

符号奠基问题及其解决策略*

夏永红 李建会

〔摘 要〕塞尔对强人工智能的批判实际上提出了这样的问题：如果没有外部观察者的解释，形式符号系统的句法属性或物理属性是否能够充分决定语义内容？虽然塞尔给出了否定的答案，但这个问题却启发了人工智能的研究者去研究物理系统内的符号如何能够自动地获得意义的问题。哈纳德把这个问题概括为“符号奠基问题”。为了解决这个问题，目前大致有内在奠基、因果奠基、指号学奠基和基于行动的语义学等几种主要方案，但它们的解法都存在不同程度的缺陷。如何将不同进路的人工智能架构糅合起来，既免于各种进路的缺点而又能发挥其优势，以更好地解决符号奠基问题，是人工智能哲学中的值得研究的方向。

〔关键词〕符号奠基问题 人工智能 内在奠基 因果奠基 指号学 语义学

〔中图分类号〕B80-0

一、引言

经典的符号人工智能建基于纽维尔和西蒙的物理符号系统假说：“对于一般智能行为，物理符号系统具有的手段既是必要的，也是充分的。”（Newell & Simon, p. 116）因为数字计算机就是典型的物理符号系统，所以从这个假说可以得出一个支持强人工智能的推论：复杂到一定程度的数字计算机可以具有通用智能。这个论断激起了很多哲学家的反驳，其中以塞尔（J. Searle）的“中文屋思想实验”和“观察者相关论证”最为有名，它们试图表明：形式符号系统的句法属性或物理属性对于语义内容是不充分的。也就是说，形式符号系统本身并不能真正理解符号的意义，所有的意义都是被外部观察者或设计者所赋予的。为了反驳塞尔的论点，强人工智能的研究者就必须构想或设计出一个可以自主地获取意义或产生语义内容的人工系统。认知心理学家哈纳德（S. Harnad）后来将这个难题概括为符号奠基问题（symbol grounding problem，以下简称SGP）。

SGP是人工智能领域的一个经典难题。后来的研究者们相继提出了一系列与SGP内容略有差异的问题版本，比如表征奠基问题（representation grounding problem）、概念奠基问题（concept grounding problem）、内在论陷阱（internalist trap）和符号锚定问题（symbols anchoring problem）等。这一问题的重要性从中可见一斑。（Taddeo & Floridi, 2005, p. 421）相比于塞尔论证的浓厚哲学思辨意味，SGP兼具实证和哲学的内涵，在人工智能领域，它与机器学习中的模式识别、锚定问题

* 本文系国家社科基金“认知科学对当代哲学的挑战”（编号11AZD120）和中央高校基本科研业务费专项资金资助（编号SKZZB2015040）的阶段性成果。

(anchoring problem)、语言奠基 (language grounding) 等密切相关; 在哲学领域, SGP 在更大范围内涉及到了认知科学哲学、心灵哲学甚至语言哲学中的众多核心问题, 引出了篇幅浩繁的研究文献。在讨论 SGP 的过程中, 大量的研究者提出了各自的解决策略, 但迄今为止并没有一种完全令人满意的策略。在本文中, 我们力图对 SGP 的解决廓清一些基本问题。我们将首先介绍哈纳德对 SGP 的分析, 继而对包括哈纳德在内的各种奠基策略作出分析。在评析各种奠基路径得失的过程中, 我们将归纳出解决 SGP 的三条标准。

二、哈纳德对符号奠基问题的回答

在塞尔的中文屋思想实验的基础上, 哈纳德假想了两个类似的例子。在第一个例子中, 一个人将中文作为第二语言来学习, 但他只能求助于一本双汉词典。这种学习类似于破译一种古代语言, 虽然难度很大但并非没有可能, 但它的可能性奠基于学习者的第一语言和在真实世界中的经验和知识。在第二个例子中, 这个人同样也只能求助双汉词典, 但他将中文作为第一语言来学习, 从而排除任何预先的语言和经验。哈纳德认为, 后一种学习过程就像从一个无意义的符号转向另一个无意义符号的连环圈, 完全不可能学会汉语了。心灵的纯粹符号模型与此类似, 虽然这些符号可能是有意义的, 但它们的物理外形和句法属性都不能提供符号的语义线索, 因此它们所处理的符号, 最终对于能动者而言都是毫无意义的。(Harnad, 1990, p. 339) 这样一来, 如何摆脱从符号到符号的无意义循环, 从而让符号与能动者所嵌入的环境关联起来, 就成为一个关键问题, 这就是所谓的 SGP:

形式符号系统的语义解释如何可以内在于系统, 而不是依赖于我们头脑中的意义? 那些仅仅基于它们任意的形状而被操作的无意义的符号标记, 如何奠基于其他事物而非另外一些无意义的符号? (ibid., p. 335)

要直接解决这个问题, 就需要让形式符号系统发展出一种自动的语义能力。然而, 哈纳德实际上也承认形式符号系统对于语义的不充分性, 所以他改换了这个问题, 没有将智能能动者设定为一个形式符号系统, 而是转而设计了一个杂合的非符号/符号系统, 在保留了传统符号系统的基础上, 补充了一个非符号的神经网络模型。在这个杂合系统中, 符号系统中的意义最终奠基于非符号系统之上。它兼具符号主义与联结主义的优点, 既可以句法地操作符号, 又可以奠基符号的意义。

基本的奠基过程可以分为三个阶段: 图像化 (iconization)、区分 (discrimination) 和辨识 (identification)。需要注意的是, 虽然哈纳德并未使用图像化这一术语, 后来的研究者一般都用它来指代图像表征形成的过程。在图像化阶段, 感觉器官接受远端对象的数据输入而形成图像表征 (iconic representations), 它保留了感觉映射的原始轮廓, 各种图像是连续和混合的。比如, 马匹及其环境在神经系统中所形成的图像表征, 仅仅是一团图像而已, 我们并不能从中识别出马匹的特征, 从而将它从周围环境中区分出来。

在区分阶段, 基于这些图像表征之间的相似度, 我们将会形成对它们的“相同/差异”的判断, 最终把握图像表征之间的不变特征, 从而形成类别表征 (categorical representations)。它是将图像表征选择过滤之后的输出, 经过了初步的分类整理, 我们可以从中区分出不同的事物。比如, 我们通过区分不同马匹之间以及马匹与其他事物之间的相同和差异, 从而在连续的感觉图像中识别出马, 形成马的类别表征。

在辨识阶段, 我们“能够对一个类别的输入指派某种独特的 (通常是专有的) 回应, 比如一个名称, 将它们全部作为对等的或某些方面不变的”。(ibid., p. 341) 因此, 辨识的功能在于将输入与

某种特定的符号关联,从而形成原子符号表征。辨识与区分的差异在于,区分并不需要知道事物是什么,它只需要把握到图像表征之间的相似与差异,并从中抽出某一类图像的不变特征;而辨识则可以将某一固定的名称与这种不变特征相关联,比如,在我们区分出马的类别之后,再将“马”这一符号赋予马的类别。

在这三种表征中,图像表征和类别表征都是非符号表征,都产生于杂合系统中的人工神经网络之中。在哈纳德看来,这两种表征都还不具有意义。图像表征相当于照相机中的照片,它们的意义都是衍生的而不是内禀的,因为最终还是依赖于观察者的解释;而类别表征只是一种消极的分类,并没有被真正命名。到了奠基的第三个阶段,类别表征经过初步辨识而被命名,形成的原子符号表征,才有了真正的意义。

然而,最初被图像和类别表征所奠基的原子符号,并不满足符号系统的要求。在哈纳德看来,符号系统中的表征需要满足复合性(compositeness)与系统性(systematicity)两个要求,从而可以被组合和重组,进而被规则化地带入命题之中。所谓复合性指的是原始的原子符号标记可以组合成符号-标记串,而系统性指的是整个符号系统及其部分,都是语义上可解释的(semanticly interpretable),也就是说,句法可以被系统地赋予意义,以代表一个对象,描述一个事态。(Harnad, 1990, p. 336)而图像和类别表征都是基于模拟的神经网络信号,无法分解、组合和重组,最终无法进行形式的句法操作,不能赋予它们一种系统的语义解释。因此,仅仅基于神经网络而形成的原子符号,并不具有符号系统所特有的复合性与系统性的特征。

为了满足这两个要求,在产生图像表征和类别表征的神经网络系统之上,还需要一个符号系统,它不仅可以将原子符号指派到特定的类别之上,还可以对其进行组合和重组。在这种符号系统之内,形成了真正的符号表征。它是由原子符号构成的复合符号,满足复合性和系统性的要求。这样,当一套基础符号被奠基之后,符号系统中的符号并不需要全部被奠基,因为其余的符号可以通过它们的组合而产生,并可以继承其内在奠基属性。比如“斑马”这个符号,是通过“条纹”和“马”组合而成的,即便我们没有见过斑马,也可以知道它的意义,只要“条纹”和“马”已经是被奠基的。

哈纳德将这种被奠基了的符号系统称之为专用符号系统(dedicated symbol systems),相比于一般符号系统,它对符号标记具有更多的约束。在一般符号系统中,只有纯粹的句法约束。但在专用符号系统中,除了句法约束之外,还存在着与非符号表征相关的模拟约束。也就是说,对符号的操作不仅基于符号标记的任意形状,而且基于非任意的奠基了基础符号的图像和类别表征的“形状”,也即对象的感觉映射的感官不变量(sensory invariants)。

因为这种双重约束,在哈纳德所设想的自下而上的符号奠基过程中,在远端对象、近端感觉映射和内部的表征生成之间就存在着充分的因果关联,符号的奠基过程存在着因果力的作用。也就是说,整个认知过程是执行依赖的,而不像形式符号系统那样是独立于物理执行的。这样,哈纳德就回应了塞尔的诘难,在他的杂合系统中,符号的意义不再是任意指定的,不是源于外部解释者的头脑,而是内在于专用符号系统。

这里需要注意的是,哈纳德的专用符号系统虽然采用了人工神经网络,但它实际上最终基于对机器人能力的设计。(ibid., p. 345)哈纳德后来进一步澄清了机器人架构对于奠基的必要性。在他看来,一个奠基的系统,它的符号能力必须奠基于机器人能力之中,而不是通过外部的解释的投射,因此专用符号系统需要同时具有机器人和符号能力,以最终通过总体图灵测试(Total Turing Test)。所谓总体图灵测试,也即一个人工系统无论在符号能力上,还是在行为能力上,与人类系统难以分辨,我们才能说这个人工系统具有与人类系统同等意义的智能。(Harnad, 1993, p. 16)哈纳德认为,只

有通过了总体图灵测试,才能免于塞尔的中文屋论证的诘难。因此,机器人感觉运动能力对于解决SGP至关重要。由此可见,哈纳德为解决SGP而设计的智能系统架构实际上融合了机器人学、神经网络和符号主义三种不同的人工智能进路。

三、符号奠基的两种路径:内在奠基与因果奠基

后来的研究者并不完全赞同哈纳德的方案,人工神经网络的支持者试图舍弃哈纳德杂合模型中的符号部分,而情境进路的支持者则否认联结主义模型可以奠基符号。从不同的人工智能进路出发,研究者们提出了种类繁多的奠基策略。塔迪奥·(M. Taddeo)和弗洛里迪(L. Floridi)根据表征在奠基过程中的不同作用,将符号奠基策略划分为表征主义路径、半表征主义路径和反表征主义路径。(Taddeo & Floridi, 2005, p. 420)而齐姆克(T. Ziemke)则基于能动者以何种方式钩连(hook to)外部世界,区分了认知主义和生成范式两种解决策略。(Ziemke, p. 179)但这两种划分都没有把握不同策略背后的哲学差异。我们在此将采纳查尔莫斯(D. J. Chalmers)基于哲学语义学中的内在论和外论之分而提出的一个划分:因果奠基(causal grounding)和内在奠基(internal grounding)。在查尔莫斯看来,关于表征如何获得真实的意义这一问题,依赖于对意义的看法是外论的还是内在论的。如果采取外论立场,就会认为符号的意义是外在的、延展的,表征只能奠基于能动者与外部环境的感受运动交互之中,这就是所谓的因果奠基。而如果采取内在论的立场,认为符号的意义最终是内禀的,我们就必须保证表征具有充分的内部结构以承载内禀内容,因此奠基的方法就在于找到一种内部介质,它可以占有比计算标记更丰富的表征内容。查尔莫斯将这种奠基称之为内在奠基,它的代表就是基于神经网络的联结主义,哈纳德的奠基策略也可以归属到这个范畴。(Chalmers, p. 43)然而,查尔莫斯并未对两种奠基进行更详细的阐述,我们在此将基于他所提供的这个类型学划分,联系一些代表性策略来分析它们各自存在的困境。

(一) 内在奠基的困境

杰克逊(S. A. Jackson)和夏基(N. E. Sharkey)基于联结主义提出了他们的奠基策略。在他们看来,符号主义的认知系统是通过句法引擎驱动的,也就是说,认知是通过执行纯粹句法的表征或符号操作而实现的。但句法引擎所操作的心理表征,却面临着一个内在论陷阱:在内在论的心灵理论中,表征不过是心理事件之间的交互,缺乏指称意义,不能与外部事物钩连。他们所说的内在论陷阱或钩连问题,是SGP的另一种表述,指涉的都是如何让最基本的原子表征与其所表征的世界相联结的问题。在他们看来,虽然句法引擎并非一定不能逃脱内在论陷阱,但为了更好地解决这个问题,就必须使用非符号表征,而这是句法引擎所不能做到的。

句法引擎诉诸的是原子表征与外部对象之间的句法结构相似性关系,而联结主义诉诸的则是原子表征与感觉信号之间的空间相似性关系(spatial structure similarity relations),所以他们也将联结主义的认知机制称之为空间引擎(spatial engine)。因为在联结主义中,表征介质就是不同神经元之间的联结权重,所以他们也将空间引擎形成的表征称之为权重表征(weight representations)。凭借与感觉信号的空间结构相似性,权重表征具有丰富的表征资源,可以被充分奠基。它们是原子式的、语境独立的,在加工过程中并不变化,可以进一步组合成分子式的、语境依赖的单元表征(unit representation)。比如,两个权重表征“马”和“条纹”可以组合成单元表征“斑马”。单元表征通过构成它们的权重表征与世界钩连。(Jackson & Sharkey, p. 77)

与哈纳德的杂合模型相比,他们所代表的这种联结主义路径,完全抛弃了符号系统,从最基础的权重表征到单元表征,其奠基过程都采用了人工神经网络系统。他们认为,在哈纳德的杂合模型中,

原子表征之间的句法结构并不足以决定由其组合而成的分子表征的语义,因为这种组合是语境依赖的,它们可以组合成多种不同语义的分子表征。比如,“条纹”和“马”不一定组合成“斑马”,也可能组合成“披着条纹马鞍的马”,这都是依具体的语境而定。(Jackson & Sharkey, p. 76)可见,仅仅凭借句法结构相似性并不足以充分决定分子表征的语义内容,它还需要外部观察者的语义赋值,因此并不能真正解决SGP。而在他们的方案中,无论是原子表征还是分子表征,都是诉诸空间结构相似性来表征世界,它们比纯粹基于句法引擎的表征拥有更多的表征资源,从而可以让表征与感觉信号充分勾连。

然而,完全采用人工神经网络真的就能解决SGP吗?我们先来看塔迪奥和弗洛里迪提出的一条符号奠基策略的标准,即零语义承诺条件(zero semantical commitment condition):一个成功的符号奠基策略应该让该人工智能系统具有自主生成语义内容的能力,而不需要依赖于预先或外部的语义植入。该系统可以预先设计好认知结构、模式和规则,但是在认知过程中,不能借助先天或外在的语义资源。(Taddeo & Floridi, 2005, p. 421)只有满足这个零语义承诺条件,一个符号奠基策略才是成功的。当然,零语义承诺条件并非唯一的标准,我们将在后续的讨论中,继续补充其他的评价标准。

在塔迪奥和弗洛里迪看来,当前几乎所有符号奠基路径,都无法满足零语义承诺条件。以人工神经网络为例,包括当前最流行的深度学习在内,它们的学习方式都可以分为监督学习和非监督学习。但按照零语义承诺的标准,两种学习都无法自主生成语义。在监督学习中,它需要一个包含了输入数据(感觉映射)和期望得到的输出值(标签)的训练样本,最终实现对数据的分类处理,这显然需要外部的语义资源;而非监督学习虽然没有任何训练样本,但仍然需要内置偏好(built-in biases)和特征探测器(feature-detectors)来获得所期待的输出,因为存在输出审核的问题,同样需要外部语义资源。(ibid., p. 424)可见,无论是监督学习还是非监督学习,都需要外部观察者或明或隐的语义赋值。最终,人工神经网络的语义内容还是取决于我们对这个网络的解释,语义内容依然是寄生于观察者的头脑中的。就此而言,包括杰克逊和夏基的方案在内的基于联结主义的符号奠基路径,因为在学习过程中依然需要外部的语义资源,起码到目前为止还并不完全成功。

(二) 因果奠基的困境

如果说内在奠基体现的是人工智能的联结主义进路,那因果奠基则代表了人工智能的具身和情境进路,它诉诸人工能动者与环境的交互,认为意义就诞生于这种交互之中。就在哈纳德提出符号奠基问题的同年,布鲁克斯提出了所谓的“物理奠基假说”(Physical Grounding Hypothesis)。在布鲁克斯看来,此前的基于计划和表征的人工智能研究已经误入歧途,有必要以新的范式取而代之。他提出,新的人工智能将基于物理奠基假说:“建立一个智能系统,必须要具有奠基于物理世界中的表征。”(Brooks, 1990, p. 5)一旦这种“表征”实现,传统的符号表征就毫无必要了。在这种系统中,能动者通过一套传感器和驱动器与世界进行联结,实时的动态交互取代了静态的输入和输出。它具有两个最为关键的特征:具身性和情境性。一方面,机器人具有身体,可以通过其感觉运动器直接经验世界,从而因果地与世界关联;另一方面,机器人是情境化于世界中的,它所处理的是世界此时此刻对系统行为的直接影响,而不是抽象的推理或对世界的建模。布鲁克斯将这种观念称之为无理性智能或无表征智能。在他看来,智能能动者无需像经典人工智能那样建立关于世界的模型,相反,“世界就是它自己的模型”,智能行为可以通过机器人与环境之间的交互来实现,符号表征完全没有必要。(Brooks, 1991, pp. 584-585)

布鲁克斯的无表征智能模型因为不使用任何符号表征,似乎避免了SGP,然而,如果涉及到高级认知行为,尤其是涉及到不同能动者之间的语言操演(language performance),那么就不可避免地会

使用符号,就必须再一次处理SGP。其他一些因果奠基进路试图对语言操演中的语言符号进行奠基,它们像布鲁克斯那样采用情境机器人学的人工智能方案,并且同样放弃了表征,只不过布鲁克斯的物理奠基诉诸的是能动者与物理环境的交互,而它们则诉诸能动者与社会环境的交互。根据弗洛里迪的归纳,这种非表征进路的符号奠基大致有两种类型:一是基于通讯的模型,它将学习能动者与教师能动者在一个共同环境中的交互,通过前者对后者的言语行为的模仿,来学习共同语言;二是基于行为的模型,它基于语言习得理论,将语言视为能动者与环境和其他能动者交互的工具,也就是说,语言是为了满足能动者的需要而操纵环境的一种方式。(Taddeo & Floridi, 2005, pp. 439 - 441)这两种奠基模型都将意义奠基于外部的通讯或行为模式中,而不是归诸于行动者的内部状态。

然而,因果奠基进路虽然逃脱了内在论陷阱,却面临着齐姆克所说的外在论陷阱(externalist trap):“如果仅仅是世界的此时此地决定一个能动者的行为,也就是,如果能动者仅仅是对当下的环境做出回应,能动者最好被描述为受环境木偶操控者控制的,而不是自主的。”(Ziemke, p. 183)这样的能动者其行为很可能是被动地由环境决定的,而不是由自身的经验引致的。这也意味着,如果这个能动者的行为涉及意义,那么其语义资源最终可能是由外部植入,而不是自主生成的。因此,纯粹的因果奠基很可能也会像内在奠基那样违反零语义承诺条件。

要摆脱这种外在论陷阱,就需要将能动者的行为奠基到其内在的经验中。为了达到这个目标,一方面可以诉诸能动者对内部状态和记忆的使用,一方面可以诉诸能动者的学习机制,可以通过它获得新的经验。这些都可以增强能动者的自主性,降低环境和外部设计者对能动者行为的决定作用。这种思路让我们不得不回到内在奠基的思路,即诉诸一个复杂的人工神经网络来奠基符号。因此,如果我们将因果奠基和内在奠基的思路结合起来,或许可以同时摆脱内在论和外在论陷阱的困扰。这就引导我们去构想一种更综合的人工智能架构。

四、符号奠基的指号学路径

除了内在奠基和因果奠基路径,还存在着一种指号学(semiotics)奠基路径,它通过引入皮尔士的指号学理论,重新界定符号这一概念,试图构造一个指号学符号系统来解决SGP。费策尔(J. H. Fetzer)较早地将皮尔士指号学引入到人工智能哲学中。纽维尔和西蒙将心灵视为一个物理符号系统,而费策尔则认为,心灵应当定义为一个指号学系统。两种系统的差异在于,符号系统中的符号对于机器的使用者(user)而言是有意义的,而指号学符号对于系统的使用(use)而言是有意义的。(Fetzer, p. 35)也就是说,指号学系统中存在符号的使用者即解释者,可以通过解释者奠基符号与对象之间的关系,因此指号学系统从定义上就是包含了奠基关系的。但费策尔的主要旨趣在于批判计算主义的观点,并没有对SGP做出直接探讨。后来的斯蒂尔斯(L. Steels)和沃格特(P. Vogt)等人独立提出了关于指号学符号的理论,并用它来解决SGP。

在沃格特看来,只要将认知过程解释为形式系统中的符号操作,就不可避免地会面临框架问题(frame problem)和SGP的困扰,而认知的具身路径虽然避免了这些问题,但却不能处理高级认知功能。沃格特致力于综合两种范式,认为符号可以从能动者与环境的交互作用中被奠基。(Vogt, 2002, p. 430)但他认为这种符号不是物理符号系统意义上的符号,而只能是皮尔士指号学意义上的符号。在皮尔士指号学中,指号分为形式(指号介质)、意义与所指三个元素。指号学符号从定义本身就包含了意义,它是形式与所指之间的功能关系,通过能动者与所指之间的交互作用而确定。如果将指号学符号系统作为心灵模型,那么它的意义就是内在的,而不是外部赋值的。这样,对于指号学符号系统,SGP实际上是不相关的,它可以转换为一个如何在工程学上实现指号学系统的问题,也就是物理

符号奠基问题。(Vogt, 2002, p. 435)

沃格特认为SGP和锚定问题(anchoring problem)一样,都是要构造和维持符号与实在之间的锚点(anchors)。对于一个指号学系统,解决SGP就是在形式与所指之间构造和维持锚点。然而,形式与所指之间的锚定是通过所指与意义、意义与形式的两个锚定而间接实现的。在能动者与所指的交互过程中,通过知觉和分类而产生了意义,从而实现了所指与意义之间的锚定;而在能动者之间的通讯过程中,约定了特定的符号来代表某种特定的意义,进一步实现了意义与形式的锚定。(Vogt, 2003, pp. 110-111)体现在人工智能工程中,所指与意义之间的锚定可以通过机器人与环境的交互来实现,这种机器人可以区分不同的感官经验,并辨识其特征;而形式与意义的锚定则是通过机器人对语言的使用来实现的。

那么,语言又是如何被使用的呢?斯蒂尔斯和沃格特采用了一种适应性语言游戏(adaptive language games)的方案来解决这个问题。在这个游戏中,两个机器人言者(speaker)和听者(hearer)共享了一个环境。其中言者给它感知的某个对象命名,通过与听者的通讯,听者获知了这个新的名称。虽然听者并不知道这个名称所指称的对象,但它可以向言者指出这个名称所可能指称的对象。通过从言者那里获知的反馈,在不断的试错之后,听者最终可以确定这个名称所指称的对象。(Steels & Vogt, p. 476)通过不断地进行这个游戏,机器人之间就可以逐渐发展出一种共同的指号学符号系统。最终,一个能动者种群可以在与世界和他者的交互作用中,自主地奠基一个符号系统。

相比于内在奠基,指号学奠基方案强调了能动者与世界的动态交互作用;相比于因果奠基,它保留了能动者对世界的表征。因此我们将它视为一种与内在奠基和因果奠基完全不同的奠基方案——虽然它在工程学采取的人工智能方案方面与因果奠基相似,但它所依据的指号学理论却不同于因果奠基所依赖的主流情境认知理论。从理论上说,指号学奠基方案是极为成功的,它吸收了前两种方案的优点,同时又没有违反零语义承诺。也因此,斯蒂尔斯认为,SGP已经解决了,余下的不过是一些技术上的改进。(Steels, p. 244)但在我们看来,它还存在着两个方面的缺陷。首先,在人类的语言游戏中,除了命名之外,还涉及到设问、反讽、戏仿、命令等难以计数的种类,但猜谜模型却无法完全捕捉到这些语言游戏的全部特征,在更为复杂的语言游戏中,机器人能否捕捉到它们的语用特征,是存在疑问的。这可能导致那些更为语境依赖的符号难以被奠基。其次,在这些被奠基的基本符号之间,也可能缺乏恰当的句法规则来将这些符号组合起来,从而无法构成为一个完整的符号系统。正如哈纳德所强调的,一个符号或表征应当可以被组合或重组,最终可以被规则化地带入命题中。由此,我们可以在零语义承诺条件之上补充第二个符号奠基的标准:被奠基的符号必须具有句法特征和语用特征,从而使其具有组合性、系统性和语境性。斯蒂尔斯和沃格特的指号学奠基策略要满足这一标准,无疑还需要在技术上进一步完善,一方面让人类生活中更多的语言游戏融入其中,另一方面也需要引入句法引擎并让其发挥更多的作用。这实际上要求我们构想一种基于指号学理论、更加情境化和具身化,但同时也融合了符号主义的人工智能架构。

五、符号奠基问题与自然化语义学

SGP实际上还可以进一步区分为两个层面,其一涉及语义内容的产生机制,其二涉及自主语义能力的实现。前者是一个语义学问题,后者则是一个工程学问题。在工程学层面上,它关涉的是设计一个什么样的智能模型或系统,从而可以让其不借助外部的解释者而自主地获得语义内容——我们前面所述的几种方案都是在设计这样的人工系统。而在语义学层面上,SGP要求我们对语义内容产生的自然机制做出更为深刻的理论解释。在克里斯蒂安森(M. H. Christiansen)和蔡特(N. Chater)看来,

SGP 实际上是在提供一种原子符号如何指称的理论, 它和意义理论都是要说明“非指称的原子标记或系统状态如何成为系统中的语义基元”。(Christiansen & Chater, p. 155) 而毕勒卡 (K. Bielecka) 也指出, 符号奠基问题实际上和指称的因果理论之间, 面临的都是一个共同的难题, 即意义的来源问题。(Bielecka, 2016, p. 77) 在这个意义上, 语义学层面的 SGP 实际上等价于自然化语义学问题, 它们最终都试图澄清意义产生的自然机制。这两个层面的问题可以相互促进, 如果我们在工程学层面上设计新的人工智能架构来检验 SGP 的解法, 那也会在语义学层面上丰富我们对语义产生的自然机制的认知, 更新传统的自然化语义学理论, 并进一步对 SGP 在工程学层面的解决构成正向反馈作用。

塔代奥和弗洛里迪意识到了 SGP 在语义学和工程学层面上的区分。他们所倡议的实践解法 (praxical solution) 首先提出了一种新的自然化语义学即所谓的基于行动的语义学 (Action-based Semantics, AbS), 随后又在工程学层面提出了一种双机人工能动者 (two-machine artificial agents, AM²) 的工程架构来实现 AbS, 从而试图解决 SGP。

AbS 是一种自然化语义学, 但它与传统自然化语义学不同, 并不是将意义奠基于任何一种自然的因果或目的论关系之中, 而是奠基于人工能动者的行动操演所引致的内部状态。弗洛里迪设计了两个抽象层次 (levels of abstraction, LoA), 它需要一个双层构架的人工能动者 (AM²) 来实现。AM² 的底层 (M1) 处理的是行动如何关联于世界 (实际上是内部状态如何关联于行动), 而顶层 (M2) 则处理符号如何关联于行动。因此, AM² 的行动对应着其内部状态, 而这种内部状态又有一个符号与其对应。符号最终被奠基于 AM² 在世界中的行动。需要说明的是, 为了免于预先的语义植入而违反零语义承诺条件, 这里所说的行动是目的论无涉的, 也就是说, 它并不涉及行动的目的, 而仅仅涉及行动的操演 (performance)。但是这种被奠基的符号有可能成为维特根斯坦所说的私人语言, 为此, 塔代奥和弗洛里迪设计用一个 AM²s 种群来发展其通讯技能, 最终创造共享的语汇。它类似于斯蒂尔斯和沃格特所提出的猜谜游戏, 通过这种社会协作, 符号在演化过程中的得以避免私人语言的困境。(Taddeo & Floridi, 2007, p. 384)

他们的这种解法虽然提出了新的自然化语义学, 但在工程学层面大体上可以归入到因果奠基的范畴。相比于其他奠基进路, 它似乎存在着更多的问题, 比如它的符号是非组合性的, 无目的的行动概念将导致粗率的行为主义错误, LoA 关系仅仅是一种简单的映射关系, 等等。(cf. Bielecka, 2015, pp. 86 – 88) 我们在此不拟对这些问题一一评论, 主要关注其中最重要的一个问题。由于 AbS 中任意的符号总是与某种实存事物相关联, 因此它就不能意指任何非实存的事物, 从而人工能动者将无法进行错误表征。这里的问题在于, 如果 AbS 语义模型缺乏错误表征的能力, 那么这将意味着它并不具有正常的表征能力。正如德雷斯基 (F. Dretske) 令人信服地指出的那样, 真正的表征能力总是意味着有进行错误表征的能力。(Dretske, p. 17) 它对于人工能动者的自主学习行为而言是一个必要条件, 如果没有错误表征的能力, 那就意味着人工能动者的学习行为只是受到一种必然性规律的驱动, 而不是出于自主的能力驱动。因此, 借鉴自然化语义学中的相关讨论, 我们可以补充符号奠基策略的第三条标准: 人工智能系统应该有错误表征的能力。

六、结语

当前基于各种人工智能进路的奠基策略, 都无法较好地解决 SGP, 这部分说明了任何一种进路都存在着理论或技术上的不足。然而, 哈纳德早在提出 SGP 的时候就已经意识到了这一点, 因此他构造了一个杂合人工智能, 既免于各种进路的缺点而又能发挥其优势。实际上, 不同进路的人工智能之间的融合, 现在已经成为一种趋势。比如, 伽米罗 (M. Garnelo) 等人提出的深度符号强化学习

(Deep Symbolic Reinforcement Learning), 它糅合深度强化学习和经典的符号人工智能的架构, 既可以避免深度强化学习缺乏抽象层次的推理的局限, 也可以通过深度学习来避免符号人工智能所必然面临的 SGP。(Garnelo et al., p. 2) 可是, 正如我们在文中所分析的, 深度学习远没有解决 SGP, 并不能满足零语义承诺条件。那么, 如果我们进一步糅合情境机器人学以及强调社会语境的指号学方案, 设计出一种新的人工智能架构, 是否可以克服现有进路的种种局限, 并满足解决 SGP 的三条标准? 这种杂合人工智能, 或许正是解决 SGP, 朝强人工智能前进的正确方向, 值得我们进一步探究。

参考文献

- Bielecka, K., 2015, "Why Taddeo and Floridi did not solve the symbol grounding problem", in *Journal of Experimental & Theoretical Artificial Intelligence* 27 (1).
- 2016, "Symbolgrounding problem and causal theory of reference", in *New Ideas in Psychology* 40, Part A.
- Brooks, R. A., 1990, "Elephants don't play chess", in *Robotics and Autonomous Systems* 6 (1-2).
- 1991, "Intelligence without reason", in *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, San Mateo, CA: Morgan Kaufmann.
- Chalmers, D. J., 1992, "Subsymbolic computation and the Chinese Room", in J. Dinsmore (ed.), *The Symbolic and Connectionist Paradigms: Closing the Gap*, Lawrence Erlbaum.
- Christiansen, M. H. & Chater, N., 1993, "Symbol grounding: the emperor's new theory of meaning?", in *Proceedings of the 15th Annual Conference of the Cognitive Science*, Society University of Colorado.
- Dretske, F., 1986, "Misrepresentation", in R. Bogdan (ed.), *Belief: Form, Content, and Function*, Oxford University Press.
- Fetzer, J. H., 1990, *Artificial Intelligence: Its Scope and Limits*, Kluwer.
- Garnelo, M., Arulkumaran, K. and Shanahan, M., 2016, "Towards deep symbolic reinforcement learning", *arXiv*: 1609.05518 [Cs], <http://arxiv.org/abs/1609.05518>.
- Harnad, S., 1990, "The symbol grounding problem", in *Physica D: Nonlinear Phenomena* 42 (1).
- 1993, "Grounding symbols in the analog world with neural nets", in *Think* 2 (1).
- Jackson S. A. & Sharkey, N. E., 1996, "Grounding computational engines", in *Artificial Intelligence Review* 10 (1-2).
- Newell, A. & Simon, H. A., 1876, "Computer science as empirical inquiry: symbols and search", in *Communications of the ACM* 19 (3).
- Steels, L., 2008, "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 University Press.
- Steels, L. & Vogt, P., 1997, "Grounding adaptive language games in robotic agents", in *Proceedings of the Fourth European Conference on Artificial Life*, MIT Press.
- Taddeo M., & Floridi, L., 2005, "Solving the symbol grounding problem: a critical review of fifteen years of research", in *Journal of Experimental & Theoretical Artificial Intelligence* 17 (4).
- 2007, "A praxical solution of the symbol grounding problem", in *Minds and Machines* 17 (4).
- Vogt, P., 2002, "The physical symbol grounding problem", in *Cognitive Systems Research* 3 (3).
- 2003, "Anchoring of semiotic symbols", in *Robotics and Autonomous Systems* 43.
- Ziemke, T., 1999, "Rethinking grounding", in A. Riegler et al. (eds.), *Understanding Representation in the Cognitive Sciences*, Springer US.

(作者单位: 夏永红, 华南师范大学公共管理学院; 李建会, 北京师范大学哲学学院)

责任编辑: 朱葆伟