

神经科学视阈下 自由意志的二阶梯度

隋婷婷

[摘要] 随着神经科学的发展,自由意志概念受到了诸多挑战。20 世纪 60 年代以来,以大脑“准备电位”实验为代表的众多神经科学研究提出,人类意识由先发的神经活动所决定,人类行为先于其本身对行为的自我意识。神经科学对大脑神经递质与脑区功能的研究也使得意志的自由性成为易被影响和易于消解的对象。但若依照自由意志的层序欲望理论,这一现象并不意味着神经科学研究对哲学概念中自由意志的否定,而是反映了自由意志当中“一阶欲望”与“二阶意志”观察梯度的参差,其本质不在于自由意志概念本身的存废,而在于自由意志当中“一阶欲望”在神经科学领域的易于影响、易于消解与“二阶意志”的难以观测及难以量化。

[关键词] 自由意志;神经递质;脑区功能存抑;层序欲望理论

一、引言

自由意志这一概念自诞生以来便饱受争议。宏观尺度上,自由意志的概念与因果决定论颇难兼容:若承认万事万物的发展遵循因果相继的递进关系,则自由意志的加入将尴尬地破坏原有因果链条的连贯性。而若转向消除因果性的量子随机论,人类的行为又将成为完全随机的结果,从而在另一个方面消除了自由意志的存在。^①

面对宏观尺度的困境,一种较为普遍的兼容性态度是将自由意志的讨论聚焦于人类意识的微观层面,认为自由取决于个体自身的意愿,而非由外在的因果决定,即使个体没有现实层面的备选可能性(alternative possibility),个体的行动依旧可以被称为自由的,因而个体需要为行为负起责任。^②这也是人能够成为道德主体的关键,“因为,当我们在自己能力范围内行动时,不行动也在我们的能力范围之内,反之亦然”^③,这也是在伦理与法律意义上,人们在结果相同的情境中用于区分有意与无意、蓄意与过失的理论依据,而且是后续嘉奖或归责具有道德正当性的前提。比如,在《墨子·耕柱》的火灾案例中,对于“一人奉水将灌之”和“一人掺火将益之”但“功皆未至”

作者: 隋婷婷,北京大学哲学系博雅博士后研究人员, suitingting@pku.edu.cn。

* 本文系中国博士后科学基金面上项目“人工智能算法的人本嵌入研究”(2021M690243)、国家社会科学基金重大项目“人工智能伦理风险防范研究”(20&ZD040)和江苏省社会科学基金重点项目“人工智能的伦理风险及防范研究”(20ZX A001)的阶段性成果。在《中国人民大学学报》编辑部举办的“实验哲学的方法、价值与应用”专题研讨会上,评审专家们对本文提供了可贵的建议与热诚的帮助,在此谨表诚挚谢意,文责自负。

① J. J. C. Smart. “Free Will, Praise and Blame”. *Mind*, 1961, 70 (279): 291 - 306.

② 李恒熙、李恒威:《二阶欲望与欲望的冲突》,载《哲学分析》,2010 (03)。

③ 亚里士多德:《尼各马可伦理学》,76 页,商务印书馆,2009。

的情境，人们从道德动机层面更赞赏“彼奉水者之意，而非夫掺火者之意”。^①

在人类意识层面上，自由意志作为认知活动的一种，天然地成为由认知神经生理过程介导的可观察对象。^② 尽管囿于意识与生理基础之间的解释鸿沟，目前尚无法解释自由意志涌现或突现的完整神经机制，但除了彻底的心脑二元论者将心理与生理完全割离开，研究者们通常承认自由意志与大脑神经生理活动密切相关，其功能实现涉及皮层下基底神经节（basal ganglia）、丘脑（thalamus）以及杏仁核（amygdala）前额叶（prefrontal）、前扣带皮层（anterior cingulate cortex）和顶叶皮层（parietal cortices）等边缘结构（limbic structure）组成的重入回路（re-entrant loops）：基底神经节主要调节主体进行运动控制、动机注意力和计划所必需的一些联想、认知和情感功能；丘脑则将感觉皮层的信息传递给调节情绪的杏仁核以及其他更高级的大脑区域，同时，丘脑通过对皮质的投射在意识觉醒和意识活动方面也发挥了关键作用；前扣带皮层作用于共情、感知和冲突解决；大脑皮层的前额叶区域则主要调节道德实践和道德推理。^③

此类研究一方面为自由意志认知神经生理机制的观测和厘清提供了理据，但另一方面，也使得自由意志面临神经决定论的消解困境，研究者们不得不思考主体是否拥有不单纯由神经生理、机械因果性所决定的自由意志。^④

二、自由意志与一阶欲望的认知神经观察梯度

随着认知神经科学对神经递质、脑区功能以及准备电位研究的深入，科学家似乎在很多方面证明了自由意志的不自由，从而挑战了自由意志的存在。然而，神经科学语境中的自由意志往往与哲学语境中的不同，借用层序欲望理论（Hierarchical desire theories）中对自由意志结构特殊性的描述，自由意志不仅涉及个体行动的自主性，而且涉及“一阶欲望”（first-order desire）、“二阶欲望”（second-order desire）以及“二阶意志”（second-order volition）的梯度区分。认知神经科学当前对于自由意志的研究多数聚焦于“一阶欲望”层面的测量与研究，这一观察梯度与自由意志的存在困境有着紧密关联。^⑤

（一）层序欲望理论与意志的结构

哲学对意志自由的研究有着悠久的历史，哈利·法兰克福（Harry Frankfurt）的层序欲望理论是当代较有代表性的理论之一^⑥，同时，也是“自律”（autonomy）理论来源当中程序性理论的代表。^⑦ 尽管该理论在欲望的冲突层次等方面颇有争议，但笔者认为可以借用其中的层序划分思想，为意志自由提供一种结构性视角，这一视角与查尔莫斯将意识问题分为“难问题”与“易问题”的思路类似。^⑧ 虽然这一划分方式并不能解决所有问题，也有相对含糊的地方，但对厘清和回应某些具有神经决定论色彩的论断有其积极意义。

层序欲望理论认为人类拥有区别于动物的意志结构，人类主体根据自身基本的喜好或欲求产生“想要”（to want）做某事的行动欲望，被称为“一阶欲望”。动物同样拥有“一阶欲望”，这一欲

① 吴毓江：《墨子校注》，657页，中华书局，1993。

② 尽管有少数学者持不承认意识与生理基础之间存在关系的极端身心二元论观点，但该观点在目前学界认可度较低，故不在下文中进行讨论。

③ W. Glannon, et al. "Neuropsychological Aspects of Enhancing the Will". *The Monist*, 2012, 95 (3): 378 - 398.

④ A. Lavazza. "Free Will and Neuroscience: From Explaining Freedom away to New Ways of Operationalizing and Measuring It". *Frontiers in Human Neuroscience*, 2016, 10 (262): 1 - 17.

⑤ H. G. Frankfurt. "Freedom of the Will and the Concept of a Person". *The Journal of philosophy*, 1971, 68 (1): 5 - 20.

⑥ 姚大志：《意志自由论的两种形态及其理论检验》，载《南京大学学报》（哲学·人文科学·社会科学版），2017（6）。

⑦ 丛杭青、王晓梅：《何谓 Autonomy?》，载《哲学研究》，2013（1）。

⑧ D. J. Chalmers. "Moving Forward on the Problem of Consciousness". *Journal of Consciousness Studies*, 1997, 4 (1): 3 - 46.

望能够直接驱使它们行动。人类则在拥有“一阶欲望”的基础上拥有“二阶欲望”和“二阶意志”。若将“一阶欲望”表述为“A想要X”，这里的X是某个特定的一阶欲望所指向的一个具体对象，那么，当把“A想要X”这一表述的内容本身视作欲求的对象，把这个表述用F来表示，则“A想要F”便体现了主体的“二阶欲望”，表达了对特定欲求之欲求，是对“一阶欲望”的延伸与反思，也是主体“二阶意志”施行时的内容和动力来源。^①

需要说明的是，“欲望”（desire）这一用词并不准确。法兰克福本人表示希望借此表达“想要”（want）做某事的动机与欲求等概念，由于没有更合适的概括性词汇，便选择了“欲望”，这也导致了“二阶欲望”概念的模糊和争议性。根据法兰克福的观点，“二阶欲望”概念的存在，很大程度上是为了划分人与动物在自由方面的区别。动物尽管拥有自主行动的自由，但没有人类的反思和自我评价能力，因而也不能通过这一能力对“一阶欲望”进行反思和挑选，形成修正后的“二阶欲望”。^② 人类拥有区别于动物的意志结构并不代表这一过程必然发生，人同样可以主动放弃反思“一阶欲望”，成为一个听凭本能行事的“放浪者”（wanton），因而也有学者借用层序欲望理论解释“意志软弱”的发生过程^③，或者将“二阶欲望”概念作为对“一阶欲望”的评价和区分，认为拥有“二阶欲望”的行动者不会被头脑里偶然出现的欲求随意驱使。^④ 总体而言，法兰克福的“二阶欲望”概念希望表达的是人类通过反思和自我评价等思维活动加工“一阶欲望”之后的产物，本质上强调了某种思考过程的存在，最终产物通常包含某个“一阶欲望”的相关内容（例如某人有“喝酒”的“一阶欲望”，反思后想要“戒酒”的念头胜出，成为“二阶欲望”的内容）以及想要施行该意欲的欲求，这一欲求被实施，便是“二阶意志”。以上过程中明显包含思考等认知活动，但在用词上并没有很好地体现这一点。^⑤

“二阶欲望”与“二阶意志”分别代表了“内容欲求”和“行动”，为了避免无穷倒退问题的出现，即在二阶之上叠加三阶、四阶等反思，“二阶欲望”通常被用来指代“二阶意志”的最终内容来源，也即是“二阶”作为一个虚指的数词，用于表达最后被意志付诸实施的那一个欲求，因而两者在某种意义上也是同一的。这会引发的一个疑惑是，既然“二阶欲望”会成为对应的“二阶意志”，那么分离两者是否必要？法兰克福的理由是存在一些只有“二阶欲望”没有“二阶意志”的对象，如成瘾者有戒断的“二阶欲望”却没有实施的“二阶意志”。^⑥ 笔者不同意这一理由，因为存在成瘾者无法依靠自己戒断转而寻求他人帮助和治疗的现象，这也体现了意志的存在。但笔者认为将“二阶欲望”与“二阶意志”分开有其正当性，因为两者分别代表知与行，参照中国传统知行观，特别是王阳明的知行观，“知”包含行的意念。^⑦ 在知行观的语境中，依然保留“知”与“行”的提法，而不是用单一的“知”或“行”来指代从认知到行动的整个过程。

（二）“一阶欲望”与神经科学对自由意志的挑战

神经科学对自由意志的存在带来了很多挑战，其中较有代表性的是对神经递质、脑区功能以及准备电位的研究。当前研究中涉及的对象与“一阶欲望”的定义较为接近，因为相比被主体反思取舍之后实施的“二阶意志”，“一阶欲望”通常较为简单，同时，其并不必然成为被实施的行为，但往往包含着“想要”做某事的动机或推动某个行为的驱动力。

①②⑥ H. G. Frankfurt. "Freedom of the Will and the Concept of a Person". *The Journal of philosophy*, 1971, 68 (1): 5-20.

③ 段素革：《意志软弱何以可能？——哈里·G. 法兰克福自主性理论框架内的道德心理学分析》，载《伦理学研究》，2011（4）。

④ 李石：《意志自由和行动自由——基于人类欲求之等级结构的分析》，载《世界哲学》，2010（1）。

⑤ 尽管“二阶欲望”的用法容易被误读为是“一阶欲望”的欲望叠加，但因为相关学者已熟悉这一词汇，为了用词的简洁性和易辨认性，下文依然采用“二阶欲望”的用法。

⑦ 姚新中、隋婷婷：《当代社会心理学视域下的知行合一》，载《江苏社会科学》，2020（1）。

1. 神经递质的影响

认知神经科学层面对于人类行动的内驱力较有代表性的研究主要涉及神经递质（neurotransmitter）。神经递质指在神经突触传递中担当“信使”的特定化学物质，神经递质经由神经元之间的传导往返，将各项功能串联起来，是实现各种认知功能的必要神经基础之一。尽管目前对神经递质的功能描述尚不统一，但很多研究指出其对人们行动的内驱力和情绪有广泛的影响。由于一些神经递质常与成瘾行为的机制相关，有时也会被蒙上神经决定论的色彩，如5-羟色胺与人的内驱力（睡眠、食欲等）和情绪密切相关，并会增强人们的“厌恶伤害”（harm aversion）倾向^①，多巴胺则与“想要”做某事的动机激励和动机唤醒相关。^②

神经递质本身具有可调节和可影响的特性，极易受外部措施的干预和扰动。在莫莉·克罗基特（Molly Crockett）的5-羟色胺“最后通牒博弈”（ultimatum game）实验中，5-羟色胺增强组（服用西酞普兰）、5-羟色胺抑制组（服用阿托西汀）以及安慰剂对照组的被试对不公平行为的容忍度产生了显著差异。克罗基特认为，这反映了5-羟色胺对被试内驱力的影响，5-羟色胺增强能使被试的心理状态更加接近理性人假设中的主体，而对其的抑制则使被试产生了“可预见的非理性”（predictably irrational）。^③

哈里斯·威斯曼（Harris Wiseman）则进一步指出这与5-羟色胺对大脑巴甫洛夫系统（Pavlovian system）的影响有关。巴甫洛夫系统是进化过程中形成的反应系统，这一系统发生作用时常常是无意识的，主要对基本的快乐和痛苦的信号作出反应，因而也代表了人们的本能冲动。^④

相比5-羟色胺，多巴胺对人类行为的影响也极为显著。早期研究常采用多巴胺介导快乐的假说，但近年来，这一理论被“激励突显”（incentive salient）以及“奖赏预测误差”（reward prediction error hypothesis）等与动机、学习相关的假说取代。^⑤

神经科学家爱德华·博伊登（Edward Boyden）认为，多巴胺对行为有影响是由于其能够通过大脑的奖赏环路（reward circuitry）发挥作用，从而推动人们的认知形成和行为固化。为证实这一假说，他采用光遗传学（optogenetics）的方法，结合光敏元素（light-sensitive elements）与遗传工程对恒河猴报偿中枢（reward center）的多巴胺神经元进行了光刺激，发现多巴胺确实能够影响恒河猴的认知、学习和行为选择并产生持续性的特定行为，在人的行为过程中也会发挥对应的作用。^⑥ 由于多巴胺具有促进学习强化的功能，一些研究又认为它与很多动机效应相关。^⑦ 因此，部分学者认为，多巴胺具有增强或戒断主体行为意愿的功能，对多巴胺的研究有助于理解成瘾、强迫、自由意志和欲望等方面的机制。^⑧

①④ H. Wiseman. *The Myth of the Moral Brain: The Limits of Moral Enhancement*. MIT Press, 2016, p. 96, p. 95.

② K. C. Berridge. "The Debate over Dopamine's Role in Reward: The Case for Incentive Salience". *Psychopharmacology*, 2007, 191 (3): 391 - 431; R. A. Wise, and M. A. Robble. "Dopamine and addiction". *Annual Review of Psychology*, 2020, 71 (1), 79 - 106.

③ M. J. Crockett, et al. "Serotonin Modulates Behavioral Reactions to Unfairness". *Science*, 2008, 320 (5884): 1739.

⑤ W. Schultz. "Dopamine Reward Prediction-error Signalling: A Two-component Response". *Nature Reviews Neuroscience*, 2016, 17 (3): 183 - 195; K. C. Berridge. "Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory". *Inquiry (Oslo)*, 2009, 52 (4): 378 - 398.

⑥ W. R. Stauffer, et al. "Dopamine Neuron-Specific Optogenetic Stimulation in Rhesus Macaques". *Cell*, 2016, 166 (6): 1564 - 1571.

⑦ J. Panksepp, and J. Moskal. "Dopamine and Seeking: Subcortical 'Reward' Systems and Appetitive Urges". In A. J. Elliot (ed.). *Handbook of Approach and Avoidance Motivation*. Psychology Press. 2008, pp. 67 - 87.

⑧ H. Wiseman. *The Myth of the Moral Brain: The Limits of Moral Enhancement*. MIT Press, 2016, pp. 107 - 109; K. C. Berridge. "Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory". *Inquiry (Oslo)*, 2009, 52 (4): 378 - 398.

2. 脑区的功能存抑

“一阶欲望”代表了个体对某种行动的自主意愿，但在大脑区域功能的层面上，主体意欲的萌发并非完全自主。脑区的抑制、损伤或功能紊乱有时会使得主体表现出间歇性削弱或丧失自由意志的状态。

在涉及脑区损伤的病理性恋童癖案例中，一些患者由于肿瘤等生理病变压迫大脑左额叶（left frontal lobe）和中央区域（central areas），产生了恋童的不道德欲望，经过肿瘤切除等针对性治疗之后患者恢复正常，恋童的欲望也消失了，数年后肿瘤复发，恋童的欲望再次产生，再次治疗后，又恢复了正常。^①从某种意义上说，患者脑区病变似乎使其短暂地丧失了对“一阶欲望”的自主管控，而痊愈之后，主体恢复了正常的自主性。在患病期间，尽管患者可能认为这些意愿是自主产生的，但意愿的产生似乎并不能被简单地判定为完全出自他们的本心。此类现象也在刑事司法层面引发了对自由意志责任认定的讨论。

克里斯·弗里奇（Chris Frith）指出人们的自主感（sense of agency）与前扣带皮层和腹内侧前额叶皮层的活动相关。^②瑞恩·德比（Ryan Darby）在脑区损伤网络成像的自由意志感知（free will perception）脑区分布研究中，进一步厘清了与主体自主感和意愿（volition）相关的脑区。通过与健康者脑区成像扫描图的对比，德比指出，自主感紊乱或缺失的主体（如异手症患者^③）在大脑楔前叶皮质（precuneus cortex）相连的区域网络有不同程度的损伤，其中较有代表性的是胼胝体（corpus callosum）损伤造成的大脑两半球信息交流受阻引发的主体感不统一。

主体的意愿缺失（如无动性缄默症患者^④）则与大脑前扣带皮层息息相关。由于前扣带皮层的活动通常对应着人类的同情、决策等认知功能，德比认为造成这一情况的原因之一是意志活动需要通过前扣带皮层相连的大脑区域网络产生作用，而脑区损伤在生理上则阻断了这一功能。^⑤

3. 准备电位的时序差异

相比神经递质和脑区存抑与意志自由的相关性，本杰明·里贝特（Benjamin Libet）的准备电位（readiness potential）实验是神经科学对自由意志最大的挑战之一。^⑥

准备电位（Bereitschaftspotential）的概念产生于20世纪60年代，源于德国神经科学家汉斯·赫尔米特·科恩休伯（Hans Helmut Kornhuber）和路德·迪克（Lüder Deecke）的研究。他们发现，当人们自发地作出动作时，通过脑电图（EEG）可以看到运动皮层出现一个缓慢的负电位变化，这一电位变化被命名为“准备电位”。^⑦

里贝特将准备电位原理应用到了预测被试意图的实验中。被试在实验当中需注视一个圆盘形示波器上顺时针运动的光点（光点转动一圈的总时间为2.56秒），以此来标记并报告自己产生弯曲手腕等动作的意图的时刻，在脑电图的即时检测中，被试大脑中运动准备电位的出现早于其活动手腕

① F. Gilbert, and A. Vranic. "Paedophilia, Invasive Brain Surgery, and Punishment". *Journal of Bioethical Inquiry*, 2015, 12 (3): 521 - 526.

② C. Frith. "Attention to Action and Awareness of Other Minds". *Consciousness and Cognition*, 2002, 11 (4): 481 - 487.

③ 异手症又称奇爱博士综合征，指患者的手不受自主意识的控制，尽管患者的手有正常的感觉，但时常作出主体命令之外的行为，如解开纽扣或敬礼等。患者有时甚至不知道异手在做什么，直到异手作出引起他们注意的行为。

④ 无动性缄默（akinetik mutism, AM）是意识障碍的一种特殊类型，基本表现尽管对刺激可有反射性的四肢运动，但无随意运动、自发言语以及任何的情绪反应。

⑤ R. R. Darby, et al. "Lesion Network Localization of Free Will". *Proceedings of the National Academy of Sciences-PNAS*, 2018, 115 (42): 10792 - 10797.

⑥ 本杰明·里贝特：《心智时间：意识中的时间因素》，75 - 76页，浙江大学出版社，2013。

⑦ H. H. Kornhuber, and L. Deecke. "Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale". *Pflügers Archiv*, 1965, 284 (1): 1 - 17.

之前的 550 毫秒，比被试所报告的意识到手腕意图的时间早了约 350 毫秒。里贝特由此提出人的动作并不受意识指挥，而是在动作发生后才意识到动作的发生。^① 这一结果也是近年来支持神经科学消解自由意志的经典论据之一。

对神经递质、脑区抑制以及准备电位的研究在不同程度上表明了大脑的神经生理基础与人类自主意愿的关系，因而有部分研究者认为自由意志不存在或人类意志的自主性十分有限，“只不过是—大群神经细胞及其相关分子的集体行为”^②，极易被神经生理基础决定、影响和改变。

在“—阶欲望”的观察梯度上，该结论似乎有一定道理。但是，在自由意志的整体概念上，考虑到“—阶欲望”和“—阶意志”的存在，“—阶欲望”尽管表现了人类的某种自主性，我们却不能单纯以“—阶欲望”代表自由意志。

三、自由意志的二阶观察梯度

人类作为一个有意志的存在者，不仅拥有“—阶欲望”，还有在其基础上萌发的“—阶欲望”和“—阶意志”。“—阶欲望”代表高阶思维对“—阶欲望”进行节制与反思的结果，当主体将“—阶欲望”诉诸意志使其成为“—阶意志”并付诸行动时，存在于思维空间的“—阶欲望”便能成为主体作出的实际选择，从而获得制约“—阶欲望”的自由。在相关实验中，“—阶意志”对“—阶欲望”的这一约束在个体行为领域也得到了佐证。

（一）“—阶意志”与自由否决权

谈到神经科学对自由意志构成挑战或者消解了自由意志时，人们的论断中往往会内在地包含两个前提：

前提—，某种神经活动或神经结构对主体行为是决定性的。

前提二，主体无法对以上活动施加影响。

前文对神经递质、脑区功能以及准备电位的论述并不完全符合以上前提，或者说至多在“—阶欲望”的范围内符合前提。因为大脑神经生理基础作为具有物理广延的实体，它与心理状态之间的相关性究竟如何定义仍有争议，很多研究者可能更愿意采用“相关性”而非“神经活动对某种行为具有决定性”这一提法，有部分研究者愿意采信“神经活动决定了某种行为”这一观点，但它们并不等于某种不能被“—阶意志”干涉的、具有不可抗力的活动。

以脑功能区损伤或抑制而产生恋童欲望的患者为例，有部分患者表示对自己产生的不道德欲望有过思想斗争，并通过求助医生或进行治疗尽力避免作出不道德行为。^③ 尽管患者产生的“—阶欲望”（恋童）是由脑区功能异常所引发，但仍保有反思欲望并形成“想要避免不道德行为”的“—阶欲望”的可能性，部分患者也成功实施了这一想法（找医生求助），这个从“想要”到“践行”的过程也是从“—阶欲望”到“—阶意志”的过程。而对于异手症等看似无法被主体控制的状态，主体同样保有对其进行治疗，或通过物理方式（捆绑）限制其动作的方式。这种自由也得到了某些坚持神经科学终将消解自由意志的学者的承认，如萨姆·哈里斯（Sam Harris）提出，尽管人是神经生理反应之下的傀儡，但若识破思想情感的种种表象，人可以有把握抓住束缚在身上的绳索，并更加明智地度过一生。^④

在神经递质的层面，多巴胺受体尽管与各种成瘾行为相关，但主体主动寻求戒断和脱敏治疗的

① 本杰明·里贝特：《心智时间：意识中的时间因素》，77-83 页，浙江大学出版社，2013。

② 弗朗西斯·克里克：《惊人的假说》，3 页，湖南科学技术出版社，2007。

③ F. Gilbert, and A. Vranic. “Paedophilia, Invasive Brain Surgery, and Punishment”. *Journal of Bioethical Inquiry*, 2015, 12 (3): 521-526.

④ 萨姆·哈里斯：《自由意志：用科学为善恶做了断》，76 页，浙江人民出版社，2013。

行为也屡见不鲜。此外,相比受到5-羟色胺影响的巴甫洛夫系统,大脑中还存在着与之竞争的目标导向系统(goal-directed system),该系统使人们能够在作出道德判断时思考社会规范,从而作出更符合社会规范和道德规范的决定和行为。巴甫洛夫系统是人类在进化过程中形成的反应系统,这一系统发生作用时常常是无意识的,而目标导向系统在发挥作用时通常是有意识的,它主要对“响应—结果关联”(response-outcome association)作出反应,借助认知理性推断从目前条件导出的最佳结果是什么,并能够帮助人们具象化自身的目标以克制本能的欲望。^①诚然,神经递质受到某些化学元素影响,可能会使得某个系统更容易胜出,但我们仍可以采用一种笛卡尔式的辩护,即:相关化学元素影响神经递质受体从而使自由意志受到削弱和影响,这一过程的前提必然包含一个存在着的自由意志。

相比神经递质和脑区功能研究,准备电位实验“动作电位出现时间先于意识”的结论对自由意志的挑战更为尖锐,但依然可以被“二阶意志”所解释。

一方面,里贝特实验尽管把动作之前的几百毫秒纳入了实验的观察域,但是,被试事前有意识思考和决策(考虑在实验中如何行动)的认知神经活动过程并不能被实验观察的时段所覆盖,也无法单纯通过时间窗定义。因此,准备电位的自相关波动对于主体的自由意志而言,可以被视为前阶段有意识的意志和指令在随机动作实施阶段这一狭窄的观察阈中的留影与余响。^②多米尼克·汤玛(Dominik Tomá)以及保罗·桑福德(Paul Sanford)在2018年以及2019年对里贝特实验的研究中表达过类似观点,认为实验前的练习能够对实验当中的反应产生影响。^③马里兹·布莱恩(Moritz Braun)的元分析研究也提出里贝特可能存在未考虑被试在实验前练习时的意识活动等疏漏。^④另一方面,在承认动作与意识有时间差的前提下,意识活动尽管在准备电位之后至少400毫秒才出现,但真实的动作仍要在意识活动发生之后的150~200毫秒才会出现,这一段时间人们的意识活动可以决定是否让运动过程得以完成,它们能够阻止或“否决”这个过程,而不使已产生准备电位的运动发生。^⑤

这一实验结果肯定了“二阶意志”有在行动上干涉、否决“一阶欲望”(活动手腕)的自由选择权。这一自由否决权的存在与自由意志概念中包含的主体拥有“另一种选择的可能性原则”不谋而合^⑥,因而也从行为角度为自由意志的“二阶意志”在认知神经基础当中划出了一席之地。

(二)“二阶意志”与神经科学的兼容性

“二阶意志”作为人类认知活动的组成部分,必然也有其对应的神经生理基础。因此,对于“二阶意志”的一个可能的问题是,“一阶欲望”能够部分被神经科学解构和消解,“二阶意志”是否会随着神经科学的发展而被逐渐消解,从而使依靠“二阶意志”论证主体拥有自由否决权的自由意志陷入尴尬处境。

在当前的研究中,神经科学对于自由意志当中涉及的意图、运动和主体感等方面有广泛的研

① F. Gesiarz, and M. J. Crockett. "Goal-directed, Habitual and Pavlovian Prosocial Behavior". *Frontiers in Behavioral Neuroscience*, 2015, 9 (5): 1-18.

② B. Libet. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action". *The Behavioral and Brain Sciences*, 1985, 8 (4): 529-539.

③ T. Dominik, et al. "Libet's Experiment: A Complex Replication". *Consciousness & Cognition*, 2018, 65 (7): 1-26; P. Sanford, et al. "Libet's Intention Reports Are Invalid: A Replication of Dominik et al. (2017)". *Consciousness and Cognition*, 2020, 77 (102836): 1-16.

④ M. N. Braun, et al. "A meta-analysis of Libet-style Experiments". *Neuroscience & Biobehavioral Reviews*, 2021, 128 (9): 182-198.

⑤ 本杰明·里贝特:《心智时间:意识中的时间因素》,85-86页,浙江大学出版社,2013。

⑥ 甘绍平:《意志自由的塑造》,载《哲学动态》,2014(7)。

究，比如，马克·简尼洛德（Marc Jeannerod）在“运动模拟理论”（motor simulation theory）中对运动与模拟网络关系的假设^①，吉亚科莫·利兹拉蒂（Giacomo Rizzolatti）对“运动”和“共情”相关的镜像神经元（mirror neuron）的研究^②，克里斯·弗里奇对意志与自主感的论述^③，海伦·格拉格（Helen Gallagher）对“心理理论”原理的探索^④等。然而，此类研究大多专注于意识、行动的功能原理及其与特定神经基础的相关性，总体属于查尔莫斯划分的“意识功能执行”“心理状态报告”“行为的自控”等意识的“易问题”范畴，不太涉及“意识怎样从大脑的物理过程中产生出来”这样的“难问题”解释。^⑤因而，研究较少涉及神经基础是否能够决定、消解或还原自由意志及其包含的“二阶意志”的心理—物理因果论断，通常也不对自由意志概念本身提出尖锐的否定，有些研究者甚至直接表达了对自由意志概念的肯定，与自由意志及其包含的“二阶意志”概念有较好的兼容性。^⑥

而通过“运动准备电位与意识的时间差”对自由意志提出挑战的里贝特实验，由于“二阶意志”涉及较复杂的意识活动，研究者们很难从庞杂的脑电图谱中对之进行完整的解读和预测。比如，在约翰·达伦·海恩斯（John Dylan Haynes）通过脑机技术实现的“人机对抗”实验中，有约66.5%的参加者可以通过有意识地控制自己的按键动作使计算机无法预测（在不让机器进行任何预测的纯概率盲猜条件下，这个数字也可能接近50%）。^⑦

同时，高阶意识活动的复杂性导致其在神经科学实验中的时间对应性较差，对于采用“事件相关电位”技术的里贝特实验而言，由于“事件相关电位”有对锁时的限制，相应的实验也难以被设计出来。罗伯特·诺斯卡特（Robert Northcott）因而也从哲学角度对自由意志的可测性提出了怀疑，并指出自里贝特实验以后，对涉及高阶认知反思的主体意识活动目前没有特别成功的神经预测实验（neuroprediction experiment）。^⑧因此，在复杂意识活动的层面上，里贝特实验中行动（动作电位）早于意识（被试意识到自己发出指令）的现象目前是难以复刻的。

对此，一种可能的假设是认为神经科学在未来发展中能够最终消解“二阶意志”，但笔者认为寄望发展是一种过于乐观的假设。首先，这涉及意识“难问题”的解决。“难问题”能否在未来被解决本身充满争议，自由意志及其包含的“二阶意志”属于意识活动的一种，其解决与否和意识问题息息相关。^⑨其次，通过发展解开所有问题本身是一个颇为可疑的前提。哥德尔曾指出：“算术系统中存在一个命题，它在这个系统中既不能被证真也不能被证伪。”^⑩这一不完备定理在某种宽泛的定义下也可表明某个形式系统中可能存在无法被证真或证伪的对象。因而笔者认为，如果“神经科学的发展能够解决‘二阶意志’的所有问题，从而能够最终消解掉这一概念”在假设层面具有正当性，那么假设“二阶意志”的研究中存在哥德尔式的不完备，也将具有同等的正当性。

① M. Jeannerod. "Neural Simulation of Action: a Unifying Mechanism for Motor Cognition". *Neuroimage*, 2001, 14 (1): 103 - 109.

② G. Rizzolatti. "Confounding the Origin and Function of Mirror Neurons". *Behavioral and Brain Sciences*, 2014, 37 (2): 218 - 219.

③ C. Frith. "The Psychology of Volition". *Exp Brain Res*, 2013, 229 (3): 289 - 299.

④ H. L. Gallagher, and C. D. Frith. "Functional Imaging of 'Theory of Mind'". *Trends in Cognitive Sciences*, 2003, 7 (2): 77 - 83.

⑤⑨ D. J. Chalmers. "Moving Forward on the Problem of Consciousness". *Journal of Consciousness Studies*, 1997, 4 (1), 3 - 46.

⑥ M. Hallett. "Physiology of Free Will". *Ann Neurol*. 2016, 80 (1): 5 - 12; 本文主要从层序视角回应里贝特实验等较为尖锐的自由意志挑战，限于篇幅，对于未对自由意志提出激烈挑战的相关神经科学研究，此处就不详细讨论了。

⑦ M. Schultze-Kraft, et al. "The Point of No Return in Vetoing Self-initiated Movements". *Proceedings of the National Academy of Sciences-PNAS*, 2016, 113 (4): 1080 - 1085.

⑧ R. Northcott. "Free Will is Not a Testable Hypothesis". *Erkenntnis*, 2018, 84 (3): 617 - 631.

⑩ K. Gödel. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. Dover Publications, 1992, p. 1.

四、结语

自由意志在众多领域中有着各不相同的定义与构架。在神学领域,自由意志通常指全知全能的“神”(或类似的存在)不能通过其所具有的神力控制人们的意志和选择权。在伦理学领域,自由意志的意义通常寓于个人在道德层面上要对自己的行为和选择承担责任,是个人能够成为一个道德主体的关键,如亚里士多德所指出的:“德性是在我们能力之内的。恶也是一样。”^①同时,自由意志在先天理性的层面上也是道德律存在的理由,如康德曾指出:“假如没有自由,则道德律也就根本不会在我们心中被找到了。”^②

而在认知神经科学领域,自由意志的概念通常被化简为主体拥有自主或自由的意志,以及该意志是否单纯由神经生理的、机械的因果性所决定。^③同时,由于认知神经科学的实验设计需求,其重点通常在于意识活动在时间层面的滞后性和可预测性。因而,神经科学家早期对自由意志概念的理解,在很大程度上延续了拉普拉斯妖的决定论观点,或主张“从意识到作出选择之前很长时间,大脑就已经作出了决定”^④,或认为“人的决定过程是一个机械的程序,结果完全由预先的机械程序所决定”^⑤,甚至直接主张自由意志的概念以及自由意志的存在已经被认知神经科学所消解。

一方面,此类论断的产生内嵌了认知神经电位与个体意识的成形之间存在因果链或零差异同步的前提,即认为人类的意识活动由先发的神经活动所决定。由于神经生理基础和意识活动之间存在着解释鸿沟,学界在普遍认同两者间拥有关联的前提下,无法明晰生理基础与意识涌现之间的因果转化链条。因此,为避免陷入可能的还原论困境或自然主义谬误,研究者们通常仅在承认两者具有同步关联性的一元论基础上对相关实验数据进行推理性解释。同时,神经科学在研究中常常存在将意识的相关概念降级、简化的现象,如将自由意志简化为不需要过多思考的手指运动,而在哲学语境中,与意识相关的概念通常指代较为复杂高阶的、涉及自身反思的活动,如在康德语境中就将先验意识解释为“作为本源的统觉对我自己的意识”^⑥。神经科学中的概念简化诚然是研究过程中为了使研究对象便于嵌入实验的一种不得已,但也极易造成在数据解读时对哲学概念的混淆不清。

另一方面,此类论断的产生也源自观察梯度的混沌不明。由于“二阶欲望”以及“二阶意志”涉及较复杂的意识活动,神经科学实验由于实验设计的限制,一些表现出近似神经决定论特征的研究往往来自对简单易重复行为的实验,而对高阶认知活动的研究常常只给出认知或行动与某脑区或某神经机制具有相关性的结果,并不表现出明显的决定论特征。鉴于目前对脑电信号的破解仍然较为初级,研究者们通常也很难完整地解读或预测复杂的意识活动。同时,必须强调的是,意识的可预测性并不等同于对自由意志的消解,因为某人自主行为的可预测性并不等于行为的非自主性,例如,一个人在下雨天出门要打伞,这是可预测的,但是这并不能说明打伞的行为是不自由的。

总体而言,目前对自由意志构成尖锐挑战的认知神经科学研究通常着眼于自由意志的“一

① 亚里士多德:《尼各马可伦理学》,76页,商务印书馆,2009。

② 康德:《实践理性批判》,2页,人民出版社,2016。

③ A. Lavazza. “Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It”. *Frontiers in Human Neuroscience*, 2016, 10 (262): 1-17.

④ K. Smith. “Neuroscience vs Philosophy: Taking Aim at Free Will”. *Nature (London)*, 2011, 477 (7362): 23-25.

⑤ J. Greene, and J. Cohen. “For the Law, Neuroscience Changes Nothing and Everything”. *Philosophical Transactions. Biological sciences*, 2004, 359 (1451): 1775-1785.

⑥ 康德:《纯粹理性批判》,97页,人民出版社,2017。

阶欲望”，有时还会将其与“二阶意志”混淆起来，使得一些对于自由意志的解释时常陷入概念性错位以及自然主义谬误当中，从而产生“自由意志被神经基础决定”或“自由意志被消解”等引发争议的论断。这一现状表明了认知神经科学在自由意志研究领域有时存在层阶性盲点问题，通过引入自由意志的层序理论，可以为解读未来的神经科学研究提供一种分层归类的视角，帮助研究者更加清晰地区别神经科学实验中涉及的意识活动的低阶与高阶，从而增加神经科学研究与哲学概念之间的可通约性，使哲学能够更好地吸收神经科学的研究成果，并最终推动两个领域的融合与发展。

Viewing the Second Hierarchical Order of Free Will from the Perspective of Neuroscience

SUI Tingting

(Department of Philosophy, Peking University)

Abstract: Following the development of neuroscience, the conception of free will is confronted with many challenges. Since the 1960s, many neuroscientific studies, represented by the brain “readiness potential” experiment, have suggested that human consciousness is determined by preemptive neural activity and that human behavior precedes the self-awareness of its own behavior. Neuroscientific research on brain neurotransmitters and the function of brain regions also makes the freedom of will an object that could be affected and resolved easily. However, according to the theories of hierarchical desire, this phenomenon does not mean that neuroscientific research denies the existence of philosophical concept of free will. In fact, it reflects an uneven observation scale for “first-order desire” and “second-order volition” of free will. Its essence does not lie in the existence or failure of the concept of free will itself, but in the “first-order desire,” which is easy to be influenced, and the “second-order volition,” which is difficult to be observed and quantified from the perspective of neuroscience.

Key words: Free will; Neurotransmitter; Inhibition and exhibition of brain function; Hierarchical desire theories

(责任编辑 李 理)