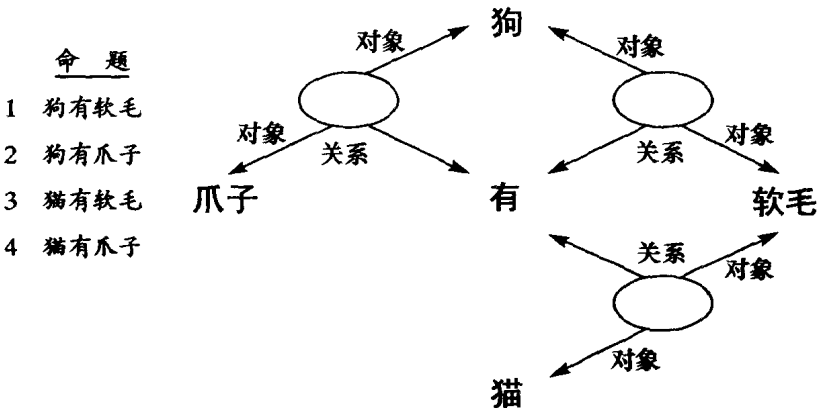


图 2

可投射的。^⑩它们被看作是这样的谓词类别,即从科学上遴选出来的名副其实类别,而不是纯偶然的集合体,因而适合于包含在关于似规律的规则的陈述之中。为了说明这一点,我们只需考察一下这些模型由关于记忆获得与失去的经验材料检验的方式。通常会作出这样的假定,即如果某被试被告知(例如)警察已逮捕了一嬉皮士,那么他就会(以一定的概率)记住警察已逮捕了嬉皮士。^⑪这个假定被认为表达了法则学上的普遍原则——它抓住了关于认知系统起作用的方式的某种似规则的东西。因此即使记住了警察逮捕了嬉皮士的那类人在心理学上的一切方面都彼此不同,但该理论仍会把他们看作是心理学上的自然类别。第三,在利用了语义网络模型的任何记忆搜索或推理任务中,追问哪一些命题被激活了、哪一些没有被激活是有意义的。因此在图1的网络中,如果没有激活猫有爪子这样的命题,那么搜索可能要终止。

IV. 联结主义假说的家族

在前一节,我们主要阐述了:常识心理学承诺了命题模块性,而认知心理学文献所倡导的许多记忆模型可与这一假定和平共处。在本节中,我们打算描述一类联结主义模型,我们要论证的是,它们并不乐意与命题模块性和平共处。我们所想到的联结主义模型具有三种属性:

(1)在联结权数和对单元的偏倚中,它们的信息编码是广泛分布的,而非局部主义的;

(2)网络中的个别隐匿单元没有适当的符号解释;用 P. 斯莫伦斯基的术语说,它们是亚符号性的(subsymbolic);

(3)那些模型像认知模型一样是有意图的,而不仅仅是认知模型的补充。

在本节稍后,我们再对这三个特征作进一步的阐述,在下一节,我们将描述一个简单的、符合我们三个标准的联结主义实例。不过,我们不奢望,我们所说的足以为我们所考虑的那类联结主义模型提供锋芒毕露的描述。这种有锋芒的描述对我们的论证是没有必要的。如果我们能使你相信:存在着一类有意义的联结主义模型,它们与民众心理学的命题模块性是不相容的,那就足够了。

在进一步说明我们所列举的三个特征之前,我们先对我们称之为“联结主义的”那类模型作个一般性的描述,然后介绍属于该领域的某种专门术语。为此,我们不妨花一定的篇幅援引 P. 斯莫伦斯基的明晰的观点:

联结主义模型是简单、并行的计算元件的巨大网络,这些元件的每一个都携带着数字激活值,而该值是它用某种更简单的数字计算从网络的其他相邻元件中计算出来的。网络元件或单元通过携带着数值强度或权数的联结相互影响各自的值
.....

在一典型的……模型中,对系统的输入是通过在网络的输入单元上增加激活值而实现的;这些数值表达了输入的某种编码或表征。对输入单元的激活与联结一同扩散,直至某组激活值出现在输出单元上;这些激活值编码在系统根据输入推

算出来的输出中。在输入和输出单元之间,可能还有其他单元,常被称之为隐藏的单元,它们不介入对输入和输出的表征。

在把活动的输入模式输送到输出模式的过程中,网络所完成的计算依赖于联结强度的集合;这些权数通常被看作是对那个系统的知识的编码(着重号为原文所加)。在这个意义上,联结强度在普通计算机中起着程序的作用。联结主义方案的主要魅力在于许多联结主义网络自我编制程序,即是说,它们有调节它们的权数的自动程序,因而最终能执行某种指定的计算。这种学习程序常常依赖于训练,正是在训练中,网络从它设定要计算的函数中得到了输入/输出的样本。在具有隐藏单元的学习网络中,该网络本身“决定了”那隐藏的单元将执行什么计算;因为这些单元既不表征输入,又不表征输出,它们从不被“告知”什么是它们应是的值,即使在训练期也是如此。(1988,1)

对斯莫伦斯基的描述必须作一点补充。在许多联结主义模型中,隐藏单元和输出单元被安排了一个数值“偏差”,它被加进了决定单元激活水平的计算中。一般而言,这种网络的学习程序确定了联结强度和偏差。因此,在这些网络中,那个系统的知识通常被当作是编码在联结强度和偏差中的。

一般性的说明就到此为止。下面让我们解释刻画了那些联结主义模型、而我们认为与命题模块性不相容的三个特征。

1. 在许多非联结主义的认知模型中,如第3节末尾所阐释的

那一个,很容易确定那模型的功能上不同的、编码系统中被表征的每一命题或事态的作用。而且,根据福多和皮利辛(1988,57)的观点,“通常的(计算机)构造要求:对它能表征的每一事态必须有不同的符号表达式。”在某些联结主义模型中,一种类似的功能定位不仅对输入、输出单元是可能的,而且对隐藏的单元也是如此。例如在某些联结主义模型中,各种个别的单元或单元的小型集合体本身就是要表征环境的特定的属性或特征。当从这样一个单元到另一单元的联结强度确实很强时,这就可能被认为是那个系统关于下述命题的表征,这命题是,如果第一个特征出现了,那么第二个也是如此。不过,在许多联结主义网络中,把命题表征定位于输入层次之上,这是不可能的。即是说,系统不具有这样的特殊特征或状态,它们为自己提供直接的语义评估。当系统作为整体因为没有以其应然的方式表征世界而未能实现其目的时,这有时可能为联结主义模型的创立者带来真正的麻烦。如果这一点碰巧发生了,那么如斯莫伦斯基所注意到的,“要为没有出现的非幻觉表征定位就一定是不可能的。任何特殊的状态都是状态的更大因果系统的组成部分,一般来说,系统未能满足目的条件不可能是根源于特殊状态或状态的组成部分”(1988,15)。在这类联结主义网络中,不可能把特定命题或事态之表征限制在节、联结强度、偏差的范围内,当我们谈论信息编码在广泛分布而非局部主义的偏差、权数和隐藏的节中时,我们所想到的正是这类网络。

2. 正如我们已注意到的那样,有这样一些联结主义模型,在其中的某些或所有单元都打算表征那系统的环境的特定属性或特

征。这些单元可以看作是这里所说的属性或特征的模型符号。不过,在其权数和偏差是由学习的规则系统调节的模型中,任何单个的单元或单元的小型集合常不会以直接的方式表征环境的特定的特征。在下节中我们将看到,即使任何隐藏的单元、权数或偏差不能恰当地看作是符号,但认为这些网络集中地、整体地编码一组命题则常常是合理的。如果出现了这种情况,那么我们会把那模型中所利用的表征策略称之为亚符号策略。一般来说(或许永远如此?),利用了亚符号表征策略的网络就将以广泛分布的方式编码信息。

3. 我们所列举的第三个项目并不是联结主义模型本身的特征,而是关于那些模型应怎样予以解释的一种观点。在阐述此观点时,我们必须以关于理论或解释层次的概念为前提条件,尽管目前的文献对此讨论很多,但这个概念远不是清晰的范式。^⑧介绍关于解释层次的概念最有把握的方式也许是在与下述熟悉的功能主义命题的背景之下进行,这命题是:心理学理论类似于可在各种非常不同的计算机上执行的程序。^⑨如果人们接受了这个类比,那么就有必要进一步探讨:一特殊的联结主义模型是否要成为在心理学层面或在基本的中枢执行层面上的一个模型。因为它们明显相似于真正的中枢构造,即使在许多方面是局部地相似,因此把联结主义模型当作是执行心理加工的模型则是很诱人的。某些联结主义模型的构造者旗帜鲜明地承认了这一点。不过,如此看来,联结主义模型并不是心理学或认知模型,就像关于认知过程怎样在量子力学层面完成的说明不是心理学说明一样。联结主义模型的构造者能够利用且经常利用的一个非常不同的观点是:他们的模型

位于心理学层面,而不在执行层面。如是解释,该模型与关于同一现象的其他心理学模型就是对立的。因此关于语词识别的联结主义模型就是取代关于语词识别的非联结主义模型的东西——而不仅仅是对非联结主义模型的一种可能的执行;关于记忆的联结主义理论常常是语义网络理论的劲敌,如此等等。对其理论持这种观点的联结主义者常常借助与其他科学的类比来阐述这样的观点。例如,斯莫伦斯基主张:联结主义模型取代了传统的认知模型(如语义网络),这极像量子力学取代经典力学。无论在哪种情况下,更新的理论对于更广泛的现象来说,是更深刻、更普遍、更准确的。但是在每种情况下,新的理论与旧的理论在同一的解释层面又是针锋相对的。如果一个正确,那么另一个就一定是错误的。

在本文中,就我们关心的而言,存在着这样一个方面,在那里,可以认为联结主义模型与量子力学之间的类比回避了一个重要问题。因为当量子力学被认为是比经典力学更好的理论时,就可能提出这样一个似乎合理的论点:从经典力学向量子力学的转化是一种本体论上保守的理论变革。无论如何,从本体论上来说,这场变革显然不是激进的。如果在本文中,我们的中心论题是正确的,那么联结主义模型与更传统的认知模型之间的关系更像关于热的热质理论与分子运动理论之间的关系。热质理论与分子运动理论处在同一的解释层面,即使从本体论上来看,从一个向另一个的转变显然是相当激进的。为了证明热质的类比是更恰当的类比,那么描述一种具体的、即使是很简单的、能符合我们一直试图说明的三个标准的联结主义记忆模型,将是大有助益的。

V. 关于记忆的联结主义模型

在建构这种模型时,我们的目的是建立一种联结主义网络,它至少能承担关于记忆的更传统的认知模型所承担的某些任务,而且能清清楚楚地完成前一节所描述的那种分布式的、亚符号的编码。首先,我们建构一个网络,并称之为网络 A,它能判断图 3 (底)横线之上所列举的 16 个命题之真假。该网络是一典型的三等线划分送进网络,由 16 个输入单元、4 个隐藏单元、一个输出单元所构成,如图 4 所示。每个命题的输入编码见图 3 的中柱。输出接近 1 便认作真,接近 0 便被认作假。反向传播*是一众所周知的联结主义学习规则系统,常用来“训练”那个网络,借以建立联结权数和偏差。当那网络对每一真命题始终如一地给予 0.9 以上的输出、对每一假命题给予低于 0.1 的输出时,训练便告结束。图 5 表明:在受过训练的网络中,联结权数位于输入单元和最左边的隐藏单元之间,与那单元的偏差存在一起。图 6 说明:联结权数和偏差仍在上游。图 7 表明:当“狗有软毛”这一命题在输入单元中被编码时,该网络用哪一种方式来计算对那命题的反应。

显然有这样的意思,即受过训练的网络 A 可以说储存着关于命题 1-16 的真或假的信息,因为当这些命题的随便哪一个显现给网络时,它就能正确地判断那命题是真还是假。在这方面,它相

* back-propagation,指通过在网络上逆向传送一个错误信息(即从产生单元开始)来训练神经网络的一种程序。

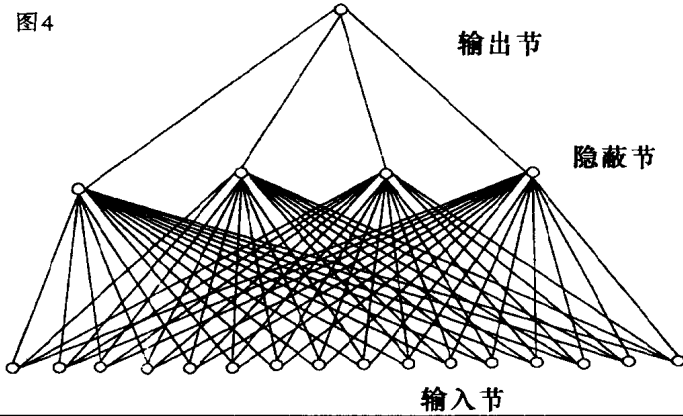
心理解释与民众心理学的地位 1053

命题	输入	输出
1 狗有软毛	11000011 00001111	1 真
2 狗有爪子	11000011 00110011	1 真
3 狗有蚤子	11000011 00110011	1 真
4 狗有腿	11000011 00111111	1 真
5 猫有软毛	11001100 00001111	1 真
6 猫有爪子	11001100 00110011	1 真
7 猫有蚤子	11001100 00111111	1 真
8 鱼有鳞	11110000 00110000	1 真
9 鱼有鳍	11110000 00001100	1 真
10 鱼有鳃	11110000 00000011	1 真
11 猫有鳃	11001100 00000011	0 假
12 鱼有腿	11110000 00111100	0 假
13 鱼有蚤	11110000 00111111	0 假
14 狗有鳞	11000011 00110000	0 假
15 狗有鳍	11000011 00001100	0 假
16 猫有鳍	11001100 00001100	0 假
附加的命题		
17 鱼有蛋	11110000 11001000	1 真

图 3

1054 第九部分

图4



网络A

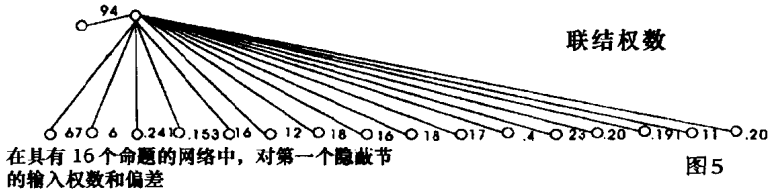


图5

网络A

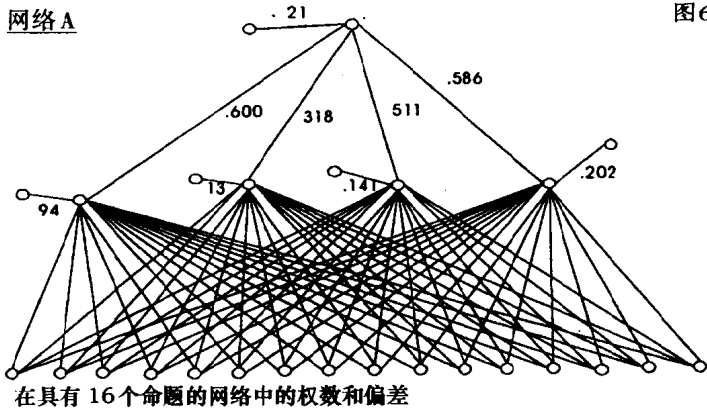


图6

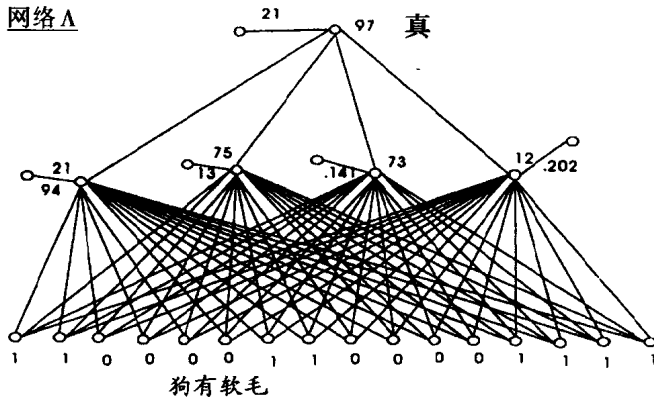


图 7

似于能完成相同任务的各种语义网络模型。不过,在网络 A 和语义网络模型之间也存在着一个鲜明的差异,如图 1 所表明的那种差异。因为正如我们先前所强调的,在语义网络中,存在着一种功能上不同的、与每一命题有联系的子部分,因此对于那网络的任何探测器来说,它就完全有理由询问:一个指定命题的表征是否起着因果作用。相比较而言,在联结主义网络中,就没有用来表征任何特殊命题的不同的网络状态或部分。在网络 A 中被编码的信息整体地储存着,并分布在整个网络中。每当信息从网络 A 中提取出来,通过给它一个输入串,看它为输出单元是算出一个高值还是一个低值,这样许多联结强度、许多偏差、许多隐蔽的单元就可在计算中发挥作用。任何特殊的权数或单元或偏差都有助于编码关于许多不同的命题的信息。如果问特殊命题之表征在网络的计算中有无因果作用,那显然是没有意义的。正是在这一方面,我们的联结主义记忆模型与常识心理学的命题模块性似乎是格格不入

1056 第九部分

的。因为正如我们在第 2-3 节中所看到的那样,常识心理学似乎预先假定:对特定信念或记忆在特定的认知进程中是否起着因果作用这一问题一般有某种答案。但是如果信念和记忆由像我们的网络那样的联结主义网络来实现,那么这样的问题似乎就没有明确的意义。

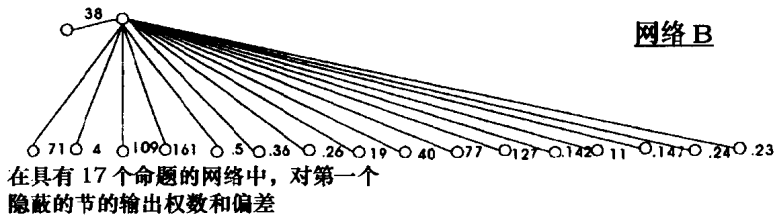


图 8

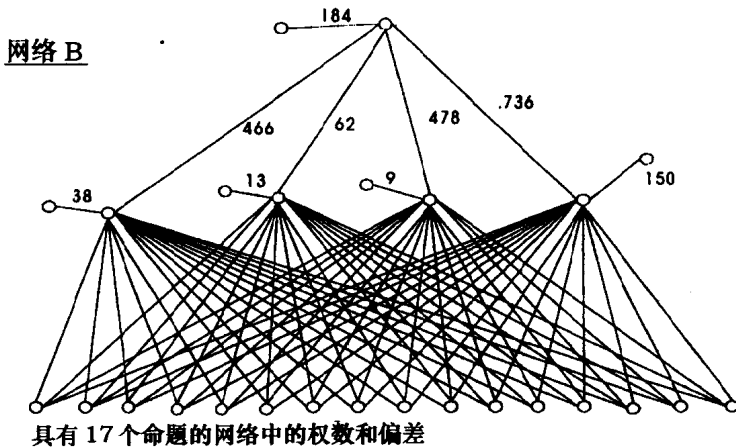


图 9

如果我们对网络 A 与我们在图 8 和 9 中所描述的第二种网

络即所谓的网络 B 作一比较,那么命题模块性与我们的联结主义模型之间的不相容性将变得更加明朗。网络 B 正像第一个网络一样,是受过训练的,所不同的是有一附加命题加在训练装置之上(像图 3 横线之下所说明的那样被编码)。这一来,网络 B 编码了所有相同的命题,就像再加了一个命题的网络 A 一样。在语义网络模型和别的传统认知模型中,要说出系统的哪一状态或特征编码了附加命题,这是一件轻而易举的事情。而且要确定附加命题的表征在该系统所模仿的特定过程中是否有作用,这也是一个简单的任务。但是在联结主义网络中,这些问题显然是没有什么意义的。这不是说两种网络之间没有差异;情形恰恰相反,差异很多且很广泛。但是这些差异并不以整体的方式与民众心理学及更传统的认知模型所假定的功能上具体、语义上可解释的状态相关联。既然信息以高度分布的方式被编码,与包含着对许多命题都很突出的信息的每一联结权数和偏差相关联,既然关于任何特定命题的信息分布在整个网络中,因此该系统就没有功能上不同的、可以加以区分的、可从语义上解释为关于个别命题的表征的子结构。

对网络 A 和网络 B 的比较能够使我们以不同的方式对常识心理学与这些类型的联结主义模型之间的不相容性作出我们的判断。在第 2 节中,我们已注意到,常识心理学把表达了命题态度语义属性的谓项当作可投射的。因此在常识心理学中,“相信狗有软毛”或“记得狗有软毛”就将是可投射的谓项。网络 A 和 B 大概可作为相信狗有软毛的认知代理人的模型;两个网络都储存或表征了关于狗有软毛的信息。这绝不是仅有的两个。如果我们用图 3 的 17 个命题连同增加的一些(或减去一些)命题去训练一个网络,

那么我们就又得到了一个系统,它不同于网络 A 和 B,就像 A 与 B 相互有别一样。这里的启示是:即使有无穷无尽的联结主义模型,它们像网络 A 那样表征了狗有软毛这样的信息,但这些网络也没有共同的、能用联结主义理论的术语描述的可投射特征。根据联结主义模型创立者的观点,可模拟相信狗有软毛的认知代理人的那类网络,并非地道的类别,而完全是一种混沌无序的分离装置。常识心理学把那类相信狗有软毛的人作为心理学上的自然类别;而联结主义心理学则不然。^⑭

VI. 驳难与答复

我们在前几节所作的论证已碰到了许多驳难。在本节中,我们将重构其中最令人感兴趣的驳难,并陈述我们的答复。

驳难 1: 模型 A 和 B 并不是关于人类信念或命题记忆的严肃的模型。

诚然,我们所建构的模型是一些为着以一种显而易见的方式说明第 3 节中所陈述的特征而制作的小玩艺儿。它们无意于模拟人类命题记忆的任何根本方面。但是人们已基于各种理由,提出了这样的怀疑,即诸如此类的模型能否被严肃地当作命题记忆的心理模型。某些批评者断言:这些模型绝不会按比例增加——因为一网络识别 15 或 20 个命题可能易如反掌,而训练一网络去识别数千个命题则是不可能的,更不用说数万个命题。^⑮还有一些人指责说:更传统的记忆模型,包括以似句子存储为基础的模型、运用了语义网络的模型、以产生式系统为基础的模型,都能为推理

或泛化提供某种战略,这种推理或泛化能使那个系统回答关于它未明确被教过的命题的问题,而我们所构想的那些模型则不能推理和泛化。还有人论证说:这些模型作为对人类记忆的说明已惨遭失败,因为它们没有提供明显的方法去说明这样的事实,即作好了适当准备的人能轻而易举地得到一个命题同时所具有的命题信息。在通常的条件下,我们可以告诉亨利:汽车钥匙在冷藏室里,接着他就能在其记忆中可靠地记录下这个事实。他用不着诸如大量的再训练之类的东西,而再训练是把一个新命题教给我们的一个联结主义网络所必不可少的。

答复:如果这是一篇旨在为联结主义的命题记忆模型作辩护的论文,那么我们就必须对这些所谓的缺陷的每一个作出细致的处理。无论从哪方面说,至少可以为联结主义说点什么。例如,可以说下述观点是错误的,即 A 和 B 这样的网络没有泛化到该网络被训练过的命题之外。比如说在网络 A 中,训练安排就包括:

狗有软毛。猫有软毛。

狗有爪子。猫有爪子。

狗有蚤。猫有蚤。

还包括:狗有腿,

但不包括:猫有腿。

不过,当该网络被给予了关于最后一个命题的编码时,它能正确地加以泛化,并肯定地作出反应。同样,该网络会对下述编码作出否定的反应:猫有重量。即使这个命题先前没有显示给它。

然而,考虑到我们在本文中所辩护的论点,对说我们的网络是关于命题记忆的不恰当的模型的这类指控逐个地作出回答就是多

1060 第九部分

余的,认识到这一点极为重要。因为我们试图建立的是一种条件命题:如果我们在第3节中所描述的那类联结主义记忆模型是正确的,那么命题态度心理学便陷入了可怕的困境。既然包含有假前件的条件句是真的,那么我们便是因错而胜,如果我们的条件句的前件真的是虚假的话。

驳难 2a: 我们的模型其实并没有违反命题模块性原则,因为那系统所学到的命题是以功能上具体的方式编码的,即使这也许不明显。

我们早有耳闻,这一驳难是按三条不同的路线而阐发的。第一条路线——让我们称之为驳难 2a——特别提到:功能上具体的编码也许常常很难被注意到,因此在因果检查中可能无望被看到。例如,试考察句子储存在典型的冯·诺伊曼式的计算机的记忆中的方式——具体地说,我们可以设想:当那计算机正运行语词加工程序时,那些句子便成了一篇英文文章的部分,并被储存起来。句子的部分可以以复杂的方式储存在物理上分散的、连在一起的记忆址中,即使对所有有关记忆址的内容能作出说明,人们也很难说一个特定的句子储存在哪里。但尽管如此,每个句子是以功能上具体的方式储存的。因此如果人们对那系统了如指掌,那么就有可能因窜改相应记忆址中的内容而删去正在储存的任何特定句子,同时使那系统正储存的其他句子成为不可企及的。同样,还有人主张:联结主义网络事实上可以以功能上具体的方式编码命题,即使在对受训练网络的偏差和联结强度的因果检查中,这一点也许不明显。

答复 2a: 要对付这一驳难的确有点困难。因为批评者所主张

的是:在像我们所建构的那些模型中,可能存在着某种隐蔽的、有待发现的、功能上具体的命题编码系统。在回答这一点时,我们必须承认:可能的确如此。我们确实还没有这样的论据,它甚至能证明:发现这种隐蔽的、功能上具体的编码是不可能的。而且我们还承认:如果这种隐蔽的系统被发现了,那么我们论证的基础将受到严重的瓦解。不过,我们倾向于认为:摆在批评者面前的论证重任就是去说明:这种系统不仅是可能的,而且是很有希望的;如果没有任何严肃的理由,而又偏要说,我们的网络是以功能上具体的方式编码命题,那么“它们可能怎样”这样的纯逻辑可能性对我们几乎不是什么严重的威胁。

驳难 2 的第二种变种——我们将称之为驳难 2b——为 A、B 之类的网络可能以之具体地、即使是隐蔽地编码命题的方式提出了一个明确的建议。据说,当一特定的命题呈现给那网络时,那种编码在隐蔽的结节激活模式中应是找得到的。既然我们的网络中有四个隐藏的结节,那么关于任何特定输入呈现的激活模式就可以表现为一个有序的 4-元组。例如当网络 A 被提供了狗有软毛这样的被编码命题时,如图 7 所示:相关的 4-元组就是(21, 75, 73, 12)。同样地,我们可以把每一激活模式看作是四维非欧几何空间中的一个点。既然每个命题对应于那多维空间中的一个点,那么那个点就可视之为命题的编码。而且,那个点表征了那个系统的功能上具体的状态。^⑥

答复 2b:上述建议不过是说:那个系统关于命题 p 的一个编码之呈现的激活模式与信念 p 是同一的。但这个主张是极其不合理的。明白这一点的最好方法也许就是注意到:在常识心理学中,

1062 第九部分

信念与命题记忆一般具有基本的持续性;而且它们是这样的东西,即哪怕是认知动因没有使用它们,但认知动因一般拥有它们中的许多成分。试看一个例子。大袋鼠是有袋目吗?多少年来你肯定相信:它们是有袋目,即使今天你很可能是第一次激活或使用你的信念。^⑩然而一种激活模式并不是网络的一种持续的状态;甚至它肯定不是网络的一种状态,除非那网络有作为输入的相关命题。而且有大量其他的信念,是你多少年来所持有的。但是假设网络可能有在长时期内绵延不绝的大量激活模式,这是没有意义的。在任何特定的时刻,一种网络充其量只表现一种激活模式。因此激活模式并不是能合理地等同于信念或它们的表征的东西。

驳难 2c:在这个节骨眼上,许多批评者主张:持久的信念不可能同一于激活模式,后者是网络的一种过渡的状态,但可同一于产生激活模式的倾向。因此在网络 B 中,关于狗有软毛的信念不能等同于激活的多维空间中的一个选址,而可同一于这样的倾向,即当命题被呈现出来时网络最终要在那个选址停留的倾向。这种倾向状态是系统的一个持续的状态;网络能进入的就是这种状态,不管其现在的激活状态是什么,正像糖块在水中有溶化的倾向一样,尽管附近并没有水。^⑪某些人还进一步设想:在作为倾向的信念和已发生的信念之间的那种流行的哲学区分,在联结主义网络中是可以予以反映的,就像产生激活模式的倾向与激活模式本身的区分那样。

答复 2c:我们对这主张的答复是:即使产生激活模式的倾向真的是系统的持续状态,但它们并不是所需要的那类持续状态——它们不是民众心理学所需要的、具体的、因果上独立的能动状

态。不要忘记,根据民众心理学关于信念和推理的概念,常存在各种各样相当不同的、基本的因果模式,它们可以导致特定信念的获取和认可。当克洛斯说男管家撒谎了时,他也许是利用火车出故障这样的长期信念作出这样的推理的;或者他还可能利用旅馆关门了这样的信念作出他的推论;或者两个长期信念都在他的那个推理中起作用。而且也有这样的可能,即克洛斯以前作出了这个推论,现在才来报告这个相对长久的信念。但难以弄明白的是:这类区分怎么可能为所谓的倾向说明所涵盖呢?在对一特定输入比如说 p 作出反应时,网络所接收的是特定的激活值。也许有接收关于别的输入如 q 和 r 的别的激活值的倾向。但没有明显的办法去解释这样的主张,即这些进一步的倾向在网络对 p 作出反应时起着作因果作用——或就那一点来说,它们没有起什么作用。我们不可能阐释这样的观点,即在一种情况下, q 的编码(即火车出故障这一命题)起着一种作用,而 r 的编码(即旅馆关门了这一命题)并没有作用,在另一情况下,事情又是另一番景象。常识心理学预先假定的命题模块性要求的是:信念标记在某些认知阶段必须是功能上具体的、能相互发生因果作用的状态,而在另一些认知阶段则仍是因果上迟钝的状态。不过,在像网络 A 之类的分布式的联结主义系统中,产生一种激活模式的倾向状态在功能上与产生另一模式的倾向状态是不可分割的。因此在某些认知过程中,就不可能把某些命题分离出来作为因果上能动的东西,而把另一些当作非能动的。我们的结论是:反应模式倾向不会作为信念标记而起作用。就我们所知,在像 A 和 B 这样的网络中,也不存在别的、能满足那种要求的状态。

VII. 结论

在本文中,我们所辩护的论点是:某种类型的联结主义模型与常识心理学所包含的命题模块性是势不两立的。这里所讨论的联结主义模型就是这样的模型,它们在认知层面作为模型而被提供出来,在其中,信息编码是广泛分布的,且是亚符号式的。在这种模型中,我们要论证的是:不存在着具体的、语义上可解释的、只在某些认知阶段、不在别的阶段发挥因果作用的状态。因此,在这些模型中,不存在民众心理学的命题态度可以合理地与之同一的东西。如果这些模型最终为人类的信念和记忆提供了最好的说明,那么我们将面临着一场本体论上的激进的理论变革——这类理论变革将支持这样的结论,即命题态度像热质、燃素一样根本就不存在。

注 释

① 例如可参阅 P.M. 丘奇兰德(1981;1986),在那里,基于对类似于我们将讨论的那些认知模型之成功的反思,得出了明确的取消论结论。

② 福多与皮利辛(1988)。

③ 我们知道,某些哲学家和科学史家实际上坚持类似于下述建议的观点,即前哥白尼宇宙学家所说的行星是不存在的。例如可参阅库恩(1962,第10章),费耶阿本德(1981,第4章)。不过,我们认为这个建议是不合理的。取消论的论证不可能那样易如反掌。关于这些问题的进一步讨论,可参阅库恩(1983)、P. 基切尔(1978;1983)。

④ 关于科学还原的某种详细的讨论,参阅内格尔(1961),P. 基切尔(1984)。遗传学的案例并非没有争议,参阅 P. 基切尔(1982;1984)。

⑤ 值得注意的是:对这一问题的诸判断在本质上是彼此有别的。一极是像费耶阿本德(1981)也许还有库恩这样的作者,对于他们来说,理论上相对小的差异足以证明这样的猜测,即存在着本体论上激进的变革。在另一极,有利康这样的作者,他写道:

我对倡导一种非常自由的观点可以说是费尽心机……我从心底里赞成放弃关于信念或愿望(或其他东西)的常识的或陈腐的理论,而且也赞成这样的判定,即我们在许多问题上是非常错误的,且不去推论说:我们再也不会谈论信念或愿望……我认为日常用语“信念”(民众心理学的理论术语)朦胧地指向了我们对之完全没有把握、有待于成熟的心理学去揭示的自然种类。我希望“信念”最终指称的是表现了有情识存在的内在状态的信息,……但是它指称的那类状态也许只有通常为常识赋予信念的一些属性。(1988a,31-32)

在我们看来,两种极端的观点都是不合理的。正如我们早就指出过的,哥白尼革命并未表明:托勒密所研究的行星是不存在的。但是拉瓦锡的化学革命也没有表明:燃素是不存在的。然而根据利康的“非常自由的观点”,闹不明白的是:我们为什么不应断言,燃素毕竟是真实存在的——它确实是一种氧气,在拉瓦锡之前,“我们在许多问题上是非常错误的。”

⑥ 常识心理学是一种理论,关于这种观点的较早的、且有影响的陈述,请看塞拉斯(1956)。近来,该观点得到了P.M. 丘奇兰德(1970;1979,第1、4章)和福多(1987,第1章)的辩护。至于对立的观点,可参阅威尔克斯(1978),马德尔(1986),夏普(1987)。

⑦ 请看斯蒂克(1983,237ff)。

⑧ 切尔尼克(1986,第3章)指出:这种无心性在文学作品和日常生活中很常见,有时会导致可怕的后果。

⑨ 关于句子模型,请看麦卡锡(1968;1980;1986),金特施(1974)。关于语义网络,请看奎利恩(1966),柯林斯和奎利恩(1972);拉梅尔哈特、林赛和诺曼(1972),安德森和鲍尔(1973),安德森(1976;1980,第4章)。关于产生式系统,请看纽厄尔与西蒙(1972),纽厄尔(1973),安德森(1983),霍兰等(1986)。

⑩ 关于可投射谓词与不可投射谓词之间的区分的经典讨论,请看古德

1066 第九部分

曼(1965)。

⑪ 例如,可参阅安德森和鲍尔(1973)。

⑫ 布罗德本特(1985),拉梅尔哈特和麦克莱兰(1985),拉梅尔哈特、麦克莱兰与 PDP 研究组(1986,第4章),斯莫伦斯基(1988),福多与皮利辛(1988)。

⑬ 这里所利用的程序概念对两种相当不同的解释是开放的。相关的读物,可参阅拉姆齐(1989)。

⑭ 这种关于联结主义模型与常识心理学之间的不相容性的观点的阐述方式得益于福多的启发。

⑮ 这个观点最先是由 D. 丹尼特提出的。

⑯ 许多人倡导这种运动,例如 G. 科特雷尔和 A. 卡辛斯。

⑰ 关于信念的常识概念,正如利康所述,即使人们睡着了,也还忘不了其中的大多数。

⑱ 这类对我们的驳难是 N. 布洛克和 F. 杰克逊提出的。

参考文献

S. 斯蒂克(1983):《从民众心理学到认知科学》,麻省理工学院出版社。

J. 福多(1983):《心灵的模块性》,麻省理工学院出版社。

A. 柯林斯与 M. 奎利恩(1972):“关于语义记忆和语言理解的实验”,载 L. 格雷格编:《学习和记忆中的认知》,约纽:威利

P. 斯莫伦斯基(1988):“论对待联结主义的正确态度”,《行为与脑科学》,11,1-74。

J. 福多和 Z. 皮利辛(1988):“联结主义与认知结构:批评性分析”,《认知》,28,3-71。

P.M. 丘奇兰德(1981):“取消唯物主义与命题态度”,《哲学杂志》,78,67-90。

——(1986):“认知神经生物学中的某些还原战略”,《心灵》,95,279-309。

T. 库恩(1962):《科学革命的结构》,芝加哥大学出版社。

P. 费耶阿本德(1981):《哲学论文集》,第1卷,剑桥大学出版社。

E. 内格尔(1961):《科学的结构》,纽约:哈考特,布雷克斯。

- W. 利康(1988):《判断与证明》,剑桥大学出版社。
- K. 威尔克斯(1978):《物理主义》,伦敦:劳特利奇与基根·保罗。
- R. 夏普(1987):“民众心理学的观念”,《探索》,30,381-393。
- C. 切尔尼克(1986):《最低限度的理性》,麻省理工学院出版社。
- A. 纽厄尔和 H. 西蒙(1972):《人类问题解决》,恩格尔伍德·克利弗斯,新泽西:普伦蒂斯·霍尔。
- J. 安德森(1983):《认知的结构》,哈佛大学出版社。
- N. 古德曼(1965):《事实、虚构和预测》,印第安拉波利斯:博布斯-梅里尔。
- D. 布罗德本特(1985):《层次问题:评麦克莱兰和拉梅尔哈特》,《实验心理学杂志:一般观念》,114,189-192。
- W. 拉姆齐(1989):“平行论与功能主义”,《认知科学》,13,139-144。

民众心理学不会消失*

特伦斯·霍根 詹姆斯·伍德沃德

民众心理学是这样一种原理之网,它构成了一种关于怎样解释人类行为的常识理论。这些原理授予某些命题态度尤其是信念和愿望以至至关重要的作用。例如该理论断言:如果某人有愿望 p , 且这愿望又不因别的愿望而失效,如果他相信:某种行动 K 将使 p 实现,如果他还相信这样的行动在他力所能及的范围之内,但不相信某种别的行动在他的能力范围之内,并且是使 p 实现的理想方式,那么在同等条件下,那个愿望和那些信念就会使他作出 K 类行动。这种理论起着重大的作用,因为它所假定的状态首先是根据它们的相互因果关系以及它们与知觉、别的环境刺激和行为的因果关系而被描述的。

民众心理学(以下简称为 FP)深深扎根于我们关于作为人的我们自身的常识概念之中。不管人是别的什么,但据假定,他首先是一有理性的(至少是很理性的)行动者——即是这样的生物,其行为整个地由他的信念、愿望和有关的命题态度所引起,并可据以得到解释。因此完全拒斥 FP 就意味着对我们的概念图式的彻底

* 该文最先发表在《哲学评论》,XCIV, No. 2(1985, 4),后收入 W.G. 利康编辑的《心灵与认知文选》一书。现根据后者译出。

修改。在面对显而易见的困难时,这一事实对我们来说就成了不能立即抛弃 FP 的初步而有力的理由。

然而,FP 近来受到了两方面的夹击。P. 丘奇兰德(1981)论证说:既然 FP 与我们一起生活了至少 25 个世纪,从而不是发展下述心理学理论的深思熟虑的和自我意识的尝试的产物,这种心理学理论与自然科学所提供的对人类的解释是一致的,因此就几乎没有理由假定:FP 是真的,或假定人类拥有信念和愿望之类。S. 斯蒂克论证说:认知科学已有的成果表明:成熟的认知心理学所假定的事件或状态绝不能等同于 FP 所假定的事件和状态;斯蒂克主张:如果真是这样,那么它就表明:FP 是极其虚妄的,人类根本不可能拥有像信念和愿望之类的心理状态。

在本文中,我们要论证的是:丘奇兰德和斯蒂克并没有为怀疑 FP 的完善性提供令人信服的理由。我们的讨论主要想说明的是:对于 FP 必须怎样与低层次的理论相融合以与之保持一致,他们持有一种不合理的、苛刻的概念。我们不否认这样的可能性,即 FP 不可能与更广泛的理论保持一致。例如如果正确的理论心理学最终是一种激进的、斯金纳式的行为主义,那么这种可能性就碰巧会发生。但我们坚持认为:没有理由假定,事实上会出现这种可能性。

在正式阐述之前,先讲几个预备程序。第一,我们将在广义上使用由特殊字体写成的“事件”一词,不仅包括标记变化,而且还指标记状态、标记过程。因此在我们的术语学中,非瞬间的民众心理学标记状态都将被看作是心理事件。

第二,我们认为 FP 由两部分构成:一组理论原理和一种存在

1070 第九部分

命题。许多或所有理论原理可能有我们开头一段的例子所例示的一般形式;即是说它们是条件公式的普遍闭合。^①就其本身而论,它们并不带有任何存在意义,因为它们大概都是空洞的真。另一方面,FP的存在命题是这样的断言,即我们关于人的日常民众心理学描述一般是真的,人们一般都经历过民众心理学的事件,我们常把这些事件归之于他们。我们承认,丘奇兰德和斯蒂克首先要反对的就是FP的存在命题;即是说,他们主张:我们日常的民众心理学描述是极其虚妄的,根本就不存在信念、愿望之类的东西。因此就我们的理解而言,他们的论证表达的是这样的可能性,即FP的理论原则是真的,当然仅仅是没有意义的真。

第三,我们没有必要断言:FP在一切方面都是完全正确的,或者说根据认知科学或神经科学的新发展,对FP作出修正或改进是没有可能性的。确切地说,我们主张的是:FP的理论原理大体上是正确的,日常民众心理学归因常常是真的。

第四,我们想把我们自己与时下很有影响的战略区分开来,后者力图把FP与潜在的科学可证伪性隔离开来——此即D. 丹尼特的工具主义(1978, 1981)。关于信念和愿望,他说:这些“所谓的……状态”可归类为“行动预言和行动解释演算中的理想化虚构角色”(1978, p. 30)。它们不是莱辛巴赫所谓的“演绎-假设的理论实在”;相反,他断言,它们是“抽象物——与计算连在一起的实在或逻辑构造”(1981, p. 13),其地位类似于力的平行四边形中的分力(1981, p. 20)。总之,他明明白白地说:它们是工具主义的虚构,因此它们与我们在认知科学或神经科学中发现的任何东西实际上是是可以和睦共处的。我们拒绝丹尼特的工具主义。我们坚持

认为:FP除了提供了有用的预言图式之外,还提供了真正的因果解释。即使对FP的意向习语的工具主义态度可与这些习语的纯粹的预言用法一致,但显然不能与它们的解释用法一致,或者说不可能与把信念和愿望作为原因看待的话语相一致。因此,有必要对FP作出比丹尼特的工具主义更强有力的辩护。

I

丘奇兰德(1981)否定FP可与神经科学相容的论证基于三方面的考虑。第一,“FP面临着大量的解释难题”(p.76)。第二,“至少在25个世纪中,它一直停滞不前”(p.76)。第三,“它的意向范畴孤芳自赏,没有还原于神经科学的任何可望可及的前景”(p.75)。不可还原性是其主要的顾虑,而且据说它又得到了其他两点的支持:“在我看来,成功的还原是不可能被排除的,但是FP的解释上的无能和长期的停滞使人对它的范畴在神经科学构架中能找到近似的反映几乎失去了信心”(p.75)。

下面让我们依次考察丘奇兰德的三个观点。在阐释第一点时,他写道:

作为核心的、重要的心理现象的例子,在FP构架中仍主要是或完全是神秘的,试考察心理疾病、创造性想象能力的本质与动力学……睡眠的本质和心理学功能……在跑动中抓住外场飞球的普通能力……三维视觉映象的内在构造……各种丰富多彩的知觉幻象……记忆的奇妙性……学习过程本身的本质(p.73)。

1072 第九部分

这一段,至少有两个重要的方面是错误的。第一,FP 本身对丘奇兰德所提及的问题几乎什么也没说,而来自于 FP 的概念为基础的理论对它们则有很多要说的。例如认知心理学家对视知觉、记忆、学习已阐发了广泛的、详尽的理论,而这些理论运用了民众心理学的信念、愿望、判断等公认的概念。^②下面要考察的归因理论和认知不协调理论与斯蒂克的名字紧密相连,对它们的阐释是这类理论的重要方面。说所有这些理论都没有解释力,这是最不合理的,而要这样说,无论如何需要丘奇兰德所没有提供的那类详细的经验论证。

第二,丘奇兰德的论证似乎强加了这样一种要求,即任何成功的心理学理论必须说明现象的某种前定的范围,且以统一的方式去说明。这类一般性的论证值得以怀疑主义和小心谨慎的态度加以对待。科学史中包含了大量的事例,根据这些事例,我们合理地希望一个理论说明或整合哪些现象这样的前理论预期已被证明是极其错误的。例如这样的要求,常常强加给早期的光学理论,以至于让它们说明我们今天认为与视觉生理学或心理学有联系的事实;这样做对早期的光学理论化具有恶劣的影响。同样的事例在化学史中也俯拾即是。^③

一般认为,对某种一般性理论可望说明哪些现象的合理判断离不开广博的理论知识;如果我们的理论知识相当原始,好像它只是关于许多心理现象的,那么这种判断就会误入歧途。指望让像 FP 这种原本是为了根据信念、愿望之类去解释日常的人类行动的理论,去说明与视知觉、睡眠或复杂的肌肉协调有联系的现象,这是没有任何可靠的、先天的理由的。关于后一类现象的真理与关

于前一类现象的真理可能是风马牛不相及的。

怎样看待丘奇兰德的第二个论证即 FP 多少世纪以来一直停滞不前呢？首先，在我们看来，至少可以论证说：在漫长的岁月中，FP 肯定经历过重要的、经验观察方面的进步，而不是停滞不前的。例如有这样一种合理的推测，即 18 或 19 世纪的欧洲人比 20 世纪的欧洲人更倾向于根据具有持久的人格特征的性格类型解释人类的行为，因为后者常常诉诸于“情境”因素。（这种差别在 18 或 20 世纪的文学作品中是显而易见的。比如，试比较简·奥斯汀与约翰·马思。）^④在当代文化中，经验观察方面的进步的另一例子也许就是人们更乐意诉诸于无意识信念和动机。

对“经验上无进步性”这一论证提出质疑的另一个理由就是：利用了似信念和似愿望事件的认知心理学理论已得到了大量新颖的、令人震惊的预言，而这些在过去常常要靠实验才能形成。[下面我们将讨论一些相关的例子。至于另一些突出的案例，建议读者读尼斯比特和罗斯(1980)。]然而丘奇兰德似乎论证说，FP(所谓的)的经验上无进步性好像是主张任何模仿了 FP 的理论皆错的一个充分理由。^⑤这几乎等于论证说，任何利用了中心力的深奥的物理理论必定是错误的，因为一般人关于推与拉的观念在经验上是没有进步的。

更有甚者，“经验上进步”的标准在评价像 FP 这样的理论时无论如何是没有一点用处的。FP 的通常的使用者对于运用预先存在的理论以作出关于人类行为的特殊事例的特殊因果判断津津乐道，而对系统阐述新的因果普遍原则则无兴趣。他是因果原则的消费者，而不是发明者。就此而言，他相似于历史学家、侦探或

1074 第九部分

对无生命对象作出了常见的、独特因果判断的人。我们认为,用明显是适于评价旨在系统阐述新的因果普遍原则的理論的标准去评价这些活动显然是不适当的。

如果人们意识到:隐藏在日常的(但却是非心理学的)具体因果判断后面的大量理论,即使不是在全部历史也是在过去数千年中,经历了缓慢的变化,那么就会得出上述结论。我们与我们的祖先都会判断:石块的撞击会使锅罐破碎,缺水会使骆驼死去,对A的头部的沉重打击会使其丧生,热使水沸腾等等。这些判断都不是经验上快速进步的理論的组成部分,然而(只根据这些就)断言它们可能是虚假的则是荒谬可笑的。对历史学家所运用的大多数(即使不是全部)没有疑义的因果理論也可以作出同样的结论。这些例子足以使我们认识到:并非所有的民众理論都能被一概斥之为极端的謬论。

这便把我们带到了丘奇兰德的第三个也是最根本的论证:FP与神经科学具有所谓的不可公度性:即前者不能还原为后者。他描述说,理論间的理想还原有两个主要特点:

第一,它为我们提供了一组规则——“对应规则”或“桥梁法则”,存在于标准的日常用语之中——它引起旧理論(T_0)的术语映射于新的或还原理論(T_n)的表达式的子集之中。这些规则规范着 T_n 的那些被挑选出来的表达式的应用,方法是:我们随意地把那些表达式单独运用于下述各种情形之下,在这些情形下,我们常常把它们对应的规则之类似物单独应用于 T_0 之中。

第二,也是同等重要的一点,成功的还原从理論上说有这

样的结果,即在由对应规则引起的术语映射的情况下, T_0 的中心原则(具有语义的、系统的重要性的原则)映射到作为 T_n 的原理的 T_n 的一般的句子上。(1979,p.81)

我们肯定赞同说:FP向自然科学的理想的或近乎于理想的还原是拯救FP的一种方式。我们还赞成说:假如FP至少经历了25个世纪的沧桑,因而显然不能为了逐字逐句地、平和地归并到第20世纪的科学这一目的而加以系统的阐述,那么这样一种还原——甚至肯定是一种特种的还原——不可能有什么前途。(非特种的还原几乎是不可能的,因为如果FP对人类是适用的,那么它同样也适用于火星人,尽管其物理-化学构成与我们的有重大的差异——如此相异,以致不可能有这样的理论上令人感兴趣的物理描述,它们囊括了在人身上“实现”FP的物理-化学属性和在火星人身上的对应的物理-化学属性。)

但是即使FP不能还原为低层次的理论,即使低层次理论本身能够对人类的本质和行为提供奇妙的说明,但也不能据此说:FP是极端荒谬的,或者说人类并不具有它所假定的意向事件。丘奇兰德的取消式唯物主义不是取代还原唯物主义的唯一可靠的自然主义选择。另一重要的选择是D.戴维森的非还原的、非取消的唯物主义(1970,1973,1974)。

戴维森提出了一个命题,它断言:每一具体的心理事件同一于某一具体的神经学事件,但它不断言(事实上否定)存在着把心理事件-类型或属性与神经学事件-类型关联起来的系统的桥梁法则。他称这种观点为异态一元论;它是一元论的一种形式,因为它假定心理物理的同一,而它又是“异态的”,因为它拒绝还原的桥梁

规则(或还原的类型-类型同一)。^⑥

作为还原唯物主义的一种替换品的异态一元论,其有效性表明:即使 FP 不能还原于神经科学,但 FP 所假定的标记心理事件很有可能是存在的,它们相互可能具有因果关系,还可能与感觉、与 FP 所说的、它们所完成的行为具有因果关系。

丘奇兰德从未提及戴维森关于同一论的观点——这是一种很奇怪的事情,因为它对他的论证有巨大的影响,并与它有明显的关联。然而他却直接从 FP 不可能还原为神经科学这一前提推论出下述结论,即 FP 可能是虚假的。因为标记-标记同一论对 FP 与神经科学之间的关系作出了一种可供选择的、可能的说明,因此他的论证是虚妄的。他正好错误地假定:FP 要与神经科学和睦共处,就必须能够还原为神经科学。

II

斯蒂克提出:FP 可能被证明与成熟的认知科学(以下简称 CS)不相容。现在我们来考察一下他的理由。不同于丘奇兰德的是,斯蒂克假定:FP 要与更广泛的、低层次的理论相容,就必须能够还原于后者。这里我们主要谈谈他认为 FP 必须与这些理论相适应的方式。

斯蒂克提出了两个否定 FP 与 CS 可相容的论证。在本节和下一节中,我们将分别予以考察。第一个论证试图说明:认知系统的所有因果组织与 FP 所归因的因果组织可能是不一致的。论证是这样展开的。符合下述特定谓词类型形式即“……是信念 P”

的事件,被认为具有言语和非言语类型的典型行为后果。就言语方面而言,属于这个类型中的事件就是这样的事件,即在适当的引导性条件下,它们一般使主体说出对 p 的断定。就非言语方面而言,这些事件是这样的事件,即在与主体的其他信念、愿望等相结合时,它们一般促使主体去完成 FP 所说的、符合于那个信念与别的命题态度的组合的那些行动。但是在斯蒂克看来,近来的实验论据表明:控制着非言语行为的心理事件必然独立于控制言语行为的那些事件——因此认知系统显然不包含这样的事件,即单个地看,它们具有 FP 赋予信念的那种因果作用。如果这些实验结果被证明可以泛化,如果 CS 在后来能按所设想的分离的、相当独立的、分别控制着言语和非言语行为的认知子系统的方向发展,那么我们将不得不像斯蒂克那样说:不存在信念之类的东西。

在他的重要的经验事例中,有一个就是斯托姆斯和尼斯比特(1970)所完成的对归属理论的一项研究。他是这样描述它的第一个方面的:

斯托姆斯和尼斯比特……要求患失眠症的被试记住上床的时间和睡着的时间。在记了几天之后,一组被试(“唤醒”组)在睡觉之前 15 分钟被给予安慰剂。他们被告知:那药丸会使心率加快,呼吸紊乱,体温升高,警觉增加,这些都是失眠症的典型症状。第二组被试(“放松”组)被告知:那些药丸会产生相反的症状:心率、呼吸放慢、体温、警觉性下降。归属理论预言:唤醒组的被试在服药丸的那个晚上会很快入睡,因为他们把他们的症状归之于药丸,而不是归之于贯穿着他们整个心灵的那些有沉重情绪负担的思想。它还预言:放松组的被试

1078 第九部分

要花更长的时间才能入睡。他们尽管吃了药以减轻症状,但症状仍会持续,既然如此他们便会如是推论:他们的有情绪重负的思想一定会特别地困扰他们。这种信念会进一步折磨他们,进而使他们更难入睡。显然,这两个预言都得到了证实。唤醒组的被试 28% 的人在服药的当晚入睡很快,而放松组中 42% 的人入睡的时间则很长。(斯蒂克,1983,p.232)

斯蒂克觉得特别有意义的是这一研究的第二阶段。在完成了失眠症实验第一步之后,唤醒组的成员被告知:他们在服药后将很快入睡,而放松组成员则被告知:他们入睡的时间将很长。当问及为什么会出现这种情况时,尼斯比特和威尔逊报告了下面的反应模式:

唤醒组被试一般回答说:他们常常发现:一周后更容易入睡,或者说他们参加了一场令他们心烦意乱的考试,但做得还不错,现在能够轻松,或者说与同寝室的人或女朋友的问题似乎还等着处理。放松组被试在解释他们的失眠加剧时也能找到相同类型的理由。在问及被试:他们在入睡之前是否想到了药丸时,他们几乎异口同声地说:在服药后他们完全忘记了它们。当问及他们是否想到药物可能产生(或阻止)唤醒症状时,他们常重复他们的看法:他们在服药后完全没有想到药物。最后,实验假说和被设想的归属过程便得到了详细的描述。被试对被假设的过程没有作出认同……几乎不假装相信:任何被试会经历这样的过程。(尼斯比特和威尔逊,1977,P.238)

考虑到第一阶段研究的材料,那么极有这样的可能,即控制着被试第二阶段语言反应的认知机制极不同于影响他们实际睡眠模式的认知机制。关于归属理论和认知不协调理论的文献中的其他大量研究,所得材料支持同样的结论:控制初始的非言语行为的机制极其不同于控制着被试随后对该行为理由的言语说明的机制。^⑦

如果我们对斯蒂克论证的理解没错的话,那么他进一步作出了三个结论。(1)在已述的那类事例中,对信念和愿望的归因不存在有说服力的和一以贯之的方式;因为 FP 一般把言语的和非言语的行为结果归之于特殊的信念与愿望,但是在这些情况下,非言语行为的认知原因是不同于言语行为的认知原因的,因此没有哪一类原因能恰如其分地同一于信念或愿望。(2)一般来说,我们的言语行为可能是由不同于控制着我们的非言语行为的认知机制的认知机制所控制的,因为斯托姆斯-尼斯比特模式是在归属理论和不协调理论的广泛研究的基础上产生出来的。从(1)和(2),他得出了结论(3):FP 可能是极其错误的;即是说,人类并不具有信念和愿望。

我们并不否认这样的论点,即在多如牛毛的事例中,如归属理论和不协调理论的研究所发现的事例,引起初始的非言语行为的心理状态和过程不同于引起了被试对那行为原因论的随后说明的状态与过程。我们要否定的是:(1)或(2)已为这一论点所确证。而没有(1)或(2),对(3)的论证便自然土崩瓦解。

试考察(1)。从斯托姆斯-尼斯比特研究所描述的现象来看,在对信念、愿望和其他民众心理学状态作出一以贯之的归因时真

的有难题吗？没有。因为我们可以求助于无意识的信念、愿望与推理。即使 FP 断言：信念和愿望在适当的唤起条件下，常常引起它们自己的言语表达式，但 FP 对无意识的信念、愿望并没有断言这一点。相反，它在说心理事件是无意识的时，要说的主要是：它对言语行为缺乏各种常见的直接的因果影响力。因此在斯托姆斯－尼斯比特研究中，我们可以得到关于被试行为的下述自然的、合理的民众心理学说明：他们最初的非言语行为是由无意识信念和推理所引起的，而他们随后的言语行为是由关于他们的最初的非言语行为的可能原因的、不同的、有意识的信念所引起的。总之，FP 面对这些事实并没有土崩瓦解，因为人们对假设的无意识民众心理学原因有一种选择权——自然的、合理的选择权。

我们意识到，存在着这样的诱饵，它促使人们把 FP 与“常识所说的东西”等同起来，并认为斯托姆斯－尼斯比特结论驳倒了我们的常识预期这样的事实自动证伪了 FP 的某种成分。但这种蛊惑应予抵制。常识并未假定有关的无意识信念与愿望。但是一当我们真的假定它们，根据关于非言语行为的相当精确的证据，FP 对斯托姆斯－尼斯比特所研究的被试怎样行动似乎也可作出正确的预言。

再者，正如我们对斯托姆斯、尼斯比特和威尔逊这些心理学家的观点所理解的那样，既然他们利用这种研究来证明非言语行为常常在分离的认知控制之下，那么这种对民众心理学的无意识原因的求助显然是他们对这类事例所做出的理论上的变革。归属理论和认知不协调理论为信念和愿望之类的民众心理学概念提供了中心舞台。因而双重控制论不过是刚陈述过的民众心理学命题而

已,即这样的主张:无意识信念和推理引起被试的初始的非言语行为,而不同的有意识信念(它构成了关于他们的原始行为的原因的假说)则引起了他们随后的言语行为。^⑧请注意斯蒂克本人在前面引证的一段话中是怎样描述斯托姆斯-尼斯比特研究的第一阶段的。他说:“归属理论预言:放松组中的被试将推论说:他们的有情绪重负的思想一定会特别地困扰他们。这种信念还将进一步折磨他们……”(着重号为我们所加)。这一来,斯蒂克可能想出重新解释这些主张的方式,以使他们所用的信念和推理概念远不同于FP-概念,但是在没有这种重新解释的时候,他的下述论点便是没有根据的,即信念及信念产生机制不能有说服力地归因于像斯托姆斯和尼斯比特研究中的被试那样的被试。

我们对双重控制论的分析所假定的自然是:它澄清了主张信念和别的心理事件是无意识的这一说法的意义。但是,斯托姆斯、尼斯比特和威尔逊则分明主张:不可能有这样的非言语行为标准,即使被试的言语行为与这样的心理事件不一致时,它也能确保对信念和别的心理事件的归因。^⑨

诉诸于这些标准——以及一般的无意识信念和推理——很可能是对传统的FP的发展和局部修正;但是即便如此,这也很难说是对民众心理学概念的全部拒斥。恰恰相反,自然而然地诉诸于民众心理学的无意识原因,正好反映了这样的事实,即FP所假定的全部原因构架即使在我们引进了有意识与无意识的区分的情况下依然完好无损。

因此结论(1)应予拒斥。这意思是说,即使接受了(2),FP也不一定就土崩瓦解了。但是结论(2)无论如何是应予抛弃的。从

1082 第九部分

无意识心理机制在大量事例中控制着非言语行为这一事实出发,人们不可能合理地推论说:我们的言语和非言语行为一般处在分离的认知控制之下。即使归属理论和不协调理论的发现警告我们不要过分相信我们认识我们自己的能力,但它们远不足以建立这样一种一扫无遗的结论。在这方面,考察一下 T. 威尔逊的说明(未发表)是大有裨益的,因为他是关于言语和非言语行为的“双重认知控制”观念的主要倡导者。斯蒂克利用的主要是威尔逊的观点,他把它当作是这样一种激进的论点,即我们自己关于引起我们的非言语行为的心理事件的陈述是绝不会为那些心理事件本身所引起的。但是依据我们的判断,这是一个错误的解释。威尔逊是这样阐述他的主张的:

这个论证从根本说就是:有两种心理系统,一是联结行为(尤其是不规则的行为)的,它在很大程度上是无意识的,用进化论术语说,这也许是两个系统中较古老的。另一系统也许是更新的系统,在很大程度上是有意识的,其功能就是试图用言语表达、解释、交流心理状态。正如先前曾论述过的,人们常有到达自己心理状态的直接通道,在这些情况下,言语的系统能作出直接而准确的报道。不过当通道受到限制时,言语系统就对这些过程和状态可能是什么样子作出推论。(pp.18-19)

从这一段不难看出,威尔逊并不主张:我们关于我们心理事件的话语一般是由认知事件产生的,而不是由那些心理事件本身产生的。相反,他承认:人们对他们的行为之心理原因常有直接的有

意识的通道,在这些时候,这些状态一般会引起对自身的准确报道。只有在通道受限制的地方,只有在事件不是有意识的地方,我们随后的话语才是由对可能的心理原因的推理引起的,而不是由那些心理事件本身引起的。^⑩

威尔逊进一步指出,一般来说,正是这样的事件,它们才是极不可能为行动者认识的重要加工的结果,而“更直接的状态”(例如前认知状态)则可能是更容易被认识的(p. 39)。而且有这样一些情况,它们似乎涉及到复杂的加工,在其中,如果控制着言语的和非言语行为的系统完全是独立的,那么人们就会以一种似乎很难理解的方式表现混杂在一起的言语和非言语行为。试考察完成某种复杂的任务,如向别的人解释你正在做什么——当某人讲课时在黑板上解释逻辑问题。令人迷惑不解的是,如果行动者没有到达引起他的行为的非言语部分的信念的通道(除了借助事后的推理),那么这种复合的行为是怎么可能的。

因此我们的结论是:结论(1)和(2)都没有得到斯蒂克所引证的那类心理研究成果的支持,因此他的反FP的“双重控制”论证未能如愿以偿。

III

为了使FP与CS协调一致,FP必须怎样相关于CS?对此,“双重控制”论证并未预先假定任何特定的概念。然而,斯蒂克关于FP与CS不相容的第二个论证依赖的正好就是这个概念。尤其是,他要求:信念、愿望等应与认知系统的“可自然分离”的部分

1084 第九部分

相同一,他把这称作模块性原则。

斯蒂克不想阐明这个原则,而让可自然分离性概念停留在直观的层面。因此我们也将不加说明地使用这个概念;我们认为,我们要提出的观点在任何合理的阐释之下都是适用的。

斯蒂克论证说:与CS相比,FP可能没有满足模块性原则,因此可能不存在信念、愿望之类的事件。他集中讨论了CS中关于人类记忆模拟的新近动向。他指出,关于记忆组织的某些早期模型为每一记忆假设了一种特殊的句子或似句子结构。他说这些模型显然是模块性的,因为不同的似句子结构可以同一于各别的信念。另一类模型极大地得益于想解释人们怎么可能确定与眼前特定任务有关的信息的愿望,它把记忆当作节结和被标记的链所组成的复杂网络,在这里,节结表征的是概念,链表征的是概念间的各种关系。然而斯蒂克则认为,网络模型“离余象的模块终端仍是相当遥远的”,因为在网络模型中,把起着特定信念所独有的因果作用的网络部分分离开来一般是没有问题的。(1983,p.239)

但是他指出:在近几年,几个颇有影响的理论家对高模块模型开始表示深深的怀疑,这主要是因为这些模型似乎不能处理大量的非演绎推理,而这些推理又与语言运用和理解密不可分。斯蒂克引用了明斯基(1981)的一个例子,写道:

在一篇……新近的论文中,明斯基解释了他所说的“心灵的社会”观点,根据这种观点,思维的机构被划分为许多独立的、很少交流的“专家”(p.95)。根据明斯基所描绘的图景,心理模型的不同单元或部分“本身都没有意义”(p.100),因此它们都不能同一于个别的信念、愿望等。模块性——我从明斯基那

里借用的术语——以一种激进的方式背道而驰,因为意义或内容只能从“结构的巨网”中派生出来(p.100),而且那系统的任何自然部分都不可能与“明确的”或可用言语表达的信念相互关联。(1983,p.241)

如果明斯基的“心灵的社会”观点是 CS 在未来将要采取的方向,那么模块性大概真的以激进的方式背道而驰了。

我们相当乐意承认:CS 很可能戏剧性地成为非模块性的,因此模块性原则最终可能受到经验的驳难。^⑩而且,如果人们考虑到 FP 与神经科学之间的关系——甚或 CS 与神经科学之间的关系,就此而论——那么人们便有望看到:模块性以一种更富戏剧性的方式背道而驰。在人的中枢神经系统中数百亿的神经元,有数千亿的突触接合处;因此如果神经科学的“可自然分离的”事件就是像神经元激活和电能的交互突触转换那样的事件,那么完全有这样的可能,即 FP 和 CS 中的可自然分离的事件就会在神经系统中与“结构巨网”相关联——即是说与可自然分离的神经事件之巨大集合相关联。

因此如果模块性真的是必需的,以便使 FP-事件存在并进入因果关系之中,那么缺乏模块性对 FP 与低层次理论之间的所谓的和谐一致性来说便会导致极大的麻烦。事实上,对认知科学与像神经科学这样的低水平的科学之间的所谓的和谐一致性来说也会导致极大的麻烦;因此斯蒂克式的论证似乎不过是要证明:他作为 CS 的一个倡导者宁愿它得到证明!更有甚者,对模块性的需求对神经科学与物理-化学之间的和谐一致性来说甚至也意味着重大的麻烦;因为如果物理-化学的自然类型的谓项是像“……是

1086 第九部分

一个电子”、“……是一个氢原子”之类的谓项,那么就完全没有这样的可能性,即包括在神经科学自然类型术语如“……是一个神经元”范围内的实在也会包括在物理-化学自然类型术语的范围。确切地说,神经元和神经元激活就是从物理-化学观点来看的、与“结构巨网”有关联的实在。

我们之所以陈述斯蒂克论证中的这些一般原则,是因为我们认为,它们说明了模块性原则作为理论间和谐一致性的条件是极不可能的。可以肯定,像神经元这样的对象,或像神经元激活这样的事件,为了与基本的物理-化学一致,并不一定要从基本的物理-化学图景中“自然地分离出来”;毋宁说,只要它们可以完全分解为自然可分离的部分就足够了。同样,认知-心理学事件为了与神经科学的图景和谐一致,也没有必要从这一图景中自然地分离出来;再者,只要这些事件可分解为自然可分离的部分就足够了。^⑩

我们认为:对与CS事件有关的民众心理学事件来说,情形实际上并无二致。明斯基也许是对的,(比如说)信念的作用一般是由CS事件的巨大的、极不公正被划定的集合实现的。这并未表明:信念是不存在的。恰恰相反,它所表明的不过是:信念是一种极其复杂的事件,由作为部分的、大量的CS-事件所构成。^⑪进而我们毕竟还可以预期,这些CS-事件又由作为部分的神经学事件所构成;再进一步,我们也有望看到,这些神经学事件由作为部分的大量的物理-化学事件所构成。

斯蒂克从来都不想把模块性原则论证成一种相容性条件,就像丘奇兰德从不打算证明对可还原性的要求一样。因此,斯蒂克

的模块性论证就碰到了同丘奇兰德的可还原性论证一样的问题：即它依赖于这样一种缺乏基础的、不可能牢靠的概念，该概念是关于 FP 必须怎样与更广泛的、低层次的理论紧密配合以与它们和谐一致的。（顺便说一句，值得注意的是：即使斯蒂克没有要求可还原性，但他关于理论之间协调的概念在一定的方面实际上仍比丘奇兰德的概念要强得多。因为甚至是还原论者也没有必要要求：属于高级层次的、自然类型的实在应能从低级层次的图景中自然分离出来。还原论者不可缺少的是：应当有双重条件桥梁法则，以把高层次的类别与低层次理论的开语句关联起来，但是这些低层次的开语句可能是相当复杂的，而不是（比如说）简单的自然类型的谓项。）

尽管斯蒂克没有为模块性原则提供明确的理由，但他也许受到了下述思想路线的影响：

FP 的命题态度涉及到认知者与似句子的“内在表征”之间的关系（福多 1975, 1978；菲尔德, 1978；利康 1981）。如果 FP 是真的，那么 CS 的部分任务就应当是解释这些内在表征的本质。但 CS 不可能做到这一点，除非内在表征被包括在它的自然类型谓项之中，或者说不知什么原因至少在认知系统中是“可自然地分离的”。如果明斯基的“心灵社会”的方案是 CS 未来要采取的方向，那么这种要求将不可能得到满足。因此，如果 FP 的事件不能像 CS 那样遵循模块性原则，那么 FP 就一定是极其荒诞的。

我们拒绝这种推论路线，理由之一是：我们对命题态是否是真的涉

1088 第九部分

及到内在表征或者说它们是否真的有“对象”(参考注①)深表怀疑。而且如果明斯基的方案真的成了CS的总的趋向,那么这个事实大概会瓦解似句子表征与命题态度有关联这一主张的基础——正如他的方案已经削弱了下述主张的基础,这个主张就是:这类表征涉及到作为语言运用和理解之基础的非演绎推理。

更有甚者,即使内在表征观是正确的,即使CS的部分任务就是说明这些表征,但像明斯基那样的方案并不一定使CS不能完成这个任务。因为CS的“原子”最终有可能是明斯基“心灵社会”的组成部分,CS也可能假定复合的、由这些“原子”所构成的似句子“分子”。这些分子可能非常复杂,而且是极不公正地被划分的。如果是这样,当从原子论观点看待认知系统时,这些分子就不能看作是认知系统的自然可分离的组成部分;然而从更高级的分子观点来看,它们则可看作是可自然分离的东西。(不过,我们认为更有可能的是:如果“心灵社会”方案证明可在CS内普遍化,那么结果就将是完全拒绝关于命题态度的心理表征观,如前所述,我们认为这是一种完全错误的观点。)

人们设法辩护模块性原则的另一方式,就是求助于戴维森式的、使规律的作用与因果性关联起来的思考。人们可能论证说:(I)FP没有包含严格的规律,而只有所谓的“受制于异律的”(heferonomic)一般原则(戴维森1970,1974);(II)只有当两个事件得到了例示一严格规律的描述时,它们才能作为原因与结果关联起来(戴维森,1967);(III)例示了一种特定理论的严格规律的事件—描述一定遴选出了这样的事件,即从那理论的观点来看它们是可自然分离的。从这三个主张出发,加上民众心理学事件从属

于因果关系这一假定,模块性原则便似乎是必然的结论。^⑭

但是假如一事件 c 引起一事件 e , 在这里从 FP 观点来看, c 和 e 都是可自然分离的。假设 c 完全可分解为这样的事件, 它们能分别满足一种基本的、受相同法则支配的理论 T 中的种类谓项 $F_1 \cdots F_m$, 而且这些作为部分的事件从 T 的观点看都是可自然分离的; 再假设: 这些事件都同时满足 T 的描述 D_1 (可能相当复杂), 而这一描述详述的是它们的结构上的相互联系。同样, 还假设, e 完全可分解为这样的事件, 它们能分别满足 T 的种类谓项 $G_1 \cdots G_n$, 而且这些作为部分的事件从 T 的观点看都是可自然分离的; 再假设: 这些作为部分的事件同时满足 T 的描述 D_2 , 它详述了它们的结构上的相互联系。这一来, 即使 c 和 e 没有这样的自然类型的描述, 在其之下, 它们本身例示了 T 的严格的规律, 但是 T 的严格规律可能同时衍推出下述形式的断言:

对于任何事件 x 而言, 如果 x 可完全分解为事件 $x_1 \cdots x_m$, 以至于 $D_1(x_1 \cdots x_m)$ 和 $F_1(x_1), F_2(x_2)$, 以及 $\cdots F_m(x_m)$, 那么 x 就会为事件 y 所伴随, 而 y 又可完全分解为事件 $y_1 \cdots y_n$, 以至于 $D_2(y_1 \cdots y_n)$ 和 $G_1(y_1), G_2(y_2)$ 以及 \cdots 和 $G_n(y_n)$ 。

为什么 c 与 e 之间的因果关系不能依赖于这种形式的规则性, 对此我们找不到任何理由。人们或许会称这种规则性为严格的规律, 如果是这样, 上述主张 (Ⅲ) 就是错误的; 别的人或许会保留“严格的规律”这一术语, 以表示受相同法则支配的理论的相对简单的、合规则的假定, 而不是这些假定的逻辑结论的集合——如果是这样, 上述主张 (Ⅱ) 就是错误的。不管怎么说, 受到戴维森启

1090 第九部分

迪的关于模块性原则的论证有一错误的前提。(顺便说一句,我们的意思不是把这种论证归之于戴维森本人,因为我们对是否赞成论断(Ⅲ)是持怀疑态度的。)

IV

我们已经证明:FP-事件很可能与任意复杂的、被极不公正划分的CS-事件是同一的,与CS相应,CS-事件本身是不能自然分离的,但完全可分解为具有这一特征的部分。当然,如果FP-事件真的存在,那么它们就一定会与FP的因果结构保持一致;即是说,它们一定会在因果上相互关联,并与感觉、行为相关联,就像FP关于它们所说的那样。而且,正如心灵哲学中的功能主义者经常强调的那样,FP的因果的或功能的原则对FP-事件的个体化至关重要;就广泛的范围而言,使一特定事件被认作是(比如说)一标记信念p的东西是这样的事实,即它拥有FP授予信念-类型的标记的那种因果作用。^⑤

因此如果我们关于FP与CS之间关系的非模块性图景是合理的,那么必不可少的是:复杂的、被不公正划分的事件可以严格地视之为原因,即使它们涉及到与低层次理论有关的“结构巨网”。在这里,详细的讨论超出了本文的范围,而对复杂事件的因果地位的简要考察则有助于阐明我们的论点。

我们不妨说:只要事件e引起事件f,而e的任何部分都不能引起f,那么e在最低限度上便可引起f。关于最低限度的因果关系我们想阐发两个主张,其中每一个都会得到下面某一点的支持。

第一,即使事件 e 是事件 f 的真正原因,但 f 也可能是由作为 e 的一个适当的部分的某事件所引起的,因此可能是 f 的真正原因,而不是 f 的最低限度的原因。第二, e 不仅引起 f ,而且还引起某种别的事件 g ,那么在最低限度上引起 f 的 e 的部分就不同于在最低限度上引起了 g 的 e 的部分。^⑥

这两个事实很重要,因为它们使能满足 FP 的因果原则的事件之存在相对容易。如果 FP 在 t 时把事件 f 和 g 归之于一个原因 e ,而且事实上又存在不同的(即使部分重合的)事件 e_1 和 e_2 ,以至于 e_1 在最低限度上引起 f (在 t 时), e_2 在最低限度上引起 g (在 t 时),那么这并没有必然地证明 FP 为假。因为 e 很可能有 e_1 和 e_2 作为其部分;而且,它还很可能有作为部分的各种事件,它们在最低限度上(在 t 时)引起下述各种事件的一个或另一个,FP 认为这些事件都是 e (在 t 时)的结果。只要这个复合事件是一切在前的事件的结果本身,FP 认为一切在前的事件都是 e 的原因,那么该事件就是(同一于) e 。

结论是:FP 可能很容易被证明为真,即使它与模块性戏剧性地背道而驰了。FP-事件不仅可能是复合的、极其不公正地被划分了,有大量可自然分离的 CS-事件作为其部分,而且任何给定的 FP-事件 e 都能以复杂的方式引起其结果,因而有不同的结果,而这些结果又有 e 的不同部分分别作为自己的最低限度的原因。^⑦

V

也许有人会驳难说:我们的分析太宽容了;除非我们采用斯蒂

1092 第九部分

克的模块性条件,再加上这样的要件:FP-事件与FP授予它们的因果构造相一致,否则我们就没法把有价值的约束加之于高层次的因果断言的真值条件之上;即是说,我们承认这些断言最终为真,而不管作为它们的基础的理论有什么特征。我们将通过考察这一反对意见而引出结论。

很显然,某些基础理论与某些高层次因果断言的真值是不一致的。例如如果世界像我们流行的物理学和化学所描绘的那样,那么魔鬼附体就不可能是任何心理失调的原因,燃素的丧失就不可能是金属在氧化时所经历的化学变化的原因。试考察一个更加明显的事例,如果我们是斯金纳式的生物——即是说,该生物的行为完全能由斯金纳式的心理学的基本原则所描述和解释——那么民众心理学把假定的信念、愿望之类作为我们行为的原因这一主张就不可能是真的。

值得注意的忧虑是:我们关于理论间和谐一致性的非模块性探讨方案是如此自由,以致它承认上述那种主张最终为真,即使它们显然与基础理论不协调。我们要论证的是:这种担忧是没有根据的。

对于高层理论和基础理论之间的极端的或不一致性或统一性的认识论作用,区分开两种不同的概念或期望是很有好处的。根据第一个概念,人们把缺乏这种一致性当作通向高层理论之虚妄性的重要认识论途径,在这里,那种虚妄性从别的方面看可能是不明显的。这也就是说,高层的直接证据即使不能清楚地证明高层理论的虚妄性(甚至有可能支持这种理论),但是只要我们注意到高层理论在某一方面与我们有可靠理由相信其为真的某种基础理

论不一致,那么我们就可以发现它的虚妄性。显然斯蒂克和丘奇兰德两人都是依据这个概念而进行论证的。

我们觉得,另一有重要区别的概念是更加合理的,它是关于高层理论与低层理论之间缺乏一致性的认识论意义的。我们当然不否认,高层理论可能与低层理论不相容。但我们深表怀疑的是:只是由于注意到高层理论与真实的基础理论的不一致性,人们就能知道高层理论是错误的,这是否是常见的或普遍的。更一般地说,当高层理论错误时,除了缺乏一致性之外,还有关于这个事实的直接证据。不相容性之出现,不是因为缺乏模块性,而是因为根本就没有这样的事件——不管是简单的还是复杂的——它们完全具有高层理论赋予它所假定的事件的那些特征。直截了当地说,这里的观点是:当关于少年犯罪或学习行为的各种理论与神经生理学理论或物理学理论不一致时,前者有这样的可能,即要么是为有利于社会学家和心理学家的那类证据所证实,要么是不能被证实。极为罕见的是,一种理论能为有利于社会家或心理学家的广泛的证据所支持,但最终又被证明是极端错误的(因为它的本体论未能严格吻合于某种基础理论的本体论)。因此与包含了模块性条件的理论间一致性概念相比,我们的概念倡导的是心理学这样的高层学科的更大的认识论自主性。

我们之所以强调这种认识论观点,是因为它的予头直接对准的是对纵容我们的非模块概念的忧虑。在我们的方案承认每一高层理论都能与每一基础理论相容的意义上,我们的方案绝不是毫无价值的,它既是宽容的,又是紧缩的,因为至少就广泛的事例而言,对一致性的看法不会像它们在模块性要件之下那样起着独立

1094 第九部分

的、规范性的作用。

记住这一点,让我们再回到我们本节开头所述的例子上来。首先看魔鬼附体。像别的从因果上进行的解释的观念一样,魔鬼附体的观念在很大程度上应根据它在因果关系网络中的作用来理解。魔鬼附体引起或可能引起各种病理学上的行为。这类结果可借助适当的宗教仪式(如祈祷或驱除仪式)来减少或排除。如果行为是出自于魔鬼附体,那么就没有理由假定:别的治疗形式(药物、改善营养和精神疗法等)对它有效。附体的状态本身就是具有许多别的超常力量的存在物的活动结果。

如果魔鬼附体的事件(简称为 d)应当是某种行为(如不连贯的胡诌)的一个原因,那么根据我们的分析,d就一定能够与可用我们的基础理论的谓项加以描述的某事件(简称为 e)相同一;而且一定有这样的情况,即基于这种同一,根据关于魔鬼附体的理论,d出现于其中的其他大多数因果泛化应当是真实的。(即使我们关于理论间相互一致的概念不赞成模块性,但它仍坚持认为:我们所作的同一保护了高层理论的“因果构架”。)我们承认:不管人们使假定同一于事件 d 的事件 e 多么重要、多么复杂,不管人们多么情愿把 e 的适当的部分当作因果上有效的,也完全不可能有注定要成为 e 的东西。考虑到我们现在的物理化学理论,这种 e 据说最终会使与魔鬼附体有联系的因果论断网络在主要的方面为真。即是说,无论其多么复杂,根本就没有 e 这样的事件,据说它依据规律关联于与附体有联系的各种行为,在使用了驱除仪式时,它就没有产生上述行为的效力,由于规律的作用,它据说是由具有魔鬼属性的动因产生的,等等。

这个例子说明了前面提出的一般认识论主张。事实上,我们业已证明:关于魔鬼附体的因果论断之所以为错误的,不是因为与模块性(或与“平稳的还原”)有关的谨慎的考虑,而是因为与魔鬼附体有联系的因果构架是极其错误的。根本就没有什么东西像魔鬼附体被假定的那样,处在与其他一切事件的因果关系网络之中。只要注意到:把因果有效性赋予魔鬼附体这一主张的虚妄性是显然易见的,即使没有与化学、物理学或生物学有关的思考,我们也能直接明白上述道理。如果人们恰到好处地进行受控实验,那么他们可能很快就会发现:驱除仪式不能影响魔鬼附体类型的行为,某些别的疗法概不例外,如此等等。^⑩

一系列相近的观察材料对于下述断言似乎是很重要的,这断言是,即使我们是斯金纳式的生物,我们的探讨方案仍承认关于信念的因果主张是真实的。FP 主张:信念、愿望和别的命题态度以多种多样的方式相互关联,并与感觉、行为发生因果关系。而另一方面,斯金纳的理论则否认:为了解释行为,我们必须假定这类有复杂相互作用的内在事件,而且它还彻底否认这些事件是存在的。确切地说,斯金纳主义者主张:主宰着从环境“刺激”到行为“反应”的因果链在很大程度上是彼此孤立的,非常类似于在光纤通信线路中的、各种平行的、无相互作用的通信信道;因此与任何特殊的刺激-反应对子有关联的一切内部事件,与在其他刺激-反应对子中牵涉到的内在事件都不具有有意义的因果关系;即是说,斯金纳主义者声称:作为一个经验事实,把刺激与行为联系起来的原则是如此简明和直截了当,以至于它们与下述内在事件之存在是势不两立的,这些事件相互作用的方式完全相似于民众心理学事件

1096 第九部分

被设想的相互作用的方式。因此如果斯金纳主义者是对的,那么在人类和别的有机体中就根本不存在这样的内在事件,它们与感觉、行为具有因果关系,而且相互之间也具有民众心理学授予信念、愿望等的那种因果关系。因此如果人类最终仅只是斯金纳式的生物,那么我们关于理论之间相互一致的非模块概念就真的是大逆不道了。这一来,这个概念就不是那么过度的宽容了。

这个例子还说明了上述各种认识论主张。通过中枢神经系统中的某些(可能很复杂的)事件组合,它满足了 FP 的“因果构架”,这对 FP 的真理来说是至关重要的。因此如果我们是斯金纳式的生物,以至于 FP 所假定的因果构架没有借助任何简单或复杂的事件在我们身上得到例示,那么这个事实在我们对克分子行为的相对粗略分析的层面就可能显露出来。与 FP 的因果构架不相容的刺激-反应规律将可能被发现,并可用于解释和预言人类的一切行为。因此根据最容易得到的粗浅证据,FP 从主要方面来说似乎是不真实的,而且仅仅因为没有与某种基础理论保持严格的一致,它最终还将被证明是荒谬的。

总而言之,我们的方案似乎像其应是的那样,实际上是宽容的,这个事实也说明了这一点;与之形成鲜明对比的是,关于理论间相互一致性的模块性概念似乎是极其苛刻的,因为它的目的是不明确的,它否认高层理论有一定程度的认识论自主性。因此,由于(I),到目前为止,受行为主义激发的心理学工作并没有明确地揭示能用来解释一般的人类行为、动摇 FP 的因果构架之基础的刺激-反应规律,(II)事实:民众心理学概念似乎正好存在于归属理论和认知不协调理论之类的认知理论的中心,(III)事实:FP 在

我们对行为的日常解释和预言中服务得非常好,因此似乎很难否认:民众心理学十之八九不会消失。^{⑩⑪}

注 释

① 实际上,我们在第一段把那个例子作为一种图式,当不同的句子为字母“p”所代替、不同的种类谓项由虚构短语“K类的……”所代替时,该图式便给出了全部的例示。我们情愿认为:“……相信P”这种谓项形式就是奎因所说的态度表达——即把形成谓项的算子(“相信…”)加在一个句子前所构成的一个复合一目谓项。根据这种观点,命题态度没有“对象”,因为它们不是关系陈述。欲知进一步的讨论,可参阅霍根 1989。

② 关于视知觉,可参阅格雷戈里(1970)。

③ 例如,18世纪的化学理论试图诉诸于据说能用来解释金属的合成行为的相同因素,把金属的属性解释为发光性和延展性。如拉瓦锡等人的化学理论强调的是合成,一开始就因其不能对金属的发光性和延展性作出统一的解释便受到了批评。

④ 情境理论从经济上说更充分,因而这种变化是一种进步,关于这一点的某种更重要的证据,请参阅尼斯比特和罗斯(1980)。

⑤ 于是,他对福多(1975)以及一般来说对认为信息以语句形式被储存的认知心理学理论作出了批评说明;试比较丘奇兰德(1981, pp. 78ff)。

⑥ 为了使异态一元论上升成为唯物主义的成熟的变种,人们必须为它增加这样的说明,即证明心理状态-类型(属性)对于物理-化学状态-类型的形而上学地位。我们认为,合理的学说就是戴维森所阐发的那一种(1970, 1974),即心理属性随附于物理属性。近来,几位哲学家发展了这种观点,论证说:唯物主义应与某种随附性理论相一致。试比较金(1978, 1982),豪格兰德(1982);霍根(1981b, 1982b)以及刘易斯(1983)。

⑦ 关于有关文献的评论,可参阅尼斯比特和威尔逊 1977。

⑧ 无论如何,这就是双重控制理论从斯托姆斯-尼斯比特研究中所得到的东西。信念、推理之外的其他类型的心理事件有时也涉及到了。

⑨ 例如可参阅威尔逊(未发表), pp. 7ff。

⑩ 不难理解,斯蒂克为什么把激进的双重控制论归之于威尔逊,即使威

1098 第九部分

尔逊实际上并不赞成这种观点。

⑪ 即使我们认为,CS在其最根本的层次上,很有可能成为非模块性的,但我们仍相信:理论心理学的某些高层次分支不仅会继续是模块性的,而且还会继续使用FP本身的概念。在这方面,归属理论就是例证。(我们所说的“高层次”心理学理论,意指这样的理论,它假定了作为整体的事件,它们的部分是“低层次”心理学理论所假定的事件。详后)

⑫ 另一方面值得注意,在这一方面,斯蒂克(和丘奇兰德)的论证似乎导致了广泛的、令人难以置信地强的结论。

⑬ 相关类型的一个复杂事件可能是更简单事件的部分学总合或合成;要么它大概是一种不同于这事件合成的实在。这里我们对此不打算多费时间。(这个争论直接关系到这样的问题,即像船这样的实在,在世界内和超世界的不同的等同条件下,是等同于某物理部分的合成,还是不同于这合成的实在。)就我们所知,关于事件之部分和整体的最清楚和最成熟的理论就是汤普森的理论(1977);合成的事件就是她明确肯定的复合事件的唯一种类。

⑭ 由于与斯蒂克本人的谈话,我们才想到了这种戴维森式的论证。

⑮ 但是正像孪生地球人这一著名案例(普特南 1975)所表明的那样,事件的因果作用并不是与其民众心理学个体化相关的唯一因素。孪生地球上酷似我们的人并未经历过这种类型的标记,即相信水适宜于饮用,即使他们经历了在功能上相同于我们自己的标记信念(水适宜于饮用)的事件。麻烦在于他们称之为“水”的材料根本就不是水。试比较,伯奇 1979。

⑯ 即使对这些主张的全面的辩护超出了本文的范围,但我们认为它们对许多因果陈述的真理是必要的。

⑰ 关于混合型的因果关系的观点也与斯蒂克反FP的双重控制论证有关。即使言语和非言语行为应有独立的、最低限度的原因,但FP仍可能为真;因为FP-事件可能是最低限度的原因的复合,这些复合的事件可能是言语和非言语行为的真正的原因(即使不是最低限度的原因)。(不过,我们应强调,我们并不主张:如果双重控制论是真的,那么每当主体在他的非言语行为系统的某种状态B中,以及在他的言语系统的某种状态V中,就总有这样的可能,即把他的言语和非言语行为的某种单独的民众心理学原因归之于他,这与FP是一致的。这是否可能,取决于特定状态B和V,以及它们所引

起的行为。例如,在斯托姆斯-威尔逊的失眠实验中,引起了唤醒组入睡的状态 B(被试把他们的症状归之于药物)不只是有别于引起了他们言语行为(否认上述归结与他们的入睡有联系)的状态 V;而且根据 FP 的因果原则,这两类状态都不能看作是某个信念的组成部分。

⑩ 某些驱除仪式可能是灵验的,但这仅仅是因为它们具有别的更世俗的理论所引证的某些特征。

⑪ 即使我们一贯假定,民众心理学事件是由作为部分的低水平事件构成的复合事件,但我们仍愿意承认:不诉诸于这种假定也可以为 FP 与 CS 之间的相容性作出辩护。金在权(1966,1973)认为:事件就是这样的实在,它每次都存在于一对象所例示的属性中,而心理事件则偶尔存在于个体所例示的心理属性中。根据这种观点,就不清楚:低水平的事件能否在感性上看作是 FP-事件的部分。尽管如此,金在权的事件理论的辩护者仍有可能论证说:FP-事件是存在的,并具有 FP 认为它们具有的各种相互因果关系。因为他可能论证说:这些事件随附于低水平事件的组合,而且随附性传递着因果效力。试比较,金 1979,1982,1984。

⑫ 在本文中,我们像通常一样,一直假定:如果日常民众心理学陈述是真实的,那么就真的存在着民众心理学事件,即标记信念、标记愿望等。我们相信,本文的基本观点可以用别的方式加以重新表述,如不要求心理事件的存在,甚至不要求任何事件的存在。但我们这里的目的,主要限于在事件的本体论构架内为 FP 作出辩护,这种本体论在当代心灵哲学中已得到了广泛的认可。

参考文献

- A. 坎贝尔等(1960):《美国的投票者》,纽约,约翰·威利。
- P. 丘奇兰德(1979):《科学实在论与心灵的可塑性》,纽约,剑桥大学出版社。
- (1981):“取消式的唯物主义与命题态度”,《哲学杂志》,78。
- D. 戴维森(1967):“因果关系”,《哲学杂志》,64。
- (1970):“心理事件”,载 L. 福斯特等编:《经验与理论》,伦敦,达克沃斯。

1100 第九部分

——(1974):“作为哲学的心理学”,载 S. 布鲁恩编:《心理学的哲学》,纽约,哈珀与罗。

D. 丹尼特(1981):“三种意向心理学”,载 R. 希利编:《还原、时间与同一性》,纽约,剑桥大学出版社。

H. 菲尔德(1978):“心理表征”,《认识》,13。

J. 福多(1975):《思想的语言》,纽约,托马斯, Y. 克罗韦尔。

——(1978):“命题态度”,《一元论》,61。

J. 豪格兰德(1982):“弱随附性”,《美国哲学季刊》,19。

T. 霍根(1978):“反事件的事实”,《哲学评论》,87。

——(1981a):“无行动的行动理论”,《心灵》,90。

——(1981b):“标记物理主义、随附性和物理学的普遍性”,《综合》,34。

——(1982a):“置换性与因果关联词”,《哲学研究》,42。

——(1982b):“随附性与微观物理学”,《太平洋哲学季刊》,63。

金在权(1966):“论心理-物理同一论”,《美国哲学季刊》,3。

——(1973):“因果性、合规则的归类和事件的概念”,《哲学杂志》,70。

——(1984):“随附性和随附因果性”,《南方哲学杂志》,22。

D. 刘易斯(1984):“宇宙论的新探索”,《澳大利亚哲学杂志》,61。

W. 利康(1981):“关于相信的小人理论”,《认知与大脑理论》,4。

M. 明斯基(1981):“K-路线:记忆理论”,戴诺曼编:《认知科学展望》,诺沃德,新泽西,埃布莱克斯。

R. 尼斯比特和 L. 罗斯(1980):《人类推理:社会判断的战略和缺陷》,恩格尔伍德·克利夫斯,新泽西。

联结主义与心灵哲学概论*

威廉·贝希特尔

联结主义近来引起了认知科学的极大关注,因为它似乎有望取代对于认知模型的传统的基于规则的研究方案。哲学家们也开始注意联结主义。在本文中,我的目的就是想简要地介绍一下由联结主义所引发的某些哲学争论。我不打算对联结主义本身作什么辩解,然而,如果我认为它不能对心灵和心理现象的理解作出某些重要的贡献,那么我也不负有如此之类的使命。更确切地说,我的目的就是引发进一步的讨论,并集中于此。因此在本文每一部分中,我将提出一个或更多的问题,并予以简要的讨论。在讨论这些问题时,我将指出我认为是发人深思的答案,不过这里我不想对它们作出肯定的论证。

认知的联结主义模型的基本特征

研究认知的联结主义方案除了探讨近几十年来引导对认知的大量思考的传统的计算机模型之外,首先要研究的就是作为认知

* 译自 W. 利康编:《心灵与认知文选》,巴兹尔·布莱克韦尔,1990。

1102 第九部分

基础的机制的特殊概念。这一替代模型的灵感来自于我们对大脑的理解。即使联结主义模型试图体现大脑结构的重要特征,但联结主义并未对中枢过程提供详尽的说明。确切地说,它倡导的是与认知有关的那类加工模型。在下一部分,我将进一步讨论联结主义模型怎样与神经模型相关联。在本部分,我只准备以联结主义者所用的那种方式抽象地说明联结主义模型。

联结主义结构的基本构成部分就是像神经元这样的简单单元,它们在任何特定的时候都被激活到某种程度。典型地说,这种激活就在于控制电荷。仍像神经元一样,这些单元与其他单元相联结(这些联结可能具有不定的强度),以便它们基于它们自己的激活促使其他单元激活的增加(兴奋)或减少(抑制)。此外,在某些联结主义系统中,这些联结的强度可能由于此系统中的活动而改变,以至于一个单元对另一单元的作用能随时改变。

即使这一概述涉及到联结主义系统的基本构成部分,但事实上,联结主义系统还可有许多不同的建构方法。基于一种方法,各种参数可以确定下来。一种参数决定了单元的可容许激活水平:这些单元可能被限制在少数分立状态(如开和关),或者它们也容许在特定的范围内变化(如从 0 到 1)。另一种参数制约了一单元的激活怎样根据它的先前状态和它从其他单元接受的输入而发生变化,例如其中可能包括一种衰退的功能,在没有新的激活的情况下,一单元的激活就将中止,一种临界值可能被用到,因而除非输入超过一定的量,一单元就不会被激活。第三种参数规定了一特定单元的输出量。输出可能与它自己的激活成比例,或者说它可能为一临界值所限制。最后,有许多学习规则,它们可用来在特定

的通路上调节联结的强度。其中两个最广泛地被采用的规则分别是赫布规则(Hebbian rule,它根据两个单元在它们的激活水平上是相同还是不同来加强或削弱通路)和 δ 规则(delta rule),它增强或减弱每一输入通路的力度,以便增加单元的目标值和电流值之间的协调。(关于联结主义系统各种可能构型的进一步详细说明,请参阅拉梅尔哈特等,1986)。

基本的加工系统就是这些被联结起来的单元的集合体。当一种激活的初始模式被给予这些单元中的一部分或全体时,这样的集合体就开始发挥作用。这种模式可以看作是给予那集合体的一个问题。当系统适应一种稳定状态(例如有一种状态,在那里,经过联结,激活的完成并不会导致单元改变它们的激活强度)时,加工就结束了。这种适应依据热力学原则而发生:单元由于它们的输入而改变它们的激活,从而改变它们的输出,直至该系统适应最高的熵的状态。当一种稳定模式得到了,总的稳定模式或关于某些单元的值便表征着系统对于那个问题的回答。

在建构像问题解决这样的系统时,我们应对系统的活动提供一种说明。一般来说,联结主义者常启用两种不同类型的说明图式。在局部化(Localized)说明图式中,每一单元表征的是某种对象或属性。例如在拉梅尔哈特和麦克莱兰的单词识别系统中,每一单元便表征着关于可能在场的特征、字母或单词的一种假定,单元的激活程度表征的是对于相应项目是部分输入的信任程度。^①

另一可供选择的说明图式是一种分布图式,在这里,一种说明被归结为一种对于整个单元的集合体的激活的模式。个体单元的活动本身可以有某种符号作用(例如表征被认知对象的特征)。但

是具有首要利益的不是特定的特征,而正是整个模式。使这种方案特别令人感兴趣的是:同一个网络具有适应被给予了不同输入的不同激活模式的能力。当一种模式实现了,另一些模式在它们可能为相应的输入所激活的意义上,仅仅是潜在地出现了。人类要理解已开发的、运用了分布表征的系统中的活动,是相当困难的,但是存在着许多使这种系统具有极高吸引力的性质,例如相关表征激活常见单元的能力(这可理解为对由模式表征的整个对象的特征的表征)。(参阅欣顿等,1986)

尽管大多数哲学家不会去从事联结主义网络的实际构想,但仍有一些关于它们的构想的基本哲学争论。有一个就是关于这样一些原则的,它们控制着这些网络的操作,并且我们根据它们解释这些网络的诸能力。

1. 有没有我们据以解释联结主义系统的行为的基本设计原则?或者说联结主义系统的行为仅只是突现的吗(在特殊的情形之下)?

有些理论家已预见到了这样的前景,即我们可以建构联结主义系统,它能模拟人类的认知行为,但不能以机械论的方式解释这种行为。机械论的解释方法在许多学科中广为流行。这些学科中的科学家试图解释复杂系统的行为,方法是把复杂系统分解为部分,然后说该系统的行为怎样来自于部分的行为。根据这种机械论观点,每一构成部分可视为对于整个系统的行为所作的一种具体贡献。^②在联结主义纲领中,部分的作用无关紧要,系统的行为主要是起因于部分的相互作用,而不是部分本身的行为。虽然有

这样一些学科(如热力学),其中的解释模型已得到发展,在解释中,几乎没有什么解释功能被归之于部分,数学理论也发展到了这样的地步,因而能表明整体的行为怎样产生出高水平的结果,然而这些解释常常是不令人满意的。不过联结主义模型发展的潜在结果是:我们在心理学中可能有必要放弃机械论的分解方案,而接受数学和统计学关于突现的解释,这种突现出现在现象的更高水平之上,而不是从另外的低水平的操作中零碎地聚集起来的。^③

下面是关于研究者在解释联结主义系统时所用的原则的争论:

2. 把认知的、意向的解释用于联结主义网络中的活动是建立在根本原则的基础上的吗?

一种可能性是:研究者们把一种认知的、意向的解释用于联结主义网络,仅仅是因为那个系统以容许它执行任务的方式行动,例如我们认为,一系统在识别一单词,因为如此被说明,它的行为才是适当的。这就是丹尼特(1978)关于意向描述怎样用之于包括我们自己在内的任何系统的观点,这就是体现了丹尼特的意向性说明的工具主义特征的东西。然而在许多人看来,这是一种令人极为不安的结果。对于他们来说,我们的心理状态真的是“关于”事物的,即是说它们具有塞尔(1980)所说的“内意向性”,正是这一点才是我们的心理状态的根本特征。在联结主义系统中,如果状态的实际特征中没有任何支撑我们意向属性的东西,那么某些批评者就会认为:联结主义系统未能说明认知状态的基本特征。对于关于心理功能,如在传统的人工智能纲领中所把握的那种功能

1106 第九部分

的传统计算解释,塞尔作出了同样的批评。但是至少是在福多(1975)关于计算理论的阐释中,程序的句法操作据说反映了语义学,因而状态的内在加工被认为反映了从语义学上加以解释的推理的种类,而我们却认为那些推理是那个系统作出的(例如,如果因为有闪电,系统就推论:将要打雷,那么内在过程就将包括这样一种推论,即从我们认作命题——“有闪电”的符号到“将要打雷”的命题的推理)。联结主义者拒绝这种句子加工模型,因而对于我们给予系统的一切语义学说明似乎不可能提供动机明确的解释。

即使有另一种观点,据此我们可以判断:对于意向性,联结主义系统实际上比传统的认知模型更好的精神食粮。但德雷福斯(1979)对传统人工智能纲领提出的反对意见之一就是:它们试图明确表征那种系统会用到的全部信息。这种企图除了潜在地使人工智能努力不可能实现外,还可以解释这种系统为什么缺乏内在的意向性。该系统的行为完全由其内在表征所决定,因此它不可能直接与这些表征打算“关于”的外部对象相关联。然而一种联结主义系统并不依赖于作为它的加工单元的内在表征。它产生作为对输入作出反应的内部表征,同时也有开发这样的系统的可能,这些系统能调节它们的联结力度,进而产生它们自己的、把输入范畴化的系统(参阅拉梅尔哈特等,1985)。^④而且为了实现它的认知操作,它也不需要表征它的环境的一切有关方面。它完全能以适当的方式对环境的那些特征作出反应。使自己与环境相协调,而且不完全根据按句法规则编码的表征进行操作,这两者的结合使人们更有理由认为,联结主义系统的活动真的能“关于”它们环境中的事物,因而享有比传统认知模型更内在的意向性。^⑤不过这只是

简述了一种方法,以此我们可以推进对联结主义系统的意向解释。当然,对这种系统中的意向性作出解释的潜力还需进一步探讨(参阅贝希特尔,1985a)。

认知的联结主义说明与以规则 为基础的说明之比较

联结主义方案的倡导者常常把联结主义说成是对于解释认知的更传统的方案的一种可供选择的東西、一种潜在的替换品。传统方案设想:认知系统按照规则处理表征。如果联结主义真的是一种替代性的选择,如果我们能明确地阐明两种方案如何不同,那将是大有裨益的。

由于乔姆斯基对语言的说明(据此,语言中的专门结构可能是由规则的递归集合应用于其他语言结构而产生的)以及在数字计算机上模拟认知的尝试,关于认知的规则和表征模型就变成了一种在认知科学中占有显著地位的模型。对于模拟人类认知行为的计算机的应用极为重要的是这样的假定:心灵和计算机就是纽厄尔(1980)称之为“物理符号系统”的东西。这就是说,符号从物理上编码在能够按照指定规则被操作的数据结构中。当这些规则能直接地具体化于系统的物理过程中(像它们在袖珍计算机中那样)时,它们一般会把自己编码在数据结构中,以至于在一数据结构中的某些因素控制了其他组成部分的操作。

围绕联结主义模型怎样区别于以规则为基础的模型,产生了一个基本问题。既然表征以及操纵它们的规则是以规则为基础的

1108 第九部分

说明中的关键因素,那么就让我们考察一下联结主义模型中是否有它们的类似物。从表面上看,联结主义系统似乎极不同于以规则为基础的系统。在前者中根本就没有操作作用于其上的固定表征,而只有被激活的单元,它们进行活动以增加或减少其他单元的激活,直至稳定的构型得以形成。然而外表可能是骗人的。表面的差异可能是我们对于两个系统所作的排列或分析的产物。在联结主义系统中,我们是在单个集合体中考察活动,我们看到的这种集合体通常所执行的任务是基本的认知操作(例如识别和分辨一特定输入)。但是联结主义系统要模拟令人感兴趣的认知活动将需要运用大量的集合体,而且利用一些集合体储存其他集合体的内容或控制其他集合体的加工也可能是必要的。当我们考察大量的集合体可能由以履行重大的认知任务的方式时,我们便可以看到:把那活动看作是以规则为基础的表征加工是再自然也不过的了。

为了弄清联结主义模型与以规则为基础的模型之间的差异,如果有这种差异的话,我们就有必要小心注意表征和对它们进行加工的规则这两个概念。先看表征,我们可以提出这样的问题:

3. 联结主义系统所用的表征不同于基于规则的系统所用的表征吗?^⑥

福多和皮利辛(1988)在批评联结主义时指出:既然联结主义者不是(在局部主义解释中)对单元就是(在分布的解释中)对激活模式提供了语义解释,那么他们就是表征主义者。这足以使两个系统中所用的表征相等同吗?其实仍存在着可观察到的某些差

异。在对联结主义系统的局部化说明中,被赋予了表征功能的单元常常出现,但它们是由那系统中的其他活动开或关的。这不同于表征在以规则为基础的系统中被操作的方式。但不清楚的是:这种差别究竟有多重要。就分布系统来说,它就是一种超出作为表征的单元的集合体的模式,因此在那里可能存在着更大的差别。在这样的系统中,当一种模式不出现于该系统中时,它就不被储存,就像在传统认知模型中那样。这种模式不能从记忆中检索出来,但可在适当条件下被重构。这里仍有差异,当然仅只是执行任务方面的,从而不至于引起两类系统之间的根本差别。更基本的问题是:联结主义系统中的表征是否能像以规则为基础的系统中的表征那样完成相同的基本功能。这是一种莫衷一是的争论。这里我只能摆出来,而不打算给予回答。

当我们集中注意控制两类系统中的操作原则时,对它们作出比较可能是大有希望的。以规则为基础的系统通过专门把规则直接关联于对表征的处理而发生作用,而联结主义理论家的一个远大目标就是想搞清楚:我们是否可以通过让全部活动由决定激活怎样在单元之间进行的联结来控制这一方式去摆脱规则。^⑦但这又促使我们考察下述问题:

4. 控制基于规则的系统的加工规则与控制联结主义系统的加工的各种联结之间有什么明显的差别(如果有的话)? 联结是规则的功能等价物吗?

在一定意义上,联结具有与规则相同的功能:它们都可使系统从一种表征状态过渡到另一种表征状态。联结和规则之间的关系

1110 第九部分

更清楚地体现在所谓的“可编制程序的联结系统”中,在此系统中,一种单元集合体的活动把联结固定在另一集合体中,从而决定它的加工能力(麦克莱兰,1986)。再者,明确对两类系统作出区分这一问题足颇令人困惑的,我这里不能予以解决。

试图把联结主义系统与以规则为基础的系统区别开来,由于下述事实而很难实现,这事实是:通常的联结主义系统不是直接用硬件建造的。确切地说,它们是由传统的计算机模拟的,在这里,程序(由规则陈述所构成)被编写出来,目的是要决定联结主义系统怎样行动,因而我们必须进一步提出这样的问题:

5. 用以规则为基础的冯·诺伊曼计算机模拟联结主义系统,能说明联结主义系统与以规则为基础的系统之间有基本的相似性吗?^⑧

问题似乎相当棘手。有些已接受了关于认知的以规则为基础的模型的人坚持认为:如果我们为一计算机编制这样的程序,让它像人类一样进行符号处理,那么我们就创造了一真正的能思的实在。在传统计算机上所作的对联结主义系统的模拟中,计算机计算那控制一联结系统中的单元激活的变化的诸方程式,进而确定什么是联结主义系统中的所有单元的激活强度。如果传统计算机能模拟联结主义系统,那么我们怎么能坚持认为两者是根本不同的呢?

然而这种消解联结主义系统与以规则为基础的系统之间差别的方法似乎是不恰当的,在以规则为基础的计算机上所作的联结主义机器模拟可与计算机上的飓风模拟相比拟。在两种情形之

下,计算机所作的就是决定什么是因果过程的结果,此过程实际上是未经过那些因果操作而发生的。根据这种观点,模拟了联结主义系统的以规则为基础的计算机并未例示联结主义系统的重要因果属性,而只是向我们说明了那系统中的因果相互作用的结果。

关于认知的联结主义方案与以规则为基础的方案之间的冲突

在回答上述问题 3 和 4 时,即使我们在联结主义系统与以规则为基础的系统之间作出区分这一点尚不明朗,但每类系统的倡导者为了他们那类模型的利益,对他们所信奉和论证的那类系统提出了相对清楚的范例。让我们假定两类系统是根本不同的,并转向下述问题:

6. 我们用什么标准来评判联结主义说明或以规则为基础的说明是否给予人类认知过程以恰当的说明呢?

如果有这样的认知任务,人类能够予以完成,而那两类系统中只有一类能完成,而另一类则不能如此,那么上述工作实属举手之劳。进而我们也就有相当充分的理由偏爱能完成人类所完成的任务的系统。许多以规则为基础的理论家对联结主义模型的初步回答就是主张:这其实是自然形成的状况。他们认为:存在着像递归这样的过程,人类能完成,而联结主义系统则不能完成,因而联结主义系统是不妥当的(福多和皮利辛,1980,平克尔和普林斯,1988,他们提出了这类反对意见)。

1112 第九部分

不可能有这样的结果,即一方想以此为基础表明:他们的系统能做另一种系统不能做的事情。以规则为基础的模型与网络-类型的模型之间的这类争论有一先于联结主义流行的历史。第一代联结主义模型类型(如塞尔弗里奇(1955)和罗森布拉特(1962)的网络模型)已被摒除,因为很显然它们对后来所考虑的那些类型的模型构成了限制。明斯基和帕波特(明斯基和帕波特,1969)已建立了说明这种限制的某些重要原则。然而时下的联结主义者一直想说明这些限制怎样才能被克服(拉梅尔哈特和麦克莱兰,1986)。70年代初,认知心理学中对于类型上有点不同的网状模型、语义网络(参阅奎兰,1968)和作为表征概念的工具的特征定向(feature set)模型的相对优点曾有过争论。史密斯(史密斯,1974)等人论证说:他们的特征定向模型有把它与网络模型区别开来的属性。而霍兰(霍兰,1975)则对这种区分提出了挑战,方法是说明史密斯等人的模型怎样相同于网络模型。里普斯(里普斯,1975)等人基本上接受了霍兰的主张,但对于区别又提出了不同的观点,后面我再讨论这种区别。在这点上,假定联结主义模型能完成的任何任务也能被以规则为基础的模型所完成,反之亦然,这似乎是合理的。

里普斯等人提出的、关于区分的替代观点就是主张:“选择一种而不是另一种表征可能导致理论之间的真实加工的差别,即使表征本身在某种水平上是同型的”(p. 156)。这一看法也适用于对于联结主义的常见争论。即使联结主义模型和以规则为基础的模型中的每一个都能重复另一个的全部行为,但它们在完成任务的方式方面则彼此不同。这些差别可以表明:哪一类模型可以提供

一种关于人类认知的更好的模拟。当许多联结主义者论证说:他们的模型运作的方式比传统的以规则为基础的模型更相像于人类,他们所倡导的事实上就是这样的方案。

例如联结主义者论证说:他们的模型在生物学上比以规则为基础的模型更现实。例如它们表现出有被称之为 100 步原则的迹象(费尔德曼和巴拉德,1982)。基于个体的神经过程大约只花几毫秒,许多基本的认知任务在几百毫秒内便可完成,他们便推论说:心/脑经历的不过约 100 个连续的步骤。而当代大多数完成这些任务的人工智能程序则需比 100 步多得多的步骤,但联结主义机器能在 100 步之内成功地给出答案。

联结主义者还提到了联结主义系统的一些行为特征,他们认为,这些特征体现了联结主义关于人类认知的模型的优越性。例如能够对变形输入或新的输入作出反应就是联结主义系统的、与模型识别有关的一个自然特征,因为这些输入明显不同于那些预先为它设定的输入,或不像它曾被训练对之作出反应的输入。它还会对已改变了的输入作出反应,就像它对它的先前的、最直接地与之对应的输入作出反应一样,即使是在没有特别地告诉它怎样对付这种新输入的情况下,它也将如此行动。而且当联结主义系统的部件被破坏或当它超负荷时,它的运行会平稳降级。例如如果联结主义系统的某些单元遭到破坏,那么该系统仍可以通过大体进入并适应相同的稳定状态而对输入作出反应。

然而,就此而言,联结主义论证可能碰到下述主张所碰到的困难,这个主张认为:一类系统能完成另一类系统不能完成的任务。以规则为基础的加工模型的倡导者能满足生物学和加工这两方面

1114 第九部分

的条件。例如以规则为基础的人工智能模型似乎违背了加之于大量加工步骤上的生物学约束,其中的一个理由就是:这些模型是为连续加工的计算机设计的。但当代许多认知理论(例如安德森的ACT*模型)用的是产生式系统设计,而产生式系统自然是用平行结构完成的。这样做将戏剧性地减少许多连续的步骤,这些步骤对认知过程的以规则为基础的模拟似乎是必不可少的。而且A. 纽厄尔(1987)在采用一种以规则为基础的方案时,也求助于加工时间补偿,他用了关于加工时间的信息,作为通向结构的符号的向导,并坚持认为:他为关于认知的统一理论(SOAR)所提供的候选结构与关于心理过程的时间限制数据是相吻合的。

以规则为基础的加工的倡导者也能接受为了联结主义模型的利益而改进的加工数据。已经出现了以规则为基础的范畴化模型,它们可处理变形的或不完善的输入。这些系统可接受具有不同特征的信息,并可运用这样的算法,即它们能以那些信息为基础来确定那对象最有可能存在于什么范畴中。为了设计能自然降级的、以规则为基础的系统,安排部分冗余度对于这种系统所用的加工程序来说将是必要的。但从根本上说,这与以规则为基础的观念不是不协调的。

关于两类系统能做什么的行为数据或关于它们的执行方式可能是什么的行为数据,对于回答问题6是不够的。唯一的补救办法似乎是找到决定人类心灵怎样起作用的某种直接方式。在以规则为基础的加工方案之内,像皮利辛(1984)这样的理论家已论证过:为了使认知模拟比照人类,我们必须考察心灵的基本构造。他主张,只有根据关于这种构造的知识,我们才能够确定这种模拟是

否完成了同样一组操作。^⑨如果有一种辨别这种构造的方法,那么我们就可能确定人类心灵的基本操作是联结主义系统的操作还是以规则为基础的系统的操作。

我们应怎样认识人类心灵的构造呢?皮利辛所提供的识别构造的标准(对于不同任务和认知渗透性的相对的反作用时间)是行为方面的,并碰到了这样的难题,即我们可以用一种构造模拟另一种构造的行为,正像我们能在计算机中把一种构造编译或解释为另一种构造一样。而且他的认知渗透性标准可能已引入了一种有利于寻找一种有规则的加工系统的偏见,因为他似乎集中在信息的命题类型的渗透性上。

究竟是联结主义者还是以规则为基础的理论家对我们的认知结构提出了正确的解释呢?如果作出决断变得如此困难,那么他们之间的冲突似乎就不会像过去那样使许多人感兴趣了。首先,联结主义似乎已经提供了一种以规则为基础的加工解释的真实的对比,但我在本部分和前一部分所主张的则是:两种解释之间的差异可能是很难辨别的。不过也许我用了错误的方式去看待这种冲突。我一直把它当作是竞争的理论假说之间的冲突,我要搞清楚的是;我们怎样判断哪一个是正确的。探讨该问题的一种更富有成效的潜在的方法就是把它看作是建构认知模型的两套不同类型的工具之间的冲突。因此两种方案之间存在着根本的不同。不同的任务执行起来或易或难,这主要取决于我们采用哪一类系统。如果我们采用以规则为基础的模型,那么古典的推理任务就很容易形成模型,因为大多数推理模型是从逻辑的工作中派生出来的,因而是以规则为基础的。另一方面,也有相当不同的任务类型,诸

1116 第九部分

如模式识别,它们似乎更容易在联结主义框架中处理。

这里要说明的是:我们应当思考冲突,不是关于哪一种模型提供了关于人类认知系统的正确解释,而是去问哪一种提供了发展认知理论的更有用的框架。这样我们就把这种冲突变成了一种关于理论发展战略的冲突,进而我们便可以提出下面的新问题:

7. 在联结主义框架与规则-加工框架内,发展认知理论的相对优点是什么?

从这种观点来看,在特定类型的系统中什么易于模拟这一问题格外突出,而我们能在对立类型的系统中完成许多相同的加工这一事实则无足轻重。在特定类型的系统中,一类操作很容易完成,这一事实可能促使用了那种系统的人比那些难以在他们的系统中加以实现的人更加注意那种认知操作。根据这种观点,联结主义模型的广泛引入堪称认知科学中真正重要的事件,因为它可以引导我们注意不同类型的认知现象。尤其是可以引导我们超出语言加工和推理任务(它们似乎适合于以规则为基础的模式处理)进到其他认知活动,如模式识别,而这些活动在以规则为基础的模式中已证明是极难处理的。顺着这种方案,让我们考察一下模式识别,因为它已成为联结主义者研究的典型任务。

模式识别在人类认知中的重要性

联结主义系统之进入不同的稳定状态取决于它所接收到的输入,这一点自然使人们认为这种系统可完成模式识别或范畴化。

既然能改变它们联结强度的系统也可改变它们对时间的反应模式,因此认为这些系统可学会对输入的分类就是顺理成章的。相应地,研究者一直在探讨这种系统的不同学习程序,因而联结主义系统就被设计出来了,它们学习把输入模式范畴化,在学会之后,就以与先前学到的模式的相似性为基础,把新的输入加以范畴化。(联结主义者在说以相似性为基础作出判断时,自然不是说联结主义系统能确定客观的相似性,能解决哲学文献中一直在讨论的关于相似性的各种难题。这些系统所作的相似性判断,是由该系统的联结强度所决定的。我们有特许权说这些系统能判断相似性,是基于这样的事实,即它们把我们认为是相同的模式归于一类模式中。)正是由于它们能作出这样的相似性判断,联结主义系统才证明它们有把新输入加以泛化以及适当地对变形输入作出反应的能力。

在联结主义系统中轻而易举地完成模式识别的任务这一点又提出了一个饶有兴味的问题:

8. 模式识别究竟是何等重要的认知任务?

认知科学研究者一直有这样一种普遍的看法,即模式识别是一种重要而又从属性的认知活动。模式识别在其中显得重要的一个条件就是按照机器观点去发挥作用,在这里艰巨的课题是让机器能通过感官输入识别对象。很明显,这是机器在它们能与世界上的事物相互作用之前要获得的技巧(如像智能机器人)。但建立具有良好模式识别能力的、以规则为基础的加工系统则是不容易的,这部分是因为这样的识别要求机器能处理大量的环境线索,而

1118 第九部分

这些线索的关联又可能因环境而异。然而,既然我们在系统中从符号上将信息编码,进而直接集中在使用了那种信息的高水平加工上,因而对于许多目的来说忽略这一难题就是可能的。与识别感性输入的任务有关的能力给人以这样的印象,即模式识别可能仅只是从属的认知功能,在说到其他认知任务时,它可能被忽略。

然而,近年来,发表了关于概念和范畴化的相当多的心理学文献。范畴化是一可与模式识别相比拟的任务,因为它需要识别相同项目的能力,以便把它们归于一类。在以规则为基础的构架内,人们自然想到范畴化是由说明某东西必须具有什么特征才属于一类范畴的规则所决定的。心理学家罗施(1975)对这种观点提出了挑战,认为范畴有一原型结构:某些范例(如知更鸟)作为一种范畴(如鸟)的例子就比作为其他范畴(如鸭子)的例子更好一些。这说明范畴并不能根据充分必要条件而加以描述。罗施本人并未让这种原型理论归并到关于范畴怎样在心理上被表征的理论中。但其他心理学家这样做了。即使范畴原型理论之采用瓦解了下述观点的基础,这观点是:范畴以说明了充分必要条件的简单规则为基础,但大多数试图发展关于心理表征的说明以包容原型数据的心理学家仍坚持规则方案,根据这一方案,被试判断一对象相似于具有不同特征的原型,进而运用规则来决定那对象是否属于那类范畴(关于这种文献的评论,参阅史密斯等,1981)。有充分的理由认为:这种过程自然能由联结主义系统完成。正如我们已指出过的那样,判断相似性是联结主义模型的基本能力。联结主义模式识别方案根据与先前输入模式的最直接匹配,而不需要特别地约定的规则,就能划分新的输入模式的种类。所以罗施式的关于范畴

化的原型理论大概已顺利地、自然而然地充实到联结主义系统中去了,并且避免了试图设计以规则为基础的范畴化加工所碰到的困难。

然而,对于那些赞美认知科学中以规则为基础的加工的人来说,范畴化任务像视觉任务一样似乎不是关键的(尽管我们怎样把事物范畴化这样的证据可以极大地影响我们怎样对它们进行思考和推论)。他们可能坚持认为:一当信息以严格的符号形式或密码被表征时,真正的认知工作才能开始。不过,我们一当把这些任务当作是模式识别任务时,我们也就可以看到模式识别怎样以更重要的认知加工出现。关于更高水平的认知活动如问题求解的最广泛地被接受的观点(一种由西蒙、纽厄尔和安德森等理论家奉献的观点)大概是这样的,即这些活动中的熟练行为依赖于获得适当的规则。而这些规则常被认为是以相当有意识的方式学到的,但随后就变成了自动的(无意识的),因而熟练的行动者一般能在没有明显的有意识思考的情况下运用它们。然而德雷福斯等人(1986)则论证说:这样的规则学习只会导致一定范围内的有效行为。他们坚持认为:专门行为依赖于能够识别很像以前某些情境的情境,以及以相同的方式对之所作出的反应。他们考察的例子之一就是下棋,亦即一种为接受以规则为基础的方案的理论家们经常研究的活动。德雷福斯等人主张:把专家与合格的棋手区别开来的东西不是附加的、以规则为基础的知识,而是非常发达的、识别当下情境怎样相似于以前碰到的情境的能力。^⑩如果他们是正确的,那么模式识别甚至可能出现在似乎是高水平的推理任务这样的事例中,这些任务以前似乎离不开以规则为基础的推理。假如在联结

1120 第九部分

主义系统中完成模式识别与比较是轻而易举的话,那么这就表明:要弄清在不求助于以规则为基础的推理过程的情况下,模式识别究竟有多大作为,就必须开发联结主义系统的这种能力。

在本部分中,我的任务就是努力说明模式识别在其中可能是重要的认知能力的某些关系。还有其他一些领域,在那里,这种能力可能也是重要的。例如经过模式识别过程,联结主义系统可能激发与先前相似的经验有联系的模式。这提供了一种探讨模拟可寻址内容记忆的任务的富有启发性的方法。这是一种体现了人类特征的记忆方式,但很难在古典的结构中成功地加以模拟。我们也可以发现模式识别在下述活动中的作用,它们一直被当作是像句法加工这样的按照规则加工的范例。我们发现:只要使用模式识别系统,我们即使不求助于复杂的语言学规则也能设计出有助于语言理解和生成的系统。这些自然只是建议,对于联结主义系统借助不求助于规则的模式识别过程究竟能获得多大的成功,我们将拭目以待。还应记住:设计能完成相同模式识别任务的、以规则为基础的系统多半也将是可能的。在联结主义系统中完成这些任务是相对容易的,这一点说明:我们应开发这种系统以确定认知在多大程度上依赖于模式识别,而不需以规则为基础的加工。如果联结主义系统有助于引导我们注意像模式识别之类的、别的认知活动,它们不是最近的以规则为基础的认知科学关注的焦点,那么这些系统将会有重要的启发功能,不管关于人类是联结主义系统还是以规则为基础的系统的争论的最终结局是什么。

联结主义模型与 规则—表征—模型的可能调解

即使联结主义者和以规则为基础的理论家常把他们相互视为敌对者,但我在最后一部分要强调的是:两种系统对许多不同类型的活动来说自然是各有千秋。既然推理常被看作是遵循规则,如像自然演绎系统中明确表达的那些规则,因而让以规则为基础的系统执行复杂的推理任务就是相对容易的。另一方面,联结主义系统已被证明长于像模式识别和运动控制这样的工作,并显示了可寻址内容记忆方面的潜力。这便促使我们追问:

9. 在单一的认知模型中,把联结主义模型与以规则为基础的模型统一起来以取两者之长是可能的吗?

既然联结主义者把自己当作是为推翻更传统的、以规则为基础的方案的统治而进行战斗的革命家,那么他们试图推广新方案、冲破旧框框以搞清他们仅用这一方案能将研究推进多远,就不足为奇了。但事实上,对联结主义所作的大多数研究一直集中在这样的任务上,即紧接着一系统——比如说能识别它所面对的特定模式,并期待它的后续输入的系统——所作的特定反应之后的任务。然而在典型的人类认知中,对模式的识别并不是那一过程的终结。它可以为其他类型的活动提供基础。当我认出了我的猫时,我大概会看看:我为它留下什么食物没有,或把它今天的药物给它没有,或者我大概还会被促使对它说点什么。

在联结主义系统中,我们可以用两种方法设想作为过渡到其他认知加工的模式识别任务的完成。被激活的特殊模式可以把它的激活传递到与它以传递激活的方式相关联的其他模块上。或者,一种模式的识别可以激发规则应用于其上的进一步的加工。例如如果模式识别程序认出了我的上司,那么这可能有利于满足当时被激活的规则的前件:“如果我看到了我的上司,便问问我的旅行费”,接着就导致那种输出。

两种方案的倡导者都为和睦状态的建立提供了条件。联结主义者如拉梅尔哈特和麦克莱兰把他们的纲领描述为专注于认知的微观结构的纲领,这似乎提供了这样的可能性,即更传统的认知模型可以描述微观结构。拉梅尔哈特、斯莫伦斯基、麦克莱兰和欣顿(1986)曾试图表明:传统的以规则为基础的系统的某些特征如图式和结构的运用,怎么可能在联结主义系统中实现。图雷茨基(1986)也说明了怎样在联结主义系统中完成传统的 LISP 操作。

另一方面,以规则为基础的理论家,如福多和皮利辛(1988)都情愿设想这样的可能性,即在人类身上,以规则为基础的加工可能在像联结主义系统这样的东西上完成。这牵涉到的大概是:让以规则为基础的系统的形式推理原则映射到联结主义网络的状态上。不过福多和皮利辛在完成的方式与实际认知结构之间作了明确的区分。他们认为:这种结构一定是一种规则-加工系统,它运用了(1)复合句和语义学,(2)结构敏感程序。他们通过描述几种认知特征论证了对结构的这些要求,认为:这些认知特征需要这样的过程:(a)“虽然”(though)的产生率(即我们能通过递归过程建立复合命题),(b)认知表征的系统性(它由这样的事实所说明,即

任何认为约翰爱玛丽的人也会认为玛丽爱约翰), (c) 表征的构成性(即心理表征的组成部分大约赋予它表现于其中的一切心理表征以同样的语义成分)。不过福多和皮利辛又强调: 不仅认知的那些特征需要规则-加工系统, 而且所有的认知或心理活动都发生于那种结构中。那种结构完成于其中的系统的特征仅仅只是完成的问题, 因此不是心理学家的问题。

即使我们暂时接受福多和皮利辛的下述主张, 即认知的某些特征离不开规则-加工系统, 但我们仍可以通过提出下述问题对他们关于联结主义的否定性结论提出质疑:

10. 如果规则-加工系统在联结主义系统中完成了, 那么这能说明以规则为基础的认知系统有认知的方面吗?

如果人类的规则-加工系统事实上在联结主义类型的系统中实现了, 而且这使该系统具有像模式识别、概念学习、可寻址内容记忆等能力, 那么实现的方式似乎就有认知的意义。有争议的不仅仅是实现的问题, 而且也涉及到关于认知特性探讨的至关重要性问题。福多和皮利辛似乎假定: 低水平的探究, 如规则-加工得以实现的那些水平, 对高水平的探究并无意义。他们认为: 心理学探讨基本上属于自主性探讨。^⑩但在其他学科中, 自主性假定则行不通。例如关于物理膜状物的结构的知识对于决定生物化学理论的形式就极为关键, 反之亦然。而且研究者有时也发现: 为了解决那些被认为仅限于一学科(如生物化学)范围内的问题(如氧化磷酸化作用), 从其他学科(如细胞学)中引进方法就极其重要。(关于福多和皮利辛提出的各类自主性主张所面临的难题的进一步讨

1124 第九部分

论,可参阅贝希特尔,1988a)尚不清楚的是:对实现的研究是否能为解释以规则为基础的系统内的认知加工的某些方面提供同样相关的材料。

为了提出一种把规则—加工与联结主义模型统一起来的更有效的方式,我将暂时转到概念和范畴化这一论题上来。概念是以规则为基础的认知加工单元的可能的候选者。早些时候,我曾论述过这样的可能性,即范畴化为概念可能在联结主义系统中完成,以致能利用该系统识别相似性、把例子与原型联结起来的能力。这样做还有进一步的利益。一般来说,甚至是在关于概念的罗施式的观点之内,概念也可被当作是相对稳定的。因而它们似乎就像思想中的原子——本身不可分或不可更改。但巴萨卢(1987,即出)为我们概念中的可变性提出了一重要的证据。罗施证明概念特性的手段之一就是:要求被试列举一个范畴的不同成员的典型性。她声称,在这种列举中已找到了被试间高相关的证据。然而巴萨卢则论证说:她的统计手段是无效的,因为当更适当的统计手段被使用时,典型性判断的被试间相关大约接近于0.4。而且他还发现:甚至就被试内部而言,典型性判断大约以一个月的时间间隔发生变化,只有0.8的相关。

基于这些结果,巴萨卢作出了下述推论:概念可能不是作为固定的单元储存在长期记忆中,而可能是由于特定的推理任务的需要而“匆忙地建构的”。在对认知的以规则为基础的说明内,阐释这种观点的意义是有点困难的,因为概念似乎是这些系统的原子。但根据联结主义观点,对之作出澄清则容易得多,因为在那里,存在于长期记忆中的东西仅只是联结。这些联结能使被试形成起着

概念作用的、并可用于需借助于规则的问题解决中的表征,但概念没有必要是固定的,即使它们在以规则为基础的大多数说明中是原子结构。这一来,联结主义系统中的“实现”概念可能容许我们以直截了当的方式解释在其他情况下很难解释的概念的特征。此外,用联结主义系统表征概念的一个重要优点就是我们能利用联结主义系统的学习能力解释新的原概念是怎样学到的,这是一种为福多(1980)这样的规则-加工理论家所否定的现象,但它是可能的(参阅贝希特尔,1985b)。

我这里只是粗略地说明了联结主义和以规则为基础の説明怎样结合的问题,其基本观点是:联结主义系统中的状态可以当作是以规则为基础的加工系统中的表征(包括规则表征)。这样一种调解办法的优点是:联结主义系统的某些吸引人的特征可能为规则-加工系统所利用,而它们本身的显现又无需借助规则的发明(贝希尔特,1988b,在此我更详细地阐发了这一点)。显然在这种综合的系统建立之前,还有许多技术性问题需要解决,但开发这种系统至少是可能的。

认知的联结主义模型与 神经科学说明的关系

另一由联结主义所引发的哲学争论是联结主义理论与其他理论的关系问题。一方面,有联结主义模型与神经科学模型的关系问题,另一方面,有联结主义模型与民众心理学的关系问题。这两个问题引起了这样的争论,即联结主义在某一方面是否是一种还

1126 第九部分

原主义纲领。在最后两部分,我将依次探讨这些课题。一开始我就曾强调:联结主义模型是受神经系统的启发而建立起来的。单元的激活程度可看作是神经元的激活率;激活转向另一些单元可比之为神经元之间的信号流动。但很显然,联结主义模型本身不是神经模型,因为联结主义者无意于说明神经网络的全部细节。总的来说,联结主义者对于把他们的模型变成实在论的神经模型也不感兴趣。所以我们有必要提出下述问题:

11. 如果联结主义模型本身不是神经模型,那么它们是关于什么的模型? 联结主义模型的组成部分怎样与大脑的组成部分相关联?

倡导以规则为基础的模型的那些人有一种阐述关于他们模型的类似问题的自然方法:以规则为基础的说明阐述了由计算机或大脑中完成的更基本加工所实现的加工。它们是对真正的因果过程的说明,而且这些过程是在比个体的神经元甚或神经的集合体的活动更高的水平上发生的。不过,这种有利于联结主义者的回答几乎没有什么作用。联结主义模型中的活动如此像神经网络中的那些活动,以至于把它们当作是在那些网络中的加工模型便再自然不过了。

联结主义者解决这种争论的另一方法就是坚持认为:他们的模型是神经加工系统的抽象模型。就此而言,他们忽略了真实的神经系统的某些细节,因而只是描述了一般系统的特征(斯莫伦斯基,1988)。当他们的观点说明了先前视之为奇迹的某类过程实际上怎样由像大脑这样的过程完成时,在这种抽象水平上所作的分

析事实上有极大的解释力。它证明,某些机制可以完成某些任务。而且这种方案还能为人们在中枢水平探讨理智指明方向。神经科学所面对的一个问题就是说明神经元活动怎样执行认知功能。联结主义模型说明了它们怎样出现在这些活动中,因而为未来的研究详细地确定神经系统怎样完成这些功能准备了方法。

在现在的条件下,把联结主义模型设想为关于神经系统的抽象说明似乎是合理的。但如果这是对的,那么联结主义模型就是对神经活动而不是对大脑中的高水平活动的说明。它们完全忽略了这种活动的某些细节。而且未来的一个合理的目标似乎应是发展对于神经系统实际上怎样起作用的更准确说明,而不是固守联结主义理论和神经理论之间的区分。神经模型与联结主义模型之间的差异似乎会融合为一种实用主义的差异——在发展我们对于大脑和认知的理解这个结合点上,考察各种类型的联结主义模型是有用的,而随着计划的发展,阐发关于中枢的更准确的解释将显得更加重要。然而,这就缩小了联结主义模型与神经模型之间的差别。这一点至少对某些联结主义者来说是重要的。

我所陈述的关于联结主义的观点似乎给予它在还原主义纲领中以显著的地位。如果联结主义模型成功地说明了全部认知现象,那么似乎可以准确地说,它们已说明了怎样把传统的认知理论化还原为神经水平的理论化。然而存在着一种不同的、而我们又可采纳的观点。认知研究(不管是联结主义的还是以规则为基础的)和神经研究之间的差别最好是当作研究者所提出的问题的类型的差别,以及他们用来回答问题的概念和研究手段的差别,而不是研究出现于其中的自然中的机体水平上的差别(参阅亚伯拉罕

森,1978,关于接受此观点的各学科之间的关系的讨论)。进一步说,即使某些更基本的操作是在简单的联结主义网络中完成的,但其中一些操作也会涉及到这些网络与其集合之间的高水平的相互作用。像在前一部分所设想的那样,这就是以规则为基础的加工可能会出现的地方。高水平的活动可以用神经科学的术语加以描述,但用认知的术语描述它们也会被证明是重要的。作为联结主义观点的结果而出现的东​​西是否定这样的观点,即存在着可为认知理论化的独一无二的水平,而不否定这样的观点,根据这一观点,认知的理论化(中枢的理论化也一样)发生在不同的水平,其中的一些比传统的认知理论化的符号水平要低一些。

联结主义模型与民众心理学的关系

哲学家考察心理活动时,常常集中在民众心理学上。而民众心理学是这样一种观点,据此,我们认为人是有信念、有愿望的,进而在信念和愿望的基础上解释他们的行为。哲学家们关心的主要是:在心理学和神经科学的实验探讨得到发展之后,民众心理学的结局将会怎样。在民众心理学中,我们描绘的人是这样的,即根据信念和愿望作出推理以决定该做什么,因而这种民众心理学就是下述观点的一个源泉,即认知是一种以规则为基础的加工过程。既然以规则为基础的认知模型使用了存在于民众心理学说明中的逻辑推理模型,还用了作为科学心理学基础的那种结构,那么他们必然会维护民众心理学的有效性。另一方面,如果联结主义者在他们坚定地主张心灵不是以规则为基础的系统时是正确的,

那么他们似乎剥夺了民众心理学以及以规则为基础的认知心理学的合法性。这种可能性为取消论的唯物论者 P.S. 丘奇兰德和 P.M. 丘奇兰德所强调,他们认为:关于大脑实际上怎样完成认知任务的联结主义及相关说明表明了拒斥民众心理学的必要性,正像我们已知道的、对民众物理学和民众医学的拒斥那样(参阅 P.S. 丘奇兰德,1986, P.M. 丘奇兰德,1986)。关于联结主义要提出的最后问题是:

**12. 如果联结主义模型是正确的,那么这说明了我们必须
剔除民众心理学吗?或者说民众心理学仍有其地位吗?**

两位丘奇兰德坚持认为:心理学(尤其包括民众心理学)要么必然还原为我们关于心/脑怎样工作的最佳理论,要么就是被剔除。所谓还原,他们想到的似乎是哲学意义上的还原,在此还原中,民众心理学的术语与来自于对心/脑的科学说明的术语是同一的,因而对民众心理学的理论说明是由我们关于心/脑的新的理论说明派生出来的。他们怀疑这样一种还原的可能性,因而呼吁剔除民众心理学。另一方面,他们把联结主义模型当作是关于心/脑操作的潜在正确的说明,因而当作民众心理学的潜在的替代品。这里我既赞同对还原的否定又赞同这样的主张,即联结主义对心/脑怎样工作可以提供基本正确的说明。我所要探讨的是:我们是否必须作出这样的选择:要么把民众心理学还原为联结主义理论,要么予以剔除。

让我简述一种可供选择的看法。即使关于认知的、以规则为基础的说明的某些支持者把民众心理学当作是关于心灵实际上怎

1130 第九部分

样工作的、以规则为基础的模型的原型,但这可能误解了民众心理学的功能。民众心理学最好被理解为关于人的说明,而不是关于人的内在心理过程的说明。人有信念和愿望,但这些并不是人的内部状态。民众心理学可以为我们提供一种描述面对环境的人的方法,但事实上人不可能内在地表征他们环境的一切方面。他们的内部状态使他们能根据他们所面对的环境的特定方面适当地行动,但绝没有构成那种知识的内部状态本身(贝希特尔,1985a)。

如果这是正确的,那么民众心理学不能还原于联结主义这一点就没有剥夺民众心理学的存在资格。我们可以一如继往地主张:民众心理学提供了对于行为主体的描述,具体说明了它所具有的、关于它的环境和它的行动目的的信息。就此而言,民众心理学(或某种关于它的修改过的看法)仍会扮演重要的角色。实验的理论化有时所面对的一个危险就是建构理论模型,以说明这样的行为(如记起瞎说过的话的能力),它们能在实验中诱发出来,但是在有机体应付它们的日常的生活时,这类行为并无重要的作用。对于人有关于他们的世界以及他们在其中所进行的认知活动的什么信息,民众心理学给了我们一种透视的方法。当然它不是获得这种方法的唯一途径。对于心理学的生态学研究就是另一条途径,它强调的是认知主体的真正生存活动(参阅尼瑟,1975,1982)。我的主张不过是:描述人及其所拥有的关于他们环境的信息的这类方法,对于那些想发展任何内在加工说明——联结主义的或以规则为基础的说明——的人有着至关重要的作用。因而这种观点不会为内在加工说明所剔除。恰恰相反,一种成功的内在加工说明将会解释:当某人面对一现实环境时,是什么使他满足民众心理学

原 书 缺 页

1132 第九部分

从它与世界的感性联系中所得到的适当的不变量(在表现了环境特征的感官媒介中的信息)。

⑥ D. 劳厄德最先向我指出了这个问题的重要性。

⑦ 联结主义者确实把改变联结强度的程序称之为“学习规则”。但它们完全不同于基于规则的系统中所出现的那种规则。简言之,学习规则是关于联结强度怎样根据那系统的活动而变化的数学说明。

⑧ 这个问题最初是 M. 科尔西特向我提出的。

⑨ 对于传统的计算机而言,这种指定了原始操作的结构观一般来说是相对的。当特定系统的机器语言直接限定了该系统能使用什么样的原始操作时,大多数编程就可用更高级的语言完成。但皮利辛认为:就人类而言,就没有这种相对性。

⑩ 自然,在新的条件下,我们并不总是做出在先前相似条件下所做的事情。至少有两种条件,在其之下,我们可以不遵守那个模式。在先前条件下所采取的行动如果没有令人满意的结果,那么我们可能不会再去尝试。这大概就是我们在其下考察规则的条件。

⑪ 事实上,福多和皮利辛所主张的不过是,心理学有“可自主陈述的原则”。但不清楚这个主张的意思究竟是什么。它可能是说,我们只需经过心理学的探究就能发现心理学的原则。但它也可能是说,一旦发现了心理学的原则,它就只能用心理学的术语加以陈述。这似乎更有可能,但正如下面所论述的,它也可能是荒谬的。

参考文献

J. 麦克莱兰(1986):“关于阅读的可编程黑板模型”,见他与 D. 拉梅尔哈特合著的(1986):《并行分布加工:认知的微结构探究》,麻省理工学院出版社。

G. 欣顿等(1986):“分布式表征”,载同上。

D. 丹尼特(1978):《脑猝病》,麻省理工学院出版社。

J. 福多(1975):《思想的语言》,纽约。

H. 德雷福斯(1979):《计算机不能做什么》,纽约。

D. 拉梅尔哈特等(1985):“竞争性学习所得的特征发现”,《认知科学》,

9,75-112。

W. 贝希特尔(1985a):“实在论-工具主义与意向观”,《认知科学》,9, 265-292。

——(1985b):“新的认知并行分布加工模型是认知主义的还是联结主义的?”《行为主义》,13,53-61。

——(1988a):《心理学的科学》,新泽西。

——(1988b):“联结主义与规则-表征系统”,《哲学心理学》,15-16。

A. 纽厄尔(1980):“物理符号系统”,《认知科学》,4,135-183。

J. 福多与 Z. 皮利辛(1988):“联结主义与认知构造:批判性分析”,《认知》,28,3-71。

D. 塞尔弗里奇(1955):“现代计算机中的模式识别”,《西部计算机联席会议论文集》。

F. 罗森布拉特(1962):《神经动力学原理》,纽约。

M. 奎兰(1968):“语言记忆”,载 M. 明斯基编:《语义信息加工》,麻省理工学院出版社。

I. 里普斯等(1975):“集合论与网络模型新探”,《心理学评论》,82,156-157。

E. 罗施(1975):“语义范畴的认知表征”,《实验心理学杂志》,104,192-233。

P.S. 丘奇兰德(1986):《神经哲学:迈向大脑统一的科学》,麻省理工学院出版社。

P.M. 丘奇兰德(1986):“神经生物学中的某些还原战略”,《心灵》,95, 279-309。

解决心灵哲学问题 的控制论方案*

肯尼思·M.K. 塞耶

在本书中,我的目的就是倡导对心身问题的一种解决方案,以避免上述困难。^①这一办法以通信理论中专门地定义过的信息概念为出发点。我们将看到,这个概念已超出了它原来的范围而有更广泛的应用。而且,它还可被应用到比我们通常知道的更广泛的现象上去,其中一些应用在哲学上具有相当大的裨益。

下面诸章要论证的是,这些概念适宜于解释物理和心理事件,它们可以为一种概念框架提供基础,在此框架内两类活动都能内在关联起来。这一框架还会在概念上使关于人的科学与物理科学密切结合,并有可能使生物学和物理学中的某些原则富有成效地推广应用于对有意识心理现象的研究之中。

这一任务既在概念上又在经验上确立了一种方法,该方法会使读者感到既熟悉又陌生。例如物理学家在关于热力学的熵的论述中大概会发现令他们惬意的根据,而生物学家对我利用进化论大概会产生兴趣;但是那些专家的轻信最好是通过将他们的概念

* 译自作者所著的《控制论与心灵哲学》(劳特利奇,1976)一书“导论”中的第4-5两节。标题为译者所加。

扩展于对心理现象的研究而加以验证。另一方面,社会学家和语言学家在我关于人类的人际行为的论述中将找到他们熟悉的结论,当然也可能为一种研究方法而恼火,因为这种方法优先考虑的是新的问题,而不是通常被认为是这两类学科的中心的其他问题。然而在其他尚未进行科学研究的领域(意识、理性、主观性)中,流行的研究方法似乎仍将富有革新精神。它不顾曲解的风险,依然故我。如果有关人的研究要深入到相互关系领域,那么只有具有坚实的基础才会避免概念重构的压力。我们所做的是这样一种工作,即在原科学层面上的综合,而不是在有争论的领域涵摄特定的正统观念的那种综合。因此,对我们的目的来说,不妨借用这样一条时髦的格言,即事实是金,确定的理论是银,而有争议的理论不过是铜。

最后,虽然(我已论述过的)原科学是一种哲学事业,但哲学家本人也许特别容易被下述章节的概念探讨弄得晕头转向。虽然概念重构过程应当是大家所熟悉的,但通信理论的原始词语对于大多数人来说也可能是新颖别致的。更有甚者,用于这类分析的材料可能异常地带有经验性。分析将在两个层次上进行,首先是对于经验科学的某些需要根据信息和反馈予以解释的原则的分析,其次是对像理性、意识这类不得不在科学构架内弄明白的现象的分析。第一个任务需要对某些科学原则作一些比通常的哲学研究更细致的讨论。另一方面,第二类分析极大地依赖于对我们所讨论的各种现象的描述,因为在概念尚不完善的情况下,概念的阐述是远远不够的。无论如何,程序将是分析性的,但仍是服从于校正和澄清这样的目的。

1136 第九部分

很明显,在设想智力是从活生生的有机体的其他能力中“派生”出来的意义上来说,这里的探讨同时又是综合性的。从这个角度看,理解智力就是设想它是进化的产物,因而在有机体层次上的“派生”就是一个生物过程。然而在概念层次上,从确定其生物发展可能性的方面来说,又必须把这样的智力能力表现为主观性和理性。在这种概念的“推演”中,主要的步骤是这样的,首先是通信理论中的信息概念。负反馈被分析为一种信息过程,据此,进化本身才能得到定义。接着,行为的条件反射作用被分析为进化过程的一种形式,知觉意识则被分析为条件反射作用的一种形式。语言被看作是人的知觉能力的放大,理性则被看作是不受刺激控制的语言。最后,主观性被说明为语言和理性的一种维度,以这个维度为标准,人们便设想自己不同于其他对象。从这种概念的推演中,可以得出如下结论:智力是一组信息加工能力,不过最好理解为处在其发展过程中的中间阶段的能力。(与此类推,在命题逻辑中,其他一切算符都可根据一单个的算符而加以定义,但是超过了最低限度的复杂性的逻辑推演决不能仅用这种算符而使之清楚明白。)

最后,哲学家们可能关心信息概念在其中起着基础作用的控制论图式的本体论意义。如果本书的设想是成功的,那么它不仅表明:信息概念为对物理的东西和心理的东西的分析提供了一种原语词,而且还能表明:信息状态(尚待详细说明的)先于心理状态而存在。既然在这个意义上,信息先于智能,而且又包含于一切心理状态之中,那么由此可以说,信息在本体论意义上也处于优先地位。因为如果A的例示先于B的例示,那么没有B,A也能存

在,但是反过来则不行。这大概就是“本体论的优先性”一词的意思。因此这一计划的成功将表明:信息状态的本体论可以解释心灵现象,^②因为它不同于物理事件的本体论。而且,在第三章所提供的材料的基础上,可以作出这样一个合理的推测:信息本体论同样是自然科学的基础,当然验证这一点则需借助量子物理学的细致的分析。

对心身问题的这种解决办法兼有还原论和一元论两者的优点,而又不会受到其伴随的弱点的惩罚。

罗素的中立一元论的主要优点在于:“感觉到的东西”(sensibilia)这一概念可以适用于物理的和心理的东西,因为它使人想到在两个领域之间存在的一种理论上的和睦状态。信息概念同样具有这种中立性。然而,在罗素那里,有一缺点伴随着这一优点,那就是没有关于感觉的东西的、据此可以增加我们对于任一类现象的理解的、独立的理论。与此相反,信息概念则是一种牢固地建立起来的正规理论的一部分,对于其他各门学科已证明具有适用性。

一般来说,心灵主义和唯物主义有一优点,即它们可使我们避免心身交感的问题。但是众所周知,还原主义的两种形式都缺乏有用的东西去说明它们的对立的领域。相比之下,控制论的解决办法则能阐明心身交感的模式,并能不断增加我们对两类现象的理解。

唯物主义具有直接与一些精密科学结盟的优越性,而这些科学在它们的解释的努力中显然是硕果累累。但是通信理论是数学的一个分支,因而比物理学或化学更严密,它的成就在近几十年的信息加工技术中鲜明地显露出来了。

1138 第九部分

然而,还原论和中立一元论作为对心身问题的解决办法有一最令人失望的方面,那就是:除了使它们受到各种本体论观点的怀疑这个技术特征以外,它们一点也没有推进我们对于人类有机体的真实的理解。由此说来,那可作为控制论观点的最重要的优点的东西便是:它对于人类行为的许多方面已形成了一些新颖独到的看法。不像哲学家的全部纲领中的那些可供选择的纲领那样,控制论的纲领并不注定仅仅停留于计划的阶段。

注 释

① 即上节中所介绍的二元论、唯心论、唯物论、中立一元论等理论在心身问题上所碰到的各种难题。——译者。

② 对此,本书“前言”说得更加明白:“在通信理论的专门意义上,生命的根本范畴是信息。本书的任务就是表明:具有有机的和心理的功能的人是这些信息过程的一种自然产物。”正是在这个意义上,作者将自己的心灵哲学理论简称为“信息实在论”。——译者。

人名译名对照

A

Abrahamsen, A. 亚伯拉罕森
 Ach. 阿赫
 Agre, P. 阿格雷
 Akin, K. 埃金
 Akins, K. 埃金斯
 Alexander, S. 亚历山大
 Allport, A. 奥尔波特
 Anderson, J. 安德森
 Anscombe, G. E. M. 安斯康帕
 Armstrong, D. 阿姆斯特朗
 Astington, D. M. 阿斯廷顿, D. M
 Astington, J. 阿斯廷顿, J
 Attneave, F. 阿特尼夫

B

Baars, B. 巴思
 Bacon, J. 培根
 Barsalou, L. 巴萨卢
 Baker, L. C. 贝克
 Baron-Conen, S. 巴伦-科恩
 Barwise, J. 巴怀斯
 Bateson, G. 贝特森
 Bechtel, W. 贝希特尔
 Beckermann, A. 贝克尔曼
 Bennett, J. 贝内特
 Berg, G. 伯格
 Besner, D. 贝斯纳

Biro, J. 比罗
 Bisiach, E. 比撒克
 Blackburn, S. 布莱克本
 Block, Ned 布洛克, 内德
 Boghossian, P. 博格森
 Bolivia 波利维亚
 Bolton, N. 博尔顿
 Bonevac, D. 博恩瓦克
 Bower, G. 鲍尔
 Bowers, S. K. 鲍尔斯
 Boyd, R. 博伊德
 Bradley, M. 布拉德利
 Brand, M. 布兰德
 Brandon, B. 布兰登
 Broad, C. D. 布罗德
 Broadbent, D. 布罗德本特
 Brooks, R. 布鲁克斯
 Brown, S. C. 布鲁恩
 Brunner, J. 布鲁纳
 Burge, T. 伯奇

C

Campbell, k. 坎贝尔
 Capitan, W. 卡皮坦
 Capranica, R. 卡普拉尼卡
 Carp, Joe. 卡普, 乔
 Carr, M. 卡尔
 Carrier, M. 卡里尔
 Carruthers 卡鲁瑟斯
 Cartwright, B. 卡特赖特

1140 人名译名对照

Casati, R. 卡萨蒂
 Causey, R. 考西
 Cephalus 西法鲁斯
 Chalmers, D. 查默斯
 Chappell, V. C. 查普尔
 Charnick, E. 查尼克
 Cheesman, J. 奇斯曼
 Chen, K. 程
 Cherniak, C. 切尔尼克
 Chisholm, R. M. 齐硕姆
 Chrysippus 克里斯波斯
 Church, R. 丘奇
 Churchland, M. P. 丘奇兰德, M. P.
 Churchland, P. S. 丘奇兰德, P. S.
 Clark, A. 克拉克
 Cleve, Van 克利夫, 范
 Clifford, K. 克利福德
 Cohen, B. 科恩
 Collett, T. 科利特
 Collins 柯林斯
 Copp, D. 科普
 Cornman, J. 柯曼
 Coltheart, M. 科尔西特
 Cottrell, G. 科特雷尔
 Crane, T. 克兰
 Crick, F. 克里克
 Cummis, R. 卡明斯
 Cussins, A. 卡辛斯
 Cussin, A. 卡辛

D

Dahlborn, B. 达尔邦
 Damasio, A. 达马苏
 Davis, R. 戴维斯
 Day, J. 戴
 Dennett, D. 丹尼特
 Deutsch, D. 多伊奇

Devitt, M. 德维特
 Diamond, C. 戴蒙德
 Dimond, S. 迪蒙德
 Donagan, A. 多纳根
 Dray, W. 德雷
 Dretske, F. 德雷特斯克
 Dreyfus, H. L. 德雷福斯
 Drühl, K. 德鲁尔
 Dupre, J. 杜普雷

E

Earman, J. 埃尔曼
 Eckardt, B. 埃克德
 Eckhorn 埃克霍恩
 Edelman, G. 埃德尔曼
 Ellard, C. 埃拉德
 Empiricus, S. 恩披里可
 Englert, B. 恩格勒特
 Evans, G. 埃文斯
 Everglades 埃弗格勒特
 Ewart, J. 尤尔特

F

Fahlman, S. 法尔曼
 Farah, J. M. 法拉赫
 Feigl, H. 费格尔
 Feldman, J. 费尔德曼
 Field, H. 菲尔德
 Flanagan, O. 弗拉纳根
 Flohr, H. 弗洛尔
 Fodor, J. 福多
 Forbes, M. 福布斯
 Foster, L. 福斯特, L.
 Foster, J. 福斯特, J.
 Fraassen, B. 弗拉森
 French, P. 弗伦奇

Frith, V. 弗里斯

Fry, D. 弗赖伊

Fuller, G. 富勒

G

Gallistel, C. 加利斯特

Garon, J. 加龙

Gibbon, J. 吉本

Ginet, C. 吉内特

Globus, G. 格洛布

Glover, J. 格罗弗

Glymour, C. 格里莫

Godfrey-smith, P. 戈弗雷-史密斯

Goldburg, S. 戈德伯格

Goldman, A. 戈德曼

Goodman, N. 古德曼

Gopnik, A. 戈帕里克

Gorden, R. 戈登

Gould, J. 古尔德, J.

Gould, S. 古尔德, S.

Greenwood, J. 格林伍德

Gregg, L. 格雷格

Gregory, R. 格雷戈里

Grice, H. 格赖斯

Grier, J. 格里尔

Griffin, D. 格里芬

Grimes, T. 格兰姆斯

Van Gulick, R. 久里克, 范

Gunderson, K. 冈德森

Guttenplan, S. 格滕普兰

H

Hacking, I. 哈金

Haldane, E. S. 霍尔丹

Hamilton, W. 汉密尔顿

Hampshire, S. 汉普希尔

Hannan, B. 汉南

Hanson, P. 汉森

Hardin, L. C. 哈丁

Hare, R. H. 黑尔, R. H.

Hare, M. R. 黑尔, M. R.

Harman, G. 哈曼

Harnad, S. 哈纳德

Harnish 哈尼什

Harris, Z. 哈里斯

Harrison, B. 哈里森

Hart, H. L. A. 哈特

Hatfield, G. 哈特菲尔德

Haugeland, J. 豪格兰德

Hayes, P. 海斯

Heath, A. F. 希思

Healey, R. 希利

Heil, J. 海尔

Hellman, G. 赫尔曼

Hempel, C. 亨普尔

Herrnstein 赫恩斯坦

Herzberger, H. 赫茨伯格

Hilbert, D. 希尔伯特

Hill, C. S. 希尔

Hintikka, J. 辛迪卡

Hinton, G. 欣顿

Hochber, J. 霍克堡

Holland, J. 霍兰, J.

Holland, R. 霍兰, R.

Holtzman, S. H. 霍尔茨曼

Honderich, T. 洪都里奇

Honore, A. M. 奥诺雷

Hooker, J. 胡克, J.

Hooker, M. 胡克, M.

Horgan, T. 霍根

Howard Sobel, J. 霍华德·索贝尔

Hughes, G. I. 休斯

Humeroff, R. S. 哈默罗夫

Humphrey, N. 汉弗莱

1142 人名译名对照

Hyslop, A. 希思洛普

I

Israel, D. 伊斯雷尔

J

Jackendoff, R. 杰肯多福

Jackson, F. 杰克逊

Jacob, P. 雅各布

Jammer, M. 雅莫尔

Jaynes, J. 杰恩斯

Jech, T. 杰克

Johnson-Laird, P. N. 约翰逊-莱尔德

Jones, T. 琼斯

K

Kahneman, D. 卡里曼

Kaplan, D. 卡普兰

Kauffman, S. 考夫曼

Kenny, A. J. P. 肯尼

Kim, J. 金

Kinsbourne, W. 金斯伯恩

Kintsch, W. 金特施

Kirchner, W. 柯克纳

Kirk, R. 柯克

Kitcher, P. 基切尔

Kline, M. 克莱因

Koch, C. 考克

Koons 孔斯

Krantz, D. 克兰茨

Kulpe, O. 屈尔佩

Künne, W. 库内

L

Lackey, D. 莱克

Land, E. 兰德

Legendre, G. 勒让德

Lepore, E. 莱波尔

Leslie, M. A. 莱斯利

Lettvin, J. 莱特文

Levine, J. 莱文

Levinson, J. 莱文森

Lewis, D. 刘易斯

Lewontin, R. 莱旺丁

Libet, B. 李比特

Lichtenberg, R. 利希腾伯格

Lindsay, P. 林赛

Lloyd, D. 劳埃德

Lloyd-Morgan, C. 劳埃德-摩根

Loar, B. 洛尔

Locke, Don 洛克, 唐

Lockwood, M. 洛克伍德

Loewer, B. 洛伊尔

Lucas, J. 卢卡斯

Luce, R. 卢斯

Lycan, W. 利康

M

MacDonald, C. 麦克唐纳

Machamer, P. 马查默

Mackay, A. 麦凯

MacKenzie, A. W. 麦肯齐

MacIntosh, J. 麦金托什

Madell, G. 马德尔

Marbe, K. 马尔比

Marcel, A. 马塞尔

Marler, N. 马勒

Marsh, R. 马什

Matthen, M. 马瑟恩

Mauve 莫夫

McCarthy, J. 麦卡锡

McClamrock, R. 麦克兰洛克

McClelland, I. J. 麦克莱兰
 McCloskey, M. 麦克洛斯基
 McGinn, C. 麦金
 McLaughlin, B. 麦克劳格林
 Meinong 梅农
 Melden, A. I. 梅尔登
 Mele, A. 米尔
 Mellick, D. 梅利克
 Mellor, D. 梅勒
 Meltzoff, A. 梅尔佐夫
 Merikle, P. W. 梅里克尔
 Merrill 梅里尔
 Mesulum, M. M. 梅索伦
 Millar, A. 米勒, A.
 Miller, I. 米勒, I.
 Miller, G. 米勒, G.
 Millikan, D. 米利肯, D.
 Millikan, R. 米利肯, R.
 Milner, A. 米尔纳
 Minsky, M. 明斯基
 Moore, G. E. 摩尔
 Morgan, C. L. 摩根
 Morton, A. 莫顿
 Moser, K. Paul 莫泽, K. 保罗
 Mostowski, A. 莫斯托斯基
 Murdock, G. 默多克
 Murphy, G. 墨菲

N

Nagel, E. 内格尔, E.
 Nagel, T. 内格尔, T.
 NatVander, Arnold 纳特·范德尔, 阿诺德
 Neander, K. 林德尔
 Neisser, V. 尼瑟
 Nemirow, L. 尼米罗
 Newell, A. 纽厄尔
 Nichols, S. 尼科尔斯

Nisbett, R. 尼斯比特
 Norman, A. D. 诺曼
 Nozick 诺齐克

O

Olson, D. 奥尔森
 Oppenheimer, Paul 奥本海默, 保罗

P

Palmer, S. E. 帕尔默
 Papert, S. 帕波特
 Papineau, D. 帕皮诺
 Pareto, V. 帕累托
 Pargetter, R. J. 帕吉特
 Pavlov. 帕夫洛夫
 Peacocke, C. 皮科克
 Penrose, R. 彭罗斯
 Pernier, J. 佩尔内
 Perry, J. 佩里
 Pessin, A. 佩辛
 Peters, R. S. 彼德斯
 Pickett 皮克特
 Pinker, S. 平克尔
 Pitcher, G. 皮彻
 Pollack, J. 波拉克
 Pollock, J. 波洛克
 Pople, H. 波朴尔
 Post, J. 波斯特
 Price, H. H. 普赖斯
 Prince, A. 普林斯
 Proust, M. 普鲁斯特
 Pylyshy, Z. 皮利辛

Q

Quillan, R. M. 奎兰

1144 人名译名对照

Quillian, M. 奎利恩
Quire 夸尔

R

Ramachandran 拉梅钱德朗
Ramsey, W. 拉姆齐
Reeves, A. 里夫斯
Rey, G. 雷伊
Richard, M. 理查德
Richards, A. I. 理查兹
Richardson, R. 理查森
Rips, L. 里普斯
Robinson, H. 鲁滨逊
Roediger, H. L. 罗蒂杰尔
Romanes 罗马尼斯
Rorty, R. 罗蒂
Rosch, E. 罗施
Rosenbery, A. 罗森堡
Rosenblatt, F. 罗森布拉特
Rosenthal, D. 罗森塔尔
Röska-Hardy, L. 罗斯卡-哈蒂
Ross, L. 罗斯
Rumelhart, D. 拉梅尔哈特

S

Salmon, T. H. 萨蒙, T. H.
Salmon, W. 萨蒙, W.
Savage, W. 萨维奇
Savelos, E. E. 萨维洛斯
Sayer, K. 塞耶
Schacter, D. 沙克特
Schiffer, S. 希夫尔
Scriven, M. 斯克里克
Scully, M. 斯卡利
Seager, W. 西格尔
Selfridge, O. 塞尔弗里奇

Sellars, W. 塞拉斯
Seyfarth, R. 赛发思
Shaffer 谢弗
Shallice, T. 沙里斯
Shannon, E. C. 香农
Sharpe, R. 夏普
Shear, J. 谢尔
Shepard, R. 谢泼德
Sheridan, G. 谢里登
Sherman 舍曼
Shoemaker, S. 休梅克
Shrobe, H. 施罗伯
Sibley, F. 西布利
Sikora, R. I. 西科拉
Sivers, S. 西尔弗斯
Simon, H. 西蒙
Sklar, K. 斯克拉
Smith, E. 史密斯, E.
Smith, P. 史密斯, P.
Smolensky, P. 斯莫伦斯基
Sobel, J. 索贝尔
Sober, E. 索伯
Stack, M. 斯塔克
Stalnake, R. 斯托纳克
Stampe, D. 斯坦伯
Stephen, L. 斯蒂芬
Sterelny, K. 斯特尼
Stevens, S. 斯蒂文斯
Stevenson, L. 斯蒂文森
Stich, S. 斯蒂克
Storrs, M. 斯托姆斯
Strawson, P. F. 斯特劳森
Suppes, P. 萨普斯
Swanson, J. W. 斯旺森
Swartz, R. 斯沃茨
Swinburne, R. 斯温伯恩
Szolovits, R. 索洛维茨

T

Tarski, A. 塔尔斯基
 Teichman, J. 泰克曼
 Teller, P. 特勒
 Thomas, L. 托马斯
 Thompson, W. F. 汤普森
 Tienson, J. 廷森
 Tinbergen, N. 廷伯根
 Tomberlin, J. 汤柏林
 Touretsky, D. 图雷茨基
 Towne, W. 汤恩
 Trout, D. J. 特劳特
 Tversky, A. 特夫斯基
 Tye, M. 泰伊
 Tyndall, J. 廷德尔

U

Umlta, C. 乌米塔

V

Vasilive, L. L. 瓦西利夫
 Vermazen, B. 弗默真

W

Wagman, M. 韦格曼
 Wagner, S. 瓦格纳

Wallace, G. 华莱士
 Waller, T. G. 沃勒
 Walther, R. 沃尔瑟
 Walton, D. 沃尔顿
 Watt 瓦特
 Wehner, C. 维纳
 Weiskrantz, L. 韦斯克兰茨
 Wells, H. 威尔斯
 Wheeler, A. J. 惠勒
 Whitehead, A. N. 怀特海
 Whiten, A. 怀登
 Whorff 沃夫
 Wiggins, D. 威金斯
 Wilkes, K. 威尔克斯
 Williams, G. 威廉斯
 Wilson, D. T. 威尔逊
 Wimmer, H. 温默
 Wimsatt, W. 温萨特
 Winch, P. 温奇
 Woodfield, A. 伍德菲尔德
 Woodward, J. 伍德沃德
 Wright, L. 赖特

Y

Yu, Paul 俞, 保罗

Z

Zemach, E. 泽马齐
 Ziff, Paul 齐夫, 保罗