

中图分类号:TP18 文献标识码:A 文章编号:1004-8634(2018)05-0066-(11)
DOI:10.13852/J.CNKI.JSHNU.2018.05.008

胡塞尔的意向性理论与人工智能关系刍议

徐英瑾

(复旦大学哲学学院,上海 200433)

摘要: 胡塞尔的意识哲学对于人工智能的意义,一向被学界严重低估。实际上,胡塞尔的“现象学悬搁”方法虽貌似对一切自然主义的认知建模工作构成了威胁,但依然能在某种变通的意义对人工智能构成启发。具体而言,“现象学悬搁”对于意识表征中意向模式的普遍存在性的提示,完全可以在人工智能语境中被转述为下面这个工作思路:建造一台具有自动“修偏”机能的智能机器,只有在为其配属了丰富的心理模式的情况下才是可能的,而心理模式的配置本身又会倒逼系统对于世界的素朴表征转化为那些悬置了外部真值判断的“内存在表征”。此外,对于胡塞尔的“Noema”概念的推论主义解释模式,也可以告诉我们为何福多所提出的关于命题态度的“盒喻”式处理方案乃是不对的。不过,目前主流AI在处理意向性问题方面的表现,依然是十分笨拙的:这些主流进路要么根本无法在符号表征的层面上触及意向性问题(如深度学习),要么不得不采用福多的“盒喻”而导致模型的不灵活性(如某些符号AI技术),要么就在过分强调环境与主体的相互关系的同时预设了胡塞尔所反对的“自然主义态度”(如某些能动主义技术路线)。由此看来,我们如果要设计一种在最低限度上符合胡塞尔精神的AI系统,我们就必须与目前的主流AI技术路线分道扬镳。

关键词: 意向性;悬搁;信念盒;推论主义;人工智能;意向相关项

一、导论

在心智哲学界,“意向性”(intentionality)一般被理解为心智关涉到特定事物的一种常见能力——通过这种能力,被关涉到的事物可以规避其与心智之间的物理时空阻隔,而成为心智的相关项(譬如,尽管曹操早在罗贯中写《三国演义》之前就死了,但这不妨碍罗贯中自己的意向活动关涉到曹操;尽管宇宙中任何一个黑洞的距离都离霍金非常遥远,但这并不妨碍霍金的意向活动

指向其中的某个黑洞)。此外,颇为有趣的是,即使是一些不存在的对象(如孙悟空),也完全可以成为意向活动的对象,尽管其无法在物理空间中存在。故此,意向对象的存在方式也在哲学文献里被称为“内存在”(“inexistence”,即“在主观意识的范围内存在”的意思),以区别于物理对象的外部存在方式。

从心智哲学的角度看,由于布伦塔诺(Franz Brentano)早就将是否具有“意向性”视为区分“有心智者”与“无心智者”的基本标准,^①那么,只要

基金项目: 国家社科一般项目“自然语言的智能化处理与语言分析哲学研究”(13BZX023);国家社科重大项目“基于信息技术哲学的当代认识论研究”(15ZDB020)

作者简介: 徐英瑾,复旦大学哲学学院教授,博士生导师,教育部长江青年学者,主要从事英美分析哲学、认知科学哲学等研究。

我们将该论题的效力从人类拓展到机器上,我们就不难得到下面的推论:一个真正意义上的强人工智能系统将不得不具有“意向性”,否则它就不是真正具有智能的。

不过,一涉及哲学与人工智能(Artificial Intelligence,简称“AI”)之间的跨学科对话,知识壁垒所带来的交流困难或许就会立即阻碍对话的深入。一个不熟悉此类哲学讨论的AI专家或许会问:我们做工程研究的,为何要在意布伦塔诺说了些什么呢?另外,我们又该如何通过工程学与数学语言来定义“意向性”呢?

面对这样的疑问,首先可以肯定的是:“意向性”的确很难通过严密的数学语言来加以界定,这就像“机器智能”这个术语也难以通过类似方式得到界定一样。毋宁说,最终确定一台机器或一个生物是否具有“意向性”,乃是来自第三方报告中的定性评估——换言之,大多数人若在定性的层面上觉得它有意向性,那么,它就是有意向性的(请参考丹尼特的“意向姿态”理论^②)。然而,此类定性描述依然在工程学描述中是不可或缺的,因为对于任何工程产品的第三方定性评估都是检测其“可被接受性”的最关键环节。至于建造一台使得大多数用户觉得其具有“意向性”的智能机器的实用价值,则主要体现在:(甲)只有在用户愿意对一台机器进行“意向性”指派的前提下,这台机器才可能通过“图灵测验”——换言之,没有人会认为一台不具有意向性的机器是与人类彼此不可分辨的(并因此是具有整全智能的);(乙)不同的意向性模式(如相信、怀疑、担心、期望,等等)的存在,会导向智能体对于所储存的信息的不同精细处理模式,而一种能够在这些处理模式之间进行恰当切换的机器,显然就具有了更强的对于环境的适应力,并因此是更智能的(举个例子说,一台能够进行适当的怀疑的机器,自然就比一台从来都不会怀疑的机器要来得更为智能)。

那么,我们又该如何建造一台使得大多数用户觉得其具有“意向性”的智能机器呢?一种最自然的方式,就是通过某种编程语言在其运作架构中嵌入某种“意向性程序”,而这种程序又与人类大脑自己执行的“意向性程序”有着某种家族相似关系(如果人类的大脑也可以被视为某种计算机的话)。或换句话说,我们应当先将人类自身的意向性结构吃透,然后再找到某种合适的数

理建模方式,将其移植到硅基体上,由此实现人造机器层面上的意向性。

而要实现上述理论目的,我们就很难不去认真复习布伦塔诺的弟子胡塞尔的意向性理论。众所周知,这一理论被公认为20世纪的西方哲学所能提供的最精细的意向性理论,因此,在不对胡塞尔哲学进行解读的前提下就去匆忙进行意向性建模,很可能就会让建模者错失很多“取经”的机会。然而,即使在国际范围内,将胡塞尔哲学与人工智能相互联系的文献还不多见。^③换言之,人工智能界目前还没有吃透“胡塞尔哲学”这碗饭。

熟悉当代英美心智哲学发展的读者或许会奇怪:意向性理论的AI化明显预设了自然主义的理论框架(根据这一框架,所有心智现象都可以被自然科学的话语方式顺化),而胡塞尔哲学的“反自然主义设定”的气息是很明显的。在这样的情况下,在人工智能的语境中强调胡塞尔的理论地位,是不是有“错点鸳鸯谱”之嫌呢?另外,在“自然主义化的意向性理论”这面大旗下,米利根(Ruth Millikan)、德瑞茨克(Fred Dretske)、福多(Jerry Fodor)的工作成就在西方也都获得了非常大的学术关注度,而为何我们的研究工作不以他们的成就为出发点呢?

对于上述疑问,笔者的简复如下:

第一,如果我们将自然主义的底线定位为“随附论论题”^④的话,那么,胡塞尔现象学所实行的“悬搁”操作并不意味着对于随附论的否定,而仅仅意味着对于随附论所涉及的外部物理状态的不关心态度(详后)。第二,现象学家对于认知活动所依赖的神经科学细节的漠视态度,在逻辑上也完全可以与作为自然主义立场之一的“非还原物理主义”相容(因为非还原物理主义者对认知活动所依赖的底层科学细节并不太关心)。第三,米利根的“生物语义学”^⑤也好,德瑞茨克的“信息语义学”^⑥也罢,都具有一种“外在主义语义学”的理论意蕴——也就是说,他们将认知系统之外的某些外在物理因素视为最终促发系统内部意向语义的源初“发动机”。但对于AI研究者来说,倘若他们预设了这种工作路径的话,他们就会预先建立对于系统所在的物理环境的整全模型,而这显然是不太现实的。第四,福多的思想语言假设——即认为“心语”(mentalese)^⑦对于语义符号的记号个例的句法操作是能够支撑起一套完整的意向活动的——或许在最一般的意义上是可以

成立的,但是,对于如何通过句法操作来实现“系统性”^⑧等基本语义属性,他尚没有给出一个具有突破性的“基于句法”的解决方案(实际上,他本人对于命题态度的“盒喻”式描述方案恰恰很难帮助我们正确理解意向内容与特定心理模式的结合方式,详后)。而这些麻烦恐怕只有通过对于胡塞尔哲学的系统复习才能得到全面避免。

本文的讨论,就将从对于胡塞尔意向性理论的重述开始。当然,这种重述本身是朝向“在AI平台上进行认知建模”这一终极目的的,因此,相关的重述所依赖的语言方式就不会过于忠于胡塞尔本人的晦涩语言风格,而会向以“明晰性”为特点的英美分析哲学叙事风格做一些适当的倾斜。

二、“现象学悬搁”对于人工智能提出的拷问

众所周知,胡塞尔的意向性理论的基本方法论预设乃是“现象学悬搁”(phenomenological epoché)。非常粗略地说,“现象学悬搁”是一种将世界中的所有对于外部世界判断(如“地球只有月球这一颗卫星”“汉献帝刘协是东汉最后一位皇帝”,等等)的真值都加以悬置,而只讨论其在意识之中呈现样态的哲学技术。用胡塞尔自己的话来说,通过悬搁“而被排除出去的东西,其实只是在记号层面上发生了一种导致价位重设的变化;而经过这种价位重设后的事态,其实是在现象学领域内重新找到了其位置。说得形象一点,被放到括号里去的东西,其实并没有从我们的现象学黑板上被擦掉,而仅仅是被放到括号里去了,并由此与一个索引产生了联系……”^⑨

如何以更为明晰的方式来理解胡塞尔的上述表达呢?笔者就此给出了两重“祛魅化”重述:

第一,基于德语语言知识的重述。现在假设我们都改用德语作为哲学研究的工作语言。有一定德语知识的读者都知道,在德语中有一种时态叫“第一虚拟态”,其用法是在间接引语中给出引文,却不对引文的真假做出断定。譬如,当我说“Sie sagte, sie sei krank”(“她说她得病了”)的时候,作为说话人的我并没有判定“她得病了”这一信息是否属实(当然,在这个德语例句中,她说了那句话的事实本身的确得到了肯定)。现在,我们不妨再做出这样一步大胆的设想:在将所有的德语语句都视为“我思”(Ich denke)这一心理模式的内容的情况下,这些语句其实都可以改造为第一虚拟态——在这种情况下,我对我所断言的

内容在外部物理世界中的真假不做任何判断,而仅仅是肯定了其在我意识中呈现为真(特别需要指出的是,这种意义上的“真”与一般人所言的带有自然态度的“真”并不是一回事,因为支持后一种“真”的证据是带有第三人称属性的,而支持前一种“真”的证据是带有第一人称属性的)。这也就是胡塞尔的“现象学悬搁”所试图达到的效果。

第二,基于语言哲学的重述。熟悉语言哲学发展的读者都知道,带有命题态度的语句是否可以被外延主义的框架处理,一直是困扰像蒯因(W.V.Quine)这样的语言哲学家的一个难题。^⑩该难题可以被简述如下:我们知道,如果“曹操在官渡打败了袁绍”是真的,那么“曹孟德在官渡打败了袁本初”也肯定是真的,因为这两句话描述的是同一个历史事件。现假设张三从历史书上读到了“曹操在官渡打败了袁绍”这一条记录,并相信之;又假设他并不知道曹操的表字是“孟德”、袁绍的表字是“本初”——在这样的情况下,他并不会相信“曹孟德在官渡打败了袁本初”这一条。那么,为何在加入“相信”这个命题态度词后,一个事件的一个真描述就推不出对于同一个事件的另外一个真描述了呢?这就说明命题态度的加入改变了语句的真值条件。说得更清楚一点,在张三的信念系统中,他是无法与“曹操本人”发生联系的,与之发生联系的乃是“曹操”这个名字,以及在他所读到的历史书中提到的那些描述。所以,曹操本人实际做了什么,乃是与张三的信念形成过程没有什么关系的,重要的是哪些关于曹操的信息被“喂”给了张三。现在我们就本着这一思路,不妨再做出一步更大胆的设想:如果所有呈现在吾辈面前的命题内容都可以加上“我相信”这样一个命题态度的话,那么由此构成的以“我相信P”为结构的新语句的真值条件就会与“P”的真值条件脱钩。由此,我们也就完成了“现象学悬搁”的操作。

读者或许会问:我们为何要跟着胡塞尔,给出一种基于第一人称的对于世界描述的改写呢?这样做对AI研究又有什么好处呢?

对于这个问题的回答其实很简单。如果你要制造出一个具有足够丰富的行为输出种类的人工智能体的话,那么它就必须要有足够丰富的心理状态——也就是说,在该智能体记忆库中储藏的信息并不是以某种机械的、外在的方式被摆放在那里的,而需要以特定心理装置相互配合的方式

而被预先加以裁减(这又好比说,进入一条炼钢厂流水线的铁矿石需要经过某种预处理,否则就会磕坏机器)。而人类心理装置的一个基本属性就是:它在单位时间内只能处理相对有限的信息,因此,它不可能将外部世界中客观存在的海量信息以一种不经裁减的方式纳入自己的“加工流水线”。从这个角度看,胡塞尔所说的“现象学悬搁”,在实质上便是一种“信息减负”作业;换言之,经过这种操作,在自然态度中对于外部对象的实存设定所需要的大量信息量(比如对于“曹操本人到底做了些什么”的系统化探究所带来的巨额信息量),便会通过对于现象学“明证性”的诉求而得到有效的压缩。

读者或许会问:在 AI 设计中进行这种的“现象学操作”,难道不会导致系统产生偏见吗?难道我们需要建造一台具有偏见的机器吗?

这里的回答是:其实我们别无更好的选择。说得更清楚一点,从逻辑上看,以“是否具有偏见”以及“是否能够自行修正偏见”为两大基本指标,我们只有如下四个选项可选:

选择一:建造一台具有偏见但无法修正自己偏见的机器。

选择二:建造一台具有自己偏见但可以自行修正自己偏见的机器。

选择三:建造一台没有偏见且不需要修正自己偏见的机器。

选择四:建造一台没有偏见且可以改进自己偏见的机器。

“选择四”显然是不符合逻辑的,因为一台在自身的知识储备中没有偏见的机器是不需要修正自己偏见的。而这就逼迫我们走向“选择三”。但“选择三”又过于理想化了,因为现有的主流的 AI 研究——无论是基于符号 AI 技术路线的专家系统还是以深度学习为代表的机器学习——都很难保证系统获得的知识是可以豁免于进一步修正的。^⑩说得更一般一点, AI 系统与人类一样,都是“有限存在者”,而这一点就使得其所获得的信念系统肯定会与作为“自在之物”的世界本身有所偏差(更何况目前 AI 系统的知识库中的信息往往是同样作为“有限存在者”的人类程序员输入的)。这种情况就又逼迫我们走向“选择一”。但“选择一”显然不能导致一种让用户满意的 AI 设计方案,因为用户自然希望系统是能够自行修正自身偏见的。而这种来自用户的要求又将我们导

向了“选择二”。该结果也恰恰是与胡塞尔“现象学悬搁”的理论指向相向而行的,因为胡塞尔的相关哲学操作既保证了现象学主体能够获得一种与客观真理的获取方式不同的真理获取方式(站在“自然态度”的立场上看,这种“真理”无疑就是偏见),又没有阻止现象学主体能够以相应的意识操作步骤来获取“偏见度”稍低一点的新信念(而这一点在对于“*Noema*”的推论主义解释中得到了明确体现,详后)。

但这里引发的新问题是:怎么来保证一个 AI 系统能够实现“选择二”提出的“具有自我修正力”这一规范性要求呢?这便是下一部分所要回答的问题。

三、从关于心理模式的“盒喻”到对于“*Noema*”的推论主义解释

很显然,如果一个认知系统能够自行地对既有的偏见进行修正的话,那么它就必须具有相对丰富的心理模式(以及作为其在语言表达中的对应物的命题态度)。比如,如果“相信”与“怀疑”都是这样的心理态度的话,那么一个系统必须先从“相信 P”走向“怀疑 P”,它才可能有动力去修正“P”的内容。但如何在认知建模中恰当地处理心理模式呢?

在回答这个问题之前,我们不妨先搁置一下胡塞尔的见解,而去看看美国哲学家福多在其“思想语言”构建中是怎么说的(之所以在此要提到福多的工作,乃是因为他的工作代表了在英美心智哲学领域内处理命题态度的一种典型做法):

思想语言假设说的是:命题态度乃是心灵与心灵表征之间的关系(正是这些表征表达出了相关态度的内容)。以俗语概括之,即:彼得相信铅块沉了这件事情,就是说:彼得有一个心语的表征——其内容就是“铅块沉了”——而该表征就处在彼得的“信念盒”(belief box)之中。^⑪

这里所说的“信念盒”,就是这样一种心灵装置:它能够对放入本盒的心语表征内容的个例记号进行某种机械操作,使得其能够锁定自身的真值。而在同样的心灵表征被放入其他命题状态盒——如“意图盒”(intention box)^⑫——后,心智又会根据各自相关的内置算法对其进行不同的二

阶操作。然而,不同的命题态度盒到底会导致心智采取具体怎样不同针对心语的算法化操作呢?对于此问题,福多的描述多少有点语焉不详。更麻烦的是,他的描述预设了不同命题态度盒之间的关系是离散的,也就是说,当整台心智机器“面板”上的某个或某些盒子被启动后,这些被激发的盒子不会导致别的盒子也与之发生“共振”。不过,有两个论证能够对这种“离散性”假设构成威胁:

第一,正如美国哲学家塞尔(John Searle)在其名著《意向性》中所指出的,“信念”与“欲望”均是可以被表征为一个连续量的,因此,讨论“信念”与“欲望”之“强”“弱”才是有意义的。而这种程度方面的可区分性,对表达某种复杂命题态度而言还是非常重要的——比如,对某事感到大喜这一点就不仅预设了相关心理主体相信相关事态发生了,并且还预设其强烈渴望这类事态之发生。^⑭然而,命题态度在程度上的可区分性却会使得不同盒子之间的界限变得模糊。譬如,我们不知道信念强度的减弱,在多少程度上会使得信念盒中的内容“溢入”另外一个命题态度盒——如“怀疑盒”——并由此破坏不同盒子之间的离散关系。

第二,亦正如塞尔指出的,对于不少复杂命题态度的分析将驱使我们引入时间因子——比如“对某事感到不满”就必须被分析为“现在相信某事发生,且过去相信此事在未来不会发生,且渴望此事不要发生”。^⑮这里所说的时间因子显然是认知系统内部的时间,而不是外部的物理时间(譬如,我们可以设想一个被“笛卡尔式精灵”完全弄混了物理时间的心智系统,依然维持着自身的内部时间系统的一致性,并在此基础上具有了种种心理意向状态)。这就引入了一个新的问题:如果系统对于其内部时间的表征也类似于某种命题态度的话,难道我们又需要为不同的内部时间表征——过去、现在与将来——提供不同的命题态度盒吗?然而,关于内部时间表征的一个基本常识性见解便是:我们其实是很难将“过去”“现在”与“将来”视为彼此离散的高阶态度的,因此,我们也就很难相信与之配套的命题态度盒是彼此离散的。

由此看来,要为一个入造的心智系统配属合用的命题态度,基于“盒喻”的福多式进路是行不通的。而在这个问题上,胡塞尔的见解又是什

么呢?

至少可以立即肯定的是,就内意识时间问题而言,胡塞尔是赞同将内意识时间意义上的“现在”“将来”与“未来”视为不同的心理模式所统摄的内容的——这三种心理模式分别是“原初印象”(original impression)、“前展”(protention)与“迟留”(retention)。不过,与福多不同,他更愿意强调这三种心理模式之间的相互渗透性与相互影响性——换言之,在他看来,时间环节A既可以视为“原初印象”的产物,也可以在相当程度上视为“前展”的产物,等等。^⑯这自然就等于否定了福多式的“盒喻”思路。有理由认为,胡塞尔对于“盒喻”的拒斥态度,也体现在他对于“Noema”这个概念的阐述方式之中。

那么,到底什么是“Noema”呢?按照胡塞尔研究专家扎哈维(Dan Zahavi)对于胡塞尔原意的概括,^⑰在经历“现象学悬搁”的操作之后,一个完整的意向性活动包括两个要素:(甲)意向性活动所涉及的感觉要素,即所谓“Hyle”(一般译为“质素”);(乙)意向性中的意义内容要素,也就是所谓的“Noema”。^⑱“Noema”也被扎哈维称为“在意向中被呈现出来的对象”(object-as-it-is-intended),以区别于物理对象自身。这里需要注意的是,至少按照扎哈维的解释,在“Noema”这一名目下,胡塞尔并没有在“意义赋予活动”与“意义内容”之间划出一条非常清楚的界限——这也就是说,“Noema”不能被简单地理解为弗雷格哲学意义上那种带有准柏拉图主义色彩的静态“意义”,而是同时应当带有动态的“意义赋予”的意味(尽管对于“Noema”的弗雷格式解释的确曾经在胡塞尔思想的诠释史上盛行过一时^⑲)。若以福多的“盒喻”为参照系,这种观点也就等于进一步否定了使得该隐喻富有意义的如下逻辑前提:“Noema”必须是某种中立于意识行为的准柏拉图式理念的物理记号,否则它就不能在被摆放到不同“盒子”里去之后还保持自身同一。

那么,我们又该如何在这种摆脱了弗雷格主义解释影响的“Noema”框架中重新理解命题态度呢?尽管对于诸种命题态度的全景式考察显然是本文力有不逮的,但一种关涉到信念强度变化的说明,却至少可以通过克劳威尔(Steven Crowell)的“Noema”重构方案而被给出。^⑳与克劳威尔的重构特别相关的胡塞尔原文乃是下面这段文字:

无论在何处,“对象”都是具有明证

性的意识关联系统所具有的名字。它最早是作为“*Noema* 式 *X*”的方式出现的,是作为意义的主词而出现的(这些意义本身又从属于完全不同类型的感觉或所与)。此外,它又作为“实在对象”的名字出现,也就是说,被其命名的,乃是以特定的可被明证的方式而被考虑到的理性的协同关联——在这种关联中,意义的形成过程也好,内在于这些关联的统一的 *X* 也罢,都得到了其理性的位置。^②

很显然,这段引文所说的“对象”,只能取扎哈维所说的“在意向中被呈现出来的对象”的意思,而这里所说的“关联”,则是指能够在现象学意识中被呈现出来的诸表征之间的协同性关系。而整段引文的意思即:所谓的“*Noema* 式对象”,其实就是诸现象之间的协同关系所构造出来的一个稳定的意义内核。克劳威尔对此进一步给出了如下例证:被感知到的颜色,作为一个“*Noema* 式对象”,是作为某种事物出现的预兆出现的;而被感知到的某个物体,作为一个“*Noema* 式对象”,正面蕴含了其没有被看到的背面的存在;被感知的一个谷仓,作为一个“*Noema* 式对象”,蕴含了诸如“农场”的相关事项的存在。^③这也就是说,任何一个“*Noema* 式对象”都必须通过与一系列其他表征的互相关联才能够确立。

克劳威尔自己曾明确表示,他对于胡塞尔的这种解释,乃是对于美国哲学家布兰登(Robert Brandom)在语言哲学层面上给出的“推论主义”(inferentialism)^④的意识哲学化版本。布兰登的推论主义的大致意思是:任何一个人在公共言谈中给出一个断言(如“曹操的表字是孟德”)时,他都需要做好心理准备,以便将该断言与一些相关推论联系在一起(譬如这样的推论:“曹操的儿子就是曹孟德的儿子”),并为可能的质疑提供好理由(譬如:说“曹操的表字是孟德”的根据,乃在于陈寿写的《三国志》)。外部的评估者则根据此人的表现给其打分,并反馈给说话人,由此构成语言游戏中的“积分系统”。而布兰登的这一理论的意识哲学化(即胡塞尔化)版本则是:任何一个在意识中呈现出来的“*Noema* 式对象”的确立,都需要预设其与一些相关表征有着某种协同关系(而且这种开放性也应当是具有一定开放性的),而意识主体在新的现象体验中对于期望中的协同关系的验证,则使得其对于意识主体的预先期望得

到了更高的积分,并由此使得从该对象之中推出新现象预测的推论力也变得更强。

这里需要特别提到的是,按照这种对于胡塞尔意向理论的新解释(下面我们就不妨称之为“胡塞尔—布兰登路线”),使得“*Noema* 式对象”的稳定意义内核得以呈现的那种处在表征之间的协调关系,不仅可以用以调整来自不同感官道的表征(如视觉表征、触觉表征、听觉表征等),而且也加以用以调整带有知觉内容的表征与抽象语义表征之间的关系。举一个克劳威尔用过的例子来说明:如果一个认知主体学会了关于“水”的分子结构的化学知识(这一知识无疑是一种抽象的语义),那么他就会将这一新表征嵌入原本由关于“水”的知觉表征如“(看起来)透明”“(闻起来)无味”等所构成的概念协同系统之中;而这样一来,对于“水”的存在性断言显然也就需要更多的明证性经验来加以支持了(顺便说一句,化学语言自身虽然是抽象的,但是对于化学证据的主观性吸纳依然是可以处在现象学视角之中的)。而这种经由概念协同关系的复杂化而导致的信念态度的变化,自然也就解释了为何一个具有中学化学知识水准的认知主体,一般不会认为一种貌似是水(却不具有水的化学结构)的物质是真正的水,而认为只是“伪水”(因为这种“伪水”的经验协同结构要比真水来得简单)。^⑤而从这一讨论中所得到的重要推论便是:按照“胡塞尔—布兰登路线”,信念的系统修正过程应当可以涵盖从日常知觉到科学描述的不同的意义领域,由此实现高度灵活的意义组合方式。

此外,如果我们采纳了按照“胡塞尔—布兰登主义的解释路数”构造出来的“意向性”概念的话,我们自然也就更无必要在讨论“相信”这个命题态度时采用福多的“盒喻”了。毋宁说,按照这种解释,“相信”这个命题态度只是一个“*Noema* 式对象”在相关表征网络中所处的地位的反思性判断:如果这样的地位被认为是更具有协同性的,则主体“相信”其存在的程度就更高,反之就更低。换言之,命题态度的性质,本身就是命题内容性质的关联物,而那种可以中立于各种“命题态度盒”而存在的命题内容,其实在胡塞尔的意向性理论中是无法得到恰当安顿的。

从 AI 的角度看,如果一个 AI 系统能够具有一种按照“胡塞尔—布兰登路线”的要求去表征意向对象的能力的话,那么它就能够以一种非常

自然的方式,根据不时进入工作记忆池的新证据,更新其对于某个信念的确证度,由此实现我们在前文中所提到的那种可能性:建造一个虽然具有偏见,却可以自行修正偏见的AI系统。反过来说,如果我们在实现这种可能的时候,放弃这种“胡塞尔—布兰登路线”,而去采纳福多所建议的“盒喻”的话,那么,程序员就很难不陷入下述工作所带来的巨大负担之中了。譬如,以公理化的方式预先设置不同的“命题态度盒”,并以一种“一劳永逸”的方式,预先规定各盒对置放于盒中的命题内容的不同操作原则,进而规定同一内容从任何一个盒子进入任何一个别的盒子后的真值变换规则。然而,这个做法显然会导致大量“削足适履”的僵化先验设计,并因此是很难应对在鲜活的日常语用环境中不时涌现出来的偶发情况的。我们将在下一部分的分析中以更多的证据来支持上述评判。^⑤

四、现有的主流AI研究能够满足胡塞尔哲学所提出的要求吗?

现在我们就将讨论的主要对象从胡塞尔哲学转移到AI。很显然,在AI的语境中讨论“如何在算法层面上实现胡塞尔的意向性理论”这一问题,我们就很难回避这样一个难题:如何在机器的层面上实现所谓的“现象学悬搁”,由此使得处在“机器意向性”之内的表征能够同时处在“机器意识”的笼罩之下呢?

考虑到在英美心灵哲学的脉络中,“意识”问题与“意向性”问题往往被分别处理,^⑥所以,上述问题似乎也应当被拆分为两个问题:(甲)“机器意识”何以可能?(乙)“机器意向性”何以可能?

乍一看,问题(甲)似乎是极难回答的,因为一个心灵二元论者会在根本上否定通过编程方式来实现“机器意识”的可能性。然而,在此立即就陷入与心灵二元论者的论战显然是不明智的,因为心灵二元论者与唯物论者之间的分歧实在是过于根本且过于“形而上学”了,以至于处在一个不那么抽象层面上的典型的人工智能哲学研究,并不能为这种争论提供恰当的场所。由于本部分的讨论主要是写给对AI研究所预设的自然主义前提抱有基本同情心的读者,所以,在此我们不妨就绕开与二元论的形而上学争辩,而讨论在经验科学的研究套路中处理“意识”的可能性。考虑到这种讨论必须与机器编程的工作具有可沟通性,

我们就不得不去寻找某种能够带有“功能主义”色彩的意识理论,以便为机器意识研究所用(顺便说一句,根据“功能主义”的立场,心智活动的实质乃在于其某种既能体现于碳基生命体又能体现于硅基元件的抽象功能。这种“本体论宽容性”显然是为AI研究者所乐见的)。按照该标准,巴爱思(Bernard Baars)的“全局工作场域论”(global workspace theory)^⑦似乎就应当被“机器意识”的研究者所偏好,因为这种理论的抽象描述形式——“意识状态”就是“工作记忆”中被注意力机制所关注到的事项——是完全可以对于“工作记忆”与“注意力”的计算建模工作而在一个计算平台上被加以复制的。^⑧至于这样的工作成果是否能够把握到“意识”的那种神秘的主观面相,则是一个牵涉到“主观面相”之本质的术语学问题,并因此并不需要AI专家在第一时间加以面对。^⑨

真正麻烦的是前述问题(乙),因为即使是我们采用了对于巴爱思的意识理论的计算化建模方式,我们依然无法由此就构造出具有命题态度与命题内容的完整的意向性结构。而试图在这个方向上做出努力的AI专家,则太容易落入福多提出的“盒喻”的窠臼了(尽管他们未必读过福多),因为“盒喻”这一表达与常识心理学所给出的意向性结构之间的相似性,的确很容易诱使人们去认真地对待该比喻。譬如,意大利人工智能专家癸翁奇利亚(Fausto Giunchiglia)和鲍奎特(Paolo Bouquet)在他们合写的长文《语境推理导论——一种人工智能的视角》^⑩中,便在“语境建模”这个题目下谈到了对于“语境”的“盒喻”化处理方式——有鉴于“命题态度”本身也可以被视为一种特殊的“语境”(如“相信语境”“怀疑语境”等),这种谈论显然具备了对于“意向性建模”的覆盖力。根据此二人的叙述,每个语境是一个盒子,每个盒子均有边界,进入和离开这盒子也都需要遵照一定的规则。而任何一个对语境敏感的句子,也只有在被放到这样的—个盒子中去之后,才能够获得确定的真值,并由于所在的盒子不同而具有不同的真值。譬如:“马年是甲午年”这个句子在“2014年”这个“盒子”里是真的,但移到“1978年”这个“盒子”中却马上就变成假的了(1978年虽是马年,却不是甲午年,而是戊午年)。同样的道理,“曹孟德在官渡打败了袁本初”这个句子在“李四相信”这个盒子里是真的,而在“张

三相信”这个盒子里却是假的了(假设李四是知道曹操与袁绍各自的表字的,而张三不知)。

那么,怎么刻画命题内容从一个盒子到另外一个盒子中的迁移规则呢?举例来说,古哈(Ramanathan V. Guha)和麦卡锡(John McCarthy)^⑧就在将每个语境加以编码的前提下,将语境之间的最重要关系界定为“提升关系”(lifting relations)。其相关提升公式示例如下:

$\forall C_1 \forall C_2 \forall p (c_1 \leq c_2) \wedge \text{ist}(c_1, p) \wedge \neg \text{ab aspect } 1(c_1, c_2, p) \Rightarrow \text{ist}(c_2, p)$

该公式读作:对于任何两个语境来说,只要其中一者包含于另一者,任一事件 p 在较小的语境中成立,且在“方面 1”这两个语境和该事件都不是反常的,那么该事件也在较大的语境中成立。

而癸翁奇利亚和鲍奎特则不喜欢麦卡锡和古哈所提出的方案,因为这样的方案必须将语境本身加以对象化,最后势必构成一个“大语境套小语境”的“俄罗斯套娃”结构,在技术上会显得非常笨拙。他们的替代方案是将任何一个信念主体对于外部世界的表征刻画为一个局域性理论,并在此基础上讨论不同的局域性理论之间的兼容性,由此完成从一个主体的信念到另外一个主体的信念的推理过程。相关的推理规则被统称为“桥律”:^⑨

$$\frac{c_1 : \Phi_1, \dots, c_n : \Phi_n}{c_{n+1} : \Phi_{n+1}}$$

其直观含义是:如果我们已经知道了在语境 C_1 中表达式 Φ_1 为真,语境 C_2 中表达式 Φ_2 为真……语境 C_n 中表达式 Φ_n 为真的话,那么我们也知道了在语境 C_{n+1} 中表达式 Φ_{n+1} 为真。

在这里,我们没有篇幅具体讨论癸翁奇利亚和鲍奎特的工作细节。但至少可以肯定的是,他们的工作依然不足以在最低限度上满足“制造一台能够自动修正偏见的智能机器”这一目标。其道理也是非常明显的:其理论模型只能预先假设不同的主体对于世界本身有着片面的或者近似(却都不包含明显谬误)的局域模型,却无法假设不同的主体对于世界本身有着可能在根本上就是错误的局域模型——尽管“具有世界的错误认知”这一点对于人类来说实在是太过平常了。此外,他们的推理模型并不包含对于突然涌入的新证据的处理方案,特别不包含对于新证据与旧信念之间矛盾的处理方案。因此,我们是无法从他们的工作基础出发来建立起一个带有布兰登推论

主义风味的胡塞尔意向性模型的。

读者可能会问:癸翁奇利亚和鲍奎特的工作毕竟是属于比较传统的“符号 AI”路数的,而目下如火如荼的“深度学习”模型,能够在逼近胡塞尔的意向性模型方面有所进步吗?

答案是否定的。以与命题态度刻画作为相关的深度学习模型——“深度信念网络”(Deep-belief networks)^⑩——为例,该技术目前的主要用途仅仅是对图像等初级材料进行貌似带有信念内容的语义标注。然而,这样的网络得到的语义标注都是作为网络训练者的人类程序员事先设定好的,而网络所做的,仅仅是通过大量的训练以便将特定的感觉材料与特定的语义标注加以联系——而作为这种训练的结果,网络不可能对训练流程规定之外的输入材料给出恰当的反应。与之相比较,胡塞尔意义上的“*Noema*”却可以成为具有不同感官道来源的感觉材料的意义统一者,甚至随时准备好接受某种相对抽象的语义。另外,我们也不知道这样的深度学习构架将如何处理丰富的命题态度之间的切换——换言之,我们很难将“怀疑”“期望”“担心”“回忆”这样的命题态度集指派给它。从这个意义上看,深度学习模型似乎比癸翁奇利亚和鲍奎特的工作成果更难成为胡塞尔意向性理论的合格的机器实现者。

有的读者或许还会说:在符号 AI 与深度学习之外,还有一个技术路数值得胡塞尔的意向性理论的计算建模者加以考量,这就是所谓的“能动性”(enactivism)。从哲学上看,作为一个认知科学纲领的“能动性”具有如下四个学术标签:^⑪“具身性”(embodiment),即认为认知不仅牵涉到了中枢神经系统,还牵涉到了其以外的整个身体运作;“嵌入性”(embeddedness),即认为认知活动是被嵌入到一个与之相关的外部环境中去的;“能动性”(enactedness),即认为认知活动不仅仅牵涉到组织体对于外部输入的被动信息加工,而且更牵涉到组织体对于外部环境的主动影响;“延展性”(extendedness),即认为认知活动所随附的物质基础是从大脑延展到外部环境中去。目前,这种被概括为“4E 主义”的学术主张已经成为一个横跨哲学、心理学、教育学与 AI 的跨学科运动,在西方获得了大量的学术关注度。

但这里的问题是:能动性是否一种有用的资源,可以被用来为胡塞尔的意向性概念进行计算建模呢?笔者的看法并不是那么乐观。从哲学

气质上看,对于肉身与环境在认知上的强调,使得能动主义的学术光谱更接近海德格尔与梅洛·庞蒂,而与胡塞尔更偏向意识哲学的风格有所分别。另外,对于环境因素的过多偏重,在相当程度上为能动主义者解释那些与环境脱钩的意向对象(如像“孙悟空”这样不存在的对象,以及像“产权”这样负载抽象语义的对象)在意识领域内的呈现增加了难度。更麻烦的是,在AI领域内对于能动主义的计算建模工作,其现象学意味会更加淡薄。譬如,在AI专家兰戴尔·贝尔(Randall D. Beer)的论文《一种动力学系统视角中的“能动者—环境”之间的交互关系》^⑤中,作者为了能够刻画出能动者与环境之间的互动,率先将两者分别刻画为两个动力学系统,并在此基础上将两者之间的互动性解释为两个动力学系统之间的耦合关系。但这种建模方式预设了建模者有某种独立于能动者与环境的“第三方视角”,而这显然会在哲学上预设胡塞尔所反对的“自然主义态度”,并因此与胡塞尔的哲学立场脱钩。

一个能够在能动主义与胡塞尔之间找到平衡的学术资源恐怕是认知语言学(cognitive linguistics),因为认知语言学对于“具身性”的强调的确构成了其与“4E主义”的亲缘关系;而其对于认知图式的直观化表现形式,则又让人联想到胡塞尔的现象学直观。但问题是:关于如何在计算机建模的平台上重现认知语言学的观点,目前学界尚没有与之配套的成熟的技术手段。^⑥所以,对于“胡塞尔的意向性理论的计算化建模”这一任务来说,认知语言学的资源可谓“远水不解近渴”。

综上所述,主流AI学界目前应当还没有能力消化胡塞尔的意向性理论,并将其付诸实际的建模工作。

五、总结

从总体上看,本文的讨论可以分为三个部分。第一部分讨论了胡塞尔的“现象学悬搁”方法对于AI研究的一般意义,特别是强调了建造一种具有现象学视域并具有错误自动修正机能的智能机器的必要性。第二部分则勾勒了一种以布兰登的推论主义为参考要素的胡塞尔式意向性理论,并对福多提出的关于命题态度的“盒喻”提出了批评。在论文的第三部分中,笔者则对主流AI在处理意向性问题方面的无力性进行了揭露——这些主流进路要么根本就无法在符号表征的层面上触

及意向性问题(如深度学习),要么不得不采用福多的“盒喻”而导致模型的笨拙性(如某些符号AI技术),要么就在过分强调环境与主体的相互关系的同时预设了胡塞尔所反对的“自然主义态度”(如某些能动主义技术路线)。由此看来,如果要设计一种在最低限度上符合胡塞尔精神的AI系统,我们就必须与目前的主流AI技术路线分道扬镳。

而在这个问题上一个值得推荐的“非主流”AI技术路线,则由王培先生发明的“非公理推演系统”(Non-Axiomatic Reasoning System)——或“纳思系统”——来加以提供。^⑦这是一个试图以“非公理的”(Non-Axiomatic)灵活方式为系统进行知识编码的通用人工智能系统,而“非公理的方式”一语在此真正的蕴意乃是:该系统的语义学知识,是能够随着系统学习经验的丰富化而不断被丰富化的,而这一点也就能使得编程者从“为系统事先编制万无一失的语义库”的繁重任务中被解放出来。同时,纳思知识库对于外部环境知识的“片面性”,也使得其能够更好地体现“现象学悬搁”的真义——即以在“意识”(在此指在机器的工作记忆中能被注意力机制“照亮”的部分)中涌现的信念的真为真,而不管其在外部世界中的真值条件如何。此外,在纳思系统中,词项之间的语义联系是可以在一种不引入“命题态度盒”的前提下而模拟命题态度的某些基本变化的,这在相当程度上就使得对于“Noema式对象”的计算模拟有了基本保障。不过,限于篇幅的关系,以及本篇论文的“哲学”性质,关于纳思系统模拟胡塞尔意向性理论的技术细节,笔者在此只能予以省略了。

注释:

①F. Brentano, *Psychology from an Empirical Standpoint*, London: Routledge and Kegan Paul, 1973, pp. 88-89.

②参看 Daniel Dennett, *Consciousness Explained*, New York: Little, Brown and Co, 1991.

③在国际上,在“现象学与认知科学”这个名目下展开的研究大多聚焦于海德格尔哲学或梅洛·庞蒂哲学,而不是胡塞尔哲学。少数涉及胡塞尔-AI关系的文献,对于“如何为胡塞尔的意向性理论进行计算建模”这一问题其实并没有太深入的讨论。参看:M. Issac (2018), “Towards a phenomenological epistemology of mathematical logic”, *Synthese* 195: 863-874; A. Beavers (2002), “Phenomenology and artificial intelligence”. *Metaphilosophy*: 33: 70-82(依据笔者浅见,前一文献对胡塞尔《逻辑研

- 究》关注太多,而没有触及胡塞尔的《观念》系列所提出的意向性理论;后一文献的问题是缺乏对胡塞尔的意向性理论与特定AI路径之间关系的比较性研究)。
- ④这里涉及的“随附性论题”的具体含义是指:所有心智活动都随附于物理事件(如特定的神经学事件)。对于这里提到的“随附性”概念,定义很多,其中的一种定义是:B层面发生的事件随附于在A层面上所发生的事件,当且仅当对于任意两个不同的可能世界W1和W2而言,若W1和W2在A层面上所发生的事件乃是彼此不可被分辨的,那么,它们在B层面上所发生的事件亦是彼此不可被分辨的。这里需要注意的是,关于“随附性论题”是否界定物理主义立场的基本界标,目前学术界有不同看法。鉴于主题所限,笔者就不展开了。
- ⑤Ruth Millikan, *Language, a Biological Model*, Oxford: Oxford University Press, 2005.
- ⑥Fred Dretske, *Knowledge and the Flow of Information*, Cambridge, Massachusetts: The MIT Press, 1981.
- ⑦指的是一种独立于自然语言而纯然在心理系统中存在的语言操作系统。如果将人类心智比作计算机并将自然语言比作“界面语言”的话,心语大约就等于“编程语言”的层面。
- ⑧指的是这样一种性质:任何一个语句都能够在一个合格的心语操作者那里被看成是与其他语句具有内涵关联的。比如这样的关联:“我的狗不咬人”这话就蕴含了“我有狗”。
- ⑨E. Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology*, translated by F. Klein. Hague: MartinusNijhoff, 1980, p. 171.
- ⑩对于该问题(特别是蒯因对于该问题的讨论)的反思式评述,请参看:J. M. Bell (1973), “What is Referential Opacity?” *Journal of Philosophical Logic* 2 (1):155-180.
- ⑪具体而言,对于符号AI来说,其最大的麻烦便是:在这些系统中被公理化方式加以固定的人类知识,是很难再通过公理化方式得到实质性修正的(与之对比,人类在任何一个经验领域内的知识都应当是随着人类认知的加深而被不断修正的)。而对于联接主义或者深度学习路径来说,最大的问题便是:一旦一个系统通过数据训练而得到了稳定的性能,其性能也会被锁定在一个固定的层面上,而不会发生实质性变化(与之相对比,人类习得的某种机能却可以在被习得后再得到精进)。
- ⑫Jerry Fodor, *LOT2: The Language of Thought Revisited*, Oxford: Oxford University Press, 2008, p.69.
- ⑬Ibid., p.39.
- ⑭John Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 1983, p.33.
- ⑮Ibid., p.32.
- ⑯胡塞尔本人的一个隐喻是颇能说明这种“相互渗透性”的。他指出,“当下”在意识场中的地位类似于在一个连续色谱中纯红的地位:纯红虽然是诸种红色色调在“纯粹性”这一点上所能够达到的极限,但我们并不是通过一系列不同的颜色把握活动把握到诸种红色之间的色差的——毋宁说,我们是通过某种统一的色感把握活动来把握到关于色彩的整个连续统的。与之相类比,我们也不是通过不同的时间把握活动把握到“现在”“过去”与“未来”的——毋宁说,我们是在某种统一的时间把握活动中把握到关于时间的整个现象学连续统的。参看:Edmund Husserl, *On the Phenomenology of the Consciousness of Internal Time*, translated by John Barnett Brough, Kluwer Academic Publisher, Dordrecht, 1991, pp.41-42.
- ⑰Dan Zahavi, *Husserl's Phenomenology*, Stanford (CA): Stanford University Press, 2003, pp. 57-58.
- ⑱“*Noema*”可以被翻译为“意向对象”,也可以被翻译为“意义相关项”,不过,此词目前在国内现象学界尚无一个完全没有争议的统一译名。为了避免陷入不必要的译名争议,在本文中笔者就不再翻译“*Noema*”,而在行文中直接展现此词的原文面貌。
- ⑲相关解读书籍有:D.Føllesdal (1969), “Husserl's notion of *Noema*”, *Journal of Philosophy*, 66: 680 - 687; R. McIntyre, (1982), “Intending and Referring”, in H.L. Dreyfus and H.Hall (eds), *Husserl, Intentionality and Cognitive Science*. Cambridge, MA: MIT Press, 1982, pp. 215-231.
- ⑳S. Crowell (2008), “Phenomenological immanence, normativity, and semantic externalism”, *Synthese* 160:335-354.
- ㉑E. Husserl, *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First Book: General Introduction to a Pure Phenomenology*, translated by F. Klein. Hague: MartinusNijhoff, 1980, p. 347.
- ㉒S. Crowell (2008), “Phenomenological immanence, normativity, and semantic externalism”, *Synthese* 160:344.
- ㉓R. Brandom, *Articulating reasons*, Cambridge, MA: Harvard University Press, 2000.
- ㉔从这个角度看,一个胡塞尔主义者完全可以在不引入普特南(Hillary Putnam)的外在主义语义学的前提下,说明为何在“孪生地球人”的思想实验中,为何对于“水”的判定可以既符合化学知识,也符合现象学的“内在性”要求。
- ㉕不过,在此我们不排除这样一种可能性:“盒喻”虽然在“心语”或者“思想语言”(其可类比为计算机的程序语言)的层面上无效,但却在公共的自然语言(其或可类比为计算机的界面语言)的层面上有效,否则我们就难以解释为何我们一般人可以在日常对话中使用命题态度词了。不过,这就引入了一个如何将缺乏明述化的命题态度的心语过渡为自然语言表达的问题。限于篇幅,我们在此暂不去细究这个问题。
- ㉖如塞尔就曾指出,有些意向性信念就是可以没有意识的(如这样的一个信念——“我的确相信我的祖父从来没有离开过银河系”。该信念可以仅仅以潜在的方式存在于我的意向系统之中,而不是我的意识的一部分);而有些意识是可以没有意向性的(如一阵突然涌现的狂喜)。因此,关于意识的话题并不就等于关于意向性的话题。参看:John Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 1983, p. 2. 不过,对于一个胡塞尔主义者来说,塞尔提出的那些关于“有意识而无意向性”或“有意向性却无意识”的例子似乎都很牵强(然而,由于篇幅与主题的限制,笔者在此就不再对塞尔给出的例句进行深入分析了)。
- ㉗Bernard Baars, in *The Theater of Consciousness*, New York, NY: Oxford University Press, 1997.
- ㉘譬如,对于“注意力”的计算建模成果就可以参看:Nicola De Pisapia, Grega Repovš, and Todd S. Braver, “Computational Models of Attention and Cognitive Control Nicola De Pisapia”, in Ron Son (ed.), *Cambridge Handbook of Computational Psychology*,

- Cambridge: Cambridge University Press, 2008, pp. 422-450。
- ②巴爱思的理论当然没有真正触及关于意识的“难问题”,即意识的主观面相问题。但是对于AI研究来说,做出一个能够在功能层面上具有“意识”的系统已经够具挑战性了。此外,对于巴爱思理论的基于“意识之主观面相的不可解释性”的批评,会威胁到对于意识的任何一种自然主义解释,并使得这种批评失去了对于巴爱思理论的针对性。顺便提一句,巴爱思的意识理论当然只是目下林林总总的意识理论中的一种而已。笔者之所以在此只提到这个理论,主要是因为该理论的功能主义色彩最强,因此其对于“机器意识”研究的覆盖性也最强。
- ③Fausto Giunchiglia & Paolo Bouquet, “Introduction to Contextual Reasoning: an Artificial Intelligence Perspective”, in *Perspectives on Cognitive Science*, edited by B. Kokinov, New Bulgarian University, 1997, pp. 138-159.
- ④Ramanathan V. Guha, & John McCarthy, “Varieties of Contexts”, in *Modeling and Using Contexts*, edited by Patrick Blackburn et al, Springer-Verlage, Berlin, 2003, pp. 164-177.
- ⑤Fausto Giunchiglia & Paolo Bouquet (1998), “A Context-Based Framework for Mental Representation”, in *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society*, pp. 392-397.
- ⑥关于深度信念网络的文献很多,比较有代表性的有:G. E. Hinton, S. O. Sindero and Y. W. Teh (2006), “A fast learning algorithm for deep belief nets”, *Neural Computation* 18:1527-1554。
- ⑦相关概括请参看 Mark Rowlands, *The New Science of the Mind*, Cambridge (MA): The MIT Press, 2010, p.3。
- ⑧Randall D. Beer (1995), “A dynamical systems perspective on agent-environment interaction”, *Artificial Intelligence* 72: 173-215.
- ⑨笔者所知道的试图为认知语言学进行全面计算建模的唯一尝试,乃见诸如下文献:Kenneth Holmqvist, *Implementing Cognitive Semantics*, Lund University, 1993-01-01。这是该文献作者完成的博士论文,但没有在任何一家出版社正式出版,而只是在其个人网页上提供了扫描件下载服务。他的基本思路是激活英国计算机专家玛尔(David Marr)的计算视觉理论,以便为具有明显空间性面相的认知图式(cognitive schemata)提供计算建模方面的支持。笔者认为该思路是颇为有趣的,但是考虑到认知语言学家所提供的认知图式的高度多样性,以及玛尔的视觉计算模型自身的繁复性,该技术路径所导致的建模结果也肯定是非常复杂的,实用价值有限。实际上,即使是作者本人,在完成博士论文后的工作领域也逐渐转入了纯粹认知心理学意义上的视觉问题研究,而与认知语言学无涉。此外,笔者也没有查到有任何别的学者沿着这一路径继续探索下去。
- ⑩关于纳思系统的文献很多,其中最重要的是 Pei Wang, *Rigid Flexibility: The Logic of Intelligence*, Netherlands: Springer, 2006。

Some Remarks on the Relationship between Husserl's Notions of “Intentionality” and Artificial Intelligence

XU Yingjin

(School of Philosophy, Fudan University, Shanghai 200433, China)

Abstract: The relevance of Edmund Husserl's notion of “intentionality” to artificial intelligence (AI) has long been underestimated. As a matter of fact, although the phenomenological methodology of “epoché” appears to be undermining any attempt to do cognitive modelling, it can still offer inspirations for AI in a deeper sense. To be more specific, the “omnipresence” of psychological modes in pure consciousness as the residue of “epoché”, if appropriately reinterpreted, can provide principles an AI-oriented systematic reconstruction of psychological modes, which could facilitate an AI system to revise its beliefs in a flexible manner. Moreover, Husserl's notion of “Noema”, if reinterpreted in accordance with Brandom's inferentialism, can be regarded as a competing theory with Jerry Fodor's “Language of Thought Hypothesis”, which embraces a theory of psychological modes, according to which there are different “boxes”, each of which specifies a certain psychological mode. Husserl's notion of “Noema”, by contrast, can explain belief revisions without presupposing the existence of these “boxes”. However, the mainstream AI approaches simply ignore Husserl's insights of intentionality. For example, the connectionist/deep learning approach cannot handle intentionality on the symbolic level, and the symbolic AI approach cannot resist the temptation of presupposing Fodorian “belief boxes” in an inflexible manner, whereas the enactivist approach overemphasizes the importance of the interplays between artificial agents and their environments by modelling both in a way not from a certain phenomenological perspective. Therefore, it is not unfair to conclude that the algorithmic realization of Husserlian intentionality has to be “suspending” most mainstream AI approaches.

Key words: intentionality, epoch, belief box, inferentialism, artificial intelligence, noema

(责任编辑:知 鱼)