

情感脑机接口如何理解情绪? ——基本情绪抑或情绪建构

吴雪梅

(兰州大学 哲学社会学院,兰州 730000)

摘要:情感脑机接口在理解动态情绪上面临的挑战,要求我们从认知科学哲学层面上澄清情绪的定义。目前,基本情绪论的支持者指出情绪建构论不仅由于在实验的程序操作上缺乏必要干预,而会使得情绪具有文化相对性这一结论失效,而且有证据表明情绪知识并非通过文化传承获得的。但根据对情绪建构论相关文本的分析与认知科学哲学的相关进展,上述质疑无法挑战其在相关论证上的有效性。因为情绪建构论之所以不干预操作程序,正是基于情绪与认知的双向作用保证结论的准确性。而考虑到情绪建构论是在关系实在论的基础上讨论情绪,因此情绪知识通过文化传承的结论是成立的。基于情绪建构论对情绪实例与文化传承的认识,有助于情感脑机接口理解动态情绪,同时提升其对于特定文化的情绪理解能力。

关键词:情感脑机接口;基本情绪;情绪建构

中图分类号:N02

文献标识码:A

文章编号:1674-7062(2025)06-0031-08

一、引言

情感脑机接口(affective brain-computer interfaces, aBCIs)^①是脑机接口技术应用于情感计算(affective computing)领域的新兴技术,致力于通过神经学测量的方式赋予机器情绪智能^{[1]19},以实现自动识别和调节人类的情感状态^[2]。基于此,情感脑机接口被认为是使得人机交互更加自然的关键技术。目前,情感脑机接口最常用的生理信号是脑电图信号^{[1]19},其关键优势在于能够通过非侵入的脑机接口,提供与情绪状态相关的实时神经活动^[3],但在涉及识别并理解动态内容的情绪反应时则面临较大挑战^[4]。通过对人工智能学界在情感脑机接口领域相关文献的分析,可以发现,之所以在识别动

态情绪上面临困境的主要原因在于,学界在如何理解情绪的定义方面还存在较大争议^[5]。有研究明确基于基本情绪论(Basic emotion theory, BET)定义情绪^{[1]1},也有研究提出情绪建构论(the theory of constructed emotion, TCE)可能会有助于情感脑机接口面临的难题得到更好的解决^[6]。

丹尼尔·丹尼特(Daniel Dennett)说过哲学家能够做的贡献之一在于强化并澄清可供选择的理论,让一些靠不住的观点变得足够清晰,从而能够更好地反驳^[7]。考虑到在情绪的定义上,人工智能学界在究竟应该选择基本情绪论还是情绪建构论对情绪的理解上仍然存在激烈争论。因此,从认知科学哲学层面上澄清情绪的定义,或许能够为情感脑机接口在理解动态内容的情绪方面实现突破提供理

【收稿日期】 2024-11-19

【基金项目】 兰州大学中央高校基本科研业务专项资金资助“基于预测心智的情绪建构论研究:一种AI增强的视角”(2025lzujbkyqk002)

【作者简介】 吴雪梅(1996-),女,甘肃武威人,兰州大学哲学社会学院讲师,研究方向为认知科学哲学、技术哲学与法国哲学。

① 情感脑机接口(Affective brain-computer interfaces)有时也被称之为情绪脑机接口(Emotional brain-computer interfaces),考虑到学界一般使用“情感脑机接口”,因此本文也沿用这一表述。除此之外,需要特别说明的是,尽管哲学界有时会对“情感”(affect)与“情绪”(emotion)进行区分,但基于认知科学哲学界与人工智能学界往往将二者作为同义词交替使用,且更多使用“情绪”这一表述,因此,本文在讨论相关情绪理论时,使用“情绪”这一表述。

论指导。尽管近年来情绪建构论得到了越来越多的支持,但也受到了来自基本情绪论的强烈批评。其中最值得重视的是基本情绪论的支持者迪萨·索特(Disa Sauter)等人与安东尼奥·达马西奥(Antonio Damasio)所提出的质疑。其中,索特等人通过实验指出情绪发声并非像情绪建构论所主张的那样具有文化相对性^{[8]354},而达马西奥的文化心智(cultural mind)理论则基于演化生物学的方法指出情绪并非通过文化传承建构出来的^{[9]62}。如果情绪建构论无法回应上述质疑,那么不仅会影响基于建构主义的科学研究,还会导致情感脑机接口的相关问题陷入困境。因此,本文将首先对基本情绪论与情绪建构论之间的争论进行分析,在此基础上对索特等人与达马西奥所提出的挑战进行分析。接着,通过相关论证对情绪建构论进行辩护,最终表明情绪建构论对情绪的定义能够为情感脑机接口的研发提供重要启发。

二、基本情绪论与情绪建构论之争

最近有研究指出,基本情绪论与情绪建构论之间存在的持续争论已经阻碍了情绪领域的研究^[10],这进一步对情绪定义的澄清提出了紧迫的要求,需要我们首先对基本情绪论与情绪建构论之争这一背景进行研究。一般认为,现代版本的基本情绪论在保罗·艾克尔曼(Paul Ekman)的研究中被广泛关注,艾克尔曼通过跨文化的实验研究表明情绪的面部表情是普遍存在的^[10]。在此基础上,基本情绪论的支持者对情绪的研究开始集中在基本情绪上^[11-14],主张基本情绪是跨物种共有的。正是由于该理论假设存在一种普遍的情绪类别分类,主张每种类型在不同文化下的生物学与心理学基础具有普遍性,因此学界将其称之为类型学(typology)^{[15]4}观点。这种类型学观点内部在理解基本情绪与非基本情绪的关系上产生了分歧,大致可以分为统一论题与不统一论题两种立场。其中不统一论题的主张者保罗·格里菲斯(Paul Griffiths)^[16]提出,基本情绪是人与非人类动物中的某些结构的同源物,而非基本情绪是人类所独有的,正因为基本情绪与非基本情绪是两种不同的自然类型,因此需要以两种不同的理论进行研究。而统一论题的代表性理论是杰西·普林茨(Jesse Prinz)提出的具身评价(Embodied Appraisals)理论^{[17]52},普林茨认为基本情绪是对自然选择的身

体变化的感知,而非基本情绪只是基本情绪的“混合”(blends)^{[17]97}或者对基本情绪进行“重新校准”(recalibrate)^{[17]99}后的状态。普林茨的具身评价主张尽管仍然坚持基本情绪论,但是这种统一论题承认很多非基本情绪在知觉上是模糊的。

丽莎·巴瑞特(Lisa Barrett)提出的情绪建构论在这一点上比普林茨走得更远。情绪建构论认为这种知觉的模糊性不仅存在于非基本情绪,也会渗透至基本情绪。特别是在情绪的自然类别上,巴瑞特就对基本情绪论的类型学观点进行了批判^{[15]4,[18]28}。她指出这种类型学观点的核心主张在于“情绪是具有严格边界的类别,可以在自然界中(意味着在大脑或身体中)观察到,因此可以被识别,不是由人类思想构建的”^{[18]32}。但目前的科学研究却在某种情绪何时存在或何时不存在这一问题上尚未达成共识,换言之,情绪并非类型学观点所主张的那样是可以被准确识别的,这种与类型学观点相矛盾的现象被巴瑞特称之为“情绪悖论”(emotion paradox)^{[19]20}。在巴瑞特所给出的批判基本情绪论的类型学观点的论据中,最具说服力的是劳伦斯·巴萨洛(Lawrence Barsalou)的团队与保拉·尼丹塔尔(Paula Niedenthal)的团队对关于情绪的概念性知识的讨论^{[20],[21]187}。这两个实证研究表明情绪并非像类型学观点主张的那样对应某种固定的生物学与心理学基础,进而提出情境概念化(situated conceptualization)^①对理解情绪的重要意义,指出对情绪的情境概念化是为了帮助主体在特定背景下处理该情绪而量身定制的^{[21]206}。而对于儿童的概念化知识的研究进一步挑战了基本情绪论的类型学观点,有研究指出儿童直到5岁才具备成人情绪概念的完整分类,在2岁时只具备使用“积极”和“消极”两种情绪类别的能力,因此这些概念对于儿童和成人来说并不具有相同的含义^[22]。这意味着情绪感知所必需的情绪概念很可能源自实际发生的情绪反应的统计结构,而并非能够通过面部表情等信号准确识别愤怒等情绪^{[18]39}。

基于此,巴瑞特提出情绪的“概念行为理论”(conceptual act theory)^{[19]21,[23-24]}有望解决“情绪悖论”,这种解决方案主张情绪并非基本情绪论所认为的那样是一种因果实体,对情绪类别的给定实例也不需要面部、身体和行为的特定配置。相反,人们的情绪是在概念化情绪实例时体验到的,在有关情

① “概念情境化”是指一些列特定情境的推论^{[20]70-71}。

绪的具身知识的引导之下对核心情感(core affect)^①进行分类^{[19]41}。由此可见,巴瑞特通过建构主义的方法对基本情绪论的情绪定义发起挑战,主张不存在自然的情绪,将情绪定义为一种基于具身概念知识所建构的对核心情感的分类行为,以此消解“情绪悖论”。随着巴瑞特在此基础上对情绪研究的不断完善,在2017年明确提出“情绪建构论”这一研究情绪的科学方法^[25-26]。这一方法尽管在情境概念化等方法上有所延续,但在理解情绪概念上却发生了重要的转变与更新。其中最大的区别在于在最新的版本中,情绪实例不是由情感分类产生的,而是像所有的心理事件一样,因为它不断地创造经验的特征,以服务于有效的能量调节和行动计划,所以被假设出现在大脑中^{[15]9}。经由对情绪建构论的相关文献的分析,我们可以将其对情绪定义的基本论点表述为以下三点:

(1)情绪实例是大脑依据过去的经验,对当前情境的情绪进行预测的结果^{[15]22}。

(2)根据种群思维(Population Thinking)^{[15]28}与预测心智,情绪实例在物理特征与功能特征上的变化是常态^{[15]33}。

(3)大脑预测所需的情绪知识是通过文化传承获得的,而非基因遗传^{[15]34}。

其中(1)中预测心智(predictive mind)论起到的作用是基础性的。因为预测心智论主张大脑是一个自上而下分层排列的预测机器,这意味着是大脑预测的行动产生了感知与经验,而非大脑探测到世界上的事件得到的感知导致了行动。基于预测心智论对行动与体验之间关系的认识,巴瑞特才改变了“概念行为理论”对情绪实例的认识,认识到情绪实例并非一种分类行为,而是一种大脑预测的心理事件。由此可见,在巴瑞特那里,事实上情绪实例作为一种情境化下的实时情绪状态,是比基本情绪所关注的“情绪类别”(如高兴)更基础的概念。正因为巴瑞特在预测心智理论的框架下对“情绪实例”的理解,“情绪类别”的定义也发生了相应变化,成为不同情绪实例因功能特征相似聚集的集合,没有特定的面部表情、生理反应或神经特征。如果说(1)主要是对基本情绪论在情绪是有固定的生物学基础这一主张的批判与颠覆,那么(2)就是对基本情绪论在

情绪具有跨物种、跨文化普遍性的这一主张的批判与颠覆。因为(2)强调一个人或动物不仅会在相同的情绪类别时表现出不同的面部表情,而且相同的身体信号与不同的心理特征关联时,也会产生不同的情绪。(3)则是在(1)(2)的基础上形成的情绪建构论在认识论上的总体性主张。具体而言,(1)和(2)对情绪实例的讨论主要是从个体的角度出发的,而(3)则扩展至社会认知层面,相应地,预测模型也扩展成为预测的社会心智观^②。具体而言,情绪建构论主张人与人之间可预测程度的增加,这种文化经过一代代传递,会减少每个人社会生活的负担^{[15]34-35}。这从认知层面进一步对基本情绪论所主张的情绪是通过基因遗传实现延续提出了挑战。

三、基本情绪论对情绪建构论的质疑

经研究发现,在基本情绪论对情绪建构论的质疑中,最值得关注的是索特的团队与达马西奥对情绪建构论所提出的挑战。具体而言:

第一个质疑主要来自索特的团队对情绪发声的研究。这项研究对情绪建构论的论点(2)中所强调的情绪实例在特征与功能上的变化发起了挑战。因为情绪建构论反对基本情绪论所主张的情绪具有跨文化、跨物种的普遍性,而是更强调情绪实例在跨文化上所表现出的相对性^[29]。基于此,巴瑞特等人才进一步主张情绪实例在物理特征与功能特征所发生的变化才是常态。但索特等人通过对情绪发声的实验研究表明,情绪建构论者之所以会通过实验得出情绪发声具有文化相对性这一结论,很可能是因为他们实验程序上,没有在确保辛巴族被试正确理解了情绪故事的前提下推进实验,需要在开始前询问被试的感受以确保他们能够用自己的语言解释相应的情绪。而在确保这一前提的基础上所得到的数据证明无论干扰因素的效价如何,非语言发声都传达了特定的情绪状态,这表明情绪发声是具有跨文化普遍性的^{[8]355}。尽管以巴瑞特为代表的情绪建构论者对此进行了回应^{[30]358},但目前对基本情绪论的这一主张进行辩护的实证研究还在不断增加^[31],可见索特等人的这一质疑的影响力之大。因此,为了研究情绪究竟是否具有跨文化普遍性,需要我们对基本情绪论与情绪建构论的相关研究进行深入分

① 巴瑞特所说的核心情感是指有机体神经生理状态中持续不断的短暂变化,代表其与不断变化的事件流的直接关系,是情绪生活的基本组成部分^{[19]30}。

② 关于预测的社会心智观的具体讨论详见文献内容^[27-28]。

析,在此基础上回应这一问题。

第二个质疑主要来自达马西奥所提出的基于基本情绪的文化心智理论。这项研究对情绪建构论的论点(3)中所强调的情绪知识并非通过基因遗传获得的,而是通过文化传承的这一观点发起了挑战。事实上如果达马西奥的文化心智理论成立,这也同时也意味着情绪建构论中的基本论点(1)中对情绪实例的定义受到了挑战。具体而言,巴瑞特明确将达马西奥归类为基本情绪论者,指出达马西奥的“躯体标记假设”主张不同类型的情绪有其特定的神经本质,对应于其特定的躯体标记^{[19]23,40}。普林茨在为其具身评价论进行辩护时也多次提及达马西奥,他明确指出“我所捍卫的理论并不是全新的,这是一种由威廉·詹姆斯(William James)和卡尔·兰格(Karl Lange)首创的变体,最近又被安东尼奥·达马西奥重新发扬光大。根据这一传统,情绪是对身体模式变化的感知。更正式地说,情绪是本能反应”^{[17]xiii}。而达马西奥也明确说过“情绪的本质是身体状态的集合”^{[32]156}。由此可见,达马西奥的情绪理论的确如普林茨所说的那样,主张情绪是可以被还原为生物性的本能反应的,从这个意义上而言,达马西奥无疑是基本情绪论者。

除此之外,普林茨在为具身评价论处理基本情绪与非基本情绪间关系的统一论题辩护时,也明确将达马西奥在情绪的神经生物学方面的相关研究作为重要论据。他指出达马西奥在对基本情绪的研究中也发现了负责高级认知情绪或非基本情绪的脑岛处于激活的状态,这证明非基本情绪与基本情绪实际上是统一的,其中非基本情绪只是基本情绪的认知阐述^{[17]96}。而达马西奥虽然没有明确提出具身评价的统一论题,但他在论证躯体标记假设时的确主张次级情绪(secondary emotions)也会以与初级情绪(primary emotions)^①相同的神经机制进行表达^{[32]156}。这意味着达马西奥也像普林茨那样,坚持基本情绪与非基本情绪的划分,而且支持二者之间的神经机制并非完全独立的。正是基于这种生物还原论的情绪主张,达马西奥进一步发展出文化心智理论。尽管达马西奥并未明确提及对巴瑞特的情绪理论的质疑,但他的文化心智理论在核心主张上也像基本情绪论那样,对巴瑞特的情绪建构论的基本论点(3)与(1)构成了较大挑战。达马西奥明确指出我们之所以从约40亿年前出现的细菌演化出人

类文化心智,其关键在于自然选择与遗传机制^{[9]62},但仅仅认识到这一点还不够,更重要的是,内稳态和感受是人类文化心智的生物根源^{[9]144,147}。对此,达马西奥提出的核心论证是从演化生物学的角度出发的,他指出地球上的生命是以细菌到神经系统的顺序演化而来的,正是由于神经系统的出现使得内稳态的神经调节得以可能。在此基础上,人类演化出了具有感受与创造性智力的有意识心智,这种心智的形成成为人类文化的诞生开辟了道路^{[9]56}。

正因如此,国内有研究者也指出达马西奥坚持从演化的角度理解人类心智与文化现象,其中感受处于其理论的中心地位^[33]。这意味着基于达马西奥的躯体标记假设,情绪知识并非像巴瑞特所主张的那样是通过文化传承获得,相反,自然选择、遗传机制,特别是内稳态与感受等生物基础对人类文化的建构起到关键作用。而如果达马西奥的文化心智理论成立,那将导致巴瑞特对文化传承的主张受到严重挑战。因此,我们有必要重新考察达马西奥的文化心智理论与巴瑞特的情绪建构论,进而对此质疑进行回应。

四、对情绪建构论的辩护

关于索特等人对情绪建构论的挑战。根据对巴瑞特情绪建构论的分析,可以发现,尽管目前有较多研究通过心理学的实证分析对这一质疑提出了反驳^{[30]358,[34-35]},但很少有研究从认知科学哲学的视角为巴瑞特对情绪的定义进行辩护。因此,本文将试图从心身关系的哲学基础出发,证明索特等人的质疑事实上预设了基本情绪论对情绪与认知的关系的理解,并未考虑到巴瑞特的情绪建构论所依托的心身关系立场是预测心智论,而根据预测心智基于大脑的实证研究对情绪与认知的关系的新解,索特等人的质疑不成立。具体而言,基于索特等人主张辛巴族只有先对情绪故事有正确的认知,才能够保证其情绪发声在表达上的准确性,可以看出,索特等人的主张实际上预设了认知过程与情绪表达过程不仅是分工明确的,而且认知会影响或者说控制情绪表达这一前提。但问题的关键在于,越来越多的实证研究表明基本情绪论所预设的认知大脑与情绪大脑的二分法过于简单,难以解释人脑功能的复杂性、交互性与自组织性^{[36]4}。

根据巴瑞特的情绪建构论,情绪与认知在神经

① 在达马西奥的情绪理论中,次级情绪与初级情绪对应于普林茨所说的非基本情绪与基本情绪。

机制层面上是无法做出明确区分的,而是存在复杂的双向互动作用^[37]。这一点得到了来自大量神经科学领域相关证据的支持,从对杏仁核^[38]、下丘脑^[39]等相关神经机制的研究都证明大脑事实上是基于认知大脑与情绪大脑相互作用的因果机制,使得大脑能够对感官信息进行自上而下的预测,从而产生统一的感觉^{[36]4-5}。基于此,我们也能够从认知哲学的层面解释为什么巴瑞特在回应中会坚持在程序上不对辛巴族对情绪故事的认知进行干预,指出如像索特等人那样干预则会违背实验目的^{[30]358}。这主要是因为预测心智论坚持大脑的低级情绪回路与高级认知回路互为因果,而非认知过程单方面影响情绪表达,因此只有在程序上不对辛巴族对情绪故事的认知进行干预,才能够保证辛巴族是基于原有文化,在认知与情绪相互作用的前提下识别与表达情绪。而山姆·威尔金森(Sam Wilkinson)等人基于预测心智论对情绪本质的研究也表明承认情绪的具身性并不意味着承认情绪具有普遍性,而是经由相关实证研究发现“具身预测心智必然是一种情绪心智”^{[40]116}。由此可见,索特等人对巴瑞特的情绪建构论提出的质疑不成立。

关于达马西奥的文化心智理论对情绪建构论中基本论点(3)和(1)的挑战。通过对巴瑞特与达马西奥的相关文本的分析,本文将证明达马西奥的文化心智理论与情绪建构论对文化的态度并非是完全对立的,二者同源心理建构主义,而之所以会对文化得出相反的结论主要是因为对“情绪”的概念界定存在较大差异。

在对文化的态度上,达马西奥在论证躯体标记假设时就明确表示尽管初级情绪似乎有很大的生物成分,但是次级情绪是相对于特定的文化进行定义的。最关键的是,达马西奥进一步指出关于文化如何影响情绪分类的证据主要参考的文献是詹姆斯·罗素(James Russell)的《文化与情绪的分类》^{[32]286}。通过考察这一文献,我们发现,罗素在这篇论文中的核心观点在于情绪这一概念并非出现在所有语言中,人们对情绪的分类是相对于特定的文化与语言的^[41]。而这一点也是罗素在此后提出情绪的心理建构主义的核心论据^[42]。特别是考虑到达马西奥主张“躯体标记是由次级情绪产生的感受的特殊实例”^{[32]174},可以看出,达马西奥所主张的躯体标记假设所主要探讨的情绪正是次级情绪或者说社会情绪,而次级情绪事实上是基于情绪的心理建构主义所提出的。除此之外,普林茨虽然也提出达马西奥

在情绪上可以看作是生物还原论者,但他也表明,更准确地说,达马西奥主张的是“有限还原论”(constrained reductionism),即只有小部分基本情绪是天生的^{[17]105},而其他高级认知情绪,即达马西奥所说的次级情绪则是由文化驱动的^{[17]134}。而巴瑞特也明确提出其概念行为理论的重要理论基础之一就在于罗素的心理建构主义,而且概念行为理论中所采取的人类学方法的重要理论依据也包括罗素的《文化与情绪的分类》一文^{[15]6,64},基于此,巴瑞特主张情绪知识是通过文化传承获得的。由此可见,在处理文化对情绪的影响这一问题上,达马西奥与巴瑞特都认同心理建构主义的主张。正因如此,在处理情绪与认知的关系问题上,达马西奥才会像普林茨所说的那样认同基本情绪与非基本情绪在神经机制上存在相互作用。但根据马克·米勒(Mark Miller)等人对预测心智论的研究^{[36]14},达马西奥对大脑旧皮层与新皮层的协同活动的论证并不意味着普林茨在具身评价论上的统一论题成立。因为有研究基于类似的神经生物学结论得到的并非是基于具身心智论的具身评价论,而是巴瑞特基于预测心智论的情绪建构论^{[40]104}。由此可见,普林茨将达马西奥的情绪理论简单定义为一种有限还原的生物还原论是不符合事实的,忽视了其中重要的心理建构主义基础。

也许有人会质疑,既然达马西奥的文化心智理论与情绪建构论在对文化的态度上同源于心理建构主义,那么该如何解释达马西奥主张文化心智源自感受,而巴瑞特却主张情绪源自文化传承呢?根据达马西奥与巴瑞特的相关文本,有证据表明这主要是由于二者对“情绪”定义差异造成的。达马西奥明确表明他所说的“情绪”(emotion)与“感受”(feeling)的定义并非正统的,二者不能交替使用^{[32]146}。具体而言,在达马西奥那里,“情绪”是具有公共性的,换言之,其身体状态的变化是可以被外部观察者所察觉到的,而“感受”则具有私人性质,表示自我对自身身体状态变化的体验^{[32]139}。而且最重要的是,“所有的情绪都会产生感受,但并非所有的感受都源于情绪”^{[32]143},因此“感受”不仅包括源自情绪的基本普遍情绪的感受(Feelings of Basic Universal Emotion)与微妙的普遍情绪的感受(Feelings of Subtle Universal Emotions),还包括与情绪无关的背景感受(Background Feelings)。其中基于微妙的普遍情绪的感受即是基于次级情绪(如悔恨、尴尬等)的感受^{[32]150}。由此可见,在达马西奥那里,“情绪”与“感受”是不同的现象,正是由于“情绪”具有公共性,所

以躯体标记假设关注的是涉及复杂的社会认知的次级情绪。而鉴于“感受”具有私人性质,因而“感受”与自我的心智相关。而根据“感受”与“情绪”在范围上的关系,可以看出,在达马西奥那里,对自我心智而言,“感受”是比“情绪”更基础的存在。至于“情感”(affect),达马西奥指出这是一个非常宽泛的概念,不仅表示所有可能的感受,还包括负责产生感受的情境与机制^{[9]87}。正是基于以上定义,达马西奥的文化心智理论才主张感受是文化心智得以产生的生物根源。但问题的关键在于,尽管达马西奥主张“情绪”与“感受”负责不同的方面,特别是次级情绪是由社会文化所驱动的,然而达马西奥基于演化生物学视角理解文化心智的诞生时,却忽视了次级情绪本身是相对于特定文化的,因为这意味着微妙的普遍情绪的感受也是由特定文化所驱动的。由此可见,事实上基于达马西奥对“情绪”与“感受”的定义,无法得到感受是文化心智的生物根源这一结论。

相比于达马西奥,巴瑞特的情绪建构论对“情绪”的定义则能够合理推出情绪知识是由文化传承获得的,而非基因遗传这一结论。具体而言,巴瑞特的情绪建构论主张“情绪”的定义由基本论点(1)(2)(3)所构成的。这三条基本论点分别侧重于“情绪”的不同特征:其中(1)表明基于预测心智理论,“情绪”并非像基本情绪论所主张的那样是对世界的反应,而是对世界的预测;(2)表明“情绪”没有普遍的固定生物基础,相反情绪实例在物理与功能特征上的变化才是常态;在(1)(2)成立的基础上,巴瑞特在(3)中对人类情绪究竟是如何跨代交流并进化的这一问题的回答,就不会是在生物学意义上讨论。这是因为(1)决定了“情绪实例的物理特征没有内在的、生物遗传的情绪意义,而是一种在当下生物建构的情绪意义。实现这一实例的神经组合也没有内在的、生物遗传的情绪意义”^{[15]22}。(2)则强调情绪实例在物理与功能特征上之所以有变化,是因为如微笑、皱眉等物理信号不具备固有的生物学意义上的情绪意义^{[15]33}。因此,巴瑞特在(3)中回答人类情绪究竟是如何跨代交流并进化时,所强调的并非基本情绪论所说的生物学意义上的情绪,而是从认识论的视角讨论“情绪知识”(emotion knowledge)^{[15]34},在此前提之下,才能够得出是文化传承的过程使得情绪知识得以跨代交流与进化这一结论。由此可见,达马西奥的文化心智理论对情绪建构论的基本论点(3)和(1)的挑战不成立。

五、结语

根据对基本情绪论对情绪建构论的质疑的分析,我们发现基本情绪论的代表索特等人与达马西奥对情绪的理解不足以对情绪建构论构成挑战。考虑到情感脑机接口目前所遇到的最大技术瓶颈之一就在于处理动态内容的情绪反应,而情绪建构论对情绪的理解不仅意识到了情绪实例是动态的,而且进一步认识到情绪知识是由文化传承获得的,因此,我们有理由认为情绪建构论能够对情感脑机接口对情绪的理解提供重要启发。具体而言,在情绪建构论中,适用于情感脑机接口的情绪定义主要涉及以下两大主张:第一,情绪建构论意识到在时间维度上理解情绪的必要性,将情绪理解为动态的心理事件。第二,情绪建构论意识到基于关系实在论理解情绪的必要性,将情绪实例的实在性理解为是只存在于由人脑、人体与世界产生的信号之间的关系之中^{[15]47}。关于这一点,格奥尔格·诺瑟夫(Georg Northoff)基于认知科学哲学的相关论证也表明我们在研究意识问题上,需要坚持一种关系实在论的本体论,而非传统的以属性为本的形而上学^[43]。

基于以上情绪定义,情绪建构论对情感脑机接口的启发主要包括以下两个方面:第一,情感脑机接口的研究对象应当是与情境相关的情绪实例,而非情绪类别。具体而言,基于情绪实例是与情境相关的实时心理事件,因此,情感脑机接口要实现对动态内容的情绪反应的处理,关注重点不应当是基于基本情绪论对情绪类别进行精确分类,而是在一定时间跨度内对某个在特定情境中的情绪实例进行研究,以此才能保证情感脑机接口理解的情绪反应的结论符合情绪具有动态性这一事实。第二,情感脑机接口需要增加经由文化传承得到的情绪知识的数据输入,专注于基于贝叶斯概率论进一步提升其预测能力。具体而言,根据情绪建构论,人类的情绪知识并非通过基因遗传获得的,而是文化传承得到的,因此,情感脑机接口要想获得情绪知识,应该基于多模态大模型输入特定文化下的情绪知识数据。除此之外,根据情绪建构论,人类之所以能够识别并理解情绪,是因为大脑可以看作是一台基于贝叶斯概率论的预测机器,因此情感脑机接口应该致力于基于贝叶斯概率论提升其预测能力,从而提升其对于特定文化的情绪理解能力。

当然,不可否认的是,上述理论建议无法保证情感脑机接口所面临的所有技术困境得到解决,但能

够为其在识别动态内容的情绪反应方面的瓶颈提供一种新思路,这一方案的可行性还有待于哲学与人工智能学界通过跨学科合作进行检验。

【参 考 文 献】

- [1] PEI G, LI H, LU Y, et al. Affective Computing: Recent Advances, Challenges, and Future Trends[EB/OL]. (2024-01-05) [2024-09-20]. <http://doi.org/10.34133/icomputing.0076>.
- [2] STEINERT S, FRIEDRICH O. Wired emotions: Ethical issues of affective brain-computer interfaces[J]. Science and engineering ethics, 2020, 26(1): 351.
- [3] ERAT K, ŞAHİN E B, DOĞAN F, et al. Emotion recognition with EEG-based brain-computer interfaces: a systematic literature review[J]. Multimedia tools and applications, 2024, 83(33): 79650.
- [4] MORENO-ALCAYDE Y, TRAVER V J, LEIVA L A. Sneaky emotions: impact of data partitions in affective computing experiments with brain-computer interfacing[J]. Biomedical engineering letters, 2024, 14(1): 104.
- [5] WU D, LU B L, HU B, et al. Affective brain-computer interfaces (abcis): a tutorial[J]. Proceedings of the IEEE, 2023, 111(10): 1316.
- [6] MÜHL C, ALLISON B, NIJHOLT A, et al. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges[J]. Brain-computer interfaces, 2014, 1(2): 66-84.
- [7] DENNETT D C. The part of cognitive science that is philosophy[J]. Topics in cognitive science, 2009, 1(2): 231-232.
- [8] SAUTER D A, EISNER F, EKMAN P, et al. Emotional vocalizations are recognized across cultures regardless of the valence of distractors[J]. Psychological science, 2015, 26(3).
- [9] 达马西奥. 万物的古怪秩序[M]. 李恒威, 译. 杭州: 浙江教育出版社, 2020.
- [10] VAN HEIJST K, KRET M E, PLOEGER A. Basic emotions or constructed emotions: insights from taking an evolutionary perspective[J]. Perspectives on psychological science, 2023, 20(3): 337-391.
- [11] EKMAN P. Universals and cultural differences in facial expressions of emotion[C]// COLE J. Nebraska symposium on motivation. Lincoln: University of Nebraska Press, 1971: 207.
- [12] EKMAN P, FRIESEN W V. Measuring facial movement[J]. Environmental psychology and nonverbal behavior, 1976: 74.
- [13] EKMAN P, FRIESEN W V, ELLSWORTH P. Emotion in the human face: guidelines for research and an integration of findings[M]. Elsevier, 2013: 3.
- [14] PRINZ J. Which emotions are basic? [C]// EVANS D, CRUSE P. Emotion, evolution, and rationality. New York: Oxford University Press, 2004: 69.
- [15] BARRETT L F, LIDA T. Constructionist approaches to emotion in psychology and related fields[EB/OL]. (2024-07-01) [2024-10-01]. <https://affective-science.org/pubs/2023/barrett-lida-routledge.pdf>.
- [16] GRIFFITHS P E. What emotions really are: the problem of psychological categories[M]. Chicago: University of Chicago Press, 2008: 9.
- [17] PRINZ J. Gut reactions: a perceptual theory of emotion[M]. New York: Oxford University Press, 2004.
- [18] BARRETT L F. Are emotions natural kinds? [J]. Perspectives on psychological science, 2006, 1(1).
- [19] BARRETT L F. Solving the emotion paradox: categorization and the experience of emotion[J]. Personality and social psychology review, 2006, 10(1).
- [20] BARSALOU L W, NIEDENTHAL P M, BARBEY A, et al. Social embodiment[C]// ROSS B. The psychology of learning and motivation. San Diego, CA: Academic Press, 2003: 81.
- [21] NIEDENTHAL P M, BARSALOU L W, WINKIELMAN P, et al. Embodiment in attitudes, social perception, and emotion[J]. Personality and social psychology review, 2005, 9.
- [22] WIDEN S C, RUSSELL J A. A closer look at preschoolers' freely produced labels for facial expressions[J]. Developmental psychology, 2003, 39(1): 114.
- [23] BARRETT L F, MESQUITA B, OCHSNER K N, et al. The experience of emotion[J]. Annu. Rev. Psychol., 2007, 58(1): 373-403.
- [24] BARRETT L F. Emotions are real[J]. Emotion, 2012, 12(3): 419.
- [25] BARRETT L F. How emotions are made: the secret life of the brain[M]. New York: Pan Macmillan, 2017: 30.
- [26] BARRETT L F. The theory of constructed emotion: an active inference account of interoception and categorization[J]. Social cognitive and affective neuroscience, 2017, 12(1): 1.
- [27] 何静. 社会心智: 描述性预测与规范性预期[J]. 哲学研究, 2021(8): 119.
- [28] 何静. 预测心智进阶中的社会认知观[J]. 自然辩证法通讯, 2021, 43(1): 28.
- [29] GENDRON M, ROBERSON D, VAN DER VYVER J M, et al. Cultural relativity in perceiving emotion from vocalizations[J]. Psychological science, 2014, 25(4): 911.
- [30] GENDRON M, ROBERSON D, BARRETT L F. Cultural

- variation in emotion perception is real: a response to Sauter, Eisner, Ekman, and Scott [J]. *Psychological science*, 2015, 26(3).
- [31] KELTNER D, SAUTER D, TRACY J, et al. Emotional expression: advances in basic emotion theory[J]. *Journal of nonverbal behavior*, 2019, 43: 133 – 160.
- [32] DAMASIO A. *Descartes' error: emotion, reason, and the human brain*[M]. New York: Avon Books, 1994.
- [33] 李恒威. 从心智到文化: 达马西奥的生命哲学[J]. *西北师大学报(社会科学版)*, 2020, 57(5): 61.
- [34] BERENT I, BARRETT L F, PLATT M. Essentialist biases in reasoning about emotions [EB/OL]. (2020 – 09 – 23) [2024 – 10 – 06]. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7538619/pdf/fpsyg-11-562666.pdf>.
- [35] BARRETT L F. Context reconsidered: complex signal ensembles, relational meaning, and population thinking in psychological science [J]. *American psychologist*, 2022, 77(8): 894.
- [36] MILLER M, CLARK A. Happily entangled: prediction, emotion, and the embodied mind [EB/OL]. (2018 – 07 – 05) [2024 – 10 – 03]. https://www.pure.ed.ac.uk/ws/portalfiles/portal/35969473/SYNT_D_16_00612_R3_2.pdf.
- [37] BARRETT L F, SIMMONS W K. Interoceptive predictions in the brain [J]. *Nature reviews neuroscience*, 2015, 16(7): 419 – 429.
- [38] PESSOA L. Understanding brain networks and brain organization [J]. *Physics of life reviews*, 2014, 11(3): 400 – 435.
- [39] KIVERSTEIN J, MILLER M. The embodied brain: towards a radical embodied cognitive neuroscience [J]. *Frontiers in human neuroscience*, 2015, 9: 237.
- [40] WILKINSON S, DEANE G, NAVE K, et al. Getting warmer: predictive processing and the nature of emotion [J]. *The value of emotions for knowledge*, 2019.
- [41] RUSSELL J A. Culture and the categorization of emotions [J]. *Psychological bulletin*, 1991, 110(3): 444.
- [42] RUSSELL J A. Core affect and the psychological construction of emotion [J]. *Psychological review*, 2003, 110(1): 145.
- [43] 诺瑟夫. 自发的大脑: 从心身问题到世界 – 大脑问题 [M]. 陈向群, 徐嘉玮, 译. 北京: 中央编译出版社, 2024: 12.

How Do Affective Brain – Computer Interfaces Understand Emotion?

——Basic Emotion or Constructed Emotion

WU Xue – mei

(School of Philosophy and sociology, Lanzhou University, Lanzhou 730000, China)

Abstract: The challenges faced by affective brain – computer interfaces (aBCIs) in understanding dynamic emotions demand clarification of the definition of emotions from the perspective of the philosophy of cognitive science. Currently, proponents of basic emotion theory argue that the theory of constructed emotion not only fails to establish the cultural relativity of emotions due to the lack of necessary procedural interventions in its experimental design, but also that there is evidence suggesting emotional knowledge is not acquired through cultural inheritance. However, based on an analysis of texts related to the theory of constructed emotion and advances in the philosophy of cognitive science, these criticisms fail to undermine the validity of its arguments. The absence of procedural interventions in the theory of constructed emotion is rooted in its reliance on the bidirectional interaction between emotion and cognition to ensure the accuracy of conclusions. Moreover, considering that the theory of constructed emotion discusses emotions on the basis of relational realism, the conclusion that emotional knowledge is culturally inherited holds true. Insights from the theory of constructed emotion regarding emotion instances and cultural inheritance can assist aBCIs in understanding dynamic emotions and enhance their ability to interpret emotions specific to particular cultures.

Key words: affective brain – computer interfaces; basic emotion; constructed emotion

(责任编辑 赵 雷)