

论布兰顿对强人工智能观的实用主义辩护*

周 靖

(上海社会科学院哲学研究所, 上海 200235)

[摘要]人工智能体能否拥有心灵,这是人工智能哲学研究中的一个重要议题。塞尔是持否定意见的代表,其反对强人工智能立场基于两个关键断言:(1)大脑导致心灵;(2)句法不是语义学的充分条件。长久以来,人们很难驳倒塞尔的论证。布兰顿哲学中蕴含着一种支持强人工智能立场的论证,该论证对塞尔论证构成了威胁。布兰顿吸收了实用主义洞见,认为心灵与世界、语义语汇和语用语汇之间不存在本体论上的界限,从而关涉实质内容的语用语汇能够构成语义语汇的充分条件。布兰顿反驳塞尔论证的关键一步在于,将句法语汇视为一种语用语汇,认为人工智能同样可以依靠引导指令从事实践活动,在此意义上,句法也能衍生语义,这反驳了断言(2)。布兰顿因而扩大了“实践”概念的内涵,祛除了认为人类活动是实践的唯一形式这种旧有立场,以及因之而来的认为人类智能是唯一智能的人类中心主义思想。布兰顿承认人工智能体可以基于非因果性的信息发展出不与人类心灵严格同一的心灵这种可能,即人工智能=人类创造的智能形式≠人类创造出的人类智能形式。这种理解在对心灵做出更为宽泛的界定的意义上,直接反驳了断言(1),从而反驳了塞尔论证。

[关键词]人工智能 句法实用主义 心灵 约翰·塞尔 罗伯特·布兰顿

[中图分类号] B152 **[文献标识码]** A **[文章编号]** 2096-983X(2022)04-0048-09

相较于仅将机器界定为工具的狭隘人工智能观,强人工智能设想机器也能获得心灵,从而成为像我们人类一样的理性主体,进而它也需要承担诸如伦理和法律等方面的责任。关于人工智能的技术、伦理、法学诸方面的研究因此蔚然成风,然而,这一切在很大程度上奠基于一强人工智能是否可能这一问题。于是,在哲学层面拷问人工智能获得心灵的可能性,乃至心灵的本质,始终是一项不容忽视的工作。本文将主要诉诸布兰顿的理论资源提出一种“宽”的强人工智能观。

塞尔对强、弱人工智能的区分,以及他对强人工智能的反驳,有着深远的影响。根据塞尔的界定,强人工智能具有意识或心灵,而弱人工智能仅体现为对人类心灵的模仿。^{[1](P417)}塞尔反强人工智能的立场基于如下两个关键断言:(1)大脑导致心灵;(2)句法不是语义学的充分条件。^{[2](P39-41)}本文第一节将简要阐明断言(1)是断言(2)的一个推论,从而聚焦于阐明断言(2)暗含的问题,即句法仅具有形式有效性,而无语义学所具有的关于心理内容(mental-content)的实质(material)有效性。基于这一阐

收稿日期:2021-08-30;修回日期:2022-04-22

*基金项目:国家社会科学基金青年项目“剑桥实用主义研究”(21CZX050)

作者简介:周靖,研究员,主要从事语言哲学、实用主义哲学、心灵哲学、符号哲学等研究。

释,第二节试图借助布兰顿实用主义的理论资源以证明,任何形式的有效性均有着实质的来源和后果,从而反对断言(2)。我们由于对内容设置了太过“人类化”的理解,从而忽略了实践样式以及由之而来的内容样式的多样性,这也为理解布兰顿式论证带来了障碍。本文第三节进一步试图淡化人工智能的“人类中心主义”思考倾向,指出“人工智能=人类创造的智能形式≠人类创造出的人类智能形式”,这是布兰顿论证带来的教训之一。基于否定断言(2)的理由,我们将认识到断言(1)也会因此失去根据。与此相关的另一个教训体现为本文讨论衍生的一个判断,即关于人工智能的类脑计算研究有着现实的重要意义,但对于元哲学问题——诸如有关心灵、意识、认知等问题探讨来说,即便机械脑能够完全复绘有机脑,这仅是一项科学成就,哲学仅可能会因之增进对相关问题的理解,但不会必然因之取得实质进展。哲学工作者应该在人工智能论题上持有清醒的问题意识,做出审慎的讨论。^①

一、塞尔论强人工智能的问题场景

期待人工智能像人类一样思考,乃至能像人类一样对糖有“甜”的感知,能在面对蒙克的画作《呐喊》时感到“压抑”和“绝望”,能够对他者感到“爱”的依系或因被爱而感到“幸福”——这种期待实际上是一种十分常识性的思考。但这种思考也抓住了反人工智能论题的关键要点:人工智能无法拥有像我们一样的诸如“甜”“压抑”“绝望”“爱”“幸福”之类的

心理内容或关于“世界”的感知。塞尔下述如所周知的推理体现了这一要点^{[2](P39)}:

(P1) 人工智能程序是形式的(句法的)

(P2) 人类心灵具有心理内容(语义内容)

(P3) 句法不是由语义学所构成,也不是构成语义学的充分条件

因而,塞尔得出了这样的结论:

(C1) 人工智能程序不是由心灵所构成,也不是构成心灵的充分条件(强人工智能不可能)

上述推理中,塞尔将(P2)中的“语义内容”视同为“心理内容”,从而将(P3)中的函项“语义学”替换为“心灵”,得出仅遵循句法逻辑的人工智能没有心灵的结论,其主要理由在于,句法是纯形式的,而语义学则须诉诸现实世界中的状况来理解意义。我们可以简要回顾他的“中文屋”论证来阐明他所做出的这一替换意味着什么:

设想一个对中文一窍不通且只会说英语的人,他被关在一间只有一个开口的封闭房间中。房间里有一本英文写成的操作手册,通过查询手册他可以知道如何处理汉语信息并对信息做出相应的回复(此时查询手册如同一个计算程序)。房外的人向房内的人递进中文写成的问题,房内的人可以根据那一手册找到合宜的指示,从而给出正确的答案(例如汉字“红”,此时对于屋内的人来说如同一幅提供信息的“图画”,手册上有着“if ‘红’, then ‘red’”的操作指示)。然而,尽管答案正确,房内的人其实对问题本身以及中文一无所知。人工智能类似于房中人,它尽管能够有效地信息做出回应,但它不具有关于问题本身以及中文意义的真实认知。^{[1](P418)}

那么,拥有心理内容意味着在运用语言表

^①或许需要事先交待一点,基于与一些学友就本文展开的讨论,可以预见有人会迷惑于本文中究竟哪些是布兰顿的立场,哪些是笔者的立场。感到困惑的学友主要质疑布兰顿是否真实提出了本文中辨明的人工智能立场,从而认为笔者对布兰顿哲学做出了过多自己的阐释。笔者的回应是,布兰顿的确对人工智能论题做出过直接讨论,并且本文正是对其立场的阐明。故而,我们可能在如何理解布兰顿相关立场上怀有分歧,但就布兰顿的人工智能立场持有怎样的立场这一具体问题而言,笔者暂无自己的定论,也未试图在本文中嵌入自己关于人工智能这一论题的立场。

达式时能够真实地 (genuinely) 知道该表达式是什么意思, 使用该表达式时意在 (intends) 什么, 即拥有关于什么内容的意向性。相较之下, 中文屋里的人并不拥有中国人在使用“红”这一概念时所拥有的意向状态, 同样地, 人工智能仅能在句法上构建“红”和“red”间的形式关系 (从而有P3), 此时人工智能并不具有关于某种 (例如) “红性”的因果意向性。这里的思路可表述为如下三段论推理:

(P4) 拥有意向性的人类智能均拥有心灵

(P5) 人工智能不拥有意向性

(C2) 人工智能不拥有心灵 (强人工智能不可能)

(P4) 和 (P5) 把问题引向了关于意向性的探究, 一些研究者认为意向性源于人脑内的状态, 从而试图通过类脑计算或模拟神经形态计算 (Neuromorphic Computing), 在对脑内的因果模式进行“模拟”的意义上试图使得人工脑无限像有机脑一样运作, 从而获得意识或心灵。就此而言, 塞尔的确认为:

(C3) 大脑状态因果地导致意识^{[2](P39)}

但是, 关于 (C3), 塞尔指出, “意识是人类和某些动物的大脑的生物特征。它由神经生物过程所产生, 就像光合作用、消化或细胞核分裂等生物特征一样, 都是自然生物秩序的一部分。这一原理是理解意识在我们世界观中的地位的第一步。”^{[4](P75)}故而, 实际上, 在塞尔看来, 即便类脑计算无限复杂, 它仅是在模拟因果性, 而始终无法获得生物所应对的真实自然, 程序模拟无法取代现实因果性, 从而人工智能无法具有真正意义上的意向性、意识, 或心灵。

基于上述阐述, 我们看到塞尔批判强人工智能的要点在于: 人工智能程序仅具有句法层面上的形式有效性[故而, (2) 句法不是语义学的充分条件], 而不具有语义学在关涉 (of/about) 实质 (心理) 内容意义上的有效性[故而, (1) 大脑导致心灵]。在此意义上, 如果我们否

定了断言 (2), 那么我们便有可能拥有一种不同于断言 (1) 的理解, 即获得另一种心灵观, 这种心灵观能够衍生出布兰顿为之辩护的新的强人工智能立场。

二、布兰顿对强人工智能的辩护

布兰顿哲学的一项重要任务在于, 以语义推论的方式阐明 (articulate) 理性主体在使用语句表达式时所意在的内容。在谈及塞尔关于意向性的区分时, 布兰顿指出, “约翰·塞尔在他的《意向性》一书中……为我们提供了理论资源。‘如果状态S是一种意向状态的话, 那么我们必须回答如下问题: S是关于 (about) 什么的? S是涉及 (of) 什么的? S陈述 (S-that) 表述了什么?’ 语义理论的一项主要任务在于, 指出 ‘that’ 意向性和 ‘of/about’ 意向性彼此间有着怎样的关联。”^{[5](P424)}与此相关, 布兰顿试图基于使用 “that” 语句的推理活动 (这项活动在人际间的社会维度中展开) 来阐明所 “关涉” 的 “内容”。

乍看之下, 布兰顿这项工作似乎在 “支持” 塞尔的反强人工智能论证: 布兰顿式的语义推理是人工智能所不能完成的, 此外, 人际间的社会维度亦是人工智能所欠缺的。然而, 细究之下, 布兰顿消除两类意向性界限的做法, 对实质的语用语汇和形式的语义语汇之间迭代发展关系的阐释, 以及将算法视为一种特殊的语用语汇的观点等, 实际上反驳了塞尔论证, 并为捍卫一种新的强人工智能立场提供了思想支持。下文中, 笔者将先提供布兰顿支持强人工智能的思想资源, 而后展开具体论证。

首先, 布兰顿承袭了消解心灵与世界之间本体论界限的实用主义洞见,^[6] 这在此处体现为, 认为在实践的做的活动 (doings) 中使用的、关于内容的语用语汇 (“of/about” 语汇) 能够充当使用、制造或承认在表达上更高阶的语义语

汇(“that”语汇)的充分条件;而后继脱离了实践语境的非指示性的语义语汇则能够构成解释实践层次的语用语汇的必要条件;并且,这两类语汇之间不存在本体论上的界限,语汇在层次上的高低仅体现为,语义语汇是用于解释语用语汇(用于“说”出所“做”的事情)的语汇,当语义语汇面对新的实践情境而需要被加以阐明时,它便变为了新的语用语汇。^{[7](P10-11,76)}这样一来,在行动和认知的回环中,语汇逐渐变得丰富。这里的“丰富性”既体现为对世界更为清晰的表达,也体现为进行表达的语义方式上的精细化。

其次,这一实用主义洞见暗含着更为深刻的世界观和语言观,即语言和世界一道生成,在这一生成过程中,意识、自我意识、以及社会规范也在互惠的相互关系一道成熟和发展。^{[5](P235-261)}初始的一类语用语汇,诸如维特根斯坦“石板游戏”中,助手在听到建筑师发出“石板”的声音而对某对象做出可靠的、有差异的反应。此时,“在语言课中发生的是这个过程:学习者命名诸对象。也即,当教师指向石料时,他说出词语。”^{[8](P12)}这里的“石板”语汇相当于一套引导指令(bootstrapping),用于将“事物作为某物”(即,将坚硬的板状石料作为“石板”)稳健地识别出来;当指令愈发清晰时,对象得到愈发确定的界定;反之,对象的确定性将原先单纯的“声音”作为某种指令或概念而清晰地固定下来。词与物,心灵与世界是相互成就的。

反强人工智能主义者此时或许仍然欣于接受布兰顿的观点,但认为布兰顿的讨论似乎与人工智能问题无关。然而,基于上述思想,布兰顿指出,“支持AI体现的是这样的主张,即原则上来说,计算机能够做所需的事情,从而使用自动语汇……说出什么。”^{[7](P70)}而这里的“什么”就是计算机的“意向内容”,诸如中文屋中的人工智能,“if ‘红’, then ‘red’”这套引导指令中,“red”便是由“红”这种实践刺激所引发的

内容。从而,布兰顿认为,句法、算法语汇是一套我们根据之从事“做的活动”的实践语汇,进而也能构成语义语汇的充分条件。

解释这一论点需要我们理清布兰顿的支持强人工智能论证,我们可以将布兰顿的推理预先陈述如下,而后再进行阐述之事:

(P6) 关涉内容的实质语用语汇和表达内容的形式语汇之间不存在本体论界限

(P7) 算法、句法或演算语汇也是一种基础的实质语用语汇

从而,布兰顿得出与上述“(P3)句法不是由语义学所构成,也不是构成语义学的充分条件”直接相抵牾的观点(P8):

(P8) 句法可由语义学构成,同时可以构成语义学的充分条件。

基于(P6)-(P8),以及塞尔的(P2)“人类心灵具有心理内容(语义内容)”,布兰顿进一步得出支持强人工智能的结论(C4):

(C4) 人工智能程序既可能由心灵所构成,也可能构成心灵的充分条件(强人工智能可能)

布兰顿的实用主义洞见为(P6)提供了支持,并为在(P8)和(P3)这对矛盾中为何支持(P8)提供了理由。此时的问题焦点在于,句法、算法等塞尔所认为的纯形式语汇是否可能具有语义内容(心理内容)?布兰顿给出了肯定的答案。在进一步阐明布兰顿思想之前,我们需要认识到心灵在布兰顿和塞尔那里已经有了截然不同的蕴意。在塞尔那里:

心灵=人类心灵

其中,人类肉身提供的因果关系构成了心灵的充分条件,意向性是心灵的本质特征,拥有这类意向性的智能体才能有意识地知道自身在做什么。相比之下,在布兰顿那里:

心灵≠人类心灵

人类心灵仅是心灵的一种具体形式,构成心灵的条件是既从事语用实践活动,又从事语义推动活动,并且语义活动阐明的是语用活

动中的“内容”，即意向性所关涉的内容，在此意义上，布兰顿将意向性理解为话语意向性（discursive intentionality），这种理解与塞尔论证中的关键前提直接相抵牾。

塞尔仅将人类心灵视为心灵，进而将这类心灵视为衡量人工智能体是否有心灵的标准。布兰顿则聚焦心灵的构成条件，容纳心灵样式有着多样性这种可能。我们可以通过“假想”塞尔可能提出的问题并作答来进一步阐明布兰顿的思想。

假想问题一：句法仅具有形式意义，这乃是因为人工智能缺乏人类漫长演化过程中所应对的具体情境，而恰在这种情境演化中意识、意向性和意向内容才一道生成。^{[4](P43-45)}

针对这一关键问题，布兰顿反问道，一定需要涉身自然情境之中才能习得初始的语用语汇么？我们对机器的“训练”同维特根斯坦“石板游戏”中助手所受的训练有何实质不同？如若执意认为机器没有助手所置身于的自然情境，那么这至少将犯下两类错误，一类是认为缺乏感性能力（sentience）的智识能力（sapience）是无法理解的，这种布兰顿归于塞尔在解释人工智能时所承诺的观点认为，动物的感性模式构成了人类理性模式的充分条件。而前文的讨论已经能让我们充分地认识到，我们可以从智识的成就开始讨论，以语义推论的方法对内容进行阐明。内容仅是我们可以对之做出断言的任何之物，它不必然是自然情境中的事项，这便意味着“具有自然情境”不是构成智能的充分条件。

另一类相关的错误则体现为“人类中心主义”的错误，即认为仅有人类智能是唯一可能的智能，从而满足该类智能形成和发展的条件和因素是唯一可能的条件和因素。布兰顿指出，我们可以支持算法实用主义（algorithmic pragmatism），因为“对人工智能模型做出实质的算法的实践阐释值得关注，这是因为这项阐

释清晰揭示了根据知道-如何（knowing-how）来解释知道-什么（knowing-that）的实用主义纲领，即具体解释使用非意向性的、非语义的语汇来必须做什么才能够算是使用某种语汇来说出什么，从而产生能够适用于行为的意向性的、语义语汇（即一种有着语用上表达效果的引导指令）。”简单地说，布兰顿认为句法、算法语汇也是一种实质语用语汇，因为人工智能同样能够使用这种语汇“做事”，就如助手在初次玩“石板游戏”时一般，初时“石板”仅体现为一套引导指令，助手此时不具有意向性，“石板”语汇亦非语义语汇。如果“石板”语汇能够成为一套初始语汇，并且这种语汇能够构成语义语汇的充分条件，那么句法、算法语汇也可以，即（P7）。而人类可以通过对人工智能的训练而使之获得这类语汇，在这种训练里，人工智能同“石板游戏”中的助手一样，需要关于“世界”的知识以制定好的决策，并将该类知识以句法表达形式存储在数据库中。

假想问题二：布兰顿的解释避开了关键问题，即人工智能没有现实情境。即便我们承认人工智能有着自身的情境，但我们不能忽略塞尔在（C3）背后的关键主张：引导指令无法取代因果性。

在笔者看来，这一追问至少未能领会布兰顿观点中的一个要点：我们是在探究直接关涉内容的最低限度的“实质语用语汇”的条件，在这种最低限度上，我们发现人类和机器具有同样性质的基础语汇。就人类情境而言，“情境”包含了外部世界施加给人类主体的影响以及主体对外部世界的主动反应，它因而是孕生意义的场景。人工智能体也有其“情境”，它同样包含了对某种外物的反应。然而，问题的关键并不在于区分人类情境或人工智能情境，因为此时的语汇是去情境化的（context-free）——初始的语用语汇体现的是理性主体与世界最小化的直接关联。在基础的层次上，借用德雷斯基的观点或许能帮助布兰顿进一步澄清问题，

即“原材料是信息”，“知识是由信息导致的信息”。^{[9](Pviii+x)}人类因为是有机体才有着因果的信息模式，而人工智能因为是机械体，但它同样可能拥有其他不同模式的信息。坚守“因果性”无异于狭隘地苛责为人工智能体制造一颗人脑，这种坚守应该转化为如下诉求：为人工智能提供如人脑般优秀的硬件和软件的科学诉求，而非完成从机械脑到人脑的“本体论之跃”。放下这一渗透人类中心主义思想的执念，我们将会接受这样的可能：由于人工智能具有不同于人类的信息模式，因而随之产生的知识表达、智能、心灵等可能亦与人类不同，即(C4)。

不过，这里的确有着一点需要澄清的地方，布兰顿没有在如下强的意义上反对塞尔论证：

(P9) 人工智能体与人类一定有着截然不同的心灵，或者两者的信息模式必然是彼此不可互相翻译的。

布兰顿承诺的仅是，存在另一种发展出心灵的方式，以及以这种方式发展出多少与人类心灵有所差别的心灵这种可能。

假设问题三：布兰顿的解释可能过分“美好”，现实的人工智能的句法语汇根本没有资格充当语用语汇。

这一问题似乎基于当下人工智能令人不满的现状而否定对“美好”未来的预期。布兰顿对此问题的回答或许可以简明而直接一些：如果科学家在某一天能够予以人工智能充分优秀的硬件和软件，以及充分的训练，那么，这份美好便可期。但条件是由科学家而非哲学家创造的，哲学的工作仅在于就心灵和意识做一般意义上的反省和探讨，而非作出科学判断——科学家负责“人工”，哲学家解释“智能”。如若布兰顿论证遭受反驳，这也在很大程度上将是于哲学范围内展开的工作。

假设问题四：如果接受了上述解释，那么我们就承认人工智能有着自己独特的，乃至我们无法理解的神秘意识。

这一问题将我们引向了布兰顿的“功能主义”解释。与“智能”相对应的布兰顿语汇是“智识能力”(sapience)，拥有智识能力的理性能动者(agent)做出的行为是合乎理性的、可为他人所理解的。^{[10](P5)}“当我们讨论智识能力时，功能主义是一种更具前景的解释策略”，^{[7](P71)}我们根据诸如“相信P”，“if ‘红’，then ‘red’”之类的信念或引导指令式的语汇在包含这套语汇的系统内所起到的作用来理解表达语汇以及相应行为的意义。理性能动者(包括我们人类和人工智能)不会自我矛盾，以信息为基础的信念因而具有可理解性。如戴维森所论证过的那样，如果人工智能有朝一日拥有发展成熟的“机械语”，那么它不会与我们的语言有着不可通约的“概念框架”，^[11]它必然是我们所能理解的语言，但不必然是与人类语言有着表面类似性的语汇。

总结而言，取消心灵和世界二元论界限的实用主义思想勾绘了布兰顿论证的底色，他取消句法语汇和语义语汇的界限亦以这一实用主义洞察为基础。布兰顿因此认可乔姆斯基式的立场，即演算句法和语义表达在转换、生成和发展的过程中不存在泾渭分明的分界，^{[7](P20)}从而他得出了(P8)“句法可由语义学构成，同时可以构成语义学的充分条件”这一与塞尔直接相悖的观点。如若塞尔仍坚持对强人工智能的界定，那么布兰顿论证同样可以支持强人工智能立场，只不过此时我们容纳了这样的一种可能，即人工智能心灵不必然与人类心灵有着严格同一性。

三、从布兰顿论证中习得的两项教训

布兰顿论证或许实际上“降低”了“智能”的准入门槛，即从“仅仅只有人类智能才是智能”降低到“如果符合智能生成的条件，那么便

可能生成智能。”但这种“降低”并非“贬低”，布兰顿论证体现了人工智能的哲学研究的另一关键面向，即对人类智能本身进行反思，祛除其中的神秘性，从而既增进了我们对于自身的理解，也在人工智能研究上避免“人类中心主义倾向”的问题，这类问题体现为仅将心灵理解为人类心灵，仅将实践理解为人类的实践，仅将生成智能的物理基础理解为人类或有机体独有的因果信息。笔者认为，这是布兰顿论证给予我们的第一个重要教训。

面对布兰顿论证，反对者们或许会喋喋不休地反驳到，无论如何，接受“红”的信息的人工智能并未“意在”“red”！这种反驳实际上隐在遵循着“心灵存在→具有本质性的意向性→拥有这种本质意向性的主体才能拥有意向内容”这种推理顺序。反驳者设定了“心灵”的独立性和实存性，并将之仅归派给人类。这种本质主义观点既为实用主义者所摒弃，也为塞尔本人所反对。塞尔指出，“……在诸位的颅腔里可没有两个彼此不同的形而上学领域：一个是‘物理的’，另一个是‘心灵的’。确切地说，只存在着那些发生在诸位大脑里的进程而已，但是其中的一些过程同时就是意识体验。”^{[12](P128)}“意识是大脑更高层次的或突显的特征……心灵是大脑的心智属性，因而也是物理属性，正如液态是分子系统的属性。”^{[4](P15)}故而，并不存在“心灵”这种有着形而上本性的东西，长久以来人们认为它独属于人类，并与“物质”形成对立关系。布兰顿同样指出，对人工智能的传统理解，即试图讨论人工智能是否具有柏拉图主义式的或理智主义式的心灵，这是一项无望的事业；相比之下，对人工智能的理解应以接受上述所叙的“算法实用主义”为基础。^{[7](P78)}在这类理解之下，心灵、意识等不能作为解释的起点和依据，相反，它们的意义在对相关具体过程的解释中才得到披露。

心灵的去本质化使得关于智能的人类中心

主义观点失去了最为根本的根据。我们只能后验地涉入具体的实践情境中来解释心灵的“兑现价值”，然而，这里仍然有着另一种人类中心主义的危险，即只有在人类的实践形式中才能产生意识。塞尔如下表述实际上支持着实用主义的观点，“我们每一个人都是由其他生物性和社会性存在物构成的世界当中的生物性和社会性存在物，周遭布满了人工物品和自然对象。现在，我一再称之为背景的东西实际上导源于由关系构成的完整的汇集，而每一个生物的社会性存在物都和围绕它的世界具有这种关系。假如没有我的生物性构造，没有我置身其中的社会关系集，我就不可能具有我实际具有的背景。但是所有这些关系，生物的、社会的、身体的，所有我们置身其中的事物都只与背景的产物有关，因为它对我产生了影响，特别是它对我的心-脑产生了影响。世界和我的背景有关，这只是因为我与我的世界的相互作用……”^{[13](P157)}恰是缘由我们周遭的各式物品和对象，以及我们生物性的构造和所置身于的社会关系，我们的心灵和意识才因此受到影响和发展。当今的人工智能无疑不具有下述要素：生物构造、通过因果链条触及的对象，以及社会关系。实用主义者的确会接受塞尔的表述。然而，上文的讨论也已经指明其中隐含的人类中心主义倾向：在最初的实践活动中，周遭世界是混沌一片的（去语境化的，而对象、社会关系等均在语境之内），布兰顿认同算法亦是一种实践的语用语汇，这一观点体现的洞见在于，人工智能可用其自身的初始语汇来敞向世界，如同我们用因果语汇触摸并敞开一样。伴随着心灵的去本质化，世界亦需要去本质化：我们并非用心灵的镜子反映世界，也非单纯地受到外间世界的限制，而是心灵和世界一道敞开、丰富和发展。这是罗蒂和麦克道威尔教导过我们的重要一课。^{[14](P24-25)}于是，人工智能在它独特的触摸世界的活动中（即根据句法或算法的引导指令“做”事），在人类精

心的“训练”和“培养”中，人工智能也有可能在其实践过程中结出一颗心灵之果。

祛除了基于心灵本质论，以及缘由仅将人类实践视为唯一实践形式而产生的人类中心主义观点，我们此时也应该在人工智能的哲学研究中抛弃塞尔“大脑导致心灵”的观点。需要澄清的是，笔者为这一判断添加了限制条件：在人工智能的哲学研究中。本文的讨论并未反驳诸如基于脑科学研究来解释心灵的研究进路，而是仅是在敦促放弃将人工智能视为对人类智能的模拟或复绘的“窄”智能观。布兰顿论证为我们提供了关于人工智能的“宽”的理解，以及一种相应的“宽”的心灵观，这种观念既为强人工智能的可能提供了哲学依据，也界定了哲学家和科学家的工作分工。

由此，我们还应从布兰顿论证中得出另一个教训，即关于人工智能的哲学探究不应迷恋于人工智能能否在神经元层次模拟大脑之类的问题，这类问题是认知科学家们的重要工作。诉诸科学成就或许能够增进我们关于心灵和意识的理解，但在哲学层面上，我们必须首先进行一道严格的理论审思：反思这类新的经验素材是否实质撼动或增进了已有的哲学理解。如若缺乏相关的学理支持，那么诸如“神经美学”“神经伦理学”之类的标签将仅暗示着一种十足的胡说。在笔者看来，脑科学以及人工智能的研究仅为我们带来了新形式的经验，面对这些经验，我们依旧面对着同心灵、意识、认知、意义、价值等相关哲学问题。设想某一天人工智能完全现实化了，甚至人工智能演化成了一个新种族——赛博坦人，他们会是我们的邻居或友人，那么这些哲学问题也将是他们的问题。他们或许会发展出一套我们可从中受益的“赛博坦哲学体系”，但无论如何，哲学工作者应该始终恪守自身的哲学责任。简言之，就人工智能做出哲学解释而非试图僭越地得出科学的结论。

四、结语

熟悉实用主义哲学的学仁或许会讶异于运用实用主义思想为强人工智能立场做辩护这种做法。实用主义在哲学观念史上的重要贡献包含了抵制心、物二元论，以及反对关于心灵的本质论观点，而强人工智能立场则承诺了机器具有心灵的观点，这里似乎产生了一个矛盾。本文的讨论试图指出，布兰顿吸纳了实用主义的洞见，消解了与心灵相应的语义语汇以及与世界相对应的语用语汇之间的界限，并认为语用语汇可以构成语义语汇的充分条件。布兰顿支持一种新的强人工智能观的关键一步在于，承认句法也是一种可用于“做”事的实践语用语汇，从而可以构成语义语汇的充分条件，按照塞尔的推理，人工智能因此也可能具有心灵。然而，不同于塞尔，伴随着对人类中心主义的祛除，人工智能心灵不必然与人类心灵严格同一。

参考文献：

- [1] Searle J. Minds, brains, and programs[J]. Behavioral and Brain Sciences, 1980, 3(3): 417-424.
- [2] John Searle. Minds, brains and science[M]. Cambridge, Mass.: Harvard University Press, 1986.
- [3] Searle J. Intrinsic intentionality [J]. Behavioral and Brain Sciences, 1980, 3(3): 450-457.
- [4] 约翰·塞尔. 心灵的再发现[M]. 王巍, 译. 北京: 中国人民大学出版社, 2005.
- [5] Robert Brandom. A spirit of trust: a reading of Hegel's Phenomenology [M]. Cambridge, Mass.: Harvard University Press, 2019.
- [6] 周靖. 再思实用主义的实践概念: 基于哈贝马斯与布兰顿之争[M]// 实用主义研究: 第二辑. 上海: 华东师范大学出版社, 2020: 100-115.
- [7] Robert Brandom. Between saying and doing: towards an analytic pragmatism[M]. Oxford: Oxford University Press, 2010.
- [8] 维特根斯坦. 哲学研究[M]. 韩林合, 编译. 北京: 商务印书馆, 2019.
- [9] Fred Dretske. Knowledge and the flow of information [M]. Cambridge, Mass.: The MIT Press, 1981.

- [10]罗伯特·布兰顿. 阐明理由: 推论主义导论[M]. 陈亚军, 译. 上海: 复旦大学出版社, 2019.
- [11]Donald Davidson. On the very idea of conceptual scheme [M]. *The Essential Davidson*. Oxford: Oxford University Press, 2006: 196-208.
- [12]约翰·塞尔: 心灵导论[M]. 徐英瑾, 译. 上海: 上海人民出版社, 2019.
- [13]约翰·塞尔. 意向性: 论心灵哲学[M]. 刘叶涛, 译. 上海: 上海人民出版社, 2007.
- [14]Richard Rorty. *Philosophy and the mirror of nature* [M]. Princeton: Princeton University Press, 1979.
- [15]John McDowell, *Mind and world* [M]. Harvard: Harvard University Press, 1996.

【责任编辑 刘红娟】

Robert Brandom's Pragmatist Argument for Strong AI

ZHOU Jing

Abstract: The possibility of strong AI that involves a certain kind of mind is a critical issue in the philosophy of AI. John Searle's influential anti-strong AI position has two key assertions: (1) the Brain causes minds; (2) Syntax is insufficient for semantics. Though Searle's argument faces many challenges, no knockdown argument that is widely accepted appears. However, Robert Brandom's idea, which supports a strong AI position, seems to be a real threat to Searle's argument. Robert Brandom's argument takes in a pragmatism insight that eliminates the ontological boundary between the mind and the world, semantic vocabulary, and pragmatic vocabulary and commits to the claim that the de re pragmatic vocabulary is sufficient for the de dicto semantic vocabulary. The essential move Brandom makes to refute Searle's position is to take the syntactic vocabulary as a sort of pragmatic vocabulary. Brandom believes that AI can also engage in practical activities by dint of bootstrapping or automata vocabulary, in which syntactic vocabulary can also be semantic; hence assertion (2) is refuted. The scope of practice in Brandom's approach has been extended enough to remove the implicit anthropocentric idea that thinks human activities are the only form of practice and the resulting prejudice that human intelligence is the only possible form of intelligence. Brandom embraces the possibility that AI will win its particular form of mind based on non-causal information one day. Still, it is not strictly identical to the human mind, which refutes the assertion (1). In conclusion, Searle's argument fails.

Keywords: artificial intelligence; syntactic pragmatism; mind; John Searle; Robert Brandom

(上接第17页)

different special zones' institution backgrounds, different historical missions require different special zones' institutional content, and different era theme nurture different institution achievements of special zones. The development of China's special zones not only adhere to combining basic principles of Marxism with China's specific reality, but also stick to the leadership of the CPC, the supremacy of the people, independence, the Chinese path, and innovation. In the new era of socialism with Chinese characteristics, pilot zones for socialism with Chinese characteristics are the best practices for XI Jinping Thought on Socialism with Chinese Characteristics for a New Era, they are endowed newer and deeper connotation. They also become reform targets, key experiments, and important exploration. Summarizing the experience of the special zones' construction in various periods will help fully understand the unique key of the governance of China over the past the century.

Keywords: the centenary of the founding of CPC; special zones; special zones' construction; pioneer and innovation; basic rules