

# 有限理性合作观:破解人类合作之谜

刘永芳 王修欣

〔摘要〕 传统合作行为研究以生物学的进化论和经济学的理性经济人假设为前提,推演出“物竞天择,适者生存”和“纳什均衡”等结论,描绘了一幅暗淡的人类合作画面,而以此为基础形成的各式各样的关于个体选择、社会选择和集体行动的理论则成为公共管理和社会治理的基本指导思想。然而,这种观点无论在理论上还是在实践上,都遇到了难以自圆其说的困境,被称为“集体行动的困境”或“合作的或社会的困境”,亦被称为“人类合作之谜”。当代认知心理学的有限理性观及近年来在此基础上兴起的关于社会启发式的研究,为摆脱此种困境提供了启示。基于这些研究,我们提出了有限理性合作观,以还原人类合作行为的真实画面,破解人类合作之谜。

〔关键词〕 合作行为;理性;有限理性;社会启发式

当今人类社会发展遇到了前所未有的问题和困难,合作而不是竞争成为人类未来进程和前景的决定性因素。从人类与自然界关系的角度来看,全球范围内的气候变化、环境恶化、资源枯竭等问题已经到了唯有共同面对、通力合作、集体行动,才有可能找到有效的应对之策的境地;从人类内部不同层次和类型的错综复杂的关系角度来看,基于网络的全球化和扁平化时代的到来,使得人类活动的空间越来越大,参与竞争与合作的范围越来越广,却又越来越相互依存和依赖;国与国之间、不同利益集团之间、组织与组织之间、组织与个人之间、个人与个人之间的关系越来越紧密;经济、政治、文化、科技的全球一体化进程,使得没有一个国家、地区、组织或个人可以无视其他国家、地区、组织或个人而谋求独立发展之路。合作成了唯一和必然的选择。仅从我国发展的历史经验教训来看,闭关锁国的不合作政策带来的是落后和挨打,而改革开放的合作战略带来的是繁荣和昌盛。内部纷争动乱带来的是灾难,而团结合作带来的是进步。在一定意义上,可以说我国改革开放40余年来取得的举世瞩目的成就正是广泛参与国际合作和国内安定团结的结果。因此,研究人类合作行为问题,探索人类合作行为的规律,解开人类合作行为之谜,寻找有效的合作之道,无论对于应对全球范围内的人类发展问题,还是对于解决我国改革发展过程中遇到的现实问题,都具有刻不容缓的重大现实意义。本文拟在此现实背景下,分析传统合作行为研究遭遇的困境,阐述当代认知心理学的有限理性理论,并以此为基础提出有限理性合作观,以还原人类合作行为的真实画面,为破解“人类合作之谜”(the puzzle of human cooperation)提供启示。

---

刘永芳,心理学博士,华东师范大学心理与认知科学学院教授、博士生导师(上海 200062);王修欣,华东师范大学心理与认知科学学院博士研究生(上海 200062)。本文受国家社会科学基金重大项目(15ZDB121)的资助。

## 一、传统人类合作行为研究遭遇的困境

纵观人类合作行为研究的历史,主要沿着两条路线展开(刘永芳、王修欣,2018)。一条路线与人性假设有关,主要关心“人们愿不愿意合作?”的问题。传统生物学以生存和繁衍竞争为主线推演出来的“物竞天择,适者生存”的论断,对这个问题的回答是否定的。另一条路线与理性假设有关,主要关心“人们能不能够合作?”的问题。传统经济学以理性经济人假设为前提建立的纳什均衡(Nash equilibrium)及在此基础上发现的囚徒困境、公共物品困境、公共资源困境(Van Lange, Joireman, Parks & Van Dijk, 2013)等合作困境,对这个问题的回答也是否定的。二者从不同的角度共同描绘了一幅暗淡的人类发展或活动画面。然而,这幅画面与当今人类社会发展的现实格格不入,也受到了最新科学研究成果的严峻挑战。

### (一) 传统生物学遇到的挑战

传统生物学以达尔文在《物种起源》中提出的进化论为前提,推演出了一幅自私竞争的人类进化和发展画面。这种理论受到了两方面的挑战。从现实的角度看,它与人类发展的现实不符,更与人类社会未来发展的需要背道而驰。如果说人们都是带着自私基因和竞争倾向来到这个世界上的,为什么现实生活中我们会看到人们经常会付出代价或成本去做有利于他人(包括集体、社会)而无利于自己的事情?事实上,人类的文明史和发展史,在很大程度上就是一部人类不断合作的历史。人类社会能有今天的文明和进步,无不得益于合作,人类社会未来的发展更需要人类的通力合作。从科学研究的角度看,它受到了当今最新研究成果的挑战。2005年,借创刊125周年之机,《Science》邀请全球不同学科领域的顶级专家提出了125个“驱动基础科学研究以及决定未来科学研究方向”的科学难题,“人类合作行为如何演化”问题被确定为其中的25个“重大问题”之一(Kennedy & Norman, 2005)。此后10余年来,围绕这个问题,不同学科领域的科学家们开展了卓有成效的研究,取得了丰硕的成果,不少成果发表在《Nature》《Science》等国际顶级刊物上,掀起了人类合作行为研究的热潮。其中,以马丁·诺瓦克(Martin A. Nowak)为代表的研究团队取得的成果格外引人注目,其基本观点反映在《超级合作者》(Super Cooperators)一书(Nowak & Highfield, 2011)及近年来发表的一系列文章中。他们从生物学、人类学、社会学、经济学、政治学等不同的角度证明,合作是人类这种具有智慧的高等动物所具备的一种超级能力,既是人类文明进化和发展的产物,又是人类未来进步和发展的动力和机制,被称为“突变和选择”之外的“第三条进化原则”,并用直接互惠、间接互惠、空间选择、群体选择和亲缘选择等机制阐释了人类合作的原理(Rand & Nowak, 2013),为解开“人类合作之谜”奠定了坚实的基础,也为摆脱现有理论遭遇的困境指明了方向,成为当代社会科学关于人类合作行为研究的前沿问题之一。

### (二) 传统经济学遇到的挑战

传统经济学以亚当·斯密在《国富论》中提出的理性经济人假设为前提,假定人是自利的和精于计算的理性个体,并提出了“看不见的手”(invisible hand)的理论,认为当市场中每个人都理性地追逐个人利益时,客观上说,市场(或社会)整体的福利就会被提升(亚当·斯密,1972)。这种理论也受到了来自现实和理论研究的挑战。从现实的角度看,如果说人们都是精于算计的自利人,为什么即便有“搭便车者”从公共资源中无本获利,人们仍然自愿为公共资源贡献自己的一己之力?西方经济发展的现实在很大程度上也否定了其合理性,不断出现的经济危机让人们意识到市场这只“看不

见的手”不是万能的,单靠单个经济人的理性是无法让市场顺利运转的。纳什均衡(Nash equilibrium)及随后的一系列博弈论研究更是直接否定了亚当·斯密的结论。纳什均衡也叫非合作均衡,意思是指人们在不与他人合作上达成均衡,导致个体的理性利益选择总是与整体的理性利益选择背道而驰,其根源在于每个人都是充分理性的、精于计算的经济人。当市场中每个人都理性地追逐个人利益时,在客观上说,市场(或社会)整体的福利是无法达到最优的,这样就衍生出了各种各样所谓的“合作困境”(Van Lange, Joireman, Parks & Van Dijk, 2013)。

一种假设导致了两种截然不同的结论,而且这两种结论均在一定程度上不符合人类社会发展的现实。实际情况是,人类社会(包括经济、市场)的发展既不像亚当·斯密设想的那么好,又不像博弈论学者设想的那么差。合作行为领域的研究者们也认识到了这一点,并提出社会偏好(如叶航、陈叶奉、贾拥民,2013)、社会价值取向(Van Lange, 1999)等概念,以修正传统的经济人假设。社会偏好概念用于说明人们不仅考虑自身利益,还会考虑他人利益。社会价值取向则旨在描述这种对他人利益考量的个体差异。修正后的理论模型可以在一定程度上描述人类的合作现状,然而由于没有摒弃理性人假设,在描绘人类合作画面及指导实践上依然困难重重。比如,在描绘人类合作行为上,从亲缘选择(可在一定程度上视为亲属间的互惠)到直接互惠,从直接互惠到间接互惠,理论不断修正变更,却难以解释为什么人们会在单次博弈这种不存在互惠可能性的场景中依然选择合作。在指导实践上,根据理性人假设,施加惩罚或奖励会促进人们的合作行为,然而该推论却屡屡遭受挑战。比如,在许多研究中发现惩罚并没有发挥增进人们合作的作用(Wu, et al., 2009;连洪泉、周业安、左聪颖、陈叶烽、宋紫峰,2013),并且常常带来许多妨碍合作的副作用,如降低人际信任(Mulder, Dijk, Cremer & Wilke, 2006;王沛、陈莉,2011)。通过奖励增进合作的方式也存在类似的问题(Irwin, Mulder & Simpson, 2014)。

总体上看,传统的进化理论及以博弈分析为主线建立起来的各式各样的关于个体竞争、公共选择、集体行动的经济学理论,即便不是错误的,至少也遭遇了无法解释和解决当代社会诸多现实问题的困境。不是人类社会遭遇了这些理论模型所谓的“集体行动的困境”(the dilemma of collective action)或“合作的或社会的困境”(cooperative or social dilemma),而是这些理论和模型遭遇了无法解释和解决社会现实问题的困境。

## 二、有限理性理论:摆脱困境的新视角

如上所述,经济学家们围绕理性经济人假设争论不休,推演出了不同的结论,却没有人怀疑这种假设本身,而发端于当代认知心理学的有限理性革命则直接对这种假设提出了质疑。人们是完全理性的吗?在任何情况下都能够精确地计算出自己的利弊得失吗?从赫伯特·西蒙(Herbert Simon)到莱因哈德·泽尔腾(Reinhard Selten),从丹尼尔·卡尼曼(Daniel Kahneman)到理查德·塞勒(Richard Thaler),众多具有心理学背景的诺贝尔经济学奖获得者的研究对此给出了否定的答案。

传统经济学的理性人假设及在此基础上演绎出来的各式各样的标准化或规范化决策模型(von Neumann & Morgenstern, 1944)均假定,人们在决策时会经过理性计算并选择效用最大化的选项。然而这种假定不仅高估了人们的认知资源和计算能力,而且忽视了环境的不确定性。“有限理性”(bounded rationality)之父赫伯特·西蒙在提出这个概念伊始就一针见血地指出,人类自身仅有有限的计算能力,而且环境具有不确定性,这使得理性假设无法成立,效用最大化的幻想无法实现(Simon, 1947)。这里的环境不确定性不仅指事件发生概率的不确定性,还指可能的选项和选项导致结果的不确定性。在认知能力有限性和环境不确定性的双重限制下,人们无法像理性经济人那样遵循标准化或规范化的决策规则选择效用最大化选项,而是遵循描述性的启发式规则选择相对满意的选项(Simon,



1955; 1978)。

自从西蒙提出有限理性概念和观点以来,研究者们围绕这种观点开展了众多有意义的探索。在这个过程中,出现了两种对有限理性概念“异化”的理解(Gigerenzer, 2016)。一种理解来自于经济学,将有限理性视为受限制条件下的最大化。有限理性研究累积的大量证据使得新古典经济学家们不得不承认信息搜索和决策等心智成本的存在,但他们不愿放弃固守已久的最大化原则,而是将心智成本等限定纳入最大化原则的范畴内加以考量。也就是说,在他们看来,有限理性充其量不过是在原来的最大化模型中增加了一个参数而已,人们会充分权衡效用和心智成本,当继续投入心智成本不能使效用增加时,人们便会放弃信息搜索。这种理解和做法似乎承认和遵循了有限理性原则,实际上却忽视了概率、选项或选项导致的结果的不确定性,也为有限的认知能力增加了额外的负担。事实上,西蒙早已指出,有限理性不是纯理性或寻求效用最大化的理性(Simon, 1947)。有限理性研究应该抛弃效用最大化的幻想,关注人们是如何在无法做到效用最大化的限定下进行决策的(Simon, 1990, p.8)。

另一种“异化”的理解来自于心理学。2002年诺贝尔经济学奖获得者卡尼曼及其同事在有限理性假设前提下深入探讨了人们的启发式决策过程,发现人们具有重失轻得的倾向,很容易受问题表述方式的影响而忽视其本质,许多时候人们会采用可得性(availability)、代表性(representation)和锚定(anchoring)启发式(而非分析性的规则)做出决策(Kahneman, Slovic & Tversky, 1982; Gilovich, Griffin & Kahneman, 2002)。但是当他们对启发式决策进行评价或价值判断时,仍然抱着效用最大化的梦想,用这种理想化的标准来评判有限理性,认为所有这些启发式都是背离效用最大化原则的认知偏差或谬误。在这种语境下,有限理性常常被等同于非理性。然而,这种以虚无的效用最大化原则作为评判标准的做法实际上是对西蒙有限理性概念的一种曲解。西蒙早就指出,有限理性不是非理性(Simon, 1985, p.297)。理性的评判标准不应该是虚无缥缈的最大化原则,而是对环境的适应性。他曾用“一把剪刀”的隐喻来说明人类认知能力与环境之间的关系,只有当两个刀片协同活动时,才能裁剪出东西(Simon, 1990, p.7)。在此基础上,吉仁泽等提出了“生态理性”(ecological rationality)的概念,认为人们做出决策时是否理性及做出的决策是否合理,应该依据其与现实环境的拟合程度来评判(Goldstein & Gigerenzer, 2002)。吉仁泽等同时指出,卡尼曼等的研究多基于假想的场景,而若将他们视为偏差或谬误的启发式放入现实生活场景,则可能产生其独特的适应价值。吉仁泽团队的一系列研究为该观点提供了可靠的证据,甚至在有些环境下,启发式决策的结果远胜于基于效用最大化原则的决策规则(吉戈伦尔、托德及ABC研究组, 2002; 吉仁泽, 2006)。

尽管还存在一些争议,但这些研究者的一个基本的结论是:人类的理性是有限的,在现实的环境下还会受到诸多客观条件的限制,因而无法在任何情况下都精确地计算出自己的利弊得失。人们是按照有限理性的启发式规则追求满意性目标的,而不是按照分析性规则追求最优化目标的。这些观点和研究证据从根本上动摇了以经济学为代表的传统社会科学中流行的“理性人假设”。用这种全新的观点和相关的发现重新审视和解读人类合作行为,有助于还原人类合作行为的真实画面,破解现有合作理论遭遇的困局。

### 三、有限理性合作观

有限理性合作观是将上述有限理性思想及相关研究成果引入和延伸至人类合作行为问题的一种尝试。它把人类合作行为理解为在有限理性前提下处理自我—他人(包括集体、社会)关系的一种决策过程,试图提出和建立合作启发式的理论体系和模型,并依据所建立的理论模型对现有的合作困境进行解析。具体而言,有限理性合作观包含以下三个基本的论断:

### （一）合作决策是有限理性的启发式过程

对于这个论断,可以从以下几个方面来论证。首先,虽然有限理性概念原本并非直接针对人类合作行为问题而提出来的,但其包含的观点和思想在很大程度上更适用于理解人类合作行为。这是因为,人类决策绕不开得和失、现在和未来、我和他三大变量框架。在有限理性思想指导下,心理学家们关于得和失、现在和未来进行权衡时的心理活动累积的文献较多(如前景理论、框架效应、参照点理论、跨期选择理论等),而关于我和他权衡时的心理活动的研究相对薄弱。然而,相对于得和失、现在和未来两大变量而言,我和他变量涉及的信息量更大,不确定性更高,且更容易引起情绪、动机等社会心理状态的变化,研究中更难以量化和操纵相关变量,却又更加重要和关键。如果说得和失、现在和未来的权衡主要考验人的理性的话,我和他的权衡则不仅考验理性,而且考验人性。因此,如果说人们在面对和处理得和失、现在和未来这样的不涉及我和他的纯粹经济问题时且理性有限的话,那么当面对和处理涉及我和他的更为复杂的社会问题时,其理性就更为有限了。而人类合作行为问题的本质恰恰就在于如何理解和权衡我和他之间复杂的利害关系。

其次,现有的关于人类有限理性决策过程的研究成果适用于理解人类合作行为,有些发现甚至可以直接推论和迁移到合作行为问题上。西蒙提出的满意性原则及抱负水平设定规则可以被用来解释人们在谈判等现实博弈过程中的行为表现。卡尼曼和特沃斯基发现的预期理论、框架效应、代表性启发式、可得性启发式、锚定启发式等(Kahneman, et al., 1982; Gilovich, et al., 2002),吉仁泽等提出的再认启发式、单一理由启发式、模仿(做他人所做)启发式等快速节俭决策规则(吉戈伦尔、托德及ABC研究组, 2002)及李纾团队提出的齐当别模型(李纾, 2016),均适用于说明人类合作决策过程中的策略选择行为。近期的一系列研究即基于框架效应,深入探讨了得失框架对合作行为的影响(de Heus, Hoogervorst & van Dijk, 2010; Evans & van Beest, 2017)。

第三,以往及近期的合作行为研究也累积了不少证据,表明合作决策是一个有限理性的启发式过程。例如,早期博弈论研究中发现的“一报还一报”(tit for tat)和“赢保持,输改变”(win-stay, lose-shift)策略就是典型的启发式规则(Nowak & Sigmund, 1993)。后来的研究表明,简单的强化学习模型可以解释重复博弈过程中人们的行为、人们对他人信誉的估算和欺骗的觉察是一个基于少量参数的启发式过程(吉仁泽、泽尔腾, 2016)。近期合作行为研究者又提出两种可能的启发式规则:基于人们以往社会互动经验的社会启发式(socialheuristic hypothesis)(Rand, Peysakhovich & Kraft-Todd, et al., 2014)和基于对他人行为预期的社会投射(social projection)(Krueger, DiDonato & Freestone, 2012)。此外,最近不少研究者分别从宏观或微观、客观和主观角度证明社会网络(Apicella, Marlowe, Fowler & Christakis, 2012)、群体规模(Barcelo & Capraro, 2015)、人口流动(Miller & Knowles, 2015)、社会处罚(Ito & Yoshimura, 2015)、接触大自然(Zelenski, Dopko & Capaldi, 2015)、吐槽(Feinberg, Willer & Schultz, 2014)、分享疼痛经验(Bastian, Jetten & Ferris, 2014)和敬畏感(Piff, Dietze & Feinberd, et al., 2015)等都会影响人们的合作行为,表明人们不是绝对理性的,至少说明在合作问题上人们的理性没有经济学家们设想的那么强大。

### （二）有限理性启发式让人更合作,但也存在个体差异和情境变异

如果人们的理性是有限的(或并非那么强大),那么传统经济学基于绝对理性人假设推导出来的纳什均衡就是站不住脚的,而基于纳什均衡演绎出来的各式各样的合作困境也就不攻自破了。有限理性假设描绘的人类合作画面要远比现有经济学描绘的画面复杂得多,基于此种假设至少可以做出两个推论:(1)相对于纳什均衡而言,人类合作行为的均衡点会向合作一端偏移;(2)相对于纳什均衡而言,人类合作行为的均衡点会呈现更加多样化的态势。

上述两个推论得到了近期研究证据的支持。兰德(Rand)及其同事通过施加时间压力或增加认知负荷来启动人们的启发式决策过程,以考察人们在无法展开理性思维条件下,会做出何种选择。结果发现,相对于没有时间压力或认知负荷的对照组而言,实验组被试在公共物品或公共资源博弈中均倾向于更合作(Rand, Peysakhovich & Kraft-Todd, et al., 2014; Rand, 2016; Everett, Ingbretsen, Cushman & Cikara, 2017)。他们提出了社会启发式假设来解释这些结果。该假设强调以往社会互动经验的作用,认为人们的启发式是以往社会互动中优势策略的沿用(Rand, Peysakhovich & Kraft-Todd, et al., 2014)。在刚开始面对合作决策时,人们也许会适当调用认知资源,进行信息搜索和认知加工,以做出决策。在随后的类似决策场景中,人们就会沿用之前的决策,以更为快速、直接地做出反应。社会启发式假设进一步指出,在现实生活中,由于人们与他人的合作往往不是单次的,而是重复博弈的过程,并且存在惩罚、奖励和声誉等有助于维持合作行为的机制,所以一般而言人们会做出合作的决定,并进而培养出合作启发式,即启发式会导致合作。社会投射假设也可以解释普遍存在的合作行为。该假设认为,人们相信他人会像自己一样行动,并将这种对他人行为的预期作为合作决策的重要线索(Krueger, et al., 2012)。像社会启发式一样,由于存在奖励、惩罚、社会规范、声誉等促进合作的机制,所以多数人相信他人会像自己一样合作,所以社会投射往往也会促进合作。

然而,后来的研究发现,有限理性启发式过程导致合作的结论可能存在个体差异性和环境变异性。所谓个体差异性,是指个体经历或个人特质及其带来后果的差异性。所谓环境变异性,是指博弈任务或决策情境的多样性。社会启发式假设认为,不同的个体或者个体在不同的情境下可能有不同的启发式,从而导致合作行为上的差异(Rand, 2016)。这些推论得到了研究证据的支持。在个体差异上,研究表明对于有过不合作受益或合作受损经历或经验的个体,时间压力不会促进合作,甚而导致不合作(Capraro & Cococcioni, 2015);时间压力会让利己的人变得更利己,让利他的人变得更利他(Chen & Krajchich, 2018);低依恋回避个体可能在之前的社会互动中有着更多的良性互动经验,我们最近的研究发现,时间压力和认知负荷促进了低依恋回避个体的合作行为,而没有改善高依恋回避个体的合作行为(Gao, Jia, Liu, Wang & Liu, 2019)。在环境变异上,对于具有相关经验的被试而言,在一次性博弈条件下,启发式可能导致人们不合作,而在多次重复性博弈条件下,启发式可能导致合作(Rand, 2016)。社会投射假设则认为不同的个体或者个体在不同的情境下可能有着不同的社会投射,从而导致合作行为上的差异。比如,亲社会个体倾向认为大多数人会像他们一样选择合作,而亲自我个体则倾向认为大多数人会像他们一样选择背叛(Dawes, McTavish & Shacklee, 1977; Messe & Sivacek, 1979),这种投射差异可能导致他们在合作行为上的差异(Krueger, et al., 2012)。社会投射也会随着个体与他人社会距离的增加而减少,从而使得人们与社会距离较远的人互动时表现出更少的合作行为(Robbins & Krueger, 2005; Krueger, Ullrich & Chen, 2016)。

### (三)有限理性导致的合作行为具有一定的社会适应性

有限理性启发式整体上让人更合作,至少不像传统经济学推导的那样难以合作。那么,有限理性启发式是否会因为不够理性而致人于不利境地或给人带来损失呢?答案是否定的。恰恰相反,基于有限理性启发式的决策结果往往增加了人们的社会适应性。

从决策过程上看,一方面有限理性启发式让合作决策变得更容易、更简单,大大节约了个体的心智成本。这里的心智成本指个体运用理性深思熟虑的“计算”成本,主要包括:一是信息成本,包括信息的收集、处理;二是认知协调成本,即在原有和现存两种不协调的认知之间做出调整,使二者相互适应,达到统一;三是决策成本,即在所有的备选项之间权衡利弊得失,最终做出选择。此外,在整个心智成本之中,都包含有时间成本,个体进行理性决策的每个环节所花费时间的机会成本也要计算在



内。人们的决策是耗费时间和心智成本的。在纷繁复杂的世界中,如果人们像理性经济人假定的那样进行决策,势必会耗费巨大的时间成本和认知资源。而启发式过程可以让个体在较短时间内很快地应对类似的情境或决策,以把更多的心智成本用到新的情境或决策问题上去。另一方面,有限理性启发式具有一定的自我调节性,增加了合作决策的灵活性。正像社会启发式假设指出的那样,现实生活中人们通常面对的是重复博弈,存在着惩罚和奖励、社会规范、风俗习惯、声誉传播等维持和促进合作的机制,使得合作成为一种优势策略。人们逐渐将这种优势策略内化为社会启发式,以应对多样化的决策场景。当以往的社会启发式不再具有适应意义时,人们会自动调整和变更启发式规则,以更好地适应当前的情境。研究者采用重复囚徒困境任务创设了两种情境,一种情境下合作是优势策略,另一种情境下合作是劣势策略,在随后的单次博弈中,来自前一种情境的被试表现得更为亲社会,更愿意惩罚自私者,有更高的信任水平 (Peysakhovich & Rand, 2015)。这表明人们会根据结果反馈这种简单的信息随时调整自己的行为。社会投射也具有这种自我调节能力,当我们根据社会投射选择合作,而对方选择背叛时,我们会实时调整自己的社会投射和合作行为;在面对社会距离较远的个体时,人们的社会投射会减少,进而表现出较少的合作 (Krueger, et al., 2016)。

从决策的结果上看,大量的研究表明,费尽心机深思熟虑做出的合作决策未必带来满意的结果,且常常带来糟糕的结果,如博弈论中发现的“赢家的诅咒”(the winner's curse) (泰勒, 2018)等现象就是明证,正所谓“机关算尽太聪明,反害了卿卿性命”。相反,采用简单启发式做出的合作决策常常会带来满意的结果,如前述的“一报还一报”策略就是典型的例证。启发式增进合作,而合作行为往往可以让个体避免惩罚、获得奖励或维护良好的声誉等,从而获得长远的利益。此外,正如最优悖论 (Maximization Paradox) 所揭示的那样,追求效用最大化目标的决策者 (maximizers) 对决策结果的满意度反而低于追求满意性目标的决策者 (satisficers) (Dar-Nimrod, Rawn, Lehman & Schwartz, 2009; Luan & Li, 2017)。所有这些都表明,基于有限理性的启发式的决策结果是具有社会适应性的。值得注意的是,这并非意味着启发式总是最佳的选择。兰德等人发现,社会启发式具有一定的溢出效应 (spillover effect),即在典型情境里习得的启发式常常延续到不太类似的非典型情境里,从而导致失误或损失 (Rand, Peysakhovich & Kraft-Todd, et al., 2014)。例如,在单次博弈中依然采用在重复博弈中习得的合作策略、不区分合作对象滥用社会投射等,都有可能使得个体被搭便车。

#### 四、总结与展望

理性经济人假设就像套在合作行为研究者脖子上的枷锁一样,让他们无法发挥想象力,揭示人类合作行为的真相。有限理性合作观以认知革命中的有限理性观点为依托,挣脱了这个枷锁,也还原了人类合作行为的真实画面。具体地,从决策过程上来看,有限理性合作观强调合作决策是有限理性下的启发式过程;从启发式过程导致的结果来看,有限理性合作观认为启发式会导致合作,但是有着个体差异和环境变异,这样复杂而精彩纷呈的画面才是对人类合作行为的客观描述;从价值判断上来看,有限理性合作观认为启发式过程及其导致的结果都具有一定的社会适应性。有限理性合作观对人类理性能力的看法并不十分乐观,却推演出了更为乐观的人类合作画面。它不仅有助于我们更好地理解人类合作行为的本质,摆脱现有合作研究遭遇的困境,而且有助于促进人们在现实生活中的合作行为。

从理性合作观到有限理性合作观的转变,也意味着合作行为研究范式从“标准化研究范式”向“描述性研究范式”的转变。理性合作观遵循标准化决策研究范式,以最大化偏好假设为基础,试图构建普适性的数学模型来预测人们的合作行为,但是由于其不切实际的理性人假定,构建的模型往往失效。有限理性合作观则遵循描述性研究范式,强调客观地认识合作决策的启发式过程及其结果,以期

真实地描述人们的合作行为。它不排斥建构数学模型,但其所推崇的模型应该是基于启发式规则的描述模型,而不是以最大化偏好假设为基础的数理模型。有限理性合作观还处于初创阶段,尚需要更加深入的研究加以丰富和发展。

首先,虽然将合作行为视为有限理性启发式过程的论断有相当的理论基础和间接的证据支持,但尚缺少更为直接的证据。未来研究可以综合采用行为实验和认知神经手段,深入探讨合作决策过程,为有限理性合作观提供更为直接的证据。此外,在不同的情境下启发式过程可能有不同的表现形式,未来研究或可对启发式的内容和形式进一步发展,以充实有限理性合作观,如公平启发式概念或许也适用于合作决策,即人们在进行合作决策时,往往会依赖公平与否这一简单的决策原则(Kriss, Loewenstein, Wang & Weber, 2011)。

其次,未来研究应该进一步关注启发式导致的结果,探讨启发式促进合作的个体或环境等调节变量。根据社会投射观点,人们认为他人会按照自己的方式行事,并以此作为自己合作决策的重要线索。然而就像上文提及的那样,社会投射可能存在一定的个体差异和情境变异,进而导致决策结果的差异,这需要未来研究进一步关注。社会启发式源自人们的经验,而人们的经验也会根据个人所处情境和个体差异而不同,这也会导致结果的变异性。如相对于信任缺失的社会,在信任程度较高的社会里,人们是否有着更为积极的社会启发式,进而会表现出更多的合作行为。未来研究可以进一步探讨可能的调节变量,以更为详细、具体地描述启发式导致的结果。

最后,应该关注有限理性合作观对现实的指导意义。构建人类合作行为的有限理性模型,并在此基础上分析现实的合作问题,对于个人参与消费、投资、社会交往,对于各级各类组织的管理、参与竞争与合作、解决冲突,对于政府的公共管理和政策制定,都具有启示和指导意义。特别是,如何规范和引导正面的集体行动、预防负面的集体行动是世界各国政府面临的一个非常现实而棘手的社会问题。虽然先前基于理性经济人假设和博弈论分析提出了解决此类问题的公共选择理论,但正如曼瑟尔·奥尔森(Olson)在《集体行动的逻辑》一书中指出的那样,他的集体行动理论较适合解释小规模群体的集体行动,而不适合解释大规模群体的集体行动(奥尔森,1995)。在当代群体活动的规模和范围越来越大的情况下,集体行动又会呈现什么样的特点和规律呢?更为重要的是,如果人们的理性本来就是有限的(加之人们并非总是那么自利的),无法在任何情况下都精确地计算出参与或不参与集体行动的利弊得失,又会怎样呢?实际上,现实生活中发生的众多的正面的或负面的集体行动并非都能够用现有的理论来解释。因此,从有限理性角度探讨人们集体行动中的合作或不合作行为问题,对于解释或解决人类合作的难题具有重要的现实社会意义。此外,基于传统理性观点,管理者主要采取奖励或惩罚方式,期望改变合作决策中的收益—损失结构,以促进合作。有限理性合作观强调人类合作决策的启发式本质,这为以“助推”方式(泰勒、桑斯坦,2015)促进人们的合作行为提供了可能性。未来研究可对此加以关注,或许可以开辟社会治理的第三条路线。

## 参考文献:

[德]哥德·吉戈伦尔、彼得·M·托德,2002:《简捷启发式—让我们更精明》,刘永芳译,上海:华东师范大学出版社。

[德]吉仁泽,2006:《适应性思维—现实世界中的理性》,刘永芳译,上海:上海教育出版社。

[德]吉仁泽、莱茵哈德·泽尔腾,2016:《有限理性:适应性工具箱》,刘永芳译,北京:清华大学出版社。

[美]理查德·泰勒,2018:《赢家的诅咒:经济生活中的悖论与反常现象》,高翠霞译,北京:中信出版集团。

[美]理查德·泰勒、卡斯·桑斯坦,2015:《助推:如何做出有关健康、财富与幸福的最佳决策》,刘宁译,北京:中信出版集团。



- [美]曼瑟尔·奥尔森,1995:《集体行动的逻辑》,陈郁、郭宇峰、李崇新译,上海:上海人民出版社。
- 李纾,2016:《决策心理:齐当别之道》,上海:华东师范大学出版社。
- 连洪泉、周业安、左聪颖、陈叶烽、宋紫峰,2013:《惩罚机制真能解决搭便车难题吗?——基于动态公共品实验的证据》,《管理世界》第4期。
- 刘永芳、王修欣,2018:《人类合作行为研究进展》,中国心理学会:《2016—2017心理学学科发展报告》,北京:中国科学技术协会出版社,第79—91页。
- 王沛、陈莉,2011:《惩罚和社会价值取向对公共物品两难中人际信任与合作行为的影响》,《心理学报》总第43期第1期。
- [英]亚当·斯密,1972:《国民财富的性质和原因的研究》,郭大力、王亚南译,北京:商务印书馆。
- 叶航、陈叶烽、贾拥民,2013:《超越经济人:人类的亲社会行为与社会偏好》,北京:高等教育出版社。
- Apicella, C., F. W. Marlowe, J. H. Fowler & N. A. Christakis, 2012, “Social networks and cooperation in hunter-gatherers”, *Letter*, vol.481, pp.497—502.
- Baecelo, H. & V. Capraro, 2015, “Group size effect on cooperation in one-shot social dilemmas”, *Scientific Reports*, vol.5, p. 7937.
- Bastian, B., J. Jetten & L. J. Ferris, 2014, “Pain as social glue: Shared pain increases cooperation”, *Psychological Science*, vol.25, pp.2079—2085.
- Capraro, V. & G. Cococcioni, 2015, “Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making”, *Proceedings of the Royal Society B: Biological Sciences*, vol.282, <https://doi.org/10.1098/rspb.2015.0237>.
- Chen F. D. & I. Krajbich, 2018, “Biased sequential sampling underlies the effects of time pressure and delay in social decision making”, *Nature Communications*, vol.9, p. 3557.
- Dar-Nimrod, I., C. D. Rawn, D. R. Lehman & B. Schwartz, 2009, “The maximization paradox: the costs of seeking alternatives”, *Personality and Individual Differences*, vol.46, pp.631—635.
- Dawes, R. M., J. McTavish & H. Shaklee, 1977, “Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation”, *Journal of Personality and Social Psychology*, vol.35, pp.1—11.
- De Heus, P., N. Hoogervorst & E. van Dijk, 2010, “Framing prisoners and chickens: Valence effects in the prisoner’s dilemma and the chicken game”, *Journal of Experimental Social Psychology*, vol.46, pp.736—742.
- Evans, A. M. & I. van Beest, 2017, “Gain-loss framing effects in dilemmas of trust and reciprocity”, *Journal of Experimental Social Psychology*, vol.73, pp.151—163.
- Everett, J. A. C., Z. Ingbretsen, F. Cushman & M. Cikara, 2017, “Deliberation erodes cooperative behavior—even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias”, *Journal of Experimental Social Psychology*, vol.73, pp.76—81.
- Feinberg, M., R. Willer & M. Schultz, 2014, “Gossip and ostracism promote cooperation in groups”, *Psychological Science*, vol.25, pp.656—664.
- Gao, Q. Y., X. Y. Jia, H. Y. Liu, X. X. Wang & Y. F. Liu, 2019, “Attachment Style Predicts Cooperation in Intuitive but Not Deliberative Response in One-shot Public Goods Game”, *International Journal of Psychology: Journal International De Psychologie*, DOI: 10.1002/ijop.12584.
- Gigerenzer, G., 2016, “Rationality without optimization: Bounded rationality”, in L. Macchi, M. Bagassi & R. Viale (eds.), *Cognitive Unconscious and Human Rationality*, Cambridge, MA: The MIT Press, pp.3—22.
- Gilovich, T., D. Griffin & D. Kahneman, 2002, *Heuristics and Biases: The Psychology of Intuitive Judgment*, New

York: Cambridge University Press.

Goldstein, D. G. & G. Gigerenzer, 2002, "Model of ecological rationality: The recognition heuristic" , *Psychological Review*, vol.109, pp.75—90.

Irwin, K., L. Mulder & B. Simpson, 2014, "The detrimental effects of sanctions on intragroup trust: Comparing punishments and rewards" , *Social Psychology Quarterly*, vol.77, pp.253—272.

Ito, H. & J. Yoshimura, 2015, "Social penalty promotes cooperation in a cooperative society" , *Scientific Reports*, vol.5, p.12797.

Kahneman, D., P. Slovic & A. Tversky, 1982, *Judgment under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Kennedy, D. & C. Norman, 2005, "What Don't We Know?" , *Science*, vol.309, pp.75—75.

Kriss, P. H., G. Loewenstein, X. H. Wang & R. A. Weber, 2011, "Behind the veil of ignorance: Self-serving bias in climate change negotiations" , *Judgment and Decision Making*, vol.6, pp.602—615.

Krueger, J. I., T. E. DiDonato & D. Freestone, 2012, "Social projection can solve social dilemmas" , *Psychological Inquiry*, vol.23, pp.1—27.

Krueger, J. I., J. Ullrich & L. J. Chen, 2016, "Expectations and decisions in the volunteer's dilemma: Effects of social distance and social projection" , *Frontiers in Psychology*, vol.7, article 1909.

Luan, M. & H. Li, 2017, "Maximization paradox: Result of believing in an objective best" , *Personality and Social Psychology Bulletin*, vol.43, pp.652—661.

Messe, L. A. & J. M. Sivacek, 1979, "Predictions of others' responses in a mixed-motive game: Self-justification or false consensus?" , *Journal of Personality and Social Psychology*, vol.37, pp.602—607.

Miller, S. & K. Joshua, 2015, "Population fluctuation promotes cooperation in networks" , *Scientific Reports*, vol.5, p. 11054.

Mulder, L. B., E. V. Dijk, D. D. Cremer & H. A. M. Wilke, 2006, "Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas" , *Journal of Experimental Social Psychology*, vol.42, pp.147—162.

Nowak, M. & K. Sigmund, 1993, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game" , *Nature*, vol.364, pp.56—58.

Nowak, M. A. & R. Highfield, 2011, *Super Cooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*, New York: Free Press.

Peysakhovich, A. & D. G. Rand, 2015, "Habits of virtue: Creating norms of cooperation and defection in the laboratory" , *Management Science*, vol.62, pp.631—647.

Piff, P. K., P. Dietze, M. Feinberg, D. M. Stancato & D. Keltner, 2015, "Awe, the small self, and prosocial behavior" , *Journal of Personality and Social Psychology*, vol.108, pp.883—899.

Rand, D. G., 2016, "Cooperation, fast and slow meta-analytic evidence for a theory of social heuristics and self-interested deliberation" , *Psychological Science*, vol.27, pp.1192—1206.

Rand, D. G., A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak & J. D. Greene, 2014, "Social heuristics shape intuitive cooperation" , *Nature Communications*, vol.5, article 3677.

Rand, D. G. & M. A. Nowak, 2013, "Human cooperation" , *Trends in Cognitive Sciences*, vol.17, pp.413—425.

Robbins, J. M. & J. I. Krueger, 2005, "Social projection to ingroups and outgroups: A review and meta-analysis" , *Personality and Social Psychology Review*, vol.9, pp.32—47.

Simon, H. A., 1947, *Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations*,

New York: Macmillan.

Simon, H. A., 1955, "A behavioral model of rational choice", *Quarterly Journal of Economics*, vol.69, pp.99—118.

Simon, H. A., 1978, "Rational decision-making in business organizations", *Nobel Prize in Economics Documents*, vol.69, pp.493—513.

Simon, H. A., 1985, "Human nature in politics: A dialogue of psychology with political science", *American Political Science Association*, vol.79, pp.293—304.

Simon, H. A., 1990, "Invariants of human behavior", *Annual Review of Psychology*, vol.41, pp.1—19.

Van Lange, P. A. M., 1999, "The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation", *Journal of Personality and Social Psychology*, vol.77, pp.337—349.

Van Lange, P. A. M., J. Joireman, C. D. Parks & E. Van Dijk, 2013, "The psychology of social dilemmas: A review", *Organizational Behavior and Human Decision Processes*, vol.120, pp.125—141.

Von Neumann, J. & O. Morgenstern, 1944, *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

Wu, J. J., B. Y. Zhang, Z. X. Zhou, Q. Q. He, X. D. Zheng, R. Cressman & Y. Tao, 2009, "Costly punishment does not always increase cooperation", *Proceedings of the National Academy of Sciences*, vol.106, pp.17448—17451.

Zelenski, J. M., R. L. Dopko & C. A. Capaldi, 2015, "Cooperation is in our nature: Nature exposure may promote cooperative and environmentally sustainable behavior", *Journal of Environmental Psychology*, vol.42, pp.24—31.

(责任编辑: 蒋永华)

## Cooperation Based on Bounded Rationality: A Perspective for Understanding Human Cooperative Behavior

LIU Yongfang, WANG Xiuxin

**Abstract:** The traditional studies on human cooperative behavior, which were premised upon the hypothesis of economic man in economics and the theory of evolution in biology, drew such conclusions as "survival of the fittest" and "the Nash equilibrium", thus painting a gloomy picture of human cooperation. The sundry theories based on such conclusions concerning the choices made by the individual and the society as well as the collective action have become the basic principle governing the public administration and social governance. However, this view faces insurmountable difficulties in both theory and practice, which are known as "dilemma of collective action", "cooperative or social dilemma" and "myth of human cooperation". The theory of bounded rationality in contemporary cognitive psychology and the social heuristics developed accordingly can well inform the settlement of such difficulties. On this basis, we propose a perspective on cooperation based on bounded rationality, from which the true nature of human cooperation can be revealed and the myth surrounding it exposed.

**Key words:** cooperative behavior; rationality; bounded rationality; social heuristic

**About the authors:** LIU Yongfang, PhD in Psychology, is Professor and PhD Supervisor at the School of Psychology and Cognitive Science, East China Normal University(Shanghai 200062); WANG Xiuxin, PhD Candidate at School of Psychology and Cognitive Science, East China Normal University(Shanghai 200062).