

· 科学技术哲学 ·

探索-利用困境再审视

——基于主动推理的视角

Reconsideration on the Exploration-Exploitation Dilemma: From the Perspective of Active Inference

王姝彦 /WANG Shuyan 柴新 /CHAI Xin

(山西大学哲学学院、科学技术哲学研究中心, 山西太原, 030006)

(School of Philosophy, Research Center for Philosophy of Science and Technology, Shanxi University, Taiyuan, Shanxi, 030006)

摘要: 探索-利用困境是当代行为心理学和人工智能等研究领域中的难点问题之一。近些年来,随着神经科学、认知心理学、智能科学、管理科学等学科的不断发展和脑成像技术的日益进步,有关探索-利用困境的探讨也得到了多向路的推展和深化。其中,神经科学家卡尔·弗里斯顿基于变分自由能原理提出的主动推理,从计算和神经层次上为探索-利用困境的重解提供了一种整合性的一元化分析立场。在主动推理框架下,可以将探索-利用困境的序贯决策问题转化为单一目标函数的优化,即预期自由能最小化。由此,探索与利用之间的平衡由预期自由能最小化所驱动,认知主体可依据外在环境的改变局部切换认知价值和实用价值从而自适应地平衡探索与利用行为。

关键词: 探索-利用困境 主动推理 预期自由能 整合

Abstract: Exploration-exploitation dilemma is a hotspot issue in contemporary research fields such as behavioral psychology and artificial intelligence. Recently, with the continuous development of disciplines such as neuroscience, cognitive psychology, intelligence science, and management science, as well as the increasing progress of brain imaging technology, the inquiry about exploration-exploitation dilemma has also been promoted and deepened in multiple ways. Among them, the active inference proposed by neuroscientist Karl Friston based on the variational free energy principle provides a monistic analytic standpoint of integration for the reinterpretation of exploration-exploitation dilemma from the computational and neural levels. Under the framework of active inference, the sequential decision problem of the exploration-exploitation dilemma can be transformed into the optimization of a single objective function, which minimizes the expected free energy. Thus, the balance between exploration and exploitation is driven by the minimization of expected free energy, and the cognitive subject can adaptively balance exploration and exploitation behavior by switching between epistemic and pragmatic values based on changes in the external environment.

Key Words: Exploration-exploitation dilemma; Active inference; Expected free energy; Integration

中图分类号: N031; B811.23 文献标识码: A DOI: 10.15994/j.1000-0763.2024.07.004

基金项目: 国家社科基金重点项目“认知整合主义研究”(项目编号: 23AZX015); 科技部、教育部“高等学校学科创新引智计划”: 当代哲学与新科学技术互动作用学科创新引智基地(项目编号: D20021)。

收稿日期: 2023年7月30日

作者简介: 王姝彦(1974-)女,山西高平人,山西大学哲学学院、科学技术哲学研究中心教授,研究方向为科学哲学、心理学哲学。Email: wshuyan@sxu.edu.cn

柴新(1996-)男,山西临汾人,山西大学哲学学院博士研究生,研究方向为认知科学哲学。Email: xinchai728@foxmail.com

探索-利用困境 (exploration-exploitation dilemma) 是当代行为心理学和人工智能等研究领域中的一个难点问题。其内涵可被概略性地描述为: 所有适应性认知主体在探求有价值的事物时都会面临一个基本的权衡, 即探索前所未有的选择以寻求更大的奖励抑或利用既往已知的信息去追求确定性的回报。可以说, 对于人类或动物如何在探索与利用之间做决策加以考察, 不仅有助于理解生物智能行为的一般机制, 更可为构拟未来通用人工智能提供一种基于自然智能的理论依据。近些年来, 随着神经科学、认知心理学、智能科学、管理科学等学科的不断发展和脑成像技术的日益进步, 有关探索-利用困境的探讨也得到了多向路的推展和深化。其中, 神经科学家卡尔·弗里斯顿 (Karl Friston) 在变分自由能原理 (variational free energy principle) 基础上提出的主动推理 (active inference), 从计算和神经层次上为探索-利用困境的重解提供了一种整合性的分析立场和一元化的方法进路。基于此, 本文在阐明传统决策理论视域下解释探索-利用困境所面临局限性的基础上, 尝试以主动推理为视角重新审视有机体决策行为中探索-利用困境, 并以近似贝叶斯推理方法为依据, 阐明认知主体可通过预期自由能 (expected free energy) 最小化从而自适应地平衡探索与利用行为。

一、探索-利用困境及其在传统决策理论下的考察

如上所述, 探索-利用困境及其相关问题的辩证日渐受到学界广泛关注, 相关探论已涉及人工智能、管理科学、行为心理学、认知科学等众多领域。具体而言, 在人工智能领域, 探索-利用困境通常源于强化学习算法本身的特点。在机器学习中, 与监督学习、无监督学习所不同的是, 强化学习强调智能体的学习是在与外在环境的交互中完成的。一方面, 为了使智能体从外在环境中获取更多的信息从而需要不断探索; 另一方面, 为了使智能体在当下获得即时的奖励也需要对已知信息加以利用,

由此, 如何权衡并在二者间保持平衡成为困扰强化学习的一个重要问题。在管理科学域面, 探索-利用困境往往与内蕴于管理实践中的决策权衡相关。例如, 在涉及对产业投资进行决策时, 投资者通常需要提早谋划是聚焦现有成熟产业还是尝试选择新兴产业。詹姆斯·马奇 (James March) 曾基于组织学习将“探索”诠释为开创、勘探与搜寻等伴随风险承担的相关活动, 将“利用”界定为执行、施用与提炼等重复性履践的相关活动。^[1] 在行为心理学中, 探索-利用困境则被认为是揭秘人类或动物决策认知机制的关键挑战。例如, 在波谲云诡的环境中, 人类或动物是选择即时获得奖励, 抑或探索可能在将来带来更大奖励的新机会。按照丰川渡 (Wataru Toyokawa) 等人的观点, 可以说“探索-利用困境是人类和非人类动物反复出现的适应性难题”。^[2] 在认知科学领域, 许多神经科学手段及动物实验也进一步推展了探索-利用困境方面的研究。例如, 借助于各种神经成像技术 (neuroimaging techniques), 关于探索与利用之间决策过程的部分神经相关性已得到初步揭示, 相关研究成果表明, 人类大脑中处理探索-利用困境的脑区主要涉及内侧前额叶皮层 (mPFC),^[3] 其中, 腹内侧前额叶皮层 (vmPFC) 的神经活动通常与预测未来行为结果、评估当下计划的可靠性以及探索新方案有关, 背内侧前额叶皮层 (dmPFC) 的神经活动则往往与评估行为结果或是放弃当前计划有关。^[4] 显然, 这也为进一步阐明认知主体探索-利用困境权衡的认知机制奠定了必要的实证基础。

由上可以看出, 基于学科领域及其相应的研究旨趣的不同, 探索-利用困境所呈现出的具体问题样态也存在一定差异, 但其基本内核在总体上并无实质上的区别。在现实决策过程中, 认知主体通常需要在探索新的行动方案或者利用当前行动方案之间做出合理决策以获得最大的回报。质言之, 探索性和利用性的区分构成了认知主体处于不确定性的环境中做决策时需要考量的两种重要状态, 即“合意性 (desirability)”与“可行性 (feasibility)”。也

就是说,“当考虑一项行动时,人们关注的要么是合意性,要么是可行性。前者指的是某一事件最终状态的价值,即某一行动本身是否具有很大的价值;后者指的是达到那个最终状态的手段,即执行该行动是否容易”。^[5]选择探索行为通常表明主体更加看重目标的合意性,而选择利用行为往往则表示主体更加看重目标的可行性。探索行为和利用行为在此构成一对矛盾选项,探索行为意味着需要投入更多资源以最大化未来奖励,但是投入更多资源并不一定会获取更大的奖励;利用行为意味着使用当前持有的资源以最大化即时回报。过多地选择探索行为可能无法获得预期的回报,而过多地选择利用行为则可能错过获取更大奖励的机遇。然而,当认知主体必须在现有方案和新方案之间做出抉择时,便会出现探索-利用两难困境。由此,探索-利用困境也逐渐成为研究认知主体如何在不确定性的环境中做出合理的决策的一个核心突破口。

在传统的决策研究视域下,可将探索-利用相关问题置于规范性范式与描述性范式两种不同的进路下加以考察。由于两种范式所依循的学科基础、解释向路与方法原则皆存在较大差异,因而对探索-利用困境的分析和求解也不相同。规范性研究范式主要遵从一种理性传统,其理论建构主要依赖于经济学原则、数学模型及统计学方法等,旨在提供既定条件下的最佳决策模型。描述性研究范式注重的则是实际的经验观察,其立论基础通常依托认知过程追踪、心理学实验以及社会学等方法,意在理解真实的认知主体在特定情境下的决策行为。

就规范性决策研究来讲,其暗含的基本假设为:决策者是完全理性的,并且可以枚举所有的备选选项进而做出最优的决策方案。其中,期望效用理论(expected utility theory)颇具影响力,该理论最早由计算机科学领域先驱冯·诺依曼(Von Neumann)和经济学家奥斯卡·摩根斯坦(Oscar Morgenstern)共同提出,他们运用数学模型和逻辑分析建立了不确定性条件下理性决策的框架。其核心在于按照“最大效用原则”去解释决策者的行为并揭示其中蕴含

的内在规律。在探索-利用困境的决策情景下,按照期望效用理论的观点,决策者应该选择效用值最高的备选方案。然而,这种基于完全理性的解释方案也受到种种质疑。其一,认知主体的内部决策模型无法仅仅通过经济效用这一单一因素来衡量,事实上,决策者复杂多元的思维方式不容忽视;其二,主体在分配决策权重(decision weight)时,对不确定性结果的主观信念并不总是符合数学概率理论的一些基本法则,它还反映了主体的非理性方面;([6], p.391)由于外在环境的不可预测性与模糊性,使得其解释效力存在很大程度的局限性。

相较而言,描述性决策研究所聚焦的并非是预设认知主体应该如何做决策,而是考究在实际多重语境下认知主体如何做决策。其中,前景理论(prospect theory)较具代表性,此理论由阿莫斯·特沃斯基(Amos Tversky)和丹尼尔·卡尼曼(Daniel Kahneman)提出,旨在缮校期望效用理论,其后又进一步推展为“累积前景理论(cumulative prospect theory)”。可以说,前景理论的提出标示着决策研究由理性主义向经验主义转换,其核心要义在于,认知主体做决策时会预设一个参照点(reference point),决策选择往往取决于结果与参照点(预期)的差异,而非结果本身。实质上,前景理论主要依赖于以下基本假设:认知主体进行决策时,其一,多数主体面对效益是风险规避的;其二,多数主体面对损失是风险偏好的;其三,多数主体对得失的判断往往是根据参照点而决定。^[7]由此出发阐释认知任务的探索-利用困境,基于认知主体面对收益与损失时风险偏好的不同,则认知主体对探索-利用问题上的决策并不直接由现实情景直接决定,而是由认知主体对当下情景的建构式的心理表征所决定。然而,前景理论也面临诸多挑战。一方面,前景理论认为多数认知主体对收益和损失的判断往往依据参照点决定,而这一参照点通常是由认知主体对于当下情境的心理表征所决定的,决策者的认知不透明性使其无法对偏好逆转现象和非传递性现象做出明确解释。([6], p.405)另一方面,尽管前景理论可以解释部分不确定

条件下的决策行为，但是认知主体内部的生成模型通常包含未来的预期信念，这也使得前景理论无法进一步解释有关认知主体具有前瞻性的未来决策规划。

通过以上分析不难看出，基于传统的决策理论对于认知主体探索-利用困境的解释仍存在较大程度的局限性。无论是期望效用理论抑或前景理论，尽管具有一定的解释效力，但在认知主体如何依据有限的感官经验在不确定性环境中对未来做出决策规划方面并未给出明确的分析。正因为如此，就需要另辟蹊径以一种新的方式对认知主体在模糊多变的环境中的探索-利用困境加以重新审视。

二、主动推理： 重解探索-利用困境的新视域

近十几年来，随着预测加工（predictive processing, PP）理论在认知科学及其相关哲学研究中的兴起，作为其核心理论构件之一的主动推理也日益受到学界的广泛关注。众多学者尝试基于主动推理的视域及其框架机制以对传统的科学与哲学难题加以再审视。由之，主动推理作为一种通用性过程理论，也为探索-利用困境的重解提供了全新的致思理路。

主动推理旨在从计算和神经层次上为有机体的感知、学习和决策提供一致性的理论阐释。“主动推理不仅仅是大脑通过感知推理减少其感官观察的不确定性。它更是涉及由具身性主体对世界进行主动的、有选择的采样。”^[8]其工作机制简单来说就是“最小化变分自由能（variational free energy）和预期自由能”。具体而言，传统的感知推理认为认知主体是被动地接受外界信息，通过持续学习并不断更新内部表征模型来最大程度地减少惊异（surprise）。而主动推理则强调，一方面，主体可以更新自己对世界潜在或隐藏状态的信念，以使其与观察到的证据相一致，从而最大程度地减少惊异；另一方面，具身主体能够积极做出规划并对外在世界采取必要的行动，以使观察与对世界状态的信念相一致，进而将惊异最小化，并将对

未来结果的不确定性降至最低。可以说，主动推理基于这样的一种假设，即大脑是一台预测性的贝叶斯推理引擎，执行感知和运动控制的闭环动力学，同时也在对世界的信念进行编码，使得认知具有概率性。^[9]显然，主动推理框架增补了一种认知主体在行动侧降低自由能的方式，从而弥补了先前理论在行动侧解释上的空缺，这也与实用主义“行动导向”的核心原则相对应，并在一定程度上推进了当前认知科学中关于感知-行动协同耦合方面的研究。

弗里斯顿认为，主动推理是有机体关于感知、行动、计划及决策等认知功能的基础规范性原则，他假设这些认知过程都可以被视为优化两个互补的目标函数，即变分自由能和预期自由能。其中，变分自由能代表认知主体对过去或现在惊异程度的上限，即“衡量内部生成模型与（当前和过去）观察之间的契合度”；（[10]，p.38）预期自由能则代表认知主体对未来结果惊异程度的上限，“衡量行动策略相对于更优的（未来）状态和观察的合理性”。（[10]，p.38）此外，变分自由能最小化侧重从感知侧降低有机体的自由能，它假使有机体的观测结果保持不变，通过改变大脑中的预测值以期契合外在环境的真实状态。而预期自由能最小化侧重从行动侧降低有机体的自由能，它假使有机体有意图地采取实际行动来改变所处环境，从而获得与大脑中预测相符合的未来观察值从而使自身内环境回归稳态。^[11]这里的预期自由能可以归结为以下三种表示形式：其一，通过把信息论中的信息熵、自由能与统计热力学中的热力熵、自由能概念相关联，将预期自由能表示为预期能量与熵的相减之差。其二，将预期自由能表示为预期模糊性（expected ambiguity）与风险（risk）的相加之和，其中，预期模糊性是似然型概率的预期熵（expected entropy），反映的是关于外部世界预期的不确定性；风险是关于观察的预期与先验观察概率之间的KL散度（Kullback-Leibler Divergence）。其三，将预期自由能表示为认知价值（epistemic value）和实用价值（pragmatic value）相加之和的负数，其中，认知价值对应

于信息增益;实用价值对应于效用值。([10], pp.33-37)可以说,主动推理在对有机体行为阐释过程中,也是运用上述三种数学公式从而将抽象的问题具体化。

依据主动推理,有机体复杂的规划、决策、行动等高级认知功能都可以转化成一种贝叶斯推理问题。在推理方面,如果说基于变分自由能的推理方式可被看成是一种“回顾性”推理的话,那么基于预期自由能的推理方式则可被看成是一种“前瞻性”推理,它可直接呈现出认知主体基于预期的未来数据的推断。也就是说,“预期自由能将主动推理扩展至包括一种典型的前瞻性认知形式:规划”。([10], p.31)这可以说是对认知主体降低自由能的另一种补充,与未来规划和决策紧密相关。在大多数不确定事件中,有机体首先规划一系列行动,通过这种预演行为来减少未来事件的随机性与不确定性。例如,为了赢得一局我方先手的象棋比赛,我们需要规划一系列行动方案,一方面我们要考量合理的行棋方式,每个可能的行棋方式被称为策略;而另一方面我们需要考量未来行棋后所观测到的结果,可能的行棋方式后的结果包括我方行棋后观测到对手“拱卒”或者“飞象”等情况,然后根据观测到的结果进一步更新先前的策略。其中,策略是每一步行棋的先前假设,行动是改变棋局的行棋方式,未来可能结果是一种可以预测的概率分布。如此一来,具有规划的有机体更有利于做出合理化决策。由此可见,“主动推理将规划和决策制定视为一个推断要做什么的过程。这将计划牢固地纳入贝叶斯推理的领域,意味着我们必须像以前一样指定先验和似然性”。([10], p.31)

进一步讲,主动推理有效地拓展了关于有机体决策和行为的解释方式,将对有机体的复杂行为方式的分析转换为近似贝叶斯推理问题。由此,自然界中具备这种推理机制的有机体在与外在环境的交互中,通过蕴含于主动推理的感知-行动循环运作方式,依据近似贝叶斯推理来推断外界环境隐藏的潜在状态,进而做出适应性决策。这种推理机制将感知与行动

两者同时进行优化,从而使有机体产生鲁棒性较强的决策模型。因此,主动推理这种隐式的贝叶斯推理规则解释了感知、学习和决策中适当的行为选择。其中,“感知被建模为后验状态估计的形成,这是通过结合先验和似然性来实现的,也就是说,主体对状态基本发生率和给定各类状态时做出种种观察的可能性的先验信念;学习被视为先验、似然性和其他模型参数的更新过程;而决策和行动被视为比较各种模型的证据的过程,这些模型是关于未来的状态和在可能的行动方案下观察到的情况”。([12], p.4)

综上可知,主动推理汲取了实用主义与贝叶斯主义的思想基础,一方面,在感知-行动循环架构中,感知与行动处于逻辑上平行的位置,其最终目标皆为实现自由能最小化,即认知过程遵循“实际目标为导向”这一核心原则;另一方面,基于变分自由能和预期自由能的形式,通过贝叶斯信念函数,可对有机体随着新的证据而不断更新初始信念的模式加以刻画。由此可见,在主动推理的视域下,认知主体的决策行为不能被置于独立的认知模块中进行理解,而是应该在一个感知-决策-行动协同的统一框架中加以把握。总而言之,神经科学领域提出的主动推理框架阐明了有机体的感知与行动如何耦合产生贝叶斯最佳行为,进而可以适用于对有机体的感知、决策和行动模型的分析。探索-利用困境无疑是一种普遍存在于有机体之中的两难决策难题,在此意义上讲,主动推理的逻辑框架也同样可对有机体在不确定性环境中的探索-利用决策行为做出恰切的说明。

三、基于主动推理的探索-利用认知架构

按照弗里斯顿的看法,有机体的决策行为与机器学习中的强化学习认知机制具有相似之处,即都是在与外界环境交互中,通过感知、学习、行动等策略以达成最佳回报,从而实现探索-利用权衡的最优策略。但二者不同之处

在于，强化学习主要强调人工智能体通过试错学习来获取某个行为的具有延迟性的外在奖励回报信号，而生物有机体并没有这种外在的回报信号。这也就引发了两方面的疑问：其一，有机体在不确定性环境中是如何表征探索-利用行为的；其二，应该如何评估探索和利用两者的优先性问题。而正是在对这两个问题的回答中，基于主动推理的探索-利用认知架构也清晰地呈现出来。

就第一个问题而言，弗里斯顿等学者尝试从价值论视角出发阐明有机体关于探索-利用行为的表征，他们将预期自由能形式化为认知价值（或内在价值（intrinsic value））和实用价值（或外在价值（extrinsic value））相加之和的负数。一方面，探索行为映射在认知价值（内在价值）中，认知价值在广义上通常被理解为知识或信念所拥有的价值，而在主动推理框架下，认知价值则可被表示为有机体探寻新奇行为的内在驱动力，同时认知价值也可以量化有机体行为的相对贡献值，这在形式上近似等同于强化学习中的外在奖励回报；另一方面，利用行为映射在实用价值（外在价值）中，这里的实用价值代表后验预测分布下预期结果的效用，“正是这种效用对偏好结果进行了编码，从而赋予行为具有目标导向的性质”。（[13], p.192）进一步讲，探索行为对应于主体的认知策略，与预期模糊性（expected ambiguity）相关联，“探索”可以帮助主体充分了解其状态空间（state space）从而降低未来预期的模糊性，这里的状态空间主要是指外在环境中难以观察的多变量因素；而利用行为对应于主体的习惯性策略，与风险（risk）相关联，“利用”可通过评估所有潜在的可能方案从而帮助主体找到效用值最大的备择选项。主动推理将认知价值和实用价值纳入在同一目标函数框架中，使得有机体可以自适应地通过预期自由能最小化来平衡探索-利用。

至于第二个问题，根据主动推理，最小化预期自由能等同于最大化信息增益（认知价值）与最大化效用价值（实用价值）。一般而言，有机体首先会执行具有认知价值的行为来

解决不确定性；之后，当认知价值被最大化到没有进一步的信息增益时，有机体则会根据先前的偏好或目标定义以最大化外在价值来选择策略。换言之，有机体可依据自身经验和外在环境的变化调整探索和利用策略，先通过探索行为降低预期模糊性，进而再选择利用行为降低预期风险。展开来讲，首先，为了保持内环境的稳定进而更好地应对复杂多变的外在环境状态，适应性有机体通常会在实际行动之前事先在心理空间中将所有潜在可能的行动方案进行预演。也就是说，有机体会预先构思或想象：“如果我存在于下一种状态下，那么我将呈现出特定类别的属性……”。^[14]其次，与人工智能体的强化学习算法依赖于设计者添设的外在奖励机制所不同的是，生物有机体的认知模式是基于知识驱动的推理方式而提高内在的认知价值，它趋向于一种自然化的自监督目标，即通过探索行为降低心理空间的认知不确定性。再后，当有机体内在的认知不确定性很低时，抑或说当其可以精确表征周围环境时，这就意味着不需要进一步获取新的价值信息，此时有机体则会依据先前的经验偏好选择期望效用最大化的利用决策。由此可见，有机体的奖励机制并非来自外部世界的反馈信号，而是来自内隐的认知驱动力的需要。总而言之，其最佳决策行为并非来自世界的真实状态，而是来自自身的信念驱动。

弗里斯顿等人进一步以小鼠T型迷宫觅食实验（[13], pp.197-208）和蚁群T型迷宫觅食实验，（[12], pp.1-11）对主动推理之于有机体在不确定性的场景中探索-利用现象的解释效力加以具体阐明。一方面，通过T型迷宫中小鼠的觅食决策行为，对有机体面临探索-利用困境时如何建立内部表征与假设做出了说明。该实验结果表明，小鼠会优先选择迷宫下臂的探索认知策略以获得对食物更多的线索信息；之后，小鼠会依据之前选择的结果更新信念，当小鼠对T型迷宫布局和环境的信心不断增强时，它便会选择左上臂或者右上臂的实用策略直接访问奖励位置。另一方面，由于蚁群觅食行为是典型的基于群体智能的分布式认知

系统(不同于基于个体智能的小鼠觅食行为),研究者们主要通过使用基于主动推理的马尔可夫模型来探讨群体中生存觅食的个体决策过程。该模拟实验表明蚁群觅食也面临着一种探索-利用权衡,蚁群对周围信息素的感知是一种探索行为,根据感知的方向来采取行动是一种利用行为,蚁群会在初始阶段去探索周围信息素来获取食物线索,当蚁群积累足够的信息素(即认知价值较高)时,他们便会根据线索采取行动从而获取目标食物。显然,这两个实验均表明,认知主体可通过预期自由能最小化从而自适应地平衡探索和利用行为。

总的来说,基于主动推理对探索-利用困境进行解释,实则是运用预期自由能的数学公式将序贯决策问题具体化。其背后的思想是:认知主体通过持续减少预期的不确定性来维持其自由能处于较低的水平。有机体的行为本质可以从推理的角度来理解,在这里,感知和行动是同一认知过程中不可分割的组成部分,通过探索以增强感知,通过利用来进行行动,二者共同遵循预期自由能最小化原则,进而可以对有机体的探索和利用加以耦合。总之,在主动推理框架下,可以将探索-利用困境的序贯决策问题转化为单一目标函数的优化,即预期自由能最小化。显然,这种规范性的处理方法可将探索与利用这两种决策方式置于一个平等的基底上,这样一来,认知主体可以依据外在环境的改变局部切换认知价值和实用价值从而平衡探索与利用的选项。

结 语

综上所述,基于主动推理框架重新审思有机体的探索-利用困境,可通过感知-行动协同的耦合架构将探索与利用加以有效整合,并将其纳入同一框架中进行表征。面对探索-利用困境,有机体必须在探索具有风险的新机会与利用熟悉的过往选项之间做出抉择,而如何对二者进行相对精确的量化成为解决该困境的一个关键点。在主动推理框架下,探索与利用这两种对立的要素便可以相对精确地评估,

最小化预期自由能这一目标使得我们能够以相同的通用形式重新定义探索与利用。在此,“策略的预期自由能由两部分组成,一部分量化了策略预期解决不确定性的程度(探索),另一部分量化了预测结果与主体目标相符合的程度(利用)”。([10], p.10)因此,主动推理解释进路下的策略选择恰当地阐明了有机体如何自适应地平衡了探索与利用两种决策方式,并揭示出两者之间的深层内在关联。

此外,主动推理突显了认知主体进行决策时信念的核心作用。传统的贝尔曼方程(Bellman Equation)可以以迭代方式求解关于认知主体行为的价值函数(value function),但其无法表征基于信念的动态规划问题,因而仅适用于无信念问题域的求解。而有机体的决策行为实际上就是在不断优化自己关于外部世界的信念。因此,“主动推理以预期(变分)自由能的形式,用(贝叶斯)信念的泛函替代了价值函数”。([15], p.713)在主动推理解释框架下,可以较好地利用预期自由能的递归形式,来补充主体自身对状态的信念的评估方案。由此可以认为,“主动推理通过将信息增益置于与奖励或价值相同的基础上,从而巧妙地解决了与先验偏好相关的探索-利用困境”。([15], p.713)

再者,主动推理明晰了认知主体在不确定性环境中的序贯决策方式。一般来讲,由于未来环境充满模糊性和风险,使得主体需要在不完全信息和不确定性因素下进行探索-利用决策。因此,能否恰当认识未来的不确定性因素对于自适应地平衡探索-利用具有重要意义。在主动推理框架下,以预期自由能最小化为核心便可以有效整合表征模糊性与风险,一方面,最小化预期模糊性旨在解决给定其原因的结果的条件不确定性;另一方面,最小化预期风险旨在解决相对于首选结果的预测结果的相对不确定性。这样一来,寻求信息和经验偏好被纳入同一个目标函数中,当主体的认知不确定性或模糊性消除时,进而根据经验偏好加以选择,由此便可在模糊多变的环境中做出最优决策。

总而言之,主动推理为探索-利用困境提

供了一种一元化的分析进路,在其解释框架下,“探索与利用之间的平衡是由预期自由能的最小化(预期惊异的上限)所驱动,并通过奖励和信息选项之间的局部选择从而得到解决”。^[16]换言之,主动推理方案的独特性就在于,它依据近似贝叶斯推理的方法基础,在一定程度上超越了感知与行动之间的界限,进而给出了一种较优的整合性立场。毋庸置疑,这种解释框架所蕴含的生物自然化的平衡策略,不仅有助于理解有机体决策行为的本质,进而可对行为心理学研究中经典的探索-利用难题给出更为恰当的说明。在更深远的意义上,也为进一步探讨更具可解释性的人工智能认知框架提供了一种可选择解释理路与方法向度。

[参考文献]

- [1] March, J. G. 'Exploration and Exploitation in Organizational Learning'[J]. *Organization Science*, 1991, 2(1): 71-87.
- [2] Toyokawa, W., Kim, H., Kameda, T. 'Human Collective Intelligence Under Dual Exploration-Exploitation Dilemmas'[J]. *PloS One*, 2014, 9(4): e95789.
- [3] Kolling, N., Behrens, T. E. J., Mars, R. B., et al. 'Neural Mechanisms of Foraging'[J]. *Science*, 2012, 336(6077): 95-98.
- [4] Domenech, P., Rheims, S., Koechlin, E. 'Neural Mechanisms Resolving Exploitation-Exploration Dilemmas in the Medial Prefrontal Cortex'[J]. *Science*, 2020, 369(6507): eabb0184.
- [5] Lu, J. Y., Xie, X. F., Xu, J. Z. 'Desirability or Feasibility: Self-other Decision-making Differences'[J]. *Personality and Social Psychology Bulletin*, 2013, 39(2): 144-155.
- [6] Smith, E. E., Kosslyn, S. M. *Cognitive Psychology: Mind and Brain*[M]. Harlow: Pearson Education Limited, 2014.
- [7] Kahneman, D., Tversky, A. 'Prospect Theory: An Analysis of Decision Under Risk'[J]. *Econometrica*, 1979, 47(2): 263-291.
- [8] Ramstead, M., Kirchhoff, M., Friston, K. 'A Tale of Two Densities: Active Inference is Enactive Inference'[J]. *Adaptive Behavior*, 2020, 28(4): 225-239.
- [9] Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., et al. 'Evidence for Surprise Minimization over Value Maximization in Choice Behavior'[J]. *Scientific Reports*, 2015, 5(1): 16575.
- [10] Parr, T., Pezzulo, G., Friston, K. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*[M]. Cambridge: The MIT Press, 2022.
- [11] Smith, R., Ramstead, M., Kiefer, A. 'Active Inference Models Do not Contradict Folk Psychology'[J]. *Synthese*, 2022, 200(2): 81.
- [12] Friedman, D. A., Tschantz, A., Ramstead, M., et al. 'Active Inference: An Active Inference Framework for Ant Colony Behavior'[J]. *Frontiers in Behavioral Neuroscience*, 2021, 15: 647732.
- [13] Friston, K., Rigoli, F., Ognibene, D., et al. 'Active Inference and Epistemic Value'[J]. *Cognitive Neuroscience*, 2015, 6(4): 187-214.
- [14] Pietarinen, A., Beni, M. D. 'Active Inference and Abduction'[J]. *Biosemiotics*, 2021, 14(2): 499-517.
- [15] Friston, K., Da Costa, F., Hafner, D., et al. 'Sophisticated Inference'[J]. *Neural Computation*, 2021, 33(3): 713-763.
- [16] Marković, D., Goschke, T., Kiebel, S. J. 'Meta-control of the Exploration-Exploitation Dilemma Emerges from Probabilistic Inference over a Hierarchy of Time Scales'[J]. *Cognitive, Affective, & Behavioral Neuroscience*, 2021, 21(3): 509-533.

[责任编辑 王巍 谭笑]