

人工智能中符号奠基问题的几种解决策略

毛郝浩 李建会

摘要:符号奠基问题(Symbol Grounding Problem: SGP)是人工智能中的一个重要理论问题。它由认知科学家哈纳德提出,意指在不借助外部中介的条件下,一个纯粹的符号系统应如何获取意义? 目前为止,SGP的解决方案主要有三种:第一,以哈纳德的杂合系统为代表的认知主义方案;第二,以布鲁克斯的物理奠基策略为代表的生成主义方案;第三,以福格特等人代表的指号学方案。但是,这些解决方案都受到了塔迪欧和弗洛里迪的批评。他们提出了“零语义承诺”,即要求这些方案不能借助任何形式的内在主义(innatism)和外任主义(externalism)解决SGP问题。不过,“零语义承诺”本身也存在诸多问题,其合理性仍需进一步阐明。

关键词:符号奠基问题;人工智能;中文屋论证

中图分类号:B80 **文献标识码:**A **文章编号:**2095-0047(2022)01-0031-44

一、人工智能中符号奠基问题的提出

纽威尔(A. Newell)和西蒙(H. Simon)于1976年提出了物理符号系统假说(physical symbol system hypothesis, PSSH):“一个物理符号系统对于普遍的智能活动来说是必要且充分的。”^①该假说认为:(1)人类的智能活动是一种符号操作的过程;(2)符号操作的过程对于智能是必要且充分的;因而,(3)机器也能产生智能——机器的计算活动同样是对符号进行操作的活动。PSSH说明了智能本身就是对符

作者简介:毛郝浩,北京师范大学哲学学院博士研究生;李建会,北京师范大学哲学学院教授。

基金项目:教育部人文社会科学研究重大课题攻关项目“人工智能的哲学思考研究”(项目编号:18JZD013);广东省普通高校重点领域专项项目。

^① A. Newell & H. A. Simon, “Computer Science as Empirical Inquiry: Symbols and Search,” *Communications of the ACM*, Vol. 19, No. 3, 1976, pp. 113—126.

号的操作,因而通过计算机实现人工智能也就成为了可能。^①建立在 PSSH 的基础上,早期的人工智能研究目标就是探索如何通过对符号操作实现人工智能。但是,PSSH 一经提出,其过强的哲学假设受到了诸多批评。1980 年塞尔(J. R. Searle)在《心灵、大脑和程序》一文中提出的“中文屋论证”指出了 PSSH 中存在的意义问题。^②1990 年机器人专家布鲁克斯(R. A. Brooks)发表的文章《大象并不下象棋》(*Elephants Don't Play Chess*)中,也对于 PSSH 加以批判,并且提出了他的非表征主义观点。^③

塞尔的“中文屋论证”实际上说明了,存在一个能够通过图灵测试的符号系统,它能够在功能上和人脑表现一致,但不具备理解符号的能力。因此,任何仅仅对无意义的符号进行操作的系统都不能称得上是智能。

在“中文屋论证”中,这种对符号进行操作的系统可以被称为符号系统。对符号系统的定义可以由以下几个命题构成:(1)一个符号是一个符号系统的一部分;(2)符号系统由一个符号集和一个规则集构成;(3)规则集的作用是对符号进行操作;(4)对符号进行操作时,仅仅依赖于符号的形状而非其意义。^④因此,符号从定义上仅仅能够被辨识的是它的形状,它本身是不携带意义的。以书面语言为例,如果它独立于人对文字的解释,文字本身是没有意义的。文字的意义则依赖于一个中介(通常就是人脑)才能被呈现。

塞尔的“中文屋论证”合理的部分在于,仅仅对符号进行操作的程序确实不涉及符号的意义,但人脑能够理解符号的意义。人工智能程序正是依赖于符号系统才能够被实现,因而人工智能不能理解符号,强人工智能也就无法实现。哈纳德(S. Harnad)更进一步对该问题进行了阐述,提出了符号奠基问题:“如何让符号被其他东西奠基,而并非仅仅是一种无意义的符号?”^⑤除此之外,符号被奠基的过程还应该是独立的,而不应依赖于人脑进行解释。符号的意义能够在人脑中被奠基,是因为我们能够利用符号的意义对外部世界进行指称。但如果这些符号被写在纸上,它们的意义源于何处?哈纳德认为,纸上的符号很显然是没有被奠基的,因为没有入脑的中介,纸上的符号和外部世界没有因果关系。因此,符号奠基的另一个问题是:“一个纯粹的符号系统,如何在不依赖外部中介的情况下,自发地和外部世

① N. J. Nilsson, “The Physical Symbol System Hypothesis: Status and Prospects,” in M. Lungarella (ed.), *50 Years of Artificial Intelligence*, Berlin: Springer, 2007, pp. 9—17.

② J. R. Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences*, Vol. 3, No. 3, 1980, pp. 417—424.

③ R. A. Brooks, “Elephants Don't Play Chess,” in P. Maes (ed.), *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, London: The MIT Press, 1991, pp. 3—15.

④ S. Harnad, “Symbol-grounding Problem,” *Encyclopedia of Cognitive Science*, 2006. http://www.scholarpedia.org/article/Symbol_grounding_problem.

⑤ S. Harnad, “The Symbol Grounding Problem,” *Physica D: Nonlinear Phenomena*, Vol. 42, Issues 1—3, pp. 335—346.

界关联?”基于该问题,塔迪欧(M. Taddeo)和弗洛里迪(L. Floridi)提出了“零语义承诺”(zero semantical commitment),要求人工智能系统不能依赖任何外部的和内部的语义资源。^①

目前为止,符号奠基问题可以被视为包含两个部分:(1)一个纯粹的符号系统如何获得意义?(2)不借助外部中介,一个纯粹的符号系统如何自发地获得意义?针对符号奠基问题的解决方案均要回答这两个问题。当问题(1)获得解决,符号实际上已获得意义,只不过这种被奠基的符号不是被自主奠基的。解决问题(1)应当被视为解决符号奠基问题的最基本要求。而当问题(2)获得解决,才能说一个符号系统真正拥有奠基符号意义的能力。除此之外,解决方案还应该说明意义是何时何地、在系统中被生成的,以及如何与符号系统进行勾连的。

二、符号奠基问题的几种解决策略

目前为止,存在多种对符号奠基问题解决方案的划分方法。齐姆克(T. Ziemke)在1991年将SGP的解决方案大致上分为认知主义的解决方案和生成主义的解决方案。^②塔迪欧和弗洛里迪在2005年将SGP的方案分成表征主义的、半表征主义的、非表征主义的,并且列举了八个具体的解决方案。大卫·查尔莫斯(D. J. Chalmers)则将符号奠基方案分为因果奠基(causal grounding)和内在奠基(internal grounding)。^③李建会教授等人也对符号奠基的解决策略作了总结。^④不过,本文更加关注的问题主要是符号奠基问题相关的哲学思想是如何演变的,因此将具有代表性解决方案分成三种:认知主义、生成主义和指号学策略。

认知主义的代表策略是哈纳德的杂合系统(hybrid system)。杂合系统的核心思想是将感知觉运动系统和符号系统结合在一起。通过感知觉和运动系统,外部世界通过图像表征和范畴表征的方式和符号系统之间产生关系。通过符号系统,范畴表征可以组合在一起用于指称外部世界。

生成主义的代表策略是布鲁克斯的物理奠基假设(physical grounding hypothesis),其相应的工程学产物是包容结构(subsumption architecture)。布鲁克斯认为,智能的定义应该不局限于对符号进行操作,而是来源于能动者和环境的交互。更加低级

① M. Taddeo & L. Floridi, "Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research", *Journal of Experimental & Theoretical Artificial Intelligence*, 2005, Vol. 17, No. 4, pp. 419—445.

② T. Ziemke, "Rethinking grounding", in A. Riegler (ed.), *Understanding Representation in the Cognitive Sciences*, New York: Plenum Press, 1999, pp. 177—190.

③ D. J. Chalmers, "Subsymbolic Computation and the Chinese Room," in J. Dinsmore (ed.), *The Symbolic and Connectionist Paradigms: Closing the Gap*, Hillsdale, NJ: Lawrence Erlbaum, 1992, pp. 25—48.

④ 夏永红、李建会:《符号奠基问题及其解决策略》,载《哲学研究》2017年第2期,第102—110页。

的生物(例如昆虫和鱼类)无需对符号进行操作,但是它们应该也是拥有智能的。因此,生成主义实际上取消了符号系统,它的目的是消解SGP问题,而不是解决它——没有符号,就无需奠基了。

在指号学策略中,斯蒂尔斯(L. Steels)和福格特(P. Vogt)都作出了相关的贡献。指号学的优势在于同时结合了认知主义和生成主义的优势,将意义的位置同时扩展到表征的内在结构和外部的环境当中。福格特对其解决方案命名为物理符号奠基(physical symbol grounding),同样可以看出其思想——福格特故意将符号奠基(symbol grounding)和物理奠基(physical grounding)组合到一起,说明其策略结合了这两种策略的优势。同时,斯蒂尔斯和福格特还引入了语言猜谜游戏,模仿了儿童的语言习得过程,这又让具体的工程学方案得到实现。

(一) 哈纳德的杂合系统

哈纳德在提出了SGP问题后,自己给出了一个解决方案,他称这一方案为“杂合符号+感知觉运动系统”(hybrid symbolic + sensorimotor system)。杂合系统分为两个部分:符号系统和感知觉运动系统。一个纯粹的符号系统本身是不包含意义的,但它负责对符号进行操作,从而形成合适的输出。感知觉运动系统能够让一个系统和外部世界进行交互,从而将外部世界的意义传递到内部。两个系统的合作才得以让整个系统获取意义。

福多(J. Fodor)也曾将认知分为输入系统和中央系统,输入系统将感知觉转换为内部的表征(例如视觉和听觉感知),而中央系统对这些表征进行操作(例如问题解决)。哈纳德的杂合系统与之也有类似之处——输入系统对应了感知觉运动系统,而中央系统对应了符号系统。如果说狭义的对智能的理解仅仅是对内部表征的操作,那么更宽泛的含义就应该同时涵盖感知觉系统和中央系统。哈纳德的做法正是扩展了智能的条件,认为智能产生于符号系统和非符号系统的结合中。

哈纳德将感觉运动系统(非符号系统)能够提供的非符号资源分为两个部分:图像表征(iconic representations)和范畴表征(categorical representations)。范畴表征的形成依赖于图像表征。图像表征指的是对近端的事物和对象的模拟——例如,一匹马投射在我们视网膜上的图像。范畴表征指的是在图像表征中提取的不变特征,这依赖于图像表征。这个世界有很多马,但是“四条腿”这样的特征是不变的,因而“四条腿”这样的不变特征就是范畴表征。

这两种资源都是非符号资源,它们的形成是由投射所引发的因果关系(从外部世界投射到内部世界)导致的,因而图像表征和范畴表征都和世界具有关联,它们是无需奠基的。接下来,最基本的范畴表征会组合成原子符号(elementary symbols),例如“马”这一名字就由“四条腿”“一条尾巴”等范畴表征组合而来。原子符号由于是已经奠基了的范畴表征的组合,因而也是被奠基的。最后,复杂的符

号表征 (symbolic representation) 由原子符号构成——例如,“斑马”这个符号就是由“马”和“条纹”这两种原子符号结合而成。这样就完成了对符号系统的奠基。

上述过程说明,只要找到能够实现这些步骤的机制,符号奠基问题就能得到解决。哈纳德给出的工程解决方案便是人工神经网络:“联结主义,因为拥有一般的模式学习能力,看起来是一个自然的解决方案——一个联结主义的网络可以学会从样本中提取正确的图像,通过不断的调整权重把图像还原成不变的特征”^①。

2002年,坎奇洛西 (A. Cangelosi) 和哈纳德应用神经网络的方法给出了具体的解决方案:“符号小偷假设”(symbolic theft hypothesis)^②。这个方案正是按照杂合系统来设计的,其中非表征的部分被他们称为“感觉运动苦工”(sensorimotor toil),与之对照的高层符号表征部分被他们称为“符号小偷”。这么称呼的原因是,在学习高级概念的时候,新的概念能够不直接依赖感知觉经验而被获得。例如,“斑马”就是“马”和“条纹”的结合,因而“斑马”这个新概念就不需要直接的感官经验,符号之间组合的运作是相当高效的,和“苦工”对比起来就是“小偷”。

但是,哈纳德的杂合系统同样受到了批判。弗洛里迪认为杂合系统有两方面没有达到SGP解决条件的要求:(1)像“四足动物”这样的概念从什么地方产生?感知运动系统确实能够从环境中学会一些不变量,但是符号系统中的概念不能凭空出现。(2)在神经网络中,如果网络是被监督的,它们学习的内容是通过预先选择的训练集产生的,因而这些内容是完全外在的。如果网络没有被监督,它们仍然需要一些内在的偏好和特征检测器来达到有效输出。这些内在偏好和特征检测器很显然是人工编码的程序。^③基于这样的考虑,杂合系统就不能满足SGP的第二个问题——它不能自发地解释符号的意义。

齐姆克则从符号的使用意义上批判了认知主义的存在的问题:“输入转换系统(对应的是感知觉运动系统),由于其本身只能得到转换的结果,因而必须在中央系统中被使用,而中央系统又必须在嵌入环境中。”^④齐姆克在这里其实想说的是,在认知主义的解决方案中,所有的符号都只具备单纯的和外部联系的能力,这种关系是一一对应的,但是这样的关系是机械的,它们没有功能价值。例如,虽然杂合系统学会了“马”的概念,但是杂合系统仅仅知道马有“四条腿”这样的不变特征。杂合系统不知道马是可以骑的,也不知道“马”这个概念能用来做什么。因而认知主义的符号缺乏了使用层面上的意义。

① S. Harnad, “The Symbol Grounding Problem”.

② A. Cangelosi, A. Greco, S. Harnad, “Symbol Grounding and the Symbolic Theft Hypothesis,” in A. Cangelosi, D. Parisi (ed.), *Simulating the Evolution of Language*, London: Springer, 2002, pp. 191—210.

③ M. Taddeo & L. Floridi, “Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research.”

④ T. Ziemke, “Rethinking Grounding”, in A. Riegler (ed.), *Understanding Representation in the Cognitive Sciences*, New York: Plenum Press, 1999, pp. 177—190.

由此看来,哈纳德的杂合系统解决了 SGP 的第一个问题——以自下而上的方式让符号获得了意义。但是没有解决第二个问题,因为杂合系统并不能自主地奠基符号,无论是程序还是训练集都是人类所设计的。此外,这样的符号也不具备任何的语用价值,在不同的语境下也无法得到正确理解。杂合系统扩展了智能的条件,从纯粹的符号系统走向了包含非符号系统的智能。在对意义的理解上,杂合系统中的符号依赖于感知觉系统中产生的内部结构,即能够用于指称外部世界的范畴表征,因而符号的意义产生于非符号系统和世界的交互中,并将意义传递到了符号系统当中。

(二) 布鲁克斯的物理奠基策略

生成主义的解决方案对应了一种不同的理解智能的方式。对于认知主义来说,智能是一种对符号进行操作的过程,下象棋、问题解决以及推理这样的认知活动都是智能的体现。对于生成主义来说,智能体现在能够和世界交互的行为,而不是对内部表征进行操作。因而,低等生物例如昆虫、鱼类也应当是有智能的。布鲁克斯对智能下过一个定义:“智能是能动者和世界的动态交互所决定的。”^①因而,在生成主义的解决方案中,SGP 问题实际上已经改头换面了——智能行为而非符号操作,是如何产生意义的?当存在符号的时候,我们才需要考虑符号的意义问题。而以行为为中心的生成主义,只需要考虑如何让行为变得智能,而并不需要考虑符号的问题。

布鲁克斯是麻省理工大学人工智能实验室研究机器人的教授。布鲁克斯对人工智能科学作出了很多贡献,例如提出了 Nouvelle AI 的概念(或者说“基于行为的机器人”, Behavior-based robots),提出了物理奠基假设(Physical grounding hypothesis),在工程学上提出了包容结构(Subsumption architecture),并且制造了很多基于行为的机器人(Allen、Herbert、Cog,等等)。^②体化认知科学起源于 20 世纪 80 年代晚期,虽然布鲁克斯自己并未利用这一进路提出体化机器人,但是福格特和齐姆克都认为布鲁克斯的成果是体化认知的典型代表。生成(enaction)这个术语由瓦雷拉(F. Valera)引进,齐姆克认为布鲁克斯的工作是生成主义的代表,我们将沿用齐姆克的分类,将以行为为中心的 SGP 解决方案归类为生成主义的解决方案。

简单来说,物理奠基假设可以这样表述:构成智能的必要条件是人工智能和外部的物理世界进行交互。在传统的认知主义中,很少关注人工智能如何对周围的世界施加行动,而仅仅关注如何将感知觉运动数据转化成表征,从而操作表征产生智

① R. A. Brooks, “Intelligence Without Reason”, in R. Myopoulos (ed.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, San Mateo: Morgan Kaufmann, 1991, pp. 569—595.

② https://en.wikipedia.org/wiki/Nouvelle_AI 参见 P. Vogt, “The Physical Symbol Grounding Problem”, *Cognitive Systems Research*, Vol. 3, No. 3, 2002, pp. 429—457.

能。物理奠基假设要求机器人和环境互动,这就要求机器人能够拥有一个能够和环境互动的机制,因而“行为”“行动”“体化”就成了必要的条件。

布鲁克斯基于进化论对这一提法进行了论证:在生命近乎 40 亿年的进化史中,人类的符号几乎是近 5000 年才诞生的。而在这漫长的历史当中,占据主流的智能并非符号化的智能,而是基于感知觉和行动的智能。如果没有符号,感觉和运动之间直接进行联系,这种更快的反应能够获得更多的生存优势。这也是为什么基于符号的机器人的表现不如基于行为的机器人好的原因。没有物理的奠基,任何符号表征与感觉和运动之间的联系都是松散的。物理的奠基能够对符号加以合适的限制,从而让这些符号真正地发挥作用。^①

基于物理奠基假设的工程学方法就是包容结构(subsumption architecture)。包容结构是一个有层级的目的结构,其中更高一级的层级能够利用更低的层级实现自己的目的(也就是对下一层进行包容)。例如,“避免碰撞”可以是最低的层级的目的,更高一层的目的就是“到处走走”,最高层级的目的是“探索世界”。每一层次都有一个固定目的的简单的有限状态机(finite state machines)的网络,这些状态机可以进行矢量和的计算。^②更晚近一些,布鲁克斯设计出了考格(Cog)机器人,目的在于更近似地模仿人类的行为。考格具备学习能力,能够检测人类的面孔和注视方向,能学会手和眼睛的协调,也可以通过眼神和人类进行社会交流。这些都让考格看起来更像人。^③

布鲁克斯认为,基于行为的机器人具有两个核心:境化性(situatedness)和体化性(embodiment)。境化性指的是机器人是被放置于环境中的,它们不处理抽象的事情,而只处理世界给予它的直接信息,这些信息能够直接影响系统的行为。体化性指的是机器人拥有身体并且能够直接地体验世界,它们的行为是和世界交互的一部分,并且能够从感觉当中获得直接的反馈。布鲁克斯说:“世界就是最好的模型。”^④

总的来看,以布鲁克斯为代表的生成主义策略具有以下几个特点:(1)从更低层次智能水平来理解智能,将智能理解为和环境的交互;(2)取消表征的概念,这也是由生成主义智能观决定的;(3)注重实时反馈,将世界作为模型,感受器作为输入变量,直接输出行为,无需利用表征再对世界进行模拟。但是生成主义策略也许面临着比认知主义更加困难的问题:(1)没有符号,是什么被奠基了?齐姆克认为,生成主义策略与其说是回答了 SGP 问题,不如说是回答了机器人奠基问题

① R. A. Brooks, “Intelligence Without Representation”, *Artificial Intelligence*, Vol. 47, Issues 1—3, 1991, pp. 139—159.

② Ibid.

③ R. A. Brooks, C. Breazeal, M. Marjanović, et al., “The Cog Project: Building a Humanoid Robot”, in C. L. Nehaniv (ed.), *International Workshop on Computation for Metaphors, Analogy, and Agents*, Berlin: Springer, 1998, pp. 52—87.

④ R. A. Brooks, “Elephants Don’t Play Chess”.

(robot grounding), 或者是身体奠基问题 (body grounding)。(2) 生成主义由于取消了表征, 那些和表征相关的智能也不可能被展现——问题解决、下象棋等依赖表征的智能不能被生成主义所体现。

在SGP的两个问题中, 生成主义成功消解了问题(1)纯粹的符号系统如何获取意义? 而对于问题(2)的回答, 生成认知的策略仍然不能令人满意。齐姆克这样评价:“布鲁克斯的策略确实能利用联结主义网络和进化算法来实现部分的自我组织, 但是由于机器人的身体, 仍然依赖外部设计者, 它们没有植根在历史当中。而生物的进化则是从历史中和环境不断的交互而形成的, 因而仍然不能说机器人是自我组织的。”^①因此, 齐姆克提出了“从奠基到根基”(from grounding to rooting)的说法, 认为仅仅物理奠基是不够的, 机器人想要拥有智能, 还需像生物一样在历史中和环境交互, 从而不断地自我组织、自我进化。

(三) 指号学奠基策略

福格特于2002年提出物理符号奠基 (Physical Symbol Grounding)。^②正如其名字所述, 物理符号奠基是物理奠基和符号奠基的结合, 其目的就是将符号奠基和物理奠基相结合起来, 让机器人既能够拥有身体, 又能够产生符号表征。其哲学资源来自皮尔斯 (C. S. Pierce) 的指号学。在工程学上实现物理符号奠基的是斯蒂尔斯, 他通过引入猜谜游戏让机器人互相沟通, 从而模仿儿童语言学习的过程。^③沙夫斯卡娃 (P. Varshavskaya) 的 KISMET 机器人, 也是通过模仿人类儿童获得语言的过程来制造的, 不过 KISMET 的理念依然基于布鲁克斯的物理奠基。^④

认知主义所关注的是如何将感知觉运动信息抽象成表征, 生成主义关注的是如何构造出和环境交互的行为机器人。福格特的指号学策略则同时关注了感知觉信息、运动以及和世界交互的能力、与其他机器人进行社会交流的能力。即使福格特的策略仍然满足不了弗洛里迪的“零语义承诺”, 它仍然是目前最有前景的策略。

在皮尔斯那里, 指号学是一种更加宽泛的符号学。皮尔斯将指号 (sign) 分成三类: 索引 (index)、图像 (icons) 和符号 (symbols)。SGP 问题中所讨论的正是符号, 它的特征是其形状任意且不具备任何意义。我们无需关注索引和图像在解决SGP中的作用, 而要关注在皮尔斯指号学中符号是如何被解释的。

皮尔斯的指号学认为一个指号的结构由三元关系构成: 表示项 (representamen)、

① T. Ziemke, “Rethinking Grounding”.

② P. Vogt, “The Physical Symbol Grounding Problem”.

③ L. Steels, “The Symbol Grounding Problem Has Been Solved. So What’s Next,” in M. De Vega, et al. (eds.), *Symbols and Embodiment: Debates on Meaning and Cognition*, Oxford: Oxford University Press, 2008, pp. 223—244.

④ P. Varshavskaya, “Behavior-based Early Language Development on a Humanoid Robot,” in *Proceedings of the 2nd International Conference on Epigenetics Robotics*, 2002, pp. 149—158.

解释项 (interpretant) 和对象 (object)。当表示项是任意的和约定的时候, 表示项就可被看作形式符号。解释项可被看作“另一种可以指称同样对象的表征”, 而斯蒂尔斯对皮尔斯的阐释就直接把解释项看作意义 (meaning)。因而, 在经过福格特和斯蒂尔斯的阐释后, 一个符号的结构就是意义、形式 (form)、和指称物 (referent) 的三元关系。例如, “马” 的形式就是无意义的形式符号, “马” 这个符号的指称物就是外部世界的马, 而“马” 的符号通过它的意义来指称外部世界的马。为了方便区分, 下文将用指号学符号 (semiotic symbol) 指代皮尔斯符号观中的符号。

福格特将皮尔斯的指号学应用到其方案当中, 首要目的是解决符号的意义来源问题。哈纳德对狭义的意义看法就是“一组用于指称对象的规则”。例如, “四足马” 的范畴就是一种规则, 用于指称那些有四只足的马, 无论现实生活中是否存在四足马, “四足马” 这个符号已经具备了一组用于指称对象的规则。再例如, “晨星” 和“暮星” 虽然指称的对象相同, 但是由于它们的规则不同, 因而也具备不同的意义和不同的指称规则。而指号学并不采取“规则集” 的方式来指称对象, 一个符号的意义在于“一个指号学过程, 它是一种形式、意义和指称之间的交互, 这就意味着指号学符号取决于指号学符号是如何被构成的, 具备怎样的功能……因而, 指号学符号的意义可以被看成形式和指称物的功能关系”^①。

福格特这里的“功能关系” 概念非常宽泛, 它既可以指代一组“规则集”, 可以指代一组动作 (和环境互动)。维特根斯坦对语言“使用” 的解释可以对这种功能关系作出一个可能的说明。在维特根斯坦的后期哲学中, 语言更像是一种游戏, “意义” 并非抽象的对象, 而是在人类社会行为中的扮演表达的作用。要认识表达的意义只需要在交流情景中知道它是怎样被使用的就可以了。^② 在福格特所阐释的三元关系中, “意义” 和维特根斯坦的“使用含义” 是比较相似的。如果再考虑一下福格特和斯蒂尔斯所制造的机器人, 情况将会更加清楚——机器人所采用的语言游戏和维特根斯坦所描绘的语言游戏几乎别无两样。

福格特的机器人建立在斯蒂尔斯机器人的基础上。在福格特的实验中, 机器人拥有非常简单的身体并且仅可以通过视觉系统和物体进行交流。在语言游戏中, 说话者 (speaker) 首先利用视觉检测到一个物体, 并赋予它一个随机的名字。而猜测者 (hearer) 将会猜测这个命名所对应的视觉对象, 如果猜中了, 语言游戏就算成功了。而在这个过程中, 指号学符号的功能仅仅是让猜测者能够正确的指称对象。因而, 意义分布在如下的过程中: (1) 机器人感知并且命名一个物体; (2) 猜测者作出正确的反应。而随着时间的变化, 机器人和环境的交互越多, 一个指号学符号的意

① P. Vogt, “The Physical Symbol Grounding Problem”.

② W. G. Lycan, *Philosophy of Language: A Contemporary Introduction*, New York: Routledge, 2012.

义就越丰富。

为了说明这样的动态系统是如何运作的,下面将以斯蒂尔斯制作的机器人 QRIO 为例。QRIO 是一种类人机器人,它们拥有两条腿、一只头、两只手以及手指,有摄像机作为视觉输入,麦克风作为音频输入。同样,每个传感器都连接到了相应的动作模块用于输出动作。QRIO 拥有了身体,也能够与环境互动,因而满足了体化的要求。实验开始时,会在一张桌子面前放置两个面对面的机器人。桌子上铺开了不同的色块,以供机器人进行识别。两个机器人中的一个将会随机成为说话者,另一个成为猜测者。说话者将选择一个色块作为样本,并且赋予随机的符号,例如“wabado”用于指代红色。听话者将会猜测“wabado”代表的是哪个色块,而当说话者同意了猜测者的猜测时,游戏就成功了。

实验采用了 20 个机器人系统(它们可能用了同一个身体)。在协作过程中,机器人之间通过两种方式来交流:(1)当符号和意义之间的联系正确的时候,它们之间的关系分数会上升,反之则会减少。于是那些正确的符号将会更快地在群体中传播开来。(2)根据交流的结果(游戏失败或者成功),每个机器人都会不断调整自身范畴表征的权重,从而更好地完成互相交流。

从上述的交流策略中可以看出,一个符号的意义在交流的过程中是不断变化的。对于单个的机器人来说,它所建构的范畴表征与其他机器人的不同。当交流开始的时候,所有的机器人都要在一次次成功和失败中重新调整符号的意义,因而意义是被机器人群体共享的,而非独立的、个体的。斯蒂尔斯的实验结果表明,当使用语言作为交流工具后,所有的机器人几乎在符号的意义上达成了一致。而不使用语言作为工具的机器人所奠基的符号的意义存在一定的差异。于是在这种交流过程中,符号的意义就在符号的自我组织中诞生了。

与维特根斯坦的符号观相同的是,符号的意义成为一种语言游戏的产物,它们是符号“使用”层面上的意义。但除了赋予符号单纯的使用意义之外,被奠基的符号同时也具有范畴表征。福格特认为,采取这样的策略有两个优点:(1)指号学符号在三元关系中是从定义上被奠基的。这是因为一个指号学符号是必然伴随其意义和对象的,没有意义和对象的符号就不能被称为指号学符号。(2)指号学符号是境化的和体化的。指号学符号是从能动者和世界和其他能动者的交互中诞生的,因而满足了体化认知和境化认知的要求。

然而,弗洛里迪依然认为福格特等人的策略不能解决 SGP 问题,理由在于:(1)指号学符号仍然只有在观察者的眼中才具备意义,人工智能自身不能理解指号学符号的意义,要理解指号学符号的意义,除非机器人已经拥有关于符号的语义。这是一个循环论证。(2)猜谜游戏对于符号奠基并非必要,在机器人进行交流之前,符号已经被机器人实现奠基了。

总的来说,福格特和斯蒂尔斯的指号学策略融合了认知主义和生成主义的长处,避免了认知主义缺乏体化性和境化性的缺点,是一个非常有前景的策略。在SGP问题上,指号学策略也成功的解决了问题(1),但针对问题(2),依然有学者(如弗洛里迪)认为由于外部设计的程序将会带来外部语义资源,因而机器人不具备自主性。

三、“零语义承诺”及其批判者

塔迪欧和弗洛里迪在2005年的一篇综述中提出了“零语义承诺”,其内容如下:(1)任何形式的内在主义都是不被允许的。语义资源不应该被预置在人工智能当中。(2)任何形式的外在主义都是不被允许的。语义资源不应该从外部世界加载到人工智能当中。(3)人工智能应当拥有它自己的能力(计算的、程序的、感知的、教育的、身体,等等)和资源去奠基它的符号。^①

塔迪欧和弗洛里迪认为,目前的解决方案要么引入外部的语义资源,要么引入内部的语义资源,本质上都是不成功的。不过,多数学者认为“零语义承诺”对人工智能的要求过于苛刻。提出批评的学者有比莱卡(K. Bielecka)^②、缪勒(V. C. Müller)^③、库柏克(R. Cubek)^④等。其中,比莱卡批判的主要是弗洛里迪的实践解决方案(praxical strategy),其批判方法在于揭示出因果奠基所采用的哲学方法中的纰漏。缪勒将“零语义承诺”比喻成查尔莫斯论述的“意识难问题”,认为SGP的难问题是不可解决的,也不应该得到解决。库柏克将批评“零语义承诺”的思想作了一个简短的总结,并指出针对弗洛里迪的批评是合理的。

弗洛里迪等人在提出“零语义承诺”之后,给出了一种基于实践的解决策略。^⑤这一解决策略和福格特的解决策略有很多相似之处,例如:同样关注和环境的互动,关注语言的进化过程,关注机器人之间的互动和交流,等等。其不同点在于,弗洛里迪的实践策略使用了一种基于行动的语义学 Abs (Action-based semantics)。

① M. Taddeo & L. Floridi, “Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research”.

② K. Bielecka, “Symbol Grounding Problem and Causal Theory of Reference”, *New Ideas in Psychology*, Vol. 40, 2016, pp. 77—85.

③ V. C. Müller, “Which Symbol Grounding Problem Should We Try to Solve?”, *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 27, No. 1, 2015, pp. 73—78.

④ R. Cubek, W. Ertel, G. Palm, “A Critical Review on the Symbol Grounding Problem as an Issue of Autonomous Agents”, in G. Randy, et al. (eds.), *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, Springer, 2015, pp. 256—263.

⑤ M. Taddeo & L. Floridi, “A Praxical Solution of the Symbol Grounding Problem”, *Minds and Machines*, Vol. 17, No. 4, 2007, pp. 369—389.

Abs 的核心在于将意义看成一个内部的行为状态,而非传统的通过感知觉数据形成的心理表征。这种行为状态和感觉输入无关,也不设立目的。因此,在最初的意义产生阶段,基于 Abs 的系统所进行的行为是对环境漫无目的的探索。在此基础上,系统通过行为结果反馈将世界映射到系统内部,从而对符号的意义进行奠基。

比莱卡认为,在弗洛里迪的方案中,语义和目标之间的关系是缺失的:“Abs 将面临一个比塔迪欧和弗洛里迪所批评的感觉运动解决方案更大的琐碎问题(trivialization problem)。此外,符号与行动的关联容易产生简单和困难的析取问题。因为在他们对主体行动的解释中没有目的论的假设,容易的析取问题就是无法解决的——就像早期行为主义中的反应一样,行动是个体化的”^①。在这里,比莱卡认为弗洛里迪的方案遵循的是一种因果奠基方法,即通过行动因果地将世界和符号勾连在一起。但是在弗洛里迪的方案中,这种方法仍然存在析取问题、琐碎问题等尚未被解决的问题。更重要的是,基于 Abs 的整个系统是无目的的,这就意味着任何内部状态都可以表征任何与之有因果关系的外部状态。

缪勒同样指出了“零语义承诺”中无目的性的问题——任何真正的人工智能都不能无目的性地诞生。自主人工智能存在的最小条件就是具备目的性,如果没有目的性,人工智能就是一具环境的傀儡——正如早期的行为主义所假设的那般。自主人工智能的诞生不可能不被设定目的——“如果没有目标,就没有‘尝试’,没有‘更好’(指的是行为),因而真正的人工智能是具备目的性的,要不然它仅仅是一个和环境互动的系统”^②。而人工智能的目的要么是外部植入的(外在主义),要么是内部预置的(内在主义),“零语义承诺”会因此陷入困境。模仿查尔莫斯对意识问题的难易分类,缪勒也给出了 SGP 的难问题和易问题。易问题是:我们如何解释和复制人工智能中意义的功能和行为能力?难问题是:意义如何从纯粹的物理现象中诞生?

实际上,SGP 中的意识问题在哈纳德那里其实已经得到阐述:“身心问题有很多伪装的说法:意识(consciousness)、知觉(awareness)、主观性(subjectivity),感受质(qualia)、意向性(intentionality)、第一人称状态(1st-person states),这些问题实质上都是身心问题……符号奠基并非意义奠基,符号奠基是一种输入/输出的功能表现……而意义,更像是一种心灵的东西。为了解决身心问题的命名游戏,我们只能引入一个更加适合的并且不需解释的术语来解释——感觉(feelings)……所有的

① K. Bielecka, “Symbol Grounding Problem and Causal Theory of Reference”.

② R. Cubek, W. Ertel, G. Palm, “A Critical Review on the Symbol Grounding Problem as an Issue of Autonomous Agents”.

心灵状态,不仅仅是自主的动态功能系统,也同样是感觉状态……因而感觉才是心灵真正的标志物。符号奠基问题仅仅能够触及身心问题的功能部分。”^①哈纳德在这里指出,符号奠基问题是一种功能问题而非意义问题,意义问题会涉及独特的心灵状态——感觉(即意识经验,哈纳德为了避免混乱而使用该术语)。感觉留给我们的问题则是难以解决的,因而SGP能够讨论的只能是功能问题。

最后,库柏克同样指出“零语义承诺”所逃避的问题:“如果人类具有内置的目的性(基因),并且从出生开始从外部接受语义资源(文化),为什么机器人必须要满足零语义承诺?”库柏克的结论是,SGP的难问题和设计目标导向的自主机器人无关。真正有需要的是研究工程学的方法,从而填补符号和亚符号表征之间的鸿沟。

总的来说,要求符号系统能够不借助外部中介自发地理解意义存在着内在的困难。对于被设计出来的人工智能来说,判定其是否拥有自主理解符号的能力涉及意识难题。当我们无法成为人工智能系统本身时,只能通过其行为来推断它是否能够自发地理解意义。因而,要判定人工智能是否自发地产生了对符号的理解,就需要我们对人工智能的行为进行解释。例如,在福格特的猜谜游戏中,机器人的确呈现出语言自主组织能力和理解能力。但机器人是否真正产生了自主行为却依赖于我们对其行为的解释,而解释行为本身就违反了人工智能自发理解意义的条件。

不过,即使符号奠基问题中存在意识难题,对其研究也不应止步于此。在米利肯(Ruth Millikan)等人建立的生物语义学(teleosemantics)当中^②,符号奠基问题可以表述为生物如何使用符号表征外部世界的问题。例如,蜜蜂的八字舞具有表征能力,其舞蹈的形式表征了采集花蜜地点的方向和距离信息。很难说类似蜜蜂这样的低级生物能够有意识地对花蜜进行表征,但其符号确能够被外部环境所奠基。米利肯认为,生物的表征能力源于进化过程中的自然选择。正是因为蜜蜂能够正确表征花蜜信息,它才能够促进蜜蜂的生存和繁衍。因此,生物的目的性是可以自然的方式来进行解释的。在不涉及意识问题的基础上,符号奠基问题仍值得进一步的研究——例如,如何设计出一种人工智能系统,其符号可以通过类似于蜜蜂的生物语义功能被奠基。

四、结论

本文介绍SGP问题的起源、发展和目前对SGP问题的哲学批判。符号奠基问

① S. Harnad, “Symbol-grounding Problem”.

② R. G. Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge: MIT Press, 1984.

题是哲学和科学互动的一个典型例子,不同的哲学思想往往会得出不同的解决策略,而新的解决策略又会推动对当前问题进一步的哲学分析。我们将从意义观、智能观的变化上对 SGP 问题解决策略的演变进行总结。

第一,意义观的变化。不同的解决策略所认可的意义理论是不同的。以哈纳德为代表的认知主义认为意义是内在的,是一组能够指称外部世界的规则集。以杂合系统为代表的各种解决方案中,意义就是各种范畴表征的组合规则。这些组合规则能够使得符号具备指代外部世界的能力。以布鲁克斯为代表的生成主义由于拒斥符号的作用,认为意义来源于感知觉和运动系统与世界的交互。以福格特和斯蒂尔斯为代表的指号学策略扩展了哈纳德的意义观,将符号的意义从个体的、内在的扩展成群体的、同时存在于心灵和环境中的动态规则。这种策略兼具其他解决策略的优点,使符号以体化的和境化的方式被奠基。

第二,智能观的变化。传统的人工智能研究对智能的定义就是对符号和表征的操作,这使得具有智能的计算机成为可能。但传统的人工智能能够模仿的仅仅是涉及符号系统的部分智能——例如,推理和问题解决,等等。哈纳德扩展了智能的含义,将以符号操作系统为基础智能观扩展到以杂合系统为基础的智能观。非符号系统成为了符号系统和外部世界产生关系的桥梁。福格特和斯蒂尔斯虽然没有明确说明他们的智能观,但是我们可以看到,他们确实融合了认知主义和生成主义的智能观——智能是体化的、境化的、基于交流的,同时也包含了对符号的操作。

除此之外,我们还可以看到人工智能的工程学形态也在不断发生变化。在认知主义那里,人工智能还没有身体,而仅仅是神经网络程序,也没有与环境的互动,而是一种基于感知的人工智能程序。在布鲁克斯提出物理奠基假设后,机器人有了身体和行动,是一种基于行为的人工智能机器人。指号学策略提出后,人工智能又增添了交流的维度,我们可以将这种机器人称为基于交流的机器人。

通过对符号奠基问题进行哲学分析,我们发现 SGP 和“自主性”或意识问题有关,这一部分难以得到解决。但仅仅将符号奠基问题中无法解决的部分归类为意识问题是一种过于简单的做法。米利肯的生物语义学说明了,即使不涉及意识问题,生物的目的性仍然可以用自然的方法来解释。如何对符号和世界的勾连问题进行进一步的自然化解释,将会进一步扩大符号奠基问题的解决空间。

(责任编辑:韦海波)

Abstracts and Key Words

• On the Ethical Principles of the Development of Artificial Intelligence

SUN Weiping LI Yang

Abstract: As a revolutionary and subversive technology, artificial intelligence is deeply changing and cultivating human beings and societies. In the field of ethics, artificial intelligence has caused many ethical problems and thus many ethical conflicts, especially challenging the dominant subjective moral position of human beings. Facing these sharp problems and challenges, some proposed ethical principles cannot generate holistic effect because they are irrelevant to each other, especially irrelevant to the positive ethical support for the development of artificial intelligence. Having analyzed the possible ethical consequences of artificial intelligence based on its development and application, this paper puts forward a system of ethical principles which includes humanistic principle (core principle), justice principle and responsibility principle. By realizing the purpose of regulating the development of artificial intelligence, this paper tries to achieve its positive value in serving humanity and promoting the development of man and society.

Key words: artificial intelligence; ethical principles; humanistic principle; justice principle; responsibility principle

• An Ethical Crossing of the “Responsibility Gap” in Artificial Intelligence Application: Taking the Self-Driving Car as a Typical Case

WANG Tianen

Abstract: Taking self-driving cars as a typical example, the “responsibility gap” of artificial intelligence application has become a new problem that needs to be solved urgently. The responsibility gap is essentially a phenomenon in which machine agents have no responsibilities at the stage when their autonomy is increasing but have not gained the position of the responsibility subject. With the more and more common application of artificial intelligence, the responsibility gap brought about by artificial intelligence is a theoretical subject that needs to be broken through urgently in the field of practice. The application of artificial intelligence will bring the biggest dividend that human beings can imagine at present, while the responsibility gap may constitute the bottleneck of artificial intelligence application. While expecting the development of artificial intelligence to bring huge dividends, it is urgent to deal with the accompanying liability dilemma. The cross of the responsibility gap in the application of artificial intelligence is related to the development and application of artificial intelligence, thus related to the successful delivery of unprecedented dividends in the development of artificial intelligence. In the face of the responsibility gap in the application of artificial intelligence, it is impossible to find a reasonable theory and method in the traditional paradigm. One can only bridge it in the ethics of world-making which means the integration of rules and laws. World-making ethics has different nature and characteristics, among which the integrity, class characteristics and commonality are the most basic elements. Taking automatic driving vehicles as a typical case, the related responsibilities and obligations mainly involve the design of intelligent devices, the responsibility of use and the ethical obligation of group benefit. If there is no person responsible for the accident caused by the self-driving car and only the beneficiary’s obligation exists, the relevant losses should be borne by the beneficiary group of the autonomous car in the form of obligation. The ethical bridge of the responsibility gap of AI application will form a virtuous circle between the application of AI and the realization of corresponding dividends.

Key words: artificial intelligence; responsibility gap; world-making ethics; self-driving car

• The Symbol Grounding Problem: Strategies from Cognitivism to Semiotics

MAO Haohao LI Jianhui

Abstract: The Symbol Grounding Problem (SGP) is an important theoretical problem. Proposed by Steven Harnad, it aims to explain how a formal symbolic system can obtain meaning without an external mediation. So far there exist three main strategies: (1) cognitivism represented by Harnad’s hybrid system; (2) enactivism

represented by Rodney Brooks's physical grounding strategy; and (3) semiotics strategy represented by Paul Vogt. However, these strategies are faced with criticisms from Mariarosaria Taddeo and Luciano Floridi. They proposed "zero semantic commitment", which requires them to solve the SGP without any form of innatism and externalism. Yet, the reasonableness of "zero semantic commitment" which itself has problems needs to be further clarified.

Key words: symbol grounding problem; artificial intelligence; chinese room argument

- "Dependent Origination and the Emptiness of Nature" and Nihilism:

A Discuss about Method in the View of Phenomenology

NI Liangkang

Abstract: Like Descartes' methodological requirement of "doubt everything", Nagarjuna's idea of "all the dharmas are empty" does not belong to ontology or epistemology, either; rather, it is first and foremost a methodological plan. If Descartes is not an exact sceptic, then Nagarjuna is also not an exact nihilist. "Empty Nature" goes beyond the ontological consideration about Being-and-Nothing, which is due to the fact that it represents a methodological operation abandoning or suspending all the wedded-ness. De facto, "Empty View" is never the ending of philosophical ideas but, exactly, an idea as such, namely, a Cartesian methodological meditation. This also corresponds to Husserl's words: The mission of philosophy is to refrain itself, via inquiries, from being wedded to traditional and pre-given pre-philosophical understandings of the world. In this connection, "Dependent Origination and the Emptiness of Nature" exactly means refusing to be wedded to the traditional and pre-given pre-philosophical, even philosophical, understandings of the world, or, in other words, suspending and Epoché.

Key words: "All the dharmas are empty"; methodological operation; suspending; Epoché

- Nishitani Keiji's Nihilism and his Critique of Heidegger

LIAO Chin Ping

Abstract: After World War II, a nihilistic atmosphere abounded in Japan. Facing this reality and working within the confines of his circumstances, Kyoto School philosopher Nishitani Keiji began to investigate the historical development of European nihilism in order to find a way to overcome it. In its historical development, Nietzschean and Heideggerian nihilism were the most important because they both cannot be separated from human existence. The only difference between them is that the former came from existence, the latter came from ontology. However, before Nishitani, it was Tanabe that criticized Nietzsche and Heidegger. From the standpoint of the philosophy of metanoetics, Nietzsche's and Heidegger's respective philosophies were regarded as wise. However, their analyses of nihilism could not escape the standpoint of rational subjectivity completely, so that they could not make it to "negativity as affirmation," i.e., "nothingness as beings." Tanabe's understanding offered a guidance for Nishitani's critique of Nietzsche and Heidegger, which was then shown in Nishitani's contention that Nietzschean and Heideggerian nihilism could be overcome through emptiness.

Key words: Nietzsche; Heidegger; nihilism; existence; emptiness

- Preliminary Study about the Level of the Body and Mind in *Zhuangzi*:

On Leveling All Things as the Center

BAO Lei

Abstract: The passage mainly discusses the level of the relationship between body and mind in the seven chapters of *Zhuangzi*. Around the numerous words about consciousness or subjectivity in *On Leveling All Things*, such as Wu, Wo, mind, Zhen Jun, Zhen Zai, etc., this article tries to find out the hierarchical relationship between them when used in *Zhuangzi*, then with the help of the level of the body and mind, interprets the thought about "butterfly dream" and "the transformation of material" in *On Leveling All Things*. The whole article will be divided into the following three parts: first of all, analysis the basic concepts in *Zhuangzi*, for example, Wu, Wo, mind, Zhen Jun, Zhen Zai; secondly, the possibility of establishing the level of body and mind with the help of the eight consciousnesses of mind-only school, and the difference between mind-only school and *Zhuangzi*; third, with the help of the level of body and mind, interpreting the thought about "butterfly dream" and "the