

皮尔士符号学及其对人工智能发展的启示

陈 曦

(深圳大学 法学院,广东 深圳 518061)

摘 要: 皮尔士将符号理解为符号化三元关系中的一元。通过阐述符号、对象和解释项之间的互动,他将思维过程转译为符号化过程,并认为所有思维皆为符号,甚至人也只是一种根据推理规则的符号。一方面,皮尔士的符号学表明了人类思维的人造性;另一方面,则因其主张人类具有机器不具备的高阶自控能力而体现出人机分野。皮尔士的符号学不仅为发展人工智能奠定了理论基础,而且划定了实践界限。其符号学本身兼顾科学精神和人文关怀的内在平衡,可视为指导人工智能未来发展的基本准则。

关键词: 皮尔士;思维;符号;符号化;人工智能

中图分类号: B804

文献标识码: A

文章编号: 1008-407X(2018)06-0107-05

一、引言

在诸多实践领域,人工智能已获得实质运用。据测算,2015年至2018年间,全世界工业机器人的增量为130万台,服务机器人的总量则在2018年达致3100万台^[1]。同时,由于概率统计方法的日趋精准、可用数据的大量增长、廉价海量计算能力更为可及等因素的助推,人工智能对人类社会的政治、经济、伦理等方面的影响不仅越发明显,而且大有不可逆转之势。对此,不仅普通大众对人工智能津津乐道,而且理论者们亦持续对人工智能的历史、现状、前景、界限等发展问题做了细致评估。事实表明:一方面,人工智能的诞生和发展有历史必然性;另一方面,人类具有掌控人工智能的内在诉求。面对这一深刻变革,人类无论是谋求发展还是抑制人工智能,都必须对其恰当评估,准确定位它与人类之间的关系。

关于人工智能的哲学讨论并不鲜见。有论者甚至认为,由于人工智能研究与哲学在思维、意义、推理、语言、行动和理性等问题上具有类似旨趣,因此,前者在很大程度上就是哲学研究^[2]。然而,人工智能哲学在本质上并非某种特定的人工智能理论。哲学并不试图像一般学科那样对其关注问题提供答案,而在于对世界现象提供值得批判反思的问题、方法和视域。哲学家必须在哲学的学科性和超越性之间找到平衡^[3]。这

意味着,人工智能哲学既要人工智能研究的特定问题有所阐发,又要对人工智能之本质有所阐述。其中,人类与机器在思维上的异同关系,涉及人工智能本质以及由此透视出的人类对其自身理解的哲学面相,值得深入研究。虽然这一问题在认知科学中已获得广泛关注,但皮尔士(C. S. Peirce)的符号学却与该问题具有更为紧密的哲学联系,而晚近认知科学的发展趋势也在某种程度上具有皮尔士符号学的“影子”。

基于此,本文拟从3个部分展开论述。第一,本文将从皮尔士符号学切入,通过阐述皮尔士眼中的符号化过程,表明人类思维的人造本质。同时,基于符号优先的这一前提,在皮尔士的理论框架下,可以得出人机趋同的结论。第二,本文将以自我控制的概念辨析为起点,以自我控制的审议性、无限性、目的性这三大特征为抓手,阐明在皮尔士的理论框架下人类与机器的本质差异是自控方式、自控目的以及自控机制上的差异。第三,本文将基于上述内容,揭示皮尔士符号学对人工智能的历史发展与未来趋势独特价值。

二、人机趋同——符号优先

笛卡尔“身心二元论”认为,心智是一种存于脑内的非物理性实体,思维则是通过心智将关于外在世界信息予以表征的精神过程,是存在于认知有机体脑内的一种结构。对此,皮尔士评论到:“当代哲学从未完

收稿日期:2017-07-25; 修回日期:2017-10-23

基金项目:国家社科基金重点项目:“全面推进依法治国的逻辑理性根基研究”(13AXZ0017)

作者简介:陈曦(1986-),男,广东湛江人,讲师,博士,主要从事法哲学、法律逻辑学、人工智能与法研究,E-mail:chenxi8601@126.com.

全摆脱笛卡尔的心智观念,即某物‘居住’在所谓的松果体之中。如今,所有人都嘲笑于此,但却又以同样的方式认为心智是内在于人之中,或是归属于人并与真实世界有关的某种东西。”^[4]

皮尔士对笛卡尔哲学的不满体现在诸多方面。比较有典型意义的是:第一,他否定第一性知识存在;第二,他否定先于符号的思维存在。在皮尔士眼中,没有符号,人类将无力思考,“那些可能被认知的思维,仅仅是符号中的思维,且无法认知的思维是不存在的。因此,所有思维必定是符号”^{[4]251}。他甚至认为,如果宇宙不是仅由符号构成,那也充斥着符号。不仅思想是符号,甚至人也如此。

符号优先于思维,思维必须借助符号实现,皮尔士的这种思想具有直接的经院哲学渊源。他对奥卡姆符号思想的借鉴,在 1867 年关于范畴分析的论文中清晰可见。一方面,这与皮尔士实用主义哲学对思维的定性——通过符号操作考量问题直接相关;另一方面,奥卡姆本身对符号的偏好则增加了两者产生理论共振的可能。皮尔士说到:“奥卡姆总认为精神概念并非存在于字面上、声音中、心智中的逻辑词项,它具有一般性的本质,即符号。概念与语词在两方面有异。第一,语词是任意强加的,而概念则是自然符号;第二,就语词所表达的任何东西,它是通过概念所直接表达的同样的东西而仅仅间接地予以符号化。”^{[4]20}

第一,符号对思维的优先是一种逻辑优先,而非时空意义上的优先。事实上,皮尔士并不关注思维究竟在时空中如何被安置,因为这种问题本身就充满了主体哲学的色彩。第二,对概念与语词的定性,既体现出符号对语言的独立性,且由于语词表达必须借助符号,因此又体现出符号对语言的优先性。第三,如果像思维这样的精神概念本质上是符号,那么对其进行逻辑分析就无需依赖于对思维究竟为何这样的本体论问题的回答,而可将研究重心转移到符号构成思维的具体运作层面。

符号对思维的优先性,不仅意味着思维并非笛卡尔意义上的实体,更意味着皮尔士眼中的符号并非思维产物,那符号究竟是什么呢?

关于这个问题,皮尔士的理解存在渐进式。在很长一段时间,他都认为符号是符号化过程三元关系中的一元。他说到:“符号是这样一种东西,一方面它与对象有关,另一方面与解释项有关……它使得解释项与对象相关,相应地它自身也与对象相关。”^{[4]332}然而,在 1906~1910 年期间,皮尔士则通过细分对象与

解释项,对其符号理论做了实质扩充。

首先,他将对象分为直接对象和动态对象。其中,直接对象是符号被使用时直接指涉的特定时空中的对象,而动态对象则是在科学探究终端“等待”人类企及的对象。前者是人类假定如是的对象,后者则是对象真实的样子^[5]。直接对象和动态对象并非两种不同的对象,毋宁说,直接对象是动态对象的不完全形态,两者处于符号化过程的不同阶段。

例如,妻子问丈夫:“今天天气怎样?”丈夫回答说:“狂风暴雨。”此处,妻子和丈夫的话语皆为符号。以丈夫的回答为考察中心,该符号的直接对象就是丈夫在回答时所认定的情形——“狂风暴雨”。然而,虽然丈夫的回答告知了妻子现在的天气是狂风暴雨,但他并未表述出具体的风向、风力、雨速、雨量等客观事态,而这就是动态解释项所代表的事物本来的样子。用皮尔士的话来表述,“直接对象是对当下天气的确定……动态对象则是对那时实际或真实天气情况的确定。”^{[4]314}

其次,皮尔士还将解释项分为 3 种。他说到“动态解释项是无论做出何种解释任何心智对某一符号的实际理解……最终解释项不在于任何心智的实际行动方式,而在于每一心智的可能行动方式。换言之,它是这一条件命题所表述的真理:‘如果这样的情形发生在任何心智上,那么该符号将决定那一心智如此这般行动……’直接解释项则在于某一符号所适宜产生的印象之属性,而非造成的任何实际反应”^{[4]315}。

第一,直接解释项是对符号的一般性理解。结合上例,丈夫话语所形成的符号的直接解释项就是常人关于狂风暴雨天的理解图示。对于一些常见现象,其图示对不同的人而言可能大致相同,但却可能并不精确。直接解释项是在排除语境后某一符号所明确表示的全部,这一概念实际上仅指符号的语义和语形维度。第二,动态解释项是由符号引起的实际效果,它是符号化过程中的阶段性理解。动态解释项与直接对象和动态对象有关:一方面,直接对象总是对给定符号所有先前动态解释项的后继解释之结果。换言之,直接对象之形成无法脱离时序在前的动态解释项;另一方面,动态解释项又对动态对象提供了某种不完全解释。第三,就最终解释项与最终对象而言,两者皆处于符号化终端,最终解释项是就动态对象所形成的最终且符合实在的理解。

显见,皮尔士眼中的符号并非静态、独立的,而是动态、过程的。由于符号根本无法脱离符号化过程而获得理解,且后者代表整个思维过程,因此,思维过程

就被随之转译为符号化过程。例言之,人类通过思维所形成的命题(最终解释项),并非是通过我们直接对客观实在(动态对象)或常人感觉(直接对象)描述所得,而是通过持续不断地解释(符号化过程)所形成的关于客观实在的一般性理解(直接解释项)和阶段性理解(动态解释项)的混合产物。

在符号化过程中,符号不仅代表某种事物,而且还支持某种事物:符号不直接指涉客观实在,而是通过另一个符号(解释项)指涉其对象。解释项虽为主体之理解,但在符号与其对象相关联的符号化过程中,它实际上只是带有“主体影子”的另一个符号。在本质上,符号使用者并非符号化过程的发动者,而仅仅是一个媒介。这种将思维理解为循环不止的符号化过程的做法,生动体现了人类思维的动态性。

总之,心智通过符号系统加以定义和学习的过程。符号亦非心智产物,反而心智取决于符号以及符号化过程。在本质上,“心智是一种根据推理规则的符号发展”,^{[4] 313}其总是在符号化过程中得以实现。

三、人机差异——高阶自控

皮尔士否定心智存有是决定人类与机器思维分野的关键,这令其看上去持有一种控制论观点。在谈及“逻辑机”时,他确实表达出这一倾向。他认为在推理层面上,人和机器无实质差别,只不过,机器是通过齿轮完成推理,而人则恰巧通过大脑完成。因此,即便大多数人都认为思维意味着生命体的推理,但皮尔士仍然坚持主张,推理实施的媒介和情境,与可同等适用于机器和人的逻辑批判无关^{[4] 59}。此外,应当注意的是,虽然很多理论家将创造性视为人机分野的核心标准,但皮尔士对此同样予以拒绝。

基于此,不少理论家将皮尔士定性为一名控制论者^[6],但反对声亦不绝于耳^[7]。无论争议如何,问题的关键是,反对笛卡尔式的主体哲学难道就意味皮尔士认为人类思维与机器完全没有差异吗?答案当然是否定的。在他眼中,差异的关键在于自我控制。

对“自我”可做两种理解。其一是指某种东西既是行动主体,又是行动客体。其二则指某种东西实施特定操作无需超越自身范围,此时自我与自动类似。如果采用第一种理解,那意味在某种实质主体概念下谈及控制,这与皮尔士哲学的基本立场不符。因此,皮尔士眼中的“自我”更类似于“自动”。这种自动可以在所有类型的有机体中发现,而人无疑当属其中。“控制”

则可理解为管控和修正。其中,“管控”是指与规范性或理想性要求保持一致、不偏离。“修正”则是指发生偏离情形时的复归。事实上,对“控制”的两种理解只是分别从正反两方面阐述了同一问题,它们是兼容的。在大多数情况下,皮尔士在第二种意义上使用“控制”。总之,自我控制并非预设自我意识,而是将实际行为与正确性标准加以比较的修正能力。在这一意义上,只要具有实现这种功能的协调反馈机制,那么无论人还是机器,都可以具有这种能力。

现实也的确如此。科学研究表明,低等动物也具有一定的控制能力。而在皮尔士眼中,逻辑机具备推理能力也表明其具有一定的自我控制能力。因此,如果人与机器两者都具有自我控制能力,那么两者的真正差异就不在于自我控制之有无,而在于控制的程度、方式以及机制。

第一,人类的自我控制是审议式的,机器则否。在皮尔士的理论框架中,审议意味着人的控制是一种习惯,这种习惯并非基于外界环境变化的自动反馈,而与想象和理性有关^{[4] 479}。很明显,这一点是机器缺乏的。即便机器可从某一单一前提中逻辑推出与该前提相关的所有结论,但它并没有在众多结论之间进行选择的能力。在人类没有对机器输入管控推理的既定目的时,机器得出的那一堆结论可能毫无意义。机器对于结论与前提之间复杂关系之展示存在局限,其无法考量推理过程的观察性和形象性部分。在皮尔士看来,机器与人类在思维上的重要差异之一正在于机器无法将某种特定、甚至动态可变的复杂目的结合经验观测来指导推理,而这种能力在人类思维中明显占据重要地位。或许是对这种能力的高度重视,使得皮尔士在人类思想史上首先提出了“溯因推理”。

第二,人类的自我控制是无限的,机器则否。机器可以进行推理,这在皮尔士的理论框架下意味机器具有一定的控制能力。可即便机器在某些方面可以趋近人类的自我反思,但它的自我控制归根到底是有限的。皮尔士说到:“控制本身可被控制,批判本身也受制于批判。理想地看,这一序列不存在明显的确定界线。”^{[4] 442}这一评论表明,人类的自我控制的无限性具有分层特征,即对自我控制本身的控制。这种分层体现在两个方面:一方面是横向的,即人类可以将自我控制纳入到一种不断增长的、异质的、在时间序列中能将过往和未来事件给予安排的有序网格之中;另一方面是纵向的,即人类的自我控制具有元表征能力,能自觉地将符号当成符号,并根据符号化过程予以不断深化。

人类思维不会拘泥于既有成果和方法,人类拥有无限批判和持续改进思想的秉性。此外,虽然皮尔士强调人类精神活动是根据习惯所进行的符号发展,但他亦承认人类自我控制无法摆脱大千世界中不确定性的介入。然而,“精神规律的不确定性不仅仅是它的一个缺陷,而正是其本质的反面。”^{[4]148} 反观机器,其只能在目的预定的前提下做出反应,过于“崇尚精确”,使得不确定性对机器而言很可能是一种“灾难”。事实上,自我控制的无限性正是源于人类面对不确定性时的自觉应对,而这一特征又与永恒不息的符号化过程如影随形。这从反面强化了皮尔士符号学的合理性,如果人类希望更好地理解世界,那关键之处正在于通过无限的自我控制洞悉符号之间永恒的复杂互动。

第三,人类的自我控制具有目的性,机器则否。皮尔士并不否认机器可以根据人类设定的目的做出反应,而科学亦表明低等动物也存在某种意义上的准目的的反应。然而,虽然机器体现出对预定目的和特定目的的精确反应,但机器并无法在目的之间发生冲突时做出恰当选择。尤其是在机器面临人类社会中的道德实践困境时,这一不足将被放大。对皮尔士而言,人之所以为人,正在于自我控制的目的属性可令人类之行动符合道德理念。无怪乎,他将推理之本质理解为一种自我控制式思维,而将逻辑学定性为伦理学之运用,而后者最终依赖于其独特的美学理论^{[4]533}。由此观之,机器与人类的另一个重要差异在于,机器无法在不确定情形下合乎道德理念地实现自我控制。机器的自我控制,更像是基于因果关系而发生的。机器过于完美,也只能在完美情形下运作。然而,不完美正是人类社会的常态,是人性不可磨灭的基因。从符号化的角度看,机器这种机械式的自我控制,看似能一蹴而就地到达了这一过程的“终点”,但由于缺乏道德上的目的性,机器所达到的“终点”是缺乏生活意义的。这种“终点”,也并非人类所期望的对“动态对象”的最终解释,而至多是人类理解世界过程中的一个“驿站”。

四、启 示

综上所述,我们可以得到一些与人工智能发展相关的启示。

第一,心智不存在于大脑这样的有机器官之中,思维也不依赖于人类所构想出来的“心智”,其本质是以人为媒介的符号化过程。思维即符号,意味着人类思维的人造本质。人类思维是人在与世界进行互动过程

中的符号发展,而非人脑固有的功能。上述要点,不正是人工智能发展之要义吗?在皮尔士的年代,人工智能体尚未诞生,皮尔士亦未“预言”人工智能时代的到来,但他的符号学已极为超前地成为了人工智能哲学的最初渊源,其不仅对人工智能何以可能做出了极具说服力的解释,而且还为人工智能的后续发展奠定了理论基础。人工智能的蓬勃发展也令其理论通过人类实践获得了验证。

第二,人类和机器思维的本质差异在于人类具有审议的、合乎目的的高阶自我控制能力,此点恰恰说明了人工智能发展的界限。这对人类应当有所警示。一方面,像专家系统这样的弱人工智能,根本没有直面道德实践的机会,因此,即便其缺乏高阶自控能力,也不会对人类导致实质性伤害。对其大力发展,仅会为人所用。另一方面,像在军事领域运用的人工智能武器以及一些强人工智能,由于其缺乏高阶自控能力,因此无论它在未来发展得多么精确和精妙,都不可能完全保障自身反应总是符合人类利益。对人类而言,此类智能体是一把“双刃剑”,人类应当对其发展保持审慎。而对于那些对人类利益已经造成实质伤害或可能造成实质伤害的智能体发展,则必须严格限制。事实上,人类对人工智能的担忧,正在于后者失控可能导致的灾难性后果。而皮尔士符号学早就告知我们:机器永远不会具有应对复杂不确定世界的符号操控能力,缺乏批判反思,使得机器永远只能成为应对不完美世界的手段。而面对相关人工智能体带来的潜在风险,人类也只有诉诸其所独有的自我控制予以应对。

第三,皮尔士的符号学与对当今人工智能发展具有重大影响的认知科学理论在哲学旨趣上相近。在当今认知科学的诸种理论流派中,与人工智能研究关系密切的基本都是反笛卡尔式的。具体而言,这类理论可分为具化(embodied)路径、内嵌(embedded)路径、行为(enacted)路径和扩展(extended)路径。其中,具化路径认为,至少某些思维无法脱离除心智之外更广泛的身体官能结构而得以直接运作,身体对思维而言同为构成^[8];内嵌路径认为,至少某些思维对外在环境存在依赖^[9];行为路径认为,至少某些思维不仅仅是神经性活动,而是由有机体面对世界时的行为以及世界对其行为的反馈的共同构成^[10];扩展路径则认为,对外在环境结构的操作、转换至少是某些思维之构成,思维是内在和外在操作同时发生作用的结果^[11]。在此,我们无法细致评估这些路径之间在细节上的异同。但可以肯定,这些理论都在肯定人类思维特殊性的同时,

努力扩展思维范围,而不只是将思维理解为大脑的某种能力。很明显,这与皮尔士符号学的哲学意涵高度吻合。在这个意义上,对当今人工智能发展具有重大影响的诸种认知科学理论,无论它们在理论目的、方法和构成上有何差异,其在哲学基础上都是皮尔士式的。

第四,某些理论家通过人脑可在机器体内运作以及“脑门植入”等论据去论证人机差异已近消失,他们试图将人类思维扩展至机器身上,甚至主张创造所谓的“半机器人”(Cyborgs)实现这一扩展。与这类理论家具有极端的科学倾向不同,皮尔士在强调人类思维人造性的同时,仍然以自我控制为着力点为人的道德属性预留空间。可以说,这一作法与其将逻辑学理解为伦理学之应用的想法异曲同工,即在科学精神与人文关怀的内在张力之间取得平衡。而这种平衡,无疑可视为人工智能未来发展的基本准则。

参考文献:

- [1] SOLAIMAN S M. Legal personality of robots, corporations, idols and chimpanzees; a quest for legitimacy [J]. *Artificial Intelligence and Law*, 2016, 25(2): 155-179.
- [2] DENNETT D C. *Brainchildren: Essays on Designing Minds* [M]. Cambridge: The MIT Press, 1998: 265-266.
- [3] 杨国荣. 学科性与超学科性: 哲学的二重品格[J]. *哲学动态*, 2012(4): 5-8.
- [4] PEIRCE C S. *The Collected Papers of Charles Sanders Peirce* [M]. Cambridge: Harvard University Press, 1994: 128.
- [5] RANSDELL J. Some leading ideas in Peirces's Semiotic [J]. *Semiotica*, 1977, 19(3): 157-178.
- [6] HOLMES L. Peirce on self-control[J]. *Transactions of the Charles S. Peirce Society*, 1966, 2(2): 113-130.
- [7] TIERCELIN C. Peirce on machines, self-control and intentionality[A]. TORRANCE S B. *The Mind and Machine: Philosophical Aspects of Artificial Intelligence*[C]. Chichester: Ellis Horwood, 1984: 100-113.
- [8] SHAPIRO L. *The Mind Incarnate*[M]. Cambridge: MIT Press, 2004: 185-191.
- [9] RUPERT R. Challenges to the hypothesis of extended cognition[J]. *Journal of Philosophy*, 2004, 101(8): 389-428.
- [10] O' REGAN J K, NOË A. A sensorimotor account of vision and visual consciousness[J]. *Behavioral and Brain Sciences*, 2001, 24(5): 973-1031.
- [11] CLARK A, CHALMERS D. The extended mind[J]. *Analysis*, 1998, 58(1): 7-19.

Peirce's Semiotics and Its Inspiration to the Development of AI

CHEN Xi

(Law School, Shenzhen University, Shenzhen 518061, China)

Abstract: According to C. S. Peirce, sign is a factor of the ternary relation of semiosis. By elaborating the interactions among sign, object and interpreter, Peirce translates the process of thinking into semiosis, and thinks that all thoughts are signs and even human being is just a semiotic development in accordance with inference rules. On one hand, his semiotics indicates that human thinking is artificial; On the other hand, it reflects the division between human and machine by claiming that man has a high-level self-control that machine lacks. His semiotics not only lays the theoretical foundation for the development of AI, but also establishes its practical limits. The intrinsic balance between scientific spirit and humanistic concern, as embedded in his semiotics, can be treated as the basic maxim for guiding the future development of AI.

Key words: C. S. Peirce; cognition; sign; semiosis; AI