

概念分析法：一个物理主义的评估

叶 峰

摘 要 物理主义作为一种整体哲学世界观应该对哲学研究方法也有所论断。在物理主义框架下分析、评估传统哲学研究中的概念分析法，可以得出：概念分析并不是阐明已有的一个哲学概念的内部结构，而是寻找一个新的、用语言定义的等价概念；物理主义可以解释概念分析法的“先天性”，以及为什么哲学研究中使用概念分析法得出的结果经常不可靠；哲学与其他科学分支一样，都是试图在观察世界的基础上构造真实地描述世界的理论，哲学中的概念分析则与科学中的纯理论研究一样，是分析、重构已有的理论；哲学中的概念分析法有其实用上的不可替代性。这些结论也表明，物理主义世界观其实支持而且能够解释传统哲学中的一些合理的方法论实践。

关键词 物理主义 自然主义 哲学方法论 概念分析

作者叶峰，首都师范大学哲学系教授（北京 100089）。

中图分类号 B0

文献标识码 A

文章编号 0439-8041(2020)05-0005-10

导言：物理主义框架下的哲学方法论反思

哲学方法论问题，尤其是概念分析法与各种哲学直觉的方法论地位问题，是近年来受到国际分析哲学界广泛关注的-一个问题。^① 本文将在（还原的）物理主义框架下分析探讨传统哲学研究中的概念分析法的方法论地位。^②

在物理主义框架下反思哲学研究方法，最重要的是要接受当代进化论与神经认知科学对我们人类自身的本质、人类思维及认知过程的本质的论断，因为哲学研究方法是人类使用的认知方法，是人类思维与认知活动的模式。神经认知科学发现，人类的思维过程其实是脑神经元活动过程，人类的认知活动是人类大脑与环境的互动过程，而不是任何非物质的-心灵、精神或所谓“主体”所拥有的非物质的、精神性的过程。因此，反思我们的认知方法，包括哲学研究方法、其他科学研究方法及日常认知方法，本质上是描述、分析、评估人类大脑的活动及其与环境的互动。特别地，它不是去描述、分析、评估某个非物质的主体（或心灵）的所谓“内在精神活动”，或这种主体对所谓“外部世界”的认知，因为世上没有这样的主体，也没有这样的内在精神活动，或与主体相对的外部世界，有的就是大脑及其与环境的互动。而且，对各种方法进行哲学反思、分析与评估的也是一些大脑，即从事方法论反思的一些哲学家的大脑，而不是任

^① 相关综述可见 AM. A. Eder, I. Lawler & R. van Riel, “Philosophical Methods under Scrutiny: Introduction to the Special Issue *Philosophical Methods*,” *Synthese* (2018) <https://doi.org/10.1007/s11229-018-02051-2>; J. H. Cappelen, T. S. Gendler & J. Hawthorne (eds.), *The Oxford Handbook of Philosophical Methodology*, Oxford University Press, 2016.

^② 关于物理主义一般性介绍可见：D. Stoljar, “Physicalism,” In *Stanford Encyclopaedia of Philosophy*, <http://plato.stanford.edu/entries/physicalism/>, 2015。在本文中，“物理主义”都指还原的物理主义，见叶峰：《为什么相信还原的物理主义》，《学术月刊》2017年第2期。

何非物质的“主体”。也就是说，方法论反思这种思维活动自身，其实是我们大脑对大脑自身及其与环境的认知互动进行观察、分析与评估。这也是人类大脑对物质世界中的一类自然现象进行观察、分析与评估，与普通科学研究无异。这是物理主义框架下的方法论反思的本质。

这意味着一些传统哲学家试图为科学奠定基础的所谓“先验思辨”没有意义，因为先验思辨预设了先验主体、先验思维过程等在物理主义看来子虚乌有的东西。也就是说，传统的“先验思辨”其实并不是没有任何预设的、“零基础”的哲学思辨。不幸的恰恰是，“先验思辨”预设了其实不存在的、大脑幻想出来的东西，即所谓先验主体、先验思维过程等。其实是一些哲学家的大脑在那里做“先验思辨”，“先验思辨”这种思维活动其实是那些哲学家的大脑中的神经元活动。

物理主义框架下的哲学探索会得出许多与传统哲学不同的结论，笔者此前的研究工作一直是围绕这个主题。就方法论来说，笔者在此前的一些论著中曾尝试论证^①，传统哲学研究所依赖的一些直觉，包括心理直觉、语义直觉、知识论直觉、形而上学直觉、虚拟条件句直觉及心智模拟（mental simulation）等^②，其实是基于大脑神经网络由进化及后天训练产生的一些复杂的模式识别能力。大脑神经网络的这些模式识别能力比大脑的普通视觉及其他知觉模式识别能力要更复杂一些，其识别结果的稳定性更低一些，但就识别我们哲学研究所关心的那些事物及属性（如心理状态、指称关系、知识、因果关系等）来说，这些复杂模式识别能力与普通的知觉能力并无实质性的差异，是相连续的。知觉能力当然是科学研究中不可缺的，大脑使用知觉能力观察环境是科学研究方法中最基本的一项。至于大脑的一项复杂模式识别能力，只要它能够足够稳定地识别环境中的事物，将事物分类，而且只要大脑在经过足够的后天训练后能够做到这一点，那种复杂模式识别能力就应该算是得到知识论辩护的（epistemically justified）获取真知识的方法。进而，基于这些复杂模式识别能力的一些直觉，就可以是得到知识论辩护的获取真知识的方法。

例如，人类大脑在进化中产生了通过观察他人的面部表情、声音及姿态特征来推测他人是喜悦还是愤怒的能力。这一般被称作一种心理直觉能力。这种能力的存在有合理的进化论解释，它是人类祖先能够组成社会群体的必要前提。这种直觉能力的执行其实包括两个步骤：第一步是使用大脑神经网络的一项复杂模式识别能力，它能识别大脑所观察到的他人的面部表情、声音及姿态等的一个综合特征，依这个综合特征将人的外表分类，即分为“喜悦的外表”和“愤怒的外表”等类别；第二步是一个最佳解释推理（inference to the best explanation），即由一个人具有“喜悦的外表”推测他的大脑内部状态是“喜悦的状态”。这里的第一步就是一项复杂模式识别能力的应用。人类祖先能够进化出这种心理直觉能力，以及能够成功地组成社会群体这一事实，保证了在正常环境下这种心理直觉能力是可靠的。因此，使用上述复杂模式识别能力及最佳解释推理的这个心理直觉是可靠的，是得到知识论辩护的获取真知识的方法。

另一方面，笔者在前述论著中还提出，这个复杂模式识别能力的使用无法实际地还原为所谓硬科学方法（hard-scientific methods）的使用，后者指的是用明确的语言尤其是数学语言，对所观察到的事物特征进行语言描述及数学建模，然后通过明确的逻辑推理与数学计算来得出对事物的判断（比如判断一个外表是属于“喜悦的外表”）。这里的不可还原性是因为，大脑识别“喜悦的外表”这项复杂模式识别能力的使用，可能牵涉大脑中数百亿个脑神经元，它是数亿年进化的产物，是极为复杂的神经元计算。我们无法实际地将这种复杂的神经元计算还原成语言中明确的逻辑推理与数学计算。这里的不可还原性，其实就是近年来著名的人工智能系统 AlphaGo 的下棋能力无法还原成基于明确的下棋规则的逻辑推理与数学计算那种不可还原性，是复杂的神经网络计算不可还原为基于规则的推理的不可还原性。注意，这里说的不可还原性都是指实际上不可还原，而不是指原则上不可还原。“原则上可还原”这个概念对于表达我们的本体论立场有意义，但讨论方法论问题时所关心的方法论还原应该都是指实际可行、可操作的还原。

所以，综合起来看，物理主义世界观完全可以容纳不可还原为硬科学方法的各种直觉，承认这些直觉

① 参见“On Extreme versus Moderate Methodological Naturalism,” *Philosophia* 45(2017), pp. 371-385.

② 后者主要指 T. Williamson 在 *The Philosophy of Philosophy* (Blackwell Publishing, 2007) 中所论述的，用于哲学研究中的思想实验的方法。

不蕴涵承认任何非物理的事物或过程，或任何其他神秘的事物。物理主义其实支持的是温和的而非极端的方法论自然主义（后者有时又被称作“科学主义”），物理主义框架其实能够为常识及传统哲学中某些看似反对科学主义但合理的方法论实践作出辩护，并提出完全符合物理主义的解释。^①

此前的这些探讨主要针对直觉，本文的目的是继续这个方向的研究，尝试在物理主义框架下分析传统哲学中常见的概念分析法。时常听人说，分析哲学研究主要是从事概念分析，而科学是要真实地描述物质世界，因此两者不同。为了讨论方便，这里称这种观点为“概念分析说”。与之相对的是“描述世界说”，它认为，分析哲学研究与普通科学研究无异，都是试图在观察世界的基础上，构造能够概括世界中的现象、描述事物的内在结构或规律的理论。本文的目的是为“描述世界说”作辩护，而且强调“描述世界”指的是描述物质世界。为此本文将澄清四个方面的问题：（1）所谓概念分析究竟是什么，究竟在做什么？（2）概念分析法是否如传统哲学所认为的，是先天的研究方法？如果是，那为什么哲家用概念分析法又总是得出有争议的甚至错误的结论？（3）概念分析法与普通科学方法有哪些异同点？哲学家是否也跟科学家一样，是在观察物质世界的基础上构造能够真实地描述物质世界的理论？如果是，那么为什么哲学研究看起来很不同于其他科学研究？（4）哲学中的概念分析法是否有其独特的、不可替代的地位？下面将先做一些预备性的分析，然后分别讨论这四个方面的问题。

一、分析概念抑或描述世界

首先，各个学科都要分析自己学科内部的专业概念，而且也只有那些学科的专家才有能力去分析那些专业概念，哲学家显然没有那样的能力。例如，物理学要分析“时空”“能量”“热量”“熵”等这些物理概念，而且这些概念分析是物理学家的的工作，不是哲学家的任务。哲学家当然也分析哲学中的概念，但如果说物理学家分析物理学概念，哲学家分析哲学概念，那么“从事概念分析”这个描述并没有将哲学与其他学科区分开。

物理学研究当然不仅仅是分析物理概念。就概念而言，物理学家所做的，其实是基于观察和物理理论的构造，去发明（或发现）能够恰当地描述各种物理对象及过程的物理概念。他们的主要目的不是分析某些已有的概念本身。如果他们发现一个现有的概念不能恰当地描述相关的物理对象，他们就会建议放弃或修改那个概念，比如，爱因斯坦建议修改牛顿的时空概念，又比如，现代物理学家完全放弃了“不可分的原子”这个古老的概念。分析那些无用的、过时的、不能恰当地描述世界的概念，如古代人的“原子”概念，对于专业科学研究来说没有意义（虽然对于思想史研究来说很有意义）。而要寻找恰当的、可以更好地描述世界的概念，显然需要观察世界及构造描述世界的理论。因此，科学研究主要不是做概念分析。

那么，有没有可能哲学家仅仅就是做概念分析，即分析现成的、人们已有的一些哲学概念，而不关心那些概念是否能够恰当地描述世界中相关的对象或过程，因此哲学家不必像科学家那样去观察世界、构造描述世界的理论？一个相关的问题是，哲学研究中是否也会出现上面所说的那种情况，即一个概念被发现不能恰当地描述世界中的事物或过程，因此被哲学家建议放弃或修改？如果有，那么哲学家就同样不仅仅是在分析那个概念本身，而且也关心能否用那个概念恰当地描述世界，后者显然也需要基于观察世界以及构造描述世界的理论。

我们看当代知识论研究中的一个典型例子。按照一种理解，葛梯尔所提出的，对知识的“JTB”（Justified True Belief）定义的反例是说明，“JTB”不是对“知识”这个概念的正确概念分析，而后来知识论专家的工作，是试图提出对“知识”这个概念更正确的概念分析。但有另一种理解，即葛梯尔的

^① 详见前面所引的作者的论著。参见 T. Williamson 与 A. Rosenberg 围绕温和与极端方法论自然主义的争论：A. Rosenberg, “Why I am a naturalist” 以及 “Can naturalism save the humanities?” 载于 M. C. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?* Routledge, 2014, pp. 32-35, pp. 39-42; T. Williamson, “What is naturalism?” 及 “The unclarity of naturalism” 载于 M. C. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?* Routledge, 2014, pp. 29-31, pp. 36-38.

反例说明的是,“JTB”没有正确地描述人类认知活动中的知识这一类对象,“JTB”就像一个被发现了反例因而被证明为错误的科学理论。后来的知识论专家的工作则是试图提出新的、更正确的理论来描述知识这一类对象。这正是普通科学家每天在做的工作。

后一种理解支持描述世界说,但既然两种理解似乎都说得通,有的人可能会坚持说,知识论专家做的仅仅就是概念分析,即分析“知识”这个概念。支持这个概念分析说的主要理由似乎是这样的,哲学家也经常用例子证明一个理论是错的,比如葛梯尔用例子证明知识的“JTB”理论是错的。这看起来类似于科学家用实验证明一个理论是错的。但是支持概念分析说的人提出,哲学家用来反驳一个理论的反例都是一些思想实验,哲学家不是真的去观察世界或动手做实验。思想实验正是用来分析概念的,借助于思想实验我们可以判断对一个概念的分析是否正确,而不必真的去观察世界或动手做实验。物理学中的思想实验也是用来分析物理概念,但要检验一个物理假说就必须观察世界或真的动手做实验。用思想实验判定一个概念分析的对错是先天的工作,从认识论的角度看与动手做实验来检验一个科学假说有本质不同。所以,哲学研究仅仅就是做概念分析,包括借助于思想实验做概念分析,是先天的研究,与科学研究不同。

已经有一些哲学家对这个理由及传统概念分析说提出了一些反驳,特别地,如威廉姆斯(T. Williamson)认为,虽然思想实验及概念分析等是合理的获取知识的方法,但它们并非传统所理解的所谓先天方法。^① 本文的结论与一些哲学家已有的观点相近,但本文的主要目的是在彻底物理主义的框架下澄清相关问题,包括回应上述支持概念分析说的理由。笔者希望能够展示,物理主义框架可以更好、更彻底地清除相关的迷雾。而且笔者的结论也与前人的结论有细微差异。例如,笔者完全支持描述世界说,认为哲学与科学在本质上(即在本体论、语义学与认识论方面)完全一样,都是在描述物质世界,而威廉姆斯等反对极端自然主义的哲学家似乎都不会这样将哲学与科学完全等同。另一方面,笔者又承认哲学研究中的直觉与概念分析法的合理性、不可还原性与不可替代性,但认为这些与最极端的物理主义立场并不冲突,相反,它们可以在物理主义框架下得到解释,这与威廉姆斯所反对的极端方法论自然主义者的观点又有所不同。

在物理主义这个前提下应该不难理解“哲学与科学一样都是要描述物质世界”这一论断。因为,给定物理主义,存在着的一切就是物质世界中的事物及过程,包括哲学研究所关心的概念、思想、信念、知识、指称关系等,都是物质世界中的事物及过程。物质世界中的事物可以在不同的层次上被描述,比如,对人类大脑的活动,可以有物理学、化学、生物学、认知科学等不同层次的描述。哲学对概念、知识、指称关系等的描述应该与认知科学属于相同的层次。所以,哲学与认知科学乃至其他科学学科不应该有什么本体论、语义学及认识论上本质的差异。另一方面,给定物理主义,不存在非物质的抽象对象或非物质的心理实体,也不存在非物质的主体及主体的所谓内在世界。在过去,哲学被理解为本质上不同于科学,正是因为哲学家们认为哲学是要研究那些非物质的、作为抽象对象或心理实体的概念、思想等。物理主义认为这些都是大脑的幻想,即人们的大脑幻想这个大脑中藏着一个非物质的主体,这个主体内有一个内在世界,其中有非物质的、属于主体的概念、思想等,或者幻想有一个独立于物质世界的抽象世界(比如弗雷格的所谓第三域),其中存在着作为抽象对象的概念、思想等。(注意,大脑的幻想活动本身是一些神经元活动,幻想过程中真实存在着的就是神经元活动过程,而没有所谓非物质的“幻想的产物”。)既然存在着的一切就是物质世界中的事物及过程,一切有真假的陈述包括哲学中有真假的陈述,都必须是对物质世界中的事物及过程的描述,那么哲学与描述物质世界的科学就不应该有什么本质差异。

这是物理主义的一般性推论。但具体到概念分析法,我们还需要澄清一些更具体的、可能令人困惑的问题,即上一小节末尾所提的那些问题,还要回应上述支持概念分析说的理由。这正是本文的任务。

^① 参见 T. Williamson: *The Philosophy of Philosophy*, Blackwell Publishing, 2007 以及 “How Deep is the Distinction between *A Priori* and *A Posteriori* Knowledge?” In Albert Casullo & Joshua C. Thurow (eds.), *The A Priori in Philosophy*, Oxford University Press, pp. 291-312.

二、概念分析究竟是在做什么

我们先追问，当哲学家在分析一个哲学概念的时候，他们究竟是在做什么？是在分析哪个概念？对那个概念说了什么？

还是以知识论专家对葛梯尔问题的研究为例。假设一个哲学家 A 提出了“JTB+X”作为对知识的“JTB”定义的改进，以回避葛梯尔式的反例。然后，哲学家 B 举出了一个新的反例来反驳 A 的“JTB+X”理论。假设 A 接受了对方的反驳，因此承认“JTB+X”不是成功的知识定义。A、B 两人在讨论中都用到了“知识”这个词，而且他们对什么算是知识似乎有共识，比如，对 B 所提出的反例有共识。“知识”这个词在 A、B 两人的大脑中分别表达了存在于那两个大脑中的两个“知识”概念。^①既然两人对什么算是知识有共识，我们不妨假设存在于这两个大脑中的两个“知识”概念基本上是一致的，我们将它们都记为 K。这样，如果说 A 提出的“JTB+X”理论是对“知识”这个概念的分析，那么它应该是对 K 的分析，它就是在说，概念 K 的结构是“JTB+X”，即“K=JTB+X”。换句话说，依这种理解，这个理论其实是在说，“JTB+X”真实存在于 A、B 两人的大脑中，而且在两个大脑中都与“知识”这个词相关联，被这个词所表达。这里“JTB+X”应该是指称一个具有某种结构的概念，它的结构成分包括语词“Justified”“True”“Belief”所表达的三个概念，以及一个很复杂的语言描述 X。但显然，B 在反思自己大脑中与“知识”这个词相关联的“知识”概念 K 时，没有“看到”“JTB+X”，尤其是没有“看到”这个很复杂的描述 X，否则他一开始也会同意 A 的理论。所以，依这种理解，A 提出的理论其实是要说，虽然人们的脑中都有 K，并都被词语“知识”表达，而且在将 K 应用于具体事例的时候都有共识，但 K 的内在结构“JTB+X”是隐蔽的，不是人们轻易“可见”的。这也就是说，当人们将 K 应用于一个具体事例的时候，他们是无意识地判断那个事例是否满足条件“JTB+X”，因为他们不是有意识地认识到自己脑中的 K 就是“JTB+X”。

暂且不管有没有针对 A 提出的“JTB+X”理论的反例，这样理解“JTB+X”理论似乎可信度很低。“JTB+X”中的 X 是一个很复杂、很长、很特异的日常语言语句。如果它真的原本保存在哲学家 B 的大脑中的“知识”概念 K 中，很难想象为什么当 B 反思他的 K 时不能发现它。也许只有通过明确的语言定义来学习一个概念时，我们大脑中的那个习得的概念才是这样的用语言表达的条件。比如，我通过教科书中的定义学到了“质数”这个概念，因此当我反思我的“质数”概念时，我的确“看到”的是那个语言定义。但显然没有人是通过定义“知识就是‘JTB+X’”学习了“知识”概念 K。很难想象大家大脑中的“知识”概念 K 中竟隐藏着 X 这么一个非常复杂的、大家想不到的语言描述。

物理主义框架下的概念理论，以及前面提到的基于脑神经网络的复杂模式识别能力的直觉，可以对 K 与“JTB+X”的关系提出一个更好的解释。依笔者提出的物理主义框架下的概念理论，A、B 或其他普通人脑中的概念 K 中，除了可能包含“Justified”“True”“Belief”这些词所表达的概念（作为概念 K 的成分）之外，还包含的是对一些知识的例子的知觉记忆。比如，A 对概念 K 的学习与使用过程是这样的：在过去的自然语言学习及知识论学习过程中，当 A 将“知识”这个词在自己脑中所表达的概念 K 应用于一个所观察到的例子时，他得到关于那个例子的一个知觉印象，并将这个知觉印象保存在记忆中，作为概念 K 的一个新的成分；在最开始的学习中，A 是模仿别人将概念 K 用于一些例子，并保存对那些例子的知觉记忆于自己习得的概念 K 中；然后，在学习充分多的例子之后，当 A 自己决定是否可以将他的概念 K 用于一个新的例子时，A 大脑中的一个复杂模式识别功能将新的例子与保存于概念 K 中作为成分的知觉记忆相匹配，判断它们是否在相关的方面相似；如果匹配成功，就判断概念 K 可用于那个新的例子。有可能 A 在对概念 K 的学习过程中明确地将“Justified”“True”“Belief”这些词所表达的概念搜拢了作为概

^① 注意，这里将概念理解为大脑中的事物，不考虑概念是独立于大脑的、抽象的、公共的实体那种观点，笔者认为后者与物理主义相冲突。

念 K 的成分。那么,当他判断一个新的例子是否知识的时候,他会先检验这些词所表达的概念是否可用于那个新例子。所以,概念 K 中除了可能包含概念“Justified”“True”“Belief”之外,还包含的是一些知觉记忆,而不是 X 那样很复杂、特异的自然语言描述。K 与“JTB+X”是结构成分不同的两个概念。

然后我们可以理解 B 举出的反例(这里记为 E)意味着什么。A 和 B 都发现,概念“JTB+X”可以用于那个反例 E,即 E 满足用自然语言表达的条件“JTB+X”。另一方面,A 和 B 脑中的两个概念 K 很相似,即它们保存的知觉记忆在相关的方面很相似,因此两人一致地判断,例子 E 与他们的概念 K 中保存的知觉记忆不在相关的方面相似,因此两人都同意概念 K 不能用于例子 E。最后两人都同意,概念“JTB+X”与他们脑中已有的概念 K 不表征相同的事物。

这里我们看到,A 提出的“JTB+X”理论不是说 K 就是“JTB+X”,或 K 的内部结构是“JTB+X”。“JTB+X”是另外一个概念,是用自然语言表达的一个概念,它的成分包括那个很复杂、很特异的自然语言描述 X。与之相对,K 是用一个简单名词“知识”表达的概念,它的成分主要是一些知觉记忆(可能还有“Justified”“True”“Belief”等概念)。所以,A 提出“JTB+X”理论,实际上是提出了一个用自然语言定义的新概念“JTB+X”,然后断言,这个新概念“JTB+X”和已有的、主要基于知觉记忆的概念 K 是在某种意义上“等价的”。这里的“等价”关系的最低要求是“JTB+X”和 K 表征了相同的对象。B 举出的反例正是说明“JTB+X”理论满足不了这个最低要求。可以认为,A 提出语言描述 X 是要用语言来概括保存在 K 中的那些知觉记忆的一些共同特征,而 B 提出的反例说明这个概括没有成功。这正说明了“JTB+X”与 K 是不同的概念。

所以我们看到,所谓概念分析并不是揭示一个已有概念的真实内在结构,而是试图提出一个新的、用语言定义的、与已有概念等价的概念。等价性至少要求新的概念与已有的概念表征相同的事物。但一般我们做概念分析的时候,我们会期待比这个最低要求更多的东西,即期待这个等价是某种意义上“先天的”。这将在下一小节用一个简单的例子来说明。注意这里用了加引号的“先天”,因为在物理主义框架下“先天”的含义与传统哲学有所不同。

三、概念分析的“先天性”与可靠性问题

我们知道,鲁迅的侧面像有一些很独特的特征,它使得人们可以相当一致地判断一个侧面像是否看起来像鲁迅。设想我看过一些鲁迅的侧面像,记住了那些视觉图像,而且以这些视觉记忆为成分形成了一个概念,这里记为 LXimg,它表征那些看起来像鲁迅的侧面像的图画。我没有学习过美术,也没有专门研究过鲁迅的侧面像的特征,我也不知道鲁迅是谁,因此我大脑中的这个概念 LXimg 没有包含任何对鲁迅侧面像特征的语言描述,它只包含那些视觉记忆。然后设想一个仔细分析过鲁迅的侧面像特征的美术教师,设想他用很长的一段文字 LXdesc 非常仔细地描述了鲁迅的侧面像特征。这里的 LXimg 类似于前面的 K, LXdesc 则类似于前面的“JTB+X”。那位美术教师脑中的一个概念与我脑中的 LXimg 很相似,都是仅仅包含对一些侧面像的视觉记忆,这里都记为 LXimg。同时他的脑中还有另一个用很复杂的语言描述表达的概念 LXdesc,而我没有。

我们期待 LXimg 和 LXdesc 表征同样的对象,即看起来像鲁迅的那些侧面像。我们设想 LXdesc 完全是那位美术教师根据自己记忆中的鲁迅侧面像写出的描述,而且设想他描述得非常完备、准确,因此他没有也不必进一步观察世界就能合理地判断,他自己的 LXimg 和 LXdesc 表征完全相同的对象。在这个意义上,他的判断 LXimg = LXdesc 可以说是概念分析的结果,而且这种概念分析活动是在物理主义的意义下“先天的”,即他仅仅是根据反思、分析自己脑中已有的概念,特别是反思、分析保存在 LXimg 中的视觉记忆,来得出这个判断,他不必再进一步观察世界就可以确认这个判断是真的。这里我们看到,概念分析的结果得到的不仅仅是表征对象相同的一个新概念,它是仅仅根据分析、反思保存在自己大脑中的一个概念中的成分就能得到的结果。因此,在理想情况下,即假设这个概念分析是完备的,两个概念之间,比如这里的 LXimg 和 LXdesc 之间,就有更密切的联系,可说是“先天的”等价关系。

类似地，我们设想，哲学家 A 是根据反思、分析自己脑中保存在概念 K 中作为概念成分的知觉记忆，即对一些知识的例子的记忆，试图用语言概括它们之间的共同点，然后尝试将共同点表达为“JTB+X”。但保存在概念 K 中的知觉记忆，要比保存在 LXimg 中的那些对简单图画的视觉记忆复杂得多。我们几乎不可能用语言完备地概括记忆中那些知识的例子之间的共同点，“JTB+X”事实上也没能做到这一点。这是 K 与“JTB+X”这个例子不同于 LXimg 与 LXdesc 这个例子的地方。

由这些分析我们可以得出两个结论。首先，当哲学家 A 提出新的知识定义“JTB+X”的时候，他的这项工作确实可以称为“纯粹的概念分析”，因为提出新概念是基于反思分析已有的“知识”概念中保存的知觉记忆，而不必进一步观察世界。而且在这个意义上，这一项概念分析工作可以说是（在物理主义意义上）“先天的”。换句话说，这种概念分析工作与科学家得出“ $\text{水}=\text{H}_2\text{O}$ ”那一类研究是有所不同的。人们常识中的“水”这个概念包含了一些对水的外表的描述，以及一些对水的知觉印象的记忆。但科学家得出“ $\text{水}=\text{H}_2\text{O}$ ”，不是仅仅靠分析自己常识中的“水”这个概念的这些成分，而是需要靠这些概念成分先确定水的例子，然后对水做实验，分析水的微观化学结构，才得出“ $\text{水}=\text{H}_2\text{O}$ ”。所以科学家得出“ $\text{水}=\text{H}_2\text{O}$ ”的那一类研究是明显的后天的研究，概念“水”与概念“ H_2O ”之间的关系是后天的等价关系。与之相对，如果 A 提出的“JTB+X”理论成功的话，概念“知识”与新的概念“JTB+X”则应该是有更紧密的“先天等价”关系。

但另一方面，这种对 K 的“先天的”概念分析的结果并不是很可靠的。这是因为大脑的复杂模式识别能力中的神经元计算过程极为复杂，虽然作为进化的结果，不同大脑之间相关的复杂模式识别能力可以相当一致地得到相同的计算结果（即对什么是知识做出相当一致的判断），但我们无法用语言明确地重构这种计算，也无法用语言总结概括大脑的复杂模式识别能力能够识别的那些特征。也就是说，我们无法用明确的语言准确完备地概括那些知识的例子的共同特征（虽然我们脑中神经网络的复杂模式识别能力可以做到这一点）。当哲学家们在这种概念分析的努力中尝试这种概括的时候，他们经常只能是以猜测的方式进行，而且一般不得不进行大量的简化。这就使得他们尝试概括的结果，比如这里 A 提出的“JTB+X”，常常并不与原来的概念 K 表征相同的事物。所以这种纯概念分析的、“先天的”工作，经常并不能可靠地得到真理。

我们看到，在物理主义框架下，概念分析依旧具有传统哲学所认为的某些特征，即某种意义上的“先天性”，但物理主义框架可以解释为什么哲学家所谓的“先天概念分析”工作结果常常出错，并不可靠。

四、概念分析与理论科学研究之异同

既然概念分析是“先天的”，那么这是不是意味着哲学家的“先天”概念分析方法与普通科学方法有很大的区别，因而哲学研究本质上还是不同于科学研究？这是不是意味着哲学家仅仅是在分析哲学概念，仅仅做“先天”的研究，而不描述世界？我的回答是否定的。

支持概念分析说的人经常说，哲学研究工作是“坐在扶手椅里的研究”（armchair researches），哲学家不必观察世界或做实验。这似乎证明哲学家不是在描述世界。但这是对哲学家工作完全错误的描述。坐在扶手椅里做研究的哲学家并非没有观察世界，相反，他们坐在扶手椅里做概念分析之前，在他们从儿童时期开始的学习和成长过程中，已经大量地观察了世界，积累了大量的观察结果，并且已经基于这些观察构造了一些适于描述世界的、基于直觉的概念和理论。

例如，他们的脑中已经创造了“知识”这个概念，并能用这个概念将世界中的事物分类。这是基于他们大量地观察了人类包括他们自身的认知活动，用他们大脑神经网络的复杂模式识别能力识别出了人类认知活动中一些非常复杂的特征。而且他们不断地调整他们的识别知识这一模式的复杂模式识别能力，最后才达到他们目前具有的“知识”这个概念，以及他们目前具有的识别知识这一类事物的能力，如他们识别一个葛梯尔反例是否为真知识的能力。而且，他们已经用“知识”“信念”“辩护”“真”等概念构造了一个简单的知识论理论，它能够将人类认知活动及其结果分类，陈述这些类别之间的关系，包括“知识≠

信念 \cap 辩护 \cap 真”这个关系，而且能用这些类别解释、预测人的行为。比如，通过判断一个人相信什么但并不真的知道它，来预测他的行为有怎样可能的后果。这个理论与普通科学理论一样，是他们概括总结他们过去对环境及自身的观察而得出的。

一些坚持概念分析说的人可能会说，习得一个概念所需要的对世界的观察是所谓“习得性经验”（enabling experiences），这种经验不同于我们验证一个判断所需的“证据性经验”（evidential experiences）。但是，当我们判断一个例子 E 不是知识的时候，我们是将 E 与保存在我们的“知识”概念 K 中的知觉记忆相比较。这意味着，这些知觉记忆其实是作为我们判断“E 不是知识”的证据，是证据性经验。我们的判断其实是：知识是这些（指保存在 K 中的知觉记忆），而 E 与这些不相似，所以 E 不是知识。这个判断是以我们之前对知识的观察为证据的，虽然这些证据是保存在“知识”概念 K 中，成为了 K 的成分。^①

哲学家用“知识”等概念构造的描述世界的理论，与普通的物理、化学等科学理论的真正不同之处在于，哲学家构成理论的“知识”等基本概念不是用语言描述定义的概念，而是由知觉记忆及神经网络的复杂模式识别功能构成的概念。我们可以称这些概念为“基于直觉的”概念，称这些概念所构成的理论为“基于直觉的”理论。注意，这并不是说他们的理论完全不用日常语言。这只是强调，在理论的核心概念中，起到最核心作用的是一些知觉记忆，以及基于这些知觉记忆的直觉，即复杂模式识别能力，而不是一些语言描述。但这些概念同样能够在一定程度上正确地描述世界，包括解释、预测世界中的一些事件。

哲学家的语言哲学、形而上学探究显然也是一样的。坐在扶手椅里做概念分析之前，哲学家已经在过去的成长过程中积累了超大量的对人类如何使用语言、如何区分个体物体、如何判断因果关系等的观察结果，并且已经在这些观察结果的基础上，依靠长期进化形成的大脑的复杂模式识别能力，构造了各种相关的基于直觉的概念及理论，来识别、描述、解释、预测相关的现象和规律。例如，当一个哲学家坐在扶手椅上提出一个思想实验，并判断其中的一个关系不是因果关系的时候，他依据的是过去对大量的因果关系和非因果关系的例子的观察，以及在这些观察的基础上发展起的识别因果关系的复杂模式识别能力。他的判断是以大量的观察为基础的。如果没有这些大量观察，他不会有“因果关系”这个概念，也无法判断那个思想实验中的例子是否因果关系。过去的观察是他判断的依据。

所以，哲学家的概念分析更像是理论科学家的工作，是对已有的观察结果及已有的理论进行分析、整理与重构。例如，当哲学家 A 提出他的“JTB+X”理论时，他是尝试构造一个用语言表达的概念“JTB+X”来替代他已有的、基于知觉记忆和复杂模式识别能力的概念“知识”。这可以跟物理学家的工作做类比：在牛顿提出了牛顿力学以后，拉格朗日重新表述牛顿力学，提出分析力学；这种研究同样是对已有理论的分析与重构，是基于此前的观察对牛顿力学的支持，即拉格朗日不必再直接观察世界或做实验来支持分析力学原理。这种理论研究同样是坐在扶手椅里做的研究。理论物理学家的大量研究工作都是坐在扶手椅里做的。

从方法论的角度来说，知识论专家与拉格朗日的工作之间真正区别在于，知识论专家所想重构的先前的理论，是由“知识”这种基于直觉的（即基于大脑的知觉记忆和复杂模式识别功能的）概念构成的，不是用明确的数学语言表达的理论。如前所述，这使得知识论专家的重构工作更为复杂，其结果也更不确定。一个重构的尝试（比如“JTB+X”）是否成功，即所提出的（用复杂语言描述表达的）新概念、新理论是否与原有的、基于直觉的概念和理论等价，他自己也不确定，直到别人提出一个反例后才意识到他的重构并不成功。与之相对，只要一个物理学家对先前的一个物理学理论的重构工作逻辑上足够严谨细致，就可以保证新理论与原理论等价。

有人可能会提出，虽然理论科学家自己做研究的时候是坐在扶手椅上做“先天的”研究，但最终他得到的新理论和旧理论一样要受对世界的观察的检验，而且如果检验失败，理论科学家就可能放弃那些新概念、新理论。那么，会不会那些传统哲学概念、哲学理论，虽然也是基于观察被构造出的，却永远不会被

^① 参见前引 T. Williamson, “How Deep is the Distinction between *A Priori* and *A Posteriori* Knowledge?” 中的类似讨论。

放弃，因此，哲学研究仅仅是分析那些概念，而不必检验那些概念是否适于描述世界，或检验基于那些概念的理论是否真实地描述了世界？

回答显然是否定的。其实我们已经放弃了一些哲学概念，认为它们不能恰当地描述世界，比如“灵魂”这个概念，又比如古代形而上学中的“不可分的原子”概念。^①而且事实上笔者认为，物理主义哲学研究中最有挑战性也最有意义的一个任务，就是大规模地修正一些重要的传统哲学概念，以适应我们对世界的新的科学知识。比如“先天性”这个概念，传统哲学所理解的“先天性”概念预设了一个关于认知主体是什么的观念。在传统哲学中，认知主体究竟是什么这一点可能不是很明确，也许是笛卡尔的灵魂，莱布尼茨的单子，或康德的所谓先验主体，也许是其他什么。但无论如何，传统哲学所理解的“主体”不是一个由生物进化与个体发育形成的大脑（与身体），不是这样一个自然事物、一个自然生物-物理系统。那么，对于一个大脑，“先天性”这个概念还有没有意义？我们在物理主义哲学世界观中还有没有必要保留这个概念？如果有必要保留，需要对它做怎样的修正？这些问题与物理学家所考虑的，在迈克尔逊-莫雷实验之后还应不应该保留牛顿的时空概念这个问题，其实是同类的问题，要依靠对世界更深入的观察才能得到答案。

还可以将传统哲学中基于直觉的一些观察与认识跟常识物理学做类比，以此看出用“知识”等概念构成的哲学理论，其实也是基于观察世界构造的描述世界的理论，也需要基于对世界更深入的观察来修正。例如，常识物理学判断说，更重的物体一般比更轻的物体下落得更快一些。从现代物理学我们知道，在真空中这其实不对。但作为对日常环境中事物的观察与认识，这个判断是可靠的。在物理学中我们是用空气阻力来解释这一现象。常识物理学对于我们生存环境中许多宏观现象的规律性的观察是非常可靠、稳定的，几乎不可错的。这是大脑长期适应环境的结果，甚至可以说是人类大脑得以在进化中产生出来的先决条件。但常识物理学对于这些宏观现象背后的微观机制，对更为遥远的宏观事物，对事物的长期规律等的推测，则经常是错误的。这在进化论框架下也容易理解，因为进化压力只能选择大脑对临近的、与生存密切相关的环境特征的适应性。

传统哲学中基于直觉的概念及理论与此很相似。作为对人类思维及认知活动的一些表面现象及其规律性的观察，传统哲学理论中的一些论断也是很稳定、可靠的，几乎不可错的，因为那也是人类高级智能能够进化出来的先决条件。例如，我们用基于直觉的“信念”“知识”“指称”“真”等概念将人类认知活动分类，这种分类是稳定可靠的，而且对于解释、预测人的行为是有效的。而另一方面，对人类思维及认知活动的深层机制及规律，传统哲学的思辨也是很无知，经常只是提出一些虚幻的想象，比如笛卡尔对没有广延但能够思想的灵魂的幻想。神经认知科学才是探索这些深层机制的更为可靠的理论。但无论如何，传统哲学中基于直觉的概念及理论是基于观察世界构造的描述世界的理论，而且需要在对世界更深入的观察的基础上得到修正，在这方面它与常识物理学并无本质差异。

所以，哲学家做概念分析时仅仅是看起来没有观察、描述世界。他们其实已经具备了对世界的大量观察，包括作为这些观察的结果的、基于直觉的概念和理论，就像常识物理学。他们的概念分析工作是试图用语言重构这些基于直觉的概念和理论，就像理论科学家的理论研究。他们同样需要在更深入观察的基础上修改或甚至放弃某些已有的概念。

五、概念分析法之不可替代性

传统哲学理论与常识物理学之间的主要区别在于，常识物理学已经完全可以被科学物理学替代，但基于直觉的哲学概念所要分类概括的那些事物及其规律性极为复杂，是人类认知活动这种高层现象及其规律

^① 的确还有一些当代分析的形而上学家不想放弃“不可分的原子”概念，但作为一个充分尊重主流科学的结论的人，笔者认为那些形而上学家的研究更像是写科幻小说，而不是严谨地求真。

性,神经认知科学目前还远远不能充分地描述这些高层现象。所以,我们今天描述、解释人类认知活动的时候,还不得不大量借助于那些基于直觉的概念。常识心理学在今天的生活中还起着重要的作用。同样,今天科学哲学家描述、解释科学家们的科学认知活动的时候,也大量使用“知识”“指称”等这些基于直觉的传统哲学概念。而且,哲学家与认知科学家们在描述分析人类认知活动时的关注点有所不同。认知科学家对于我们已有的、基于直觉的“知识”“指称”等这些概念究竟意味着什么这个问题不那么感兴趣,他们习惯于将我们常识中已有的这些非常复杂的概念放在一边,另起炉灶用认知科学的语言直接为大脑的认知活动建模。哲学家则对于用语言描述来重构常识中已有的基于直觉的概念特别感兴趣。这种重构工作如果成功(或部分成功),显然也有助于我们更清晰地认识世界。因此,哲学家的概念分析工作同样是对理解世界的贡献。

然后,考虑到大脑经数亿年进化产生的复杂模式识别能力的有效性和难以替代性,哲学研究中的概念分析这种方法可能也是难以替代的。这并不是否认神经认知科学的进展会对哲学家的工作有很大的影响。这只是说,因为大脑的基于神经网络的复杂模式识别能力是如此之复杂,神经认知科学的未来进展可能也无法完全地用神经科学的语言描述它们,因此一个大脑直接动用自己的复杂模式识别能力,并以内省的方式用日常语言描述自己这种能力的使用(而不是用神经科学的词汇来描述),这很可能是实际上不可或缺的。这里的不可或缺性是源于本文前面已经提到的,我们的大脑由长期进化获得的直觉能力的不可还原性。至少,在神经认知科学还远远不能充分描述大脑的认知活动的今天,再考虑到前面所说的,我们的直觉能力的存在性完全可以在物理主义框架下得到解释,是得到知识论辩护的获得真知识的方法这些事实,我们没有理由拒绝基于直觉的一些传统哲学概念,没有理由拒绝以内省的方式对这些概念进行概念分析。

至于哲学中的概念分析的结果经常不可靠这一问题,这不会完全取消哲学中的概念分析的价值。例如,虽然所提出的“JTB+X”及类似的尝试都与基于直觉的“知识”概念不等价,但“它们不等价”这一认识也是对世界的真实的认识。而且,有时我们可能可以得到“在……这类情境下满足JTB+X的都是知识”这种在某个限定条件下的关于知识的充分条件。这可以算是对“知识”概念的部分概念分析。这些都增加了我们对基于直觉的“知识”概念所表征的对象(即知识)更清晰的认识。

我们看到,传统哲学研究中常见的所谓概念分析,是试图用明确的语言描述来重构一个基于直觉的哲学概念,即基于大脑的知觉记忆及大脑神经网络的复杂模式识别能力的概念。这种概念分析工作可以说是“先天的”,即它是大脑对已有的概念成分的分析,无需进一步观察世界。但由于基于直觉的概念的复杂性,这种概念分析的结果并不可靠。大脑的基于直觉的哲学概念,以及由这些概念构成的基于直觉的哲学理论,与常识物理学一样,是在观察世界的基础上构造的描述世界的理论,而且对世界中一定范围内的宏观现象是真实可靠的,这是人类大脑得以进化出来的前提条件。传统哲学研究中的概念分析工作,则是尝试用语言对这些基于直觉的哲学理论进行分析、重构,与科学中的纯理论研究本质上一样。但是,还是由于基于直觉的概念的复杂性,哲学中的概念分析法有其实用上的不可替代性。所以,物理主义可以保留和解释常识及传统哲学中的一些合理的方法论实践。

(责任编辑:盛丹艳)

(下转第 38 页)