

预测认知模型：认知科学的新统一范式？

叶浩生 苏佳佳

【摘要】半个多世纪以来，“认知科学”依然是一个多学科汇集而成的研究领域，并没有可担当共同科学话语体系的“研究范式”。21世纪，在“复杂科学”的时代精神下，“动力学”的发展引领了认知科学的“范式转移”，“预测认知模型”集心理学、物理学和认知神经科学等众多学科之力，有望夺得“认知科学大一统范式”的桂冠。但是，预测认知模型诞生于神经科学，能否成功嫁接于认知科学还存在争议。因此，本文的问题意识就是：预测认知模型如何能成为认知科学的新统一范式？文章首先以认知科学的理论演变“认知主义”“联结主义”和“具身认知”与预测认知模型的结合为主线，梳理出三种预测认知模型，发现三者皆因表征问题而未能突破“信息加工范式”；其次，又以“表征之战”“生成认知”和“意义建构”为脉络，论证出传统预测认知模型均站在表征主义立场上，实际上并未解决表征之战，而崇尚“以意义建构取代表征”的生成认知似乎能够与预测认知模型完美契合而应对挑战；最后，在生成认知强调“意义建构”的启发下，我们提出了意义建构的三个视角，并且尝试建立了“生成认知的预测心智模型”，澄清了意义建构的动力学机制。虽然，预测认知模型被质疑是否“可证伪”，但是，与“达尔文演化论”异曲同工，预测认知模型的确担当了认知科学的“最佳解释推理”。有鉴于此，文章结论认为，预测认知模型与生成认知的结合将推动认知科学的彻底“动力学化”，从而有望成为认知科学新统一范式。

【关键词】 预测认知模型；表征之战；生成认知；意义建构；新范式

一、认知科学究竟有没有统一的范式？

1978年，“斯隆报告”首次提出了认知科学的学科定义：“认知科学研究智能实体与外界环境互动的原理。理所当然，这种研究必将超越学科界限，将神经科学、计算机科学、心理学、哲学、语言学和人类学等学科的研究者们所做的工作整合在一起”。因此，“广义上的认知科学”主要是由上面六大支撑学科组成的“认知科学六边形”，其中又交叉形成了控制论等十一个新兴学科。但是，如果要

叶浩生，心理学博士，浙江师范大学教师教育学院教授、博士生导师（金华321004）；苏佳佳，通讯作者，浙江师范大学教师教育学院博士生（金华321004）。本文系国家社科基金后期资助重点项目“身体运动与心理发展研究”（20FTYA002）的阶段性成果。

①Google Scholar在2005—2022年期间（截至2022年8月17日）发表了3750000条以“predictive coding”和“predictive processing”为关键词的论文，可见国际上认知科学的预测研究已经如火如荼，盖因其可对知觉、注意、记忆、动作、幻觉、自闭症甚至意识等认知现象提供大一统解释，故Clark（2013）也将“预测认知”称为“大一统理论”（Grand Unified Theory, GUT）。本文中，我们将预测认知的不同取向统称为“预测认知模型”，关于该领域的详实综述，可参考Hohwy（2020）的附录Supplementary Table S1：Overview of recent philosophically oriented predictive processing research。

将认知科学视作独立学科,不仅需要一个多学科汇集而成的研究领域,更需要一个可以担当共同科学话语体系的研究范式。

在此背景之下,“狭义上的认知科学”出现了。当时,在计算机科学的时代精神下,第一代认知科学“离身认知”,将人比作计算机,采用认知主义和联结主义所共同倡导的“表征计算主义”,认为认知过程就是“知觉(输入)—认知(表征计算)—行动(输出)”的“认知三明治”模式。在此视域下,认知科学的范式一直是“信息加工范式”,而到了认知心理学中就被降格为“S-O-R”范式,这个范式的特点就是“认知”的“意义”不被关注,而只关注“认知”的“表征”。这其实是源自“笛卡尔主义”以来的一种心智研究传统,即在机器功能主义立场中,默认了认知的形式和内容是二分的,并且将“认知”定义为“认知形式”(表征),但是却忽视了“认知内容”(意义),存在着无法解决的“符号落地问题”(symbol grounding problem)。第二代认知科学“具身认知”,试图将认知的意义奠基于“身体”之上,强调“身体”在认知过程中扮演的不仅仅是因果性角色,更是构成性角色。虽然,具身认知将身体引入认知舞台的中央,但是,依然是在为传统认知科学做着修修补补的工作,并没有摆脱“信息加工范式”。经由近三十年的发展,第二代认知科学已演变为“4E 认知”,即具身认知、嵌入认知、生成认知和延展认知(Newen, et al., 2018)。其中最具争议性的“生成认知”(Varela, Thompson & Rosch, 1991, p. 8),强调大脑、身体和环境构成了一个“耦合”的动态系统,认知是“非表征”的,因对表征持有全盘否定的立场,而有望成为具有认知科学的革命性取向(De Jaegher & Di Paolo, 2007; 叶浩生等, 2019)。遗憾的是,生成认知虽然在理论上试图对旧有的认知科学范式进行革命,但没有从根本的机制上做出合理解释,这样,新一代认知科学的诞生就亟需建立一种新的研究范式。

20 世纪 90 年代以来,在“复杂科学”的时代精神下,“动力学”的发展引领了认知科学的“范式转移”(Constant, Clark & Friston, 2021): 认知活动逐渐被视为主体变量与环境变量彼此耦合、交互共变的过程。21 世纪,英国皇家学会会员、伦敦大学学院教授 Karl Friston(2005, 2009, 2010)融合心理学、物理学和认知神经科学等多学科研究成果,首次提出了“自由能原理”,并进一步演变为“预测认知模型”,与传统认知科学的信息加工范式相反,该模型认为大脑其实是一台主动的预测引擎,“预测”是一切认知活动背后的基本原理,这将一劳永逸地解决人工智能乃至人类心智的终极奥秘,从而有望夺得认知科学领域“大一统范式”的桂冠。

但是,预测认知模型终究是诞生于神经科学,是否成功嫁接于认知科学还存在争议。因此,本文的问题意识就是:预测认知模型如何能成为认知科学的新统一范式?

二、传统的预测认知模型

(一) 认知主义的预测模块假说

20 世纪 60 年代起,“认知主义”一直是认知科学的主流范式,Fodor 甚至戏称其为“城里的唯一游戏”(the Only Game in Town)。在计算机科学的时代精神下,认知主义核心假设是“计算隐喻”,认为人的心智就类似于计算机的符号加工过程,只需要考察大脑的“输入—处理—输出”过程就能理解心智。因此,在认知主义视域下,大多数学者聚焦于描绘大脑的被动推理过程,并没有对主动预测过程做出细致刻画,不过,我们仍然可以借助 Fodor、Pylyshyn 和 Marr 等人的理论来梳理出一些基本特征:①认知是被动的;②认知是表征计算的;③认知是颅内的;④预测加工模块也是被动的、表征计算的和颅内的;⑤以上推理可得,认知主义与预测认知模型也可以兼容。

基于此推理,我们尝试勾勒出了“认知主义的预测模块假说”^①:虽然认知主义强调,颅内信息加工过程需要各种功能模块互相配合共同完成认知过程,但是,预测模块仅仅是一种被动的、颅内的和表征计算的推理模块,仍然需要以大脑中的“表征计算单元”为核心中介才能进行信息加工处理。如此看来,该假说只不过是传统的“知觉—认知—行动”模型刻画到极致的“新认知三明治模型”,认知依然是纯粹的“脑活动”,那么,身体与环境自然就被排除在外。难怪哲学家 Putnam 用“缸中之脑”(brain in vat)的隐喻来批评认知主义者:离开了身体和环境,人类的大脑与培养皿中的大脑又有什么区别呢?

(二) 联结主义的预测编码理论

20 世纪 80 年代,在认知神经科学的时代精神下,“联结主义”开始挑战认知主义传统,宣称“我即我脑”(We are our brains)。认知科学家开始借鉴大脑中的神经网络模型来模拟人类的心智,其核心假设是“神经元隐喻”,认为认知过程就是神经元之间的复杂联结,可以通过设定输入层、隐藏层和输出层的方式,将信息处理过程理解为分布式、多层级、计算性的表征过程。在联结主义视域中,预测认知模型出现了以 Friston 和 Hohwy 等人为代表的“半具身”(semi-embodied)版本——“预测编码理论”。我们将基于 Hohwy 在 2013 年出版的专著 *The Predictive Mind* 来总结该模型^②:

1. 预测的源头:自由能原理(Free-Energy principle, FEP)

Friston(2005, 2009, 2010)借助物理学中的“热力学第二定律”,提出了认知科学中的“自由能原理”:从演化的角度来看,任何生物体为维持自身的生存繁衍,都必须使其自由能(意外)最小化。人作为一个以皮肤为边界的生物系统,同样需要维持一种内稳态(homeostasis)。换言之,为了活下去,人类就必须不断“预测”即将发生的危险,减少可能遇到的意外(surprise)。而这个预测过程就需要借助“贝叶斯推理”来实现。

2. 预测的工具:贝叶斯推理(Bayes theorem)

Hohwy(2013, pp. 13 – 38)认为,对人类而言,现实世界中的因果关系并非“一一对应”,大脑不是被动的感知推理,而是主动的贝叶斯推理:主体能根据(1)或然率,即采用与某个外部原因相对应的模型(信念)时,接收到当前输入信息的可能性,和(2)前述信念的先验概率,对当前采用的模型进行调整,对调整后的信念的概率(即后验概率)做出正确估计。而正确的估计就需要“预测误差最小化”作为目的。

3. 预测的原则:预测误差最小化(Predictive errors minimization, PEM)

Hohwy(2013, pp. 41 – 56)认为,贝叶斯推理无法回答“预测如何成功”,即如何判断先验假设和感知输入之间是否成功匹配。具体来看,一方面神经系统不断进行自上而下的预测,另一方面真实世界不断进行自下而上的输入,二者之间的“失匹配”构成了预测误差。因此,预测误差最小化就是预测加工模型的目的。而要实现此目的,需要依赖“多层级生成模型”。

4. 预测的架构:多层级生成模型(Hierarchical generative models)

Hohwy(2013, pp. 13 – 38)认为,现实世界中的因果关系不仅不是一一对应的,而且拥有不同的时空尺度,充满了非线性相互作用的原因。为了推理这些原因,就需要多层级生成模型。因此,与认知主义中“刺激—反应—推理”的线性预测模型不同,联结主义借助分布式预测模型,认为人类不仅具有“自上而下”与“自下而上”的双向预测架构,而且还具有多层级的隐藏信息处理层。而若要实现多层级的预测误差最小化,就必然需要“精度加权”的参与。

^①“认知主义的预测模块假说”详细论述可参见朱林蕃、刘闯,2022:《预测心智的三种理论模型:一个比较性研究》,《哲学分析》。尤见论文中图 2:后福多大规模认知单元对人类预测过程示意图。

^②“联结主义的预测编码理论”图可参见 Hohwy(2013, p. 68)中的 Figure 7. The perceptual hierarchy, second version.

5. 预测的机制:精度加权(Precision weighting)

Hohwy(2013, pp. 59 – 74)认为,世界的不同状态(比如环境)也会对人类的感官输入造成极大的不确定性。因此,预测误差最小化就需要考虑“精度”,即通过多层生成模型中双向传递的信号所含有的相应的权值来对信号进行可靠性评估。精度越高,意味着预测误差的不确定性越小,才能做出最优决策。但是,如果想要获得精度,仅通过大脑的推理是不可能的,必然要涉及身体的行动,那么,大脑是否就是“预测边界”,就尤为重要。

6. 预测的位置:自证的预测边界(Self-evidential boundary)

在 Hohwy(2013, pp. 224 – 241)看来,预测的位置肯定位于脑内,因为人类的所有推理都需要基于大脑内部的表征(证据)来实现贝叶斯计算,而身体和行动仅仅作为一种“选择性采样”的方式为大脑提供推理证据。正是在这个意义上, Hohwy(2016, p. 259)宣称:“相反,PEM 揭示了思维在推理上与世界隔绝,它似乎更多地受到大脑神经中枢的束缚,而不是具身化的或延展化的,行为本身更多的是感官输入的推理过程,而不是与环境的主动耦合”。因此,预测认知模型就不得不以大脑为“自证的预测边界”,与身体和行动无关。

这就是为什么我们称以 Hohwy 为代表的预测认知模型是“半具身”的。虽然,与认知主义的预测模块假说相比, Friston 与 Hohwy 承认,预测编码理论既需要感官为大脑提供信息输入(被动推理),也需要身体和行动对信息进行选择(主动推理)。但是, Hohwy 更强调,身体和行动处于预测边界之外,预测的主体依然是大脑,身体和行动的使命只是为大脑内部的预测过程提供信息输入。归根结底,这仍然是一种“改良了”的认知主义的预测认知模型,或者是“不彻底”的“具身”的预测认知模型。

(三) 具身认知的预测加工模型

20 世纪 90 年代,认知科学出现了一个新取向,“一个幽灵盘旋在认知科学实验室的上空,这个幽灵就是具身认知”。受控制论、系统科学和机器人学的启迪,“具身认知”拒斥通过计算机或神经网络来模拟人类心智,其核心隐喻是具身动力系统,认为“身体”在认知过程中扮演了重要角色。因此,在“预测边界”这个问题上,以 Clark 为代表的一众学者批判性地发展了“半具身”的预测认知模型,将预测认知模型与身体、环境更紧密地联系起来,提出了“预测加工模型”。我们基于 Clark 在 2016 年出版的专著 *Surfing Uncertainty: Prediction, Action and the Embodied Mind* 来总结该模型特点^①:

1. 大脑:预测引擎(Predictive engine)

具身认知的预测加工模型实际上是在强调预测大脑的基础之上,更强调身体和环境的作用。因此,该模型如何看待大脑与联结主义的预测编码理论相同,不再赘述。

2. 身体:行动导向的预测(Action-oriented predictions)

Clark(2016, p. 122)认为,预测认知模型强调认知与身体和行动紧密联系在一起,被称为“积极推理”(Friston, 2009; Hohwy, 2013),但是更清晰的表述应该是“行动导向的预测加工”。因为积极推理是指在预测认知过程中,有机体有两个途径来降低预测误差:其一,改变预测,即修正先验假设匹配当前感知输入;其二,采取行动,即进行选择采样使感知输入尽可能匹配先验预测。这一观点的洞见在于,知觉和行动是预测误差最小化这枚硬币的两面。

(1) 知觉:动允性(Affordance)

Clark(2016, p. 124)认为,根据积极推理,知觉和行动遵循同一深层逻辑:预测误差最小化。如此,知觉与其说是一种“对世界的行动中立”的假设,不如说是一种“选择行动以应对世界”的认识方

^①“具身认知的预测加工模型”图可参见 Clark(2016, p. 46)中 Figure 1.2 The Basic Predictive Processing Schema.

式。知觉的唯一目的就是指导行动。这种“以行动为导向”的知觉就是“动允性”：在这个意义上，世界是被人们的欲望需要和行动能力“定制”的，心智将世界反映为一系列行动的可能性和施加干预的机会，不同动允性之间的竞争决定了主体的知觉和行动。

(2) 行动：自证预言 (Self-fulfilling prophecy)

Clark (2016, p. 124) 认为，根据积极推理，知觉无法解释消除所有的预测误差，行动占据了重要地位。对知觉而言，这意味着当自上而下的先验预测与自下而上的感知输入无法成功匹配时，主体需要采取行动，知觉产生行动；而对行动来说，这意味着，主体采取行动的过程就是一个选择性采样的过程，换言之，行动的本质就是主体以“主动寻求大脑所期望的感知后果”的方式移动身体，行动证明知觉。因此，可以将行动视为一种对知觉的“自证预言”。

3. 环境：生成世界 (Enacting a world)

在 Clark (2016, pp. 125 – 126) 看来，积极推理将行动重新定义为选择特定的感知输入以证实这些期望的采样方式，这使行动导向的预测加工具有“生成主义” (enactivist) 色彩：根据生成主义，心智积极地建构其所揭示的世界。在较长的时间尺度上，我们借助主动推理这一过程建构了“设计者环境”和“文化实践”，以安置那些决定我们如何行动（如何对该环境采样）的新预测。最终，主体根据自己的期望，通过采取行动，改造了周围的环境，人类实际上是生活在自己所建构的“生成世界”中。

(1) 设计者环境 (Designer environments)

Clark (2016, pp. 276 – 279) 认为，行动是预测主体“积极建构世界”最为简单的途径。但是，除行动外，人类还能借助智慧的操纵 (manipulation)，为预测大脑变换问题空间，巧妙地利用世界本身来卸载神经网络加工过程的复杂性。这些操纵形式多种多样，人类就建立了设计者环境，包括算盘、笔记本、智能手机等人造物，也包括口头语言和书面文字等符号，这是一个将感知加工不断“外部化”的长期过程。

(2) 文化实践 (Cultural practices)

Clark (2016, pp. 276 – 279) 认为，强大的预测加工机制可致力于接收统计学内涵更为丰富的设计者输入，发现并提炼新的生成模型，这种对环境的塑造一次又一次、一代又一代地重复进行下去，我们（而不是其他动物）关于世界真实状态的最佳预测模型逐渐成为了稳定的、可重复检验的。人类所建造的环境中产生了大量可传播的结构，如哲学、科学、艺术等文化实践，它们训练、触发并反复改变预测大脑的活动，最终，人类得以建构出一个文化世界。

综上，与联结主义的预测编码理论相比，具身认知的预测加工模型有两点进步：其一，对待身体，更加强调与具身认知精神的契合，注重身体和行动在预测认知过程中的重要性，预测大脑不是孤立的推理引擎，而是一台行动导向的对接机器。其二，对待环境，也具有生成主义特色，强调身体和行动在设计者环境和文化实践中的作用，“身体作为一种行动”积极地建构自己的生成世界。但是，该模型的革命性意义始终存在争议。因为，Clark 依然在信息加工范式中坚持“表征主义”：“预测加工理论抛开‘表征’会怎么样？……我们就只剩下涵盖大脑、身体和世界的复杂的动态循环本身了……它无法解释（无法让我们理解）一种能够支持感知、思维、想象和行动的有意义的结构领域是怎样产生出来的”（Clark, 2016, p. 293）。

三、从表征之战到生成认知，再到意义建构

(一) 表征之战

认知科学的理论演变和范式转移都是围绕着“表征之战”展开的。传统认知科学的基本原则就是表征计算主义，这实际上预设了一个基本理论承诺——“认知包含内容” (Content Involving Cogni-

tion, CIC) (Hutto & Myin, 2013), 这些内容是由大脑内部的表征所承载的, 而不论这些表征是大脑中的符号(符号主义)还是神经网络复杂的权重模式(联结主义)。换言之, 表征位于大脑内部, 是认知的意义来源。第二代认知科学则反对将意义完全归于大脑内部的表征, 主张“身体”在认知过程中具有重要意义。由于反对表征的程度不同, 第二代认知科学阵营内部分化成具身、延展、嵌入与生成认知四大取向, 即“4E 认知”。其中前三者并没有推翻传统认知科学的表征主义框架, 实质上是一种“保守的认知”; 唯有“生成认知”对表征持有全盘否定的立场, 被称为“激进的认知”。但是, 生成认知过分强调了现象学描述, 而掩盖了认知科学的理性分析, 大脑、身体和环境之间的动态循环关系一机制(mechanism), 依然处于黑箱状态。恰好, 预测认知模型反传统信息加工范式而行之, 强调: 大脑是主动的预测引擎, “大脑”借助“预测”的能力指导“身体”行动来认识外部“世界”。如此, 预测认知模型所提供的“动力学机制”似乎有望提出对生成认知“反表征”的科学解释。但是, 无论是认知、联结还是具身的预测认知模型却都站在传统认知科学“表征主义”立场上, 限制了其潜在的强大解释力。

为了应对“表征之战”, 我们认为预测认知模型已经形成了“新谱系”: ①“保守派”: 坚持“内容结构表征”; ②“温和派”: 承认“行动导向的表征”; ③“激进派”: 否认“有意义的表征”。

1. 联结主义的预测编码理论: 保守派

与“认知主义的预测模块假说”坚持“内容表征”相比, Hohwy (2013) 认为, “生成模型”在预测认知架构中扮演了“结构表征”的角色, 虽然身体和行动对认知也有影响, 但这种影响是因果性的, 而非构成性的, 只有大脑内部的过程才是认知的构成成分。这是因为 Hohwy 将自由能原理解读为: 一方面马尔科夫毯两侧的状态具有“条件独立性”, 因此, 脑内过程与外部过程是彼此区隔的; 另一方面, 这种区隔是一种“推理隔离”, 通过这种推理性的区隔, 马尔科夫毯扮演了一道“证据边界”, 让生成模型得以作为知觉经验隐藏原因的统计模型。重要的是, Hohwy 认为这种统计模型是个体水平(personal level)的, 并决定了知觉的内容。这无疑是站在“表征主义”立场上坚持“结构表征”的解释, 正如 Hohwy (2013, p. 220) 声称: “大脑中保存的世界的生成模型是自然的内在镜映, 它概括了世界的因果结构……大脑必须从颅骨内部推断其感官输入的隐藏原因, 而这一点至关重要地依赖于与生物身体的互动……但这是因为推理任务本质上是间接的”。

因此, 根据 Gładziejewski (2017) 的总结, Hohwy 其实是“保守派”: “其一, 认同表征主义; 其二, 认为知觉与行动基于某种推理; 其三, 认同内部主义, 即认知的构成元素不会延展到中枢神经系统以外”。简言之, “保守派”的理论假设就是“大脑—科学家类比”, 即大脑的运行方式就好比科学家的科研流程, 主要工作就是如实地反映外部世界。

2. 具身认知的预测加工模型: 温和派

Clark (2016) 认为表征是“行动导向的”, 行动让我们预测的感知刺激以一种有选择性的方式展现出来。这意味着: ①“选择性采样”; ②创建“设计者环境”。驱动采样和行动的预测在功能上类似于“虚拟语气”, 它们并不是由同构于外部世界的内部模型产生的单纯的期望, 而是关系到“在何种情况下(包括在做出何种行动时), 我将会经验到何种本体觉、内感觉和感知刺激模式”。在这个意义上, 单纯的内部模型无法决定知觉乃至任何认知经验的内容, 要决定认知的内容, 就要将身体过程与外部事物纳入考量, 也就是说, 根据 Clark 的观点, 身体过程(和某些外部事物)对认知的影响是构成性的, 而不只是因果性的。这说明, 在表征问题上, Clark 较 Hohwy 而言进了一步: “大脑被揭示为(主要)不是理性的引擎……而是作为环境控制行动的器官。廉价、快速、探索世界的行动, 而不是追求真理……现在是关键的组织原则”(Clark, 2016, p. 250)。但是, Clark 依然强调, 生成模型对认知对象的概率评估是“有内容”的, 它们表征了外部世界, 即便这些内容只是“随附性”的。

因此,根据 Piekarski (2021) 的总结,Clark 可以视为“温和派”：“首先,行动与知觉紧密相连;其次,同时也更需要强调的是,多层生成模型的某些层级是“直接表征”的,另一些则只是间接的:它们以一种生成性的方式与世界关联——换言之,表征‘旨在与世界对接,而非以一种行动中立的方式描绘后者’(Clark,2015,p.4)”。可见,如果接受 Clark 的解释,就等于放弃了“大脑—科学家类比”,预测认知模型的任务并不是追求真理,正确表征外部世界也并非活着的必然前提。

3. 预测认知模型与生成认知的联姻:激进派

传统的预测认知模型虽然有保守派和温和派之分,却都站在表征主义立场之上,承认“表征”是“有意义”的,这就招致了缪勒莱耶等诸多“错觉”现象的挑战:因为即使自上而下的知识(表征)告诉我们,两条线段是相等的,我们的感知似乎无法根据知识(表征)来改进知觉假设,预测误差似乎无法最小化,因此,预测认知模型的表征主义解释在错觉现象上行不通了。为此,Gallagher 和 Hutto 等人认为势必要建立一种立足于“生成认知”的“激进派”预测认知模型。Gallagher 等人(2021)提出了一个 4E 解决方案,尤其强调“非表征主义承诺”:表征是没有意义的,认知并非生成于大脑内部,而是生成于大脑、身体与环境的耦合之中。而 Hutto 和 Hipólito (2021) 认为知觉是一种“习惯”,知觉经验中并不含有“内容”,更强调“文化主义承诺”:大脑—身体—环境的动态系统受到文化的影响,认知具有习惯性和互动性的特点。

历史地来看,在对待表征之战这个棘手问题上,预测认知模型和生成认知似乎注定要产生分歧。我们认为,预测认知模型终归是一个“动力学理论”,并无固定的表征立场,但是表征主义的解读会推出一系列不符合动力学理论基本运行逻辑的结论。^① 归根结底,表征主义的解读认为:①在有机体水平定义预测误差最小化,以预设分析(ad hoc)的方式界定认知系统;②视生成模型为结构表征,暗示了认知系统的静态属性;③重视积极推理对客观环境的反映,坚持认知系统与环境彼此隔离。而动力学理论强调:①在系统立场上,借助内稳态这一概念以事后分析(post hoc)的方式界定认知系统;②视生成模型为控制系统,强调认知活动的内容在内外状态协调中涌现,体现了认知系统的动态属性;③重视积极推理维系主体与环境间动态交互的作用,倡导认知系统与环境间的耦合。因此,从认知科学动力学化的角度来看,将预测认知模型视为生成认知阵营完全具有理论合法性。

(二) 生成认知

20 世纪 90 年代,受现象学、演化生物学、动力学和东方智慧(尤其是佛教思想)的启发,认知科学家 Varela、Thompson 和 Rosch 在 1991 年出版的《具身心智:认知科学与人类经验》一书中提出了“生成认知”的主张,与第二代认知科学阵营中其他取向(具身、延展和嵌入认知)的暧昧立场不同,生成认知坚决反对“表征”,因而被寄望成为认知科学的革命性取向(De Jaegher & Di Paolo,2007;叶浩生等,2019)。而生成认知论证的方式就是,试图融合存在主义现象学和以有机体为中心的生物学,重新建立“生命观”——“生命与心智连续性命题”,来弥合生命与心智之间存在着难以逾越的“认知鸿沟”,以建立一个统一认知科学的出发点。

生成认知认为,生命既是自创生的,又是认知的(Maturana & Varela,1980)。自创生描述了生命系统自我生产的组织形式。以细胞为例,作为一个热力学开放的系统,细胞与环境进行着持续的物质与能量的交换。有些分子通过细胞膜进入并参与了细胞内部的活动,而另一些分子作为废弃物被

^①预测认知模型作为动力学理论与表征主义立场关系的详细论证可参见刘林澍、叶浩生,2022:《生成主义预测加工:一种动力系统立场》,《心理科学》。尤见论文表一:对 PP 的表征主义解读与动力学解读。

排泄掉。伴随着这一交换过程,细胞制造了大量物质,这些物质又让细胞本身的功能得以实现,使细胞与环境的边界(即细胞膜)得以维系。简言之,细胞就是一个持续不断地自我生产的过程。认知则描述了生命系统与环境有关的行为,这种认知关系反映并服从于自创生的维系。Weber 和 Varela(2002)认为,自创生系统具有“双重目的性”,其一是维持同一性(自我生产),即面对物质成分的变化蕴涵了动态同一性的产生和维持;其二是建构意义(认知):一个自创生的系统必须理解(make sense of)这个世界,才能存续下去。对生命系统而言,这种双重目的性最终将一个中性的物理化学世界转变为一个有生物学意义的环境。简言之,自创生系统就是认知系统,也就蕴含了“意义建构”。

因此,生命与心智天然连续,“生命预示着心智,而心智属于生命”(Thompson,2007,p.128)。这种生命心智强连续性要求,有机体不需要在大脑中表征一个中性世界,而是根据生存需要“生成”一个意义世界:“我们建议以‘生成的’(enactive)为名,旨在强调一个日益增长的信念:认知不是一个对既定世界的表征,它毋宁是在‘在世存在’(being-in-the-world)所施行的多样性动作之历史基础上的世界和心智的生成”(Varela,et al.,1991,p.8)。

(三) 意义建构

意义建构(sense-making)是生成认知的核心概念。以细菌与蔗糖的关系(Thompson,2007,pp.74-75,157-158)为例:蔗糖作为一种化学物质本身并无意义可言,但当环境中以蔗糖分子为食的细菌,蔗糖分子对细菌就具有了作为“食物”的重要意义。细菌要维系自身的存在,需从环境中摄取蔗糖分子以支持新陈代谢,为此,它们会主动向蔗糖分子浓度较高的水域移动。换言之,正是细菌维系生存的“努力”令其所处的环境——含有蔗糖的水域与“中立”的物理化学世界相比具有了一种“意义盈余”(surplus of significance)。细菌的这种“自治”过程就是生成认知所强调的“意义建构”。

生成认知强调,意义建构与有机体的生存繁衍一体两面。意义建构的过程就是有机体参照其“自我保存需要”,将事物区分为积极与消极,并通过行动达成目的的过程。在此过程中,有机体生成了一个视角,意义世界也得以呈现,“就像我们看到的那样,意义建构是意义世界的生成。意义世界并不仅仅是有机体与之互动、分化出自身的物理环境……意义世界是从有机体视角所看到的世界。通过意义建构,物理世界被转换为有机体的意义世界”(Colombetti,2018,p.574)。正是这个“视角”,让事物呈现出泾渭分明的意义,有机体眼中的世界,不再是非黑即白、非好即坏,而是充满了与其生存繁衍息息相关的丰富意义。

四、生成认知的预测心智模型:认知科学的新统一范式

究其根本,“生成认知”强调的并非信息加工范式下的认知,而是“动力学”背景下的认知:意义是通过身体行动,在认知主体与世界的耦合中建构出来的,而不是寓于某种内部表征中的。不过,生成认知虽然在理论解释上高度动力学化,但缺乏一个解释意义的动力学机制。因此,我们将从“意义如何动力学化”的角度,来进一步论证二者如何结合,并尝试建立起“生成认知的预测心智模型”。Froese(2011)区分了意义建构的三种视角:①第一人称视角(即主观经验);②第二人称视角(即主体间互动);③第三人称视角(即客观观察)。受此启发,我们认为“生成认知”也应从三个视角来进行认知活动的意义建构:①第三人称视角意义建构:知觉;②第一人称视角意义建构:情绪;③第二人称视角意义建构:主体间性。因此,我们尝试提出了“生成认知的预测心智模型”(如图1所示),试图说明如何运用“动力学机制”来澄清“意义建构”。

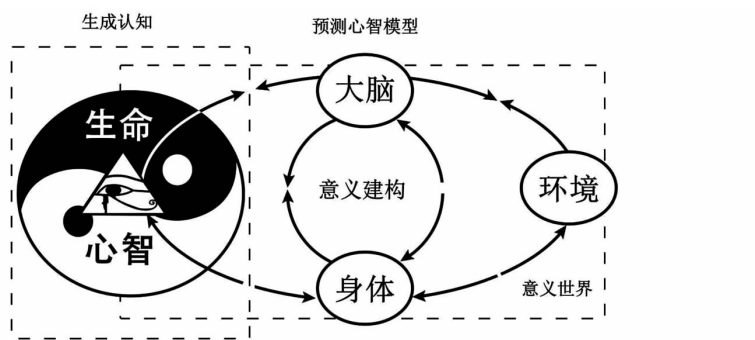


图1 “生成认知的预测心智模型”^①

（一）第三人称意义建构：知觉

意义建构是认知的。当有机体以旁观者的身份,对认知对象进行客观观察,它所从事的就是第三人称视角的意义建构活动。“知觉”就是典型的第三人称意义建构活动。

1. 生成认知:“生态知觉理论”通过“意义建构”取代“表征”

与“传统知觉理论”主张的“间接知觉观”观点相反,“生态知觉理论”(Gibson,1979/2014)作为最具有生成主义色彩的知觉理论,强调“知觉其实是一种意义建构”,“动允性”是其核心概念:对知觉主体而言,知觉对象的“意义”体现为“动允性”,即“行动的可能性”。Gibson 曾经明确指出,发明动允性这个术语就是为了取代价值这个词,因此,动允性不仅是行动的机会或可能性,更是知觉对象对知觉主体的行动价值和实践意义。但是,这些价值和意义只体现在环境和有机体的互动关系中,而这种关系是“非表征”的,即事物的价值和意义不再是对内部结构或表征进行总结、概括和推理的结果,而是直接蕴含在行动中,也是直接在知觉中获得的。总而言之,知觉是一种应对世界的过程,而不是表征世界的过程,事物是作为一种有用和有意义之物而为我们所知觉的。

2. 预测认知模型:“循环因果流程”机制可以解释“意义建构”

预测认知模型认为,知觉处于“循环因果流程”(circular causal flows)之中:在预测误差最小化这一整合机制中,知觉、认知和行动一度看似分明的界限变得模糊了:行动产生自知觉,知觉预测感知信号,随后,某些实际传入的感知信号将促成新的行动,而新的行动又会调动新的知觉……如此循环往复。这种对循环因果流程的强调,使知觉具有了浓厚的生成主义色彩。这种对循环因果逻辑的强调让预测认知模型具有了浓厚的生成主义色彩,也将知觉过程还原为非表征的第三人称意义建构。在知觉活动中,“表征”概念可为生成主义的“感知运动权变”(sensorimotor contingency)所取代(Kirchhoff & Kirvenstein,2019,pp.83-86),感知运动权变是指知觉主体在主动推理的过程中对其自身行动可能导致的知觉经验的预期,对感知运动权变的理解等同于知觉主体对“动允性”的把握。根据这种见解,知觉经验的源头既非某种“内在表征”,又非某种“行动导向的表征”,“知觉体验是通过生成过程与生成模型的‘耦合’产生的”(Kirchhoff & Kirvenstein,2019,p.85)。换言之,如果说知觉主体果真表征了什么,其表征的也是动允性——对周围环境采取行动和干预的可能性。而“行动”被置于知觉的中心,恰恰反映了知觉作为一种意义建构的实质,就是通过采取适应环境的行动,来维持知

^①实际上,生成认知与预测认知模型对中国传统文化非常友好,生成认知的经典代表人物 Varela 等人(1991,p.22)曾有言:“亚洲哲学,尤其是佛学西渐,乃是西方文化史上的第二次文艺复兴,其对西方文化的重要性堪比欧洲文艺复兴时希腊思想的再发现”。此图诞生的机缘是,我在开讲具身认知时经常使用太极图诠释笛卡尔身心二元论,我的学生苏佳佳因此备受启发。最终,我们决定以中华文化中的太极图来描绘生成认知,并与预测认知模型结合,命名为“生成认知的预测心智模型”。

觉主体的生存繁衍。这让我们得以从第三人称视角建构一系列认知对象的意义,包括只有人类才能理解的抽象概念的意义。

(二) 第一人称意义建构:情绪

意义建构不仅是认知的,还是情绪的。当有机体以当事人的身份,直接感受认知对象时,它所从事的就是第一人称视角的意义建构活动。“情绪”就是第一人称视角的意义建构活动。

1. 生成认知:“生成情绪理论”通过“意义建构”取代“表征”

与“传统情绪理论”主张情绪发生于“有机体内部”的观点相反,“生成情绪理论”(Sánchez, 2019)从“意义建构”的视角看待情绪的动力作用。首先,情绪与认知一体两面,认知过程中弥漫着情绪感受,同时情绪本身就带有“意向性”,能辅助对特定对象的认知。其次,情绪是身体的情绪,身体是有情绪的身体。身体对情绪的影响不仅是“因果性”的,更是“构成性”的。最后,也是最重要的是,有机体意义建构的过程既是认知的,也是情绪的,二者共同确立了有机体活动的“视角”,从这个视角看世界,世界就不再是中性的物理世界,而是有意义的生存世界,是为有机体所“在乎”、“关心”和“感动”的,世界的丰富意义就体现在有机体的喜、怒、哀、乐等多种多样的情绪之中。

2. 预测认知模型:“内感受推理”机制可以解释“意义建构”

预测认知模型认为,情绪的本质就是“内感受推理”(interoceptive inference)(Seth & Friston, 2016):情绪和主观感受的产生是多层推理的结果,多层加工架构的高层整合了内感觉、本体觉与外感觉线索,在下行预测与上行误差的匹配(或失匹配)过程中,生成了一个能够反映相关情境的综合体,这个综合体被个体体验为情绪。我们的情绪因此整合了身体唤醒状态的基本信息、对这些状态可能原因的高层预测以及对可能采取的行动的的准备,其不可避免地受到认知的形塑,同时认知活动也总是带有某种情绪色彩。这种解释拒绝将认知和情绪割裂开来,强调大脑、身体和环境间持续的交互作用,自然从属于生成认知阵营,同时也将情绪还原为非表征的第一人称意义建构。Barrett 和 Bar(2009)等人提出的“感受性预测假设”(affective predictive hypothesis)也主张人类并非在认知到外部事物之后再评估其价值和意义,任何认知对象对认知主体的价值和意义都是一种“前反思”的情绪感受反应,在相关感知信息作用于感受器之时,这些信息就已经进入认知加工过程了,“有意识的知觉的确融入了感受,因此特定对象在情感意义上的价值、显著性或意义并非一种事后的计算”(Barrett & Bar, 2009, p. 1328)。这种解释突出了情绪在意义建构过程中扮演的角色,情绪感受反映了认知对象对情绪主体的价值和意义,特别是对其生存与繁衍的意义,以一种最切身、最直观、最有温度的方式引导着个体的适应性行为。

(三) 第二人称意义建构:主体间性

意义建构不仅是认知的、情绪的,还是主体间的。当有机体在社交情境中与认知对象进行互动时,它所从事的就是第二人称视角的意义建构活动。“主体间性”就是一种典型的第二人称意义建构活动。

1. 生成认知:“互动理论”通过“意义建构”取代“表征”

与“传统他心理论”侧重于在“表征主义”上来理解社会认知相反,生成认知视阈下的“互动理论”(Gallagher, 2020)试图通过第二人称视角(互动),强调主体间性最终是通过互动参与式地构建出意义来的。其有三个核心概念:一是“自治”,人类作为自治系统,能够在自然环境不稳定条件下最大限度地维持自身主体性以实现生存与繁衍的意义;二是“参与式意义建构”,人类在社会认知过程中,能够通过与他人的积极交往主动地创造出超越自然生存的意义,即社会意义;三是“耦合”,意义建构

的双方通过眼神、面部表情等有意的身体行动进行互动,从而获得一种直接感知的意义。正如 Gallagher 所指出的:主体间性不是两个“笛卡尔心灵”的对话,而是两个“现象学身体”的相遇。

2. 预测认知模型:“直接感知”机制可以解释“意义建构”

预测认知模型认为,“主体间性”就是“直接感知”(direct perception)(苏佳佳,叶浩生,2021):既然知觉既是行动的原因,又是行动的后果,我们就能使用那些让自己产生相应行动的生成模型,来预测他人的行动。具体而言,互动中的一方在另一方执行身体动作的情境中会自发调低本体觉精度,同时继续基于对他人动作的感知对自身(实施同一动作时)的下一步身体行动进行预测,但无需真的做出符合预测的动作。通过在预测架构中调整“精度”的分布,我们就可以仅通过观察来预测他人的下一步行动,进而直接理解他人行动背后的意图。在这个过程中,认知概念“表征”可以被数学工具“精度”所解释,同时,也将“主体间性”还原为第二人称意义建构活动。Gallagher 和 Allen(2018)站在生成认知立场之上,提出了“预测交互”(predictive engagement),强调生物体与环境之间是动态耦合非表征的关系,行动者(agent)没有模型,而行动者本身就是模型(model),简单说,我们必须从最包容的意义上理解“模型”,模型嵌入在具身生物体和延展环境之间高度耦合的动态系统之中。社会互动就是我们与他人展开互动的意义建构活动,而且这种意义建构活动“涉及大脑内部过程与外部因素的复杂整合,后者包括一系列动作、姿势以及与他人富有表现力的其他身体姿态的交流,同样也涉及对人造物、工具和技术的使用,这一切都寓于物质性的环境之中,由社会角色和文化实践所定义”(Gallagher & Allen,2018)。

综上,与传统的预测认知模型相比,我们可以总结出生成认知的预测心智模型的优势:①从“表征之战”来看,重塑了认知科学的生命观,通过倡导生命心智天然连续,强调身心一元,明确了未来认知科学的本体论立场;②从“生成认知”来看,建构了认知科学新的概念体系,用“自治”的“意义建构”取代了传统认知科学中“他治”的“表征”,奠定了未来认知科学的认识论基调;③从“意义建构”来看,提出了认知科学的动力学机制,借助预测认知模型解释清楚了意义建构的三个视角,指明了未来认知科学的方法论发展方向。基于此三点,预测认知模型与生成认知的结合将有望引领认知科学朝向动力学的新一轮范式转移。

五、预测认知模型是可证伪的吗?

认知科学曾经作为横跨哲学、心理学、神经科学、语言学、计算机科学和人类学的“认知科学六边形”而被视为 21 世纪最有前景的大一统学科,但是,Núñez 等人(2019)指出,经过半个多世纪的发展,认知科学并没有形成具有“概念凝聚力”、“方法论兼容性”和“理论整合性”的研究范式。预测认知模型以其强大的解释力,展现了统一认知科学各个学科的潜力,但是,仍然有诸多反对意见,认为断言该模型已成为认知科学新统一范式言之尚早,因为其还是一个无法“被经验完全证明”的科学假说。

Friston 等人(2012)曾率先指出,预测认知模型的理论基础——自由能原理“不是一种对生命机制的见解或‘有魔法的配方’;只是对现存生物系统的一种描述”。Hohwy(2015)将预测认知模型类比为“达尔文演化论”,并且指出可证伪性对于真正的科学解释而言,既非充分条件也非必要条件。在达尔文本人看来,演化论并不是一个科学定理,而仅仅是一种科学假说,自然选择不是可证伪的,而是一个可以被重复发现的原理:物竞天择,适者生存。所以,波普尔(1987a, p. 177)虽然曾认为,“我已得出这样的结论:达尔文主义不是一个供可检验的科学理论,而是一个形而上学研究纲领——一个供可检验的科学理论用的可能的框架”。然而,从这一原理出发,我们可以统一许多相互竞争的

科学假设,这些科学假设终将被实证研究所证实或证伪。因此,波普尔(1987b,p.438)又修正了自己的观点:“自然选择理论可以作这样的表述,它绝非是同义反复的。在这种情况下,它非但是可以检验的,而且严格地说,它并不是普遍地真的”。不过,当无法断定一个科学模型是否可证伪时,仍然可以通过推理得到最佳解释,来展示一个理论的合理性。因此,我们认为,预测认知模型的确可以担当认知科学的“最佳解释推理”(inference to the best explanation)。

预测认知模型推动了认知科学朝向“动力学”的范式转移,这给予我们的启发就是:认知科学的根本使命就是要找到人类认知的意义来源,而“意义”的科学机制必然要在“动力学”中找到。从“范式转移”的意义上来说,“生成认知的预测心智模型”作为一种新范式具有两个优势:首先,认知科学,特别是心理学,一直有科学主义和人文主义两种文化之分,难以相融。而生成认知的预测心智模型赋予了“意义”重要的地位,用科学的话语体系将人的认知与人的生存以及人类社会特有的社会、历史与文化联系起来,弥合了科学和人文两种文化的解释鸿沟。其次,认知科学领域各学科因缺乏共同的话语体系,而不得不借助学科内部研究方法的科学性来确立科学共同体的边界,此种方式虽然维护了学科内部的团结,却也阻碍了不同学科之间的对话与合作。生成认知的预测认知模型通过强调“动力学”,将人类体验(第一人称视角)、科学观察(第三人称视角)和社会生活(第二人称视角)整合在同一个“动力学”视域中,为认知科学领域内的各学科重新建构了共同的科学话语体系,为跨学科的对话合作提供了可能性。因此,我们坚定认为,预测认知模型与生成认知的结合将有望成为真正意义上的“认知科学新统一范式”。

参考文献:

- [英]卡尔·波普尔,1987a:《无穷的探索——思想自传》,邱仁宗、段娟译,福州:福建人民出版社,第177页。
- [英]卡尔·波普尔,1987b:《科学知识进化论》,纪树立编译,北京:生活·读书·新知三联书店,第438页。
- 苏佳佳、叶浩生,2021:《“生成认知”、“互动理论”与“预测加工模型”——“他心问题”的新思考》,《心理科学》第6期。
- 叶浩生、曾红、杨文登,2019:《生成认知:理论基础与实践走向》,《心理学报》第11期。
- Barrett, L. F. & M. Bar, 2009, “See it with feeling: Affective predictions during object perception”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 364, No. 1521, pp. 1325 – 1334.
- Clark, A., 2013, “Whatever next? Predictive brains, situated agents, and the future of cognitive science”, *Behavioral and Brain Sciences*, Vol. 36, No. 3, pp. 181 – 204.
- Clark, A., 2015, “Predicting peace: Reply to Madary”, In T. Metzinger & J. M. Windt (eds.), *Open MIND: 7(R)*, Frankfurt am Main: MIND Group.
- Clark, A., 2016, *Surfing Uncertainty: Prediction, Action and the Embodied Mind*, Oxford: Oxford University Press.
- Colombetti, G., 2018, “Enacting affectivity”, In A. Newen, L. de Bruin & S. Gallagher (eds.), *The Oxford Handbook of 4E Cognition*, Oxford: Oxford University Press, pp. 571 – 588.
- Constant, A., A. Clark & K. Friston, 2021, “Representation wars: enacting an armistice through active inference”, *Frontiers in Psychology*, Vol. 11, Article 598733.
- De Jaegher, H. & E. Di Paolo, 2007, “Participatory sense-making: An enactive approach to social cognition”, *Phenomenology and the Cognitive Sciences*, Vol. 6, No. 4, pp. 485 – 507.
- Friston, K. J., 2005, “A theory of cortical responses”, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, Vol. 369, pp. 815 – 836.
- Friston, K., 2009, “The free-energy principle: A rough guide to the brain?”, *Trends in Cognitive Sciences*, Vol. 13, No. 7, pp. 293 – 301.

- Friston, K., 2010, "The free-energy principle: A unified brain theory?", *Nature Reviews Neuroscience*, Vol. 11, No. 2, pp. 127 – 138.
- Friston, K., C. Thornton & A. Clark, 2012, "Free-energy minimization and the dark-room problem", *Frontiers in Psychology*, Vol. 3, Article 130.
- Froese, T., 2011, "Breathing new life into cognitive science", *Avant*, Vol. 2, No. 1, pp. 113 – 129.
- Gallagher, S., 2020, *Action and Interaction*, Oxford: Oxford University Press.
- Gallagher, S. & M. Allen, 2018, "Active inference, enactivism and the hermeneutics of social cognition", *Synthese*, Vol. 195, pp. 2627 – 2648.
- Gallagher, S., D. Hutto & I. Hipólito, 2021, "Predictive processing and some disillusiones about illusions", *Review of Philosophy and Psychology*, <https://doi.org/10.1007/s13164-021-00588-9>.
- Gibson, J. J., 1979/2014, *The Ecological Approach to Visual Perception: Classic Edition*, New York: Psychology Press.
- Gładziejewski, P., 2017, "Just how conservative is conservative predictive processing?", *Internetowy Magazyn Filozoficzny Hybris*, No. 38, pp. 98 – 122.
- Hohwy, J., 2013, *The Predictive Mind*, Oxford: Oxford University Press.
- Hohwy, J., 2015, "The neural organ explains the mind", In T. Metzinger & J. M. Windt (eds.), *Open MIND*; 19 (T), Frankfurt am Main: MIND Group.
- Hohwy, J., 2016, "The self-evidencing brain", *Noûs*, Vol. 50, No. 2, pp. 259 – 285.
- Hohwy, J., 2020, "New directions in predictive processing", *Mind & Language*, Vol. 35, No. 2, pp. 209 – 223.
- Hutto, D. D. & E. Myin, 2013, *Radicalizing Enactivism: Basic Minds Without Content*, Cambridge, MA: The MIT Press.
- Hutto, D. D. & I. Hipólito, 2021, "Updating our theories of perceiving: From predictive processing to radical enactivism", https://www.academia.edu/52360163/Updating_Our_Theories_of_Perceiving_From_Predictive_Processing_to_Radical_Enactivism.
- Kirchhoff, M. & J. Kiverstein, 2019, *Extended Consciousness and Predictive Processing: A Third-wave View*, New York: Routledge.
- Maturana, H. & F. Varela, 1980, *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht: D. Reidel Publishing Co.
- Newen, A., L. de Bruin & S. Gallagher (eds.), 2018, *The Oxford handbook of 4E Cognition*, Oxford: Oxford University Press.
- Núñez, R., M. Allen & R. Gao, et al., 2019, "What happened to cognitive science?", *Nature Human Behaviour*, Vol. 3, No. 8, pp. 782 – 791.
- Piekarski, M., 2021, "Understanding predictive processing: A review", *Avant*, Vol. XII, No. 1, <https://doi.org/10.26913/avant.2021.01.04>.
- Sánchez, C. V., 2019, "The oscillating body: An enactive approach to the embodiment of emotions", *Revista de Filosofía Aurora*, Vol. 31, No. 54, pp. 712 – 727.
- Seth, A. K. & K. J. Friston, 2016, "Active interoceptive inference and the emotional brain", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 371, No. 1708, 20160007.
- Sloan Foundation., 1978, *Cognitive Science, 1978: Report of the State of the Art Committee to The Advisors of The Alfred P. Sloan Foundation*, New York. <http://csjarchive.cogsci.rpi.edu/misc/Suchman>, L. A., 1987, *Plans and Situated Actions: The Problem of Human-machine Communication*, Cambridge: Cambridge University Press.
- Thompson, E., 2007, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Cambridge, MA: Harvard University Press.

Varela, F., E. Thompson & E. Rosch, 1991, *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: The MIT Press.

Weber, A. & F. J. Varela, 2002, "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality", *Phenomenology and the Cognitive Sciences*, Vol. 1, No. 2, pp. 97 – 125.

(责任编辑: 蒋永华)

Predictive Cognition Model: A New Unified Paradigm of Cognitive Science?

YE Haosheng, SU Jiajia

Abstract: For more than half a century, cognitive science has remained a research field consisting of multiple disciplines, and there is no research paradigm serving as a common scientific discourse system in this field. In the 21st century, under the new zeitgeist of "complex science", the development of "dynamics" leads to a "paradigm shift" in cognitive science, in which the "predictive cognition model" (PCM) integrates the forces from psychology, physics and neuroscience, and is expected to win the title of "the unified paradigm of cognitive science". However, the PCM was born in neuroscience, and whether it can successfully be grafted into cognitive science is still under controversy. The question of concern is whether the PCM can be a new unified paradigm of cognitive science. In this article, we try to answer this question from the following three aspects. Firstly, taking the combination of the theoretical evolution of cognitive science, Cognitivism, Connectionism and Embodied cognition, and the PCM as the main line, we sort out three traditional PCMs from the theoretical perspective, which are all found to be unable to break the "information processing paradigm" because of the representation problem. Secondly, by focusing on the skeleton of "representation wars", "enactive cognition" and "sense-making", we find that the traditional PCMs all stand on the standpoint of "representationism", and it have not solved the problem of "representation wars", while the "enactive cognition" that advocates "sense-making replaces representationism" seems to be able to well fit the PCM and cope with the challenge. Finally, inspired by the emphasis on "sense-making" of enactive cognition, we further establish the PCM of enactive cognition by proposing the three perspectives of sense-making and clarifying the dynamic mechanism. Although the question of whether the PCM can be falsified remains unanswered, comparing the PCM with the "Darwinian evolution theory" reveals that there they have something in common: the PCM does indeed serve as "inference to the best explanation" of cognitive science. In view of this, we conclude that the combination of the PCM and enactive cognition will promote the full development of the "dynamics" of cognitive science, and thus can hopefully become a new unified paradigm of cognitive science.

Keywords: predictive cognition model; representation wars; enactive cognition; sense-making; new paradigm

About the authors: YE Haosheng, PhD in Psychology, is Professor and PhD Supervisor at College of Teacher Education, Zhejiang Normal University (Jinhua 321004); SU Jiajia, Corresponding Author, is PhD Candidate at College of Teacher Education, Zhejiang Normal University (Jinhua 321004).