

偏好的争论、拓展与融合：理论假说与初步探讨

周小亮 笪贤流*

摘要：建立在偏好自涉、偏好外生给定、偏好同质且稳定不变的偏好假定基础上的主流经济学，加之对真实世界中个体行为解释的牵强与矛盾，从而面临着理论困惑与现实困境，偏好的拓展成为必须。偏好的拓展包含自涉偏好与他涉偏好的统一、偏好异质与内生对偏好同质与外生的扬弃、偏好稳定的特殊性 with 偏好演化的一般性三个方面，从而融合后的个体偏好是自涉偏好与他涉偏好的统一；是情景依存的，内生于社会制度之中；是异质的、演化的。

关键词：偏好 他涉偏好 偏好异质 偏好内生 偏好演化

一、引言

偏好，可以说是经济学中最核心最基础的概念之一。主流经济学偏好理论经历了基数效用论、序数效用论与显示偏好理论的三大发展阶段之后便走向了终结，对应于科斯革命之后蓬勃发展的厂商理论，偏好理论发展缓慢，陷入了停滞。不论是否因为实用主义原因，偏好理论在主流经济学史上这种不平衡状况是令人费解的。然而一个令人更加费解与困惑的问题是，主流经济学一直缺乏一个关于偏好本身的理论。

对偏好理论的漠视是经济学理论与实践的一个重大缺憾。在真实的世界中，例如个体行为的多样性与多变性、陌生个体间单次交往的合作策略与囚徒困境纳什均衡策略的背离、理性选择模型的理论困惑与现实困境以及普遍存在的利他行为等等，建立在自涉与稳定的偏好假定下的主流经济学在解释这些问题时显得力不从心；在实践中，以自涉与稳定的偏好假设为基础的相关经济政策的制定与实施不能达到预期效果甚至适得其反。我们认为，偏好的拓展与融合将有助于这一问题的解决。幸运的是，来自主流经济学外部的非主流经济学以及心理学、神经科学等社会科学的学者们的努力，进一步从各自不同的研究立场拓展了偏好理论。这些为偏好的拓展与融合提供了依据，从而加强了经济学对真实世界的解释力与指导力。

本文试图从主流经济学对偏好的漠视与简化处理开始阐释，然后综合其他科学的最新研究成果诠释偏好拓展的三条途径，从而提出偏好融合的一个理论假说，最后就偏好如何影响决策以及融合后的偏好的意义进行了展望。

二、主流经济学对偏好的漠视与简化

效用与偏好的发展史表明：随着序数效用论逐渐替代基数效用论，偏好也成为了效用的现代名词，完成了偏好对效用的初次不完全替代；当显示偏好理论日益成熟，使得现代主流经济学效用理论普遍使用偏好范畴来取代效用范畴，把效用仅仅看作偏好的一个函数，完成了偏好对效用的二次完全替代。纵观主流的偏

* 周小亮，福州大学经济管理研究所，邮政编码：350002，电子信箱：xlfzu@163.com；笪贤流，福州大学经济管理研究所，邮政编码：350002。

汪丁丁认为这种情况出现的一个很大原因是实用主义，参见汪丁丁、罗卫东、叶航：《偏好、效用与经济学基础范式的创新》，载《浙江社会科学》，2003（3）。

偏好对效用的两次替代及效用与偏好的发展史，我们已做过初步探讨，参见周小亮、笪贤流：《效用、偏好与制度关系的理论探讨——反思消费者选择理论偏好稳定之假设》，载《学术月刊》，2009（1）。

好理论,可以归纳为:(1)偏好关系,即满足自反性、完备性、传递性和连续性的假定。(2)工具性的效用函数理论。主流经济学为了使其理论便于数学处理,进一步定义了效用函数。在处理确定情形下的选择时,偏好的强度即效用被处理成序数变量;而在处理风险选择时,效用却被处理成基数变量。(3)偏好与选择范式。主流理论模糊了偏好与选择之间的区别与联系,显示偏好遵循“我选择,我偏好;我偏好,我选择”的逻辑,隐约地将偏好与选择同一起来,二者可以自然顺畅地进行过渡。事实上,偏好与选择之间存在着过渡问题。“偏好是种心理状态以及结果或前景,而选择指一个行动、一个决策或一种策略。只有通过决定论的、或然论的或者博弈论的方法,才能把对结果的偏好翻译成对策略的偏好。”因此,偏好可以进行选择也可以不进行选择;选择反映偏好也可以不反映偏好。(4)偏好外生、偏好同质、偏好自涉及偏好稳定。我们将看到,当前学术界正在兴起的研究大大拓展了这些偏好假定,为偏好的拓展与融合奠定了基础。(5)合作悖论。自涉偏好及其向选择的自然过渡,从而使得搭便车式的机会主义行为成为个体理性之选择,这解释不了现实社会中个体间的合作行为。然而,贯穿着人类社会以及社会性动物社会的历史的一类秩序——通常被称为合作——不论从行为学角度审视还是从伦理学角度审视,它对社会现象而言,都具有根本的意义(汪丁丁,2005)。因此,对社会何以可能,何种结构导致了合作这一社会科学根本问题的解答,主流经济学表现得力不从心。我们将看到,这一问题的解决思路,则正是基于偏好的拓展与融合。

三、偏好拓展的初步探讨

20世纪中期的边际主义大论战对经济学的研究与发展产生了巨大影响,这也为经济学者对偏好的漠视提供了辩护。自涉与稳定不变的偏好,成为主流经济学一个共鸣的假设,被反复地应用于经济理论与实践。

一方面,尽管主流经济学的偏好模型不排斥利他行为,但主流经济学家对利他行为却有意地进行忽略与漠视,将其排除在分析视野之外,诸如 Hirshleifer(1977)、Nakayama(1980)、Arrow(1982)、Lindbeck和 Weibull(1988)等,坚持认为利他主义对经济学是一个多余的假设,它的存在可能导致经济活动的帕累托无效率,从而拒绝利他偏好的存在。直到实验经济学的一系列简洁而有说服力的实验以及社会偏好(Social preference)理论的诞生,才引起主流经济学对他涉偏好的关注。另一方面,通过技术性处理维护偏好的外生给定、稳定不变假设的合理性,这已越来越令人感到怀疑和不安。后面将看到,个体偏好是内生的、演化的。因此,建立在纯粹自涉与稳定不变的偏好理论基础上的主流经济学,尽管对社会中的部分个体行为能够进行比较可行的分析,然而,越来越多的经济现实使得主流理论的解释面临困境,迫使经济学从自身寻找出路。本文综合近些年来西方学界的研究成果,尝试对偏好进行拓展与融合。

(一)自涉偏好与他涉偏好的统一

本文认为,个体的偏好并非如经济人或理性人那种完全意义上的自利,而是既包含自涉偏好,又包含他涉偏好即自涉偏好与他涉偏好共存。我们将看到,利他偏好在行为与实验经济学、神经元经济学(Neuroeconomics)等领域有了令人兴奋的发展。

1. 行为与实验经济学对他涉偏好的验证

综观当前对他涉偏好的研究情况,我们可以归结为三类:互惠偏好、公平偏好与利他偏好。实验经济学者通过一些设计的博弈实验从而对主流经济学的偏好理论进行强有力的质疑与反驳。根据 Camerer和 Fehr(2001)的划分,这些博弈实验为囚徒困境博弈、公共品博弈、最后通牒博弈、信任博弈、独裁者博弈、礼物交换博弈以及第三方惩罚博弈实验七种,其均对自涉偏好构成威胁与挑战。我们将这些博弈实验归纳在表1中。

丹尼尔·W. 布罗姆利:《经济利益与经济制度》,中文版,103页,上海,上海人民出版社,1996。

社会偏好理论将公平、互惠等一些与自利假设有着明显涵义辨识度的人类社会性情感引入经济学的分析框架,借助博弈论的分析工具构建新的博弈均衡来解释实验经济学所揭示的一系列悖论。社会偏好表达的是个体不仅关心自身的物质收益,也要会关心他人的利益。下文的他涉偏好(other-regarding preferences)与社会偏好涵义相近。现有的文献还提及亲社会性偏好(prosocial preferences)、互动偏好(interdependent preferences)等,其涵义均相近。

指将涉及偏好的变动问题解释成约束条件的变化以及在主流框架中内生地解决偏好变动问题,如 Stigler(1977)、Becker(1966, 1977, 1998)给传统的效用函数附加了一个生产性变量,比如个人阅历、文化教育、风俗习惯、社会关系等能够改变人们偏好的因素,从而把偏好的改变转化为偏好的生产问题(叶航等,2007)。

表 1

行为与实验经济学验证他涉偏好的博弈实验

博弈	博弈内容	资料来源	主流偏好理论预测	实际博弈结果	博弈结果的解释
囚徒困境博弈	两个个体,策略为合作与背叛,支付为: 合作 背叛 合作 (H, H) (S, T) 背叛 (T, S) (L, L) $H > L, T > H, L > S_0$	Dawes(1980)	(L, L) 为纳什均衡策略,合作悖论。	50%比例选择合作;交流越多,合作的比例越高。	互惠偏好
公共品博弈	n 个参与个体,赋予其最初禀赋 y_i ,然后每个人同时选择一定的数量 c_i $[0, y_i]$ 作为对公共品的贡献。每人支付为: $p_i = y_i - c_i + m_j c_j$ ($j = 1, \dots, n$); $m < 1 < m n_0$	Ledyard(1995)	不对公共品做任何贡献是占优策略均衡,即 $g_i = 0_0$ 。	一次性博弈中平均贡献达到了初始禀赋 y_i 的50%;多期博弈中开始时贡献大约有50%,然后开始下降,最后一轮大多数人都不再做出任何贡献。	互惠偏好
最后通牒博弈	一定数量的金钱 s 在提议者与回应者之间分配。提议者提出 x ,如果回应者拒绝 x ,则两者获得的金钱均为0;如果接受 x ,则提议者获得 $S - x$,回应者获得 x_0 。	Güth, Schmittberger和Schwarze(1982),Camerer和Thaler(1995)	$x = \frac{s}{2}$ 为最小的正数。任何一个 $x > 0$ 的提议都将被接受。	大部分提议数量在30%~50%之间;提议在小于20%时一半被拒绝。	公平偏好 互惠偏好 消极互惠
独裁者博弈	在最后通牒实验框架内取消了回应者拒绝的权利,即回应者只能任凭提议者“独裁”,只是一个接受者。提议者、回应者支付分别为: $S - x$ 和 x_0 。	Kahneman等(1986),Andreoni和Miller(1993),Forsythe等(1988),Camerer和Thaler(1995)	子博弈纳什均衡为独裁者占据所有份额,接受者所得为零,即 $x = 0_0$ 。	提议者平均提供给回应者20%的金钱。	利他偏好
信任博弈	投资者获得支付 s ,将其中的一部分 y ($y \in [0, S]$)转移给保管人。保管人会收到 $3y$,然后回赠 x ($x \in [0, 3y]$)予投资者。投资者支付 $S - y + x$,保管人获得 $3y - x_0$ 。	Berg, Dickhaut和McCabe(1995),Fahr和Irlenbusch(2000),Cox(2000)	保管人不会返还任何数量的钱,投资者也不会赠送任何数量的钱,即 $x = 0, y = 0_0$ 。	平均赠送 y 为 s 的50%,返还的数量 x 略小于 s 的50%,并且赠送量 y 和返还量 x 之间有显著的正相关关系。	互惠偏好 积极互惠
礼物交换博弈	企业提供固定工资 w ,如果员工拒绝则双方收益都为0;如果员工接受,则他再决定付出多少努力 e ($1, 10$)。企业的收益为 $10e - w$,员工的收益为 $w - c(e)$, $c(e)$ 为努力成本,是 e 的增函数。	Fehr, Kirchsteiger和Riedl(1993),Offeman(1999),Falk,Fehr和Fischbacher(2000)	员工面对任何固定工资 w 都只会付出最低努力 $e = 1$,企业也只会提供最低工资 w ,并且工资和努力之间没有相关关系。	企业提供的工资 w 远高于最小值,员工付出的努力也高于 e ,并且工资和努力之间有显著的正相关关系。	互惠偏好
第三方惩罚博弈	A 和 B 参与独裁者博弈, C 观测 s 的数量分配给 B , C 可以惩罚 A 但是需要自己花费成本。	Fehr和Fischbacher(2001)	A 不分配任何 s 给 B , C 也不会进行惩罚。	A 分配给 B 越少, C 的惩罚越多。	公平偏好 强互惠

资料来源:作者根据相关资料整理。

挑战主流经济学偏好理论的这些博弈实验结果必然受到批判性的质疑与审问,综观各方面的责难,主要体现在:其一,博弈实验中的注码数目太小,因而实验结果不能代表他们自身的选择情况;其二,参与者对博弈实验有所误解,因而结果不具可信性;其三,实验对象可能没使他们的行为适应应该交往的非重复性质,而是遵循了来自一些更常见的重复交往的经验法则;其四,实验结果是否可以作为真实世界中行为特性的强有力指标有待检测;其五,实验参与者多为大学生,能否代表一般个体值得商讨。针对各方的质疑,行为与实验经济学者试图通过合理的设计剔除相关因素的干扰因而重新进行博弈实验来验证自身的假说回应质疑:Cameron(1998)在印度尼西亚的大学生中间进行的实验中报酬相当于三个月的平均支出,Hoffman, McCabe和Smith(1996),Fehr和Fischbacher(2001)在美国学生中做的实验注码大小是100~400美元,在俄罗斯的礼物交换博弈中所得相当于其收入的二到三倍,然而结果依然与先前一致。因此,对博弈实验中的注码数

Cameron, L., 1998. "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia" *Economic Inquiry*, Vol 37, pp. 47 - 59.

Hoffman, E.; McCabe, K and Smith, V., 1996. "On Expectations and Monetary Stakes in Ultimatum Games" *International Journal of Game Theory*, Vol 25, pp. 289 - 301; Fehr, E and Fischbacher, U., 2001. "Why Social Preference Matter" *Stockholm, Nobel Symposium on Behavior and Experimental Economics*

目太小的怀疑没有理由; Fehr和 Fischbacher(2001)将实验对象与不同的配对方式进行了多次博弈,使得第二种质疑也不足以构成威胁;第三种质疑也是站不住脚的,因为参与者很容易区分重复交往与非重复交往从而相应地调整自己的行为,退一步即使使用了某些经验法则也是与主流偏好理论不相容; Glaeser等(1999)研究发现,那些在博弈实验中表现值得信赖的参与者在真实世界的社会生活中表现的行为也值得信赖; Henrich等(2001)在五大洲12个国家,15个经济和文化环境迥异的小规模社会做的大规模跨文化最后通牒博弈实验中没有使用大学生但同样表明自利模型没有得到任何一个被研究的社会的的支持,相反,规范模型失效的范围更大了。

博弈实验能够很好地说明对于真实世界中的个体,除了仅仅关注自身利益的自涉偏好之外,还有关注他人的他涉偏好。然而这些真实的博弈实验对主流理论的挑战还是有限的,因此,自涉偏好与他涉偏好的统一还需要从理论上寻找契合。借用近些年来诸多经济学者如 Rabin(1993), Dufwenberg和 Kirchsteiger(2004), Fehr和 Schmidt(1999), Bolton和 Ockenfels(2000), Charness和 Rabin(2002), Kohler(2003)以及 Falk和 Fischbacher(2006)对效用函数重新构造的成果,我们可以初步地得到解答。正如前面互惠偏好、公平偏好与利他偏好的划分,这些效用函数也对应着他涉偏好的互惠、公平与利他偏好理论模型。篇幅限制,我们仅对 Rabin(1993)的互惠偏好模型进行说明。

Rabin(1993)将他模型中的博弈称为心理博弈(psychological games),而不是一般的物质博弈(material games)。在心理博弈中,个人的反应函数不仅取决于对方的战略选择,而且取决于自己的二阶信念。参与者根据自己的信念结构判断对方的意图,进而决定反应战略,所以善意函数的设定是互利模型的关键。以一个双人博弈为例,与只关心自己收益的博弈方不同,如果博弈方*i*(博弈方*j*亦然)觉得博弈方*j*的动机“好”,那么博弈方*i*采取的策略将有利于博弈方*j*;如果博弈方*i*觉得博弈方*j*的动机“坏”,那么博弈方*i*采取的策略将有损于博弈方*j*。总之,博弈方*i*的效用不仅取决于自己的收益,也取决于他对博弈方*j*的动机的信念。博弈方*i*的期望效用函数取决于三个变量:他自己采取的策略 $a_i(a_i, s_i)$ 、对方采取的策略 $b_j(b_j, s_j)$ 、他的信念中对方认为他所采取的策略 $c_i(c_i, s_i)$ 。在博弈方*i*相信博弈方*j*选择策略 b_j 的情况下, $h_j(b_j)$ 为 Pareto有效下博弈方*j*可能得到的最高收益,令 $l_j(b_j)$ 为 Pareto有效下博弈方*j*可能得到的最低收益,那么对博弈方*j*而言,博弈方*i*采取的策略带给他的公平收益为: $f_j = [h_j(b_j) + l_j(b_j)]/2$ 。另 $min_j(b_j)$ 为博弈方*j*可能得到的最低收益,则“善意函数”:

$$f_i(a_i, b_j) = \frac{j(b_j, a_i) - e_j(b_j)}{h_j(b_j) - min_j(b_j)} \quad (\text{界定 } h_j(b_j) - min_j(b_j) = 0 \text{ 时 } f_i(a_i, b_j) = 0)$$

显然,当博弈方*i*的收益恰好等于他的公平收益时, $j(b_j, a_i) - e_j(b_j) = 0, f_i(a_i, b_j) = 0$;当博弈方*i*的收益小于他的公平收益时, $j(b_j, a_i) - e_j(b_j) < 0, f_i(a_i, b_j) < 0$;当博弈方*i*的收益大于他的公平收益时, $j(b_j, a_i) - e_j(b_j) > 0, f_i(a_i, b_j) > 0$ 。“善意函数”衡量了博弈方*i*对博弈方*j*的善意程度。

同理,在博弈方*i*相信博弈方*j*认为他选择策略 c_i 的情况下,令 $h_i(c_i)$ 为 Pareto有效下博弈方*i*可能得到的最高收益,令 $l_i(c_i)$ 有效下博弈方*i*可能得到的最低收益,那么对博弈方*i*而言博弈方*j*采取的策略带给他的公平收益为: $f_i = [h_i(c_i) + l_i(c_i)]/2$;

另令 $min_j(b_j)$ 为博弈方*j*可能得到的最低收益,则“善意函数”:

$$\bar{f}_j(b_j, c_i) = \frac{i(c_i, b_j) - e_i(c_i)}{h_i(c_i) - min_i(c_i)} \quad (\text{界定 } h_i(c_i) - min_i(c_i) = 0 \text{ 时 } \bar{f}_j(b_j, c_i) = 0)$$

与“善意函数”类似,“善意函数”衡量了博弈方*i*认为的博弈方*j*对博弈方*i*的善意程度。

因而博弈方的效用函数为: $U_i(a_i, b_j, c_i) = i(a_i, b_j) + f_j(b_j, c_i) [1 + f_i(a_i, b_j)]$ 。

从这个效用函数可以看出:(1)当 $f_j(b_j, c_i) < 0$ 时,这意味着当博弈方*i*认为博弈方*j*对他不友善,那么

Glaeser, E.; Laibson, D.; Scheinkman, J. and Soutter, C., 1999. "What is Social Capital: The Determinants of Trust and Trustworthiness" NBER Working Paper, No. 7216

Henrich, J.; Boyd, R.; Bowles, S.; Camerer, C.; Fehr, E.; Gintis, H. and McElreath, R., 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies" American Economic Review, Vol 91, pp. 73 - 78

Rabin, M., 1993. "Incorporating Fairness into Game Theory and Economics" American Economic Review, Vol 83, pp. 1281 - 1302

博弈方 i 使自己效用最大化的策略是以怨报怨, 也就是 $f_i(a_i, b_j) < 0$; (2) 当 $f_j(b_j, c_i) > 0$ 时, 这意味着当博弈方 i 认为博弈方 j 对他友善, 那么博弈方 i 使自己效用最大化的策略是以德报德, 也就是 $f_i(a_i, b_j) > 0$; (3) 当 $f_i(a_i, b_j)$ 较大的时候, $f_j(b_j, c_i) / [1 + f_i(a_i, b_j)]$ 在效用函数中的权重减小了, 这说明物质回报高的时候, 人们对公平的关注会相应有所减少。

由于 Rabin 模型局限于两人标准式博弈, 而且由于该模型并没有对非均衡路径中的信念形成做出设定, 应用于序贯博弈时就会产生不合理的均衡, 因此, Dufwenberg 和 Kirchsteiger 对 Rabin 模型进行了拓展。

他涉偏好的公平偏好模型主要包括 F&S 模型 和 B&O 模型 以及利他偏好模型也得出了各自的效用函数。他涉偏好的互惠偏好、公平偏好与利他偏好三类偏好理论模型存在相通之处, 但更大程度上是各自相互独立地存在。为了解决三类社会偏好类型的互补问题, 后来的学者试图将这些偏好模型进行综合, 试图得出最一般意义上的偏好理论模型。如 Charness 和 Rabin (2002) 综合了利他偏好模型和互惠模型, Kohler (2003) 综合了利他偏好与公平偏好模型, Falk 和 Fischbacher (2006) 综合了互惠模型和公平偏好模型等等。

2 利他偏好: Santa Fe Institute 的深化与拓展

桑塔费学派发展了社会偏好理论中的利他主义偏好。他们认为, 一方面, 利他偏好理论对利他的界定不清晰, 往往使人造成错觉, 那些表面看起来利他的行为本质上却是基于自利偏好的动机; 另一方面, 利他偏好理论不足以给人类的经济行为提供一般性的解释。诸如互惠利他主义 (reciprocity altruism) 与血缘利他主义 (kin altruism) 都可看成是利己主义的自利偏好, 然而像纯粹利他偏好如匿名慈善捐款、战争中牺牲自己等等以及尤其当面对陌生人所表现出的利他行为时, 利他偏好理论并不能做出心悦诚服的解释。因此, 行为与实验经济学下的利他偏好理论尽管能够对人类普遍存在的一些慷慨行为提供解释, 但是作为人类社会普遍存在的合作与互助的形成, 利他偏好理论的解释远远不够。

他们将人类所具有的这些偏好称为“亲社会情感” (prosociality), 主要包括同情心、愧疚感、感激心和正义感等等。“人类心理结构的首要特征是亲社会情感, 包括移情、同情、羞愧、骄傲和敌意。这些情感——让我们珍惜他人福利如同珍惜自己的福利一样——是人类共有的情感, 不管他人对我们个人的福利做了多大贡献……在这个意义上, 人们会对合作努力做贡献, 即使这个行为包含了个人成本, 因为这样做使他们感觉很好。而且, 做贡献花费的成本越高, 他们越可能朝着有更高个人支付的活动行为靠拢, 不管是自涉还是他涉。”亲社会性情感会损害个体, 使个体为此支付较高的成本。因此, 为了保证亲社会性情感在演化中的遗传优势, 必须借助强互惠 (strong reciprocity) 的系统安排。强互惠者即群体中的这样一些个体, 他们要求合作的对等性, 积极惩罚那些不合作的人, 哪怕自己付出高昂的代价。当一个带着合作的倾向进入一个新的社会环境的强互惠者, 被预先安排通过维持或提高他的合作水平来对其他人的合作行为做出回应, 并对其他人的“搭便车”行为进行报复, 即使会给自己带来成本, 甚至是不能理性预期这种报复能在将来给个人带来的收

他们将 Rabin 的分析框架延伸到具有 N 人的扩展型博弈中, 引入了序贯互惠 (sequential reciprocity) 均衡的概念。他们定义善意函数为实际支付与公平支付之间的差异, 因而博弈方 i 对博弈方 j 的善意程度为: $f_i(a_i, b_j) = x_j(a_i, b_j) - f_j^{fair}(b_j)$, 其中的公平支付为 $f_j^{fair}(b_j) = [\max_j(a_i, b_j) + \min_j(a_i, b_j)] / 2$, 效用函数则为 $U_i(a_i, b_j, c_i) = x_i(a_i, b_j) + Y_i \frac{f_i(a_i, b_j)}{f_j(a_i, b_j)} f_j(b_j, c_i)$ 。参见 Dufwenberg, M. and Kirchsteiger, G, 2004 “A Theory of Sequential Reciprocity” *Games and Economic Behavior*, Vol 47, pp. 268 - 298。

F&S 模型提出参与者 i 的效用函数为: $U_i(x) = x_i - \frac{a_i}{n-1} \max(x_j - x_i, 0) - \frac{i}{n-1} \max(x_i - x_j, 0)$; 特别地, 双人博弈中的具体效用函数为: $U_i(x) = x_i - i \max\{x_j - x_i, 0\} - i \max\{x_i - x_j, 0\}$ ($i = j$), 其中, 第一项为物质收益直接效用, 第二项为嫉妒负效用, 第三项为同情负效用。详细请见 Fehr, E and Schmidt, K, 1999. “A Theory of Fairness, Competition, and Cooperation” *Quarterly Journal of Economics*, Vol 114, pp. 817 - 868。

B&O 模型的效用函数称为激励函数 (Motivation Function), 表示为: $v_i = v_i(y_i, c_i)$, 其中 y_i 为博弈方 i 的绝对收益, $c_i = \begin{cases} y_i/c & (c > 0) \\ 1/n & (c = 0) \end{cases}$ 为参与者 i 支付的相对份额, $c = y_i$ 为总金钱收入, n 为博弈方的个数。详细请参见 Bolton, G and Ockenfels, A., 2000. “ERC - A Theory of Equity, Reciprocity and Competition” *American Economic Review*, Vol 90, pp. 166 - 193。

赫伯特·金迪斯:《新古典经济学与行为博弈理论的交锋》, 见汪丁丁等主编:《走向统一的社会科学》, 156页, 上海, 上海世纪出版集团, 2005。

益。强互惠者将“以朋友的方式对待朋友,以微笑迎接微笑,以谎言应对背叛”。之所以称之“强互惠”是为了把它与互惠的利他、间接互惠和要求重复互动并可以用自利行为模型解释的其他形式的互惠区分开来。强互惠的重要形式是利他主义惩罚,即个体自身承担实施惩罚的成本,以让违背族群利益规范的那些人付出代价,表1中的第三方惩罚实验便给强互惠的存在提供了有力的支持;而且,最近的研究对强互惠的演化也做了很好的说明与论证(Gintis, et al., 2003; Bowles and Gintis, 2004; Sánchez and Cuesta, 2005)。

3. 他涉偏好的神经基础:神经元经济学的贡献

然而,仅仅从行为与实验经济学的理论成果来看,偏好自涉与他涉的统一似乎并不能使人信服。因为一方面,行为与实验经济学试图通过博弈规则的设计与调整剔除诸如文化、心理、情感等因素的影响,但同时又借用心理学与社会学的原理来解释实验结果,给主流经济理论带来的仅仅是学科外部的批判与挑战;另一方面,他涉偏好理论大多数秉承显示性偏好的传统,从个体的行为反应拟合效用函数。但仅仅用行为显示对偏好进行描述必然会因为行为的易变而无法得出系统、可靠的偏好特征。这也是实验经济学和目前新古典经济学之所以在偏好理论上存在重大争议的原因。幸运的是,随着神经元经济学的诞生与发展,自涉偏好与他涉偏好的统一便有了坚实的神经科学基础:一方面,神经元经济学把偏好的生成和变化置于自然和环境双重选择的基础上,通过演化效率使这个问题内生,从而避免了实验经济学和行为经济学从学科逻辑外部来解释这类现象的做法;另一方面,通过科学的实证手段揭示出偏好及其偏好在决策过程中的内在机制,比如镜像神经元的工作机理和神经系统的自激励机制等。Rustichini(2005)就最新对镜像神经元的作用机理研究的进展做了概括:当镜像神经元观察到一个特定动作时,它们便会做出反应——放电,而且镜像神经元是否放电与做这个动作的主体相距的远近无关;镜像神经元是对动作而不是对动作的对象做出反应;存在不同活动的镜像,也存在心理状态的镜像。Quervain等(2004)利用神经实验验证了利他行为存在的合理性。在IC、IF、IS与NC的对比实验中,他们发现显示出更为活跃的大脑区域是尾核(caudate nucleus),其在预计的与奖励相关的活跃区域的五个对比中被激活。而且,尾核血流峰值的有效范围分析表明了不同条件对其活跃程度所做的贡献:在IC与IF条件下超过平均水平的活跃程度,受试者便显出强烈的惩罚愿望并使这个愿望可以得到满足;在IS和NC条件下活跃水平低于平均水平,受试者要么不能满足惩罚的愿望要么没有惩罚的愿望。而且,通过计算IC条件下不同受试者的实际货币惩罚和脑区兴奋之间的相关性,他们发现尾核活跃度与惩罚投入度之间存在明显的正相关关系,从而表明了尾核在利他惩罚中起着决定性的作用,这验证了利他偏好的神经基础。

(二)偏好异质与内生对偏好同质与外生的扬弃

真实世界中个体行为的多样性与多变性使得建立在偏好同质的假定上的主流解释陷入尴尬的境地,偏好异质则是个体的一般特征。偏好异质包括个体偏好自身的异质与个体间的偏好异质。个体偏好有经济偏好如消费偏好与生产偏好、道德偏好、宗教偏好甚至信仰偏好等等,如Harsanyi(1997)将人的偏好分成“知情偏好”(informed preference)和“非知情偏好”(uninformed preference),黄有光提出偏好的关怀效应(concerning effect)与关情效应(minding effect)等。在博弈实验中一个典型的例子便是个体在作为提议者时往往将较大比例分配留给自己以期望回应者接受,而当他作为回应者时则表现出公平偏好对不平等分配

参见叶航、汪丁丁、贾拥民:《科学与实证——一个基于“神经元经济学”的综述》,载《经济研究》,2007(1)。

Rustichini, 2005. "Neuroeconomics: Present and Future." *Games and Economic Behavior*, Vol 52, pp. 201 - 212

参见 Quervain, D.; Fischbacher, U.; Treyer, V.; Schellhammer, M.; Schnyder, U.; Buck, A. and Fehr, E., 2004. "The Neural Basis of Altruistic Punishment." *Science*, Vol 305, pp. 1254 - 1158。其中, IC、IF、IS与NC分别表示有意且代价高、有意但无代价、有意但是象征性的与无意却有代价。

Harsanyi(1997)指出,社会福利评价的终极尺度应当是人们的知情偏好而不是实际偏好或者快乐。一个人的实际偏好是由其选择行为与语言表达所显示出来的偏好,知情偏好是在他掌握并充分利用了所有相关信息的情况下他本应持有的假想的偏好,其一个必要条件是个体的偏好应该是真实的而不是虚假的偏好。黄有光(2005)认为,Harsanyi用知情偏好代替实际偏好的努力是不彻底的,在进行社会评价时会造成重复计算,因而区分关怀效应与关情效应是必要的。关怀效应也称非情感性利他主义,关情效应也称情感性利他主义,指因看到或帮助他人获得快乐而自身感到快乐和心理现象。二者是不同的:一个只对自己关情的人就是其快乐不受他人快乐水平影响的人。一个人在获得非洲难民的悲惨景况后也感到难过,这是种关情效应;一个只关怀自己的人则是指其偏好不直接受他人快乐水平影响的人,但是这并不排除他人快乐可能通过关情效应而间接影响其偏好。详见黄有光:《社会福祉与经济政策》,21~25页,北京,北京大学出版社,2005。

予以拒绝。阿罗不可能性定理 (Arrow's impossibility theorem) 以及其后的肯普 - 黄有光 - 帕金斯的不可能性命题 (the impossibility propositions of Kemp - Ng - Parks)、吉尔伯德的寡头定理 (Gibbard's oligarchy theorem) 等已经从某一方面证实了个体间偏好异质性的存在,而这种异质性的社会效应通过简单的社会平均加总并不能很好地表征。“在一个由同等数量的圣人和无情的人组成的种群中,交往的结果一般不会是两种单一类型种群的结果的平均,因为种群中类型分布的极小差异就会对每个人的行为方式造成很大的影响。”

弗里德曼的一段话肯定了经济学的外生偏好假定。“尽管有一些限制,经济理论基本上将欲望(偏好——引者注)看作是固定的。这是分工的一个例子。经济学家不讨论欲望的形成。这是心理学家的领域。”弗里德曼强调的是,虽然偏好是变化的,但是对于偏好变化的分析是心理学家的事情,从而将偏好内生排除在经济分析之外。正如本文一直坚持的立场,我们承认偏好的外生给定在一些具体的经济分析中是合理的,然而更应该多关注真实世界中普遍存在的偏好内生的情况。在强调偏好他涉的诸多博弈实验中,通过设计与调整实验方案,可以对内生的偏好做充分的阐释。博弈实验可以按照表 2 的方式进行:

表 2 最后通牒博弈与内生偏好

博弈	博弈内容	资料来源	实际博弈结果	博弈结果的解释
标准玩法	抛硬币确定提议者,提议者提出分配方案,回应者选择接受或拒绝	Guth, Schmittberger 和 Schwarze (1982)	提议在小于 20% 时被拒绝	互惠偏好 公平偏好
随机出价	抛两次硬币:第一次抛硬币决定提议者,第二次抛硬币决定提议者分配的方案	Bount (1995)	低报价会被接受很少被拒绝	情景依存偏好
通过测试选择角色	提议者通过现场测试结果而定,得分高者担任	Hoffman, McCabe, Shachat 和 Smith (1994)	提议者将较大比例留给自己,因而出现很多低报价,很少遭拒	情景依存偏好
交换博弈	仅将上述博弈名称改为“交换博弈”,其他不变	Hoffman, McCabe, Shachat 和 Smith (1994)	出现很多低报价,很少遭拒	情景依存偏好
没有“公平”出价的的可能	博弈规则约定提议者提不出公平的分配方案	Falk, Fehr 和 Fischbacher (2003)	低报价,没有遭拒	情景依存偏好
标准玩法	博弈在巴布亚新几内亚的 Au 和 Gnau 人中进行	Henrich, Bowles, Boyd, Camerer, Fehr, Gintis 和 McElreath (2004)	多于 1/2 的分配比例很普遍,过高或过低的分配比例以相同概率被拒	内生偏好 情景依存偏好
标准玩法	博弈在秘鲁亚马逊流域的 Machiguenga 人中进行	Henrich, Bowles, Boyd, Camerer, Fehr, Gintis 和 McElreath (2004)	3/4 数量的分配提议是 1/4 或更少,但很少遭拒	内生偏好 情景依存偏好

资料来源:作者根据相关资料整理。主要内容来源于萨缪·鲍尔斯:《微观经济学:行为、制度与演化》,中文版,84~86页,北京,中国人民大学出版社,2006。

一旦将偏好内生化,偏好稳定不变的神话便宣告破灭,偏好演化问题应运而生,这便是本文偏好拓展的第三个方面:偏好稳定的特殊性与偏好演化的一般性。

(三) 偏好稳定的特殊性与偏好演化的一般性

如前所述,本文并不排斥偏好具有稳定的性质,在某些经济分析中偏好稳定的假设很合理也很必然。诸如 Robson (2001) 认为基于生物学的个体行为假设能够为经济学个体偏好假设提供坚实的基础,生物演化支持了主流经济学中个体偏好外生给定、稳定不变的假设。在他看来,稳定不变的偏好至少包括个体摄取能量

Bowles, S., 1998. "Endogenous Preference: The Cultural Consequences of Markets and other Economic Institutions" *Journal of Economic Literature*, Vol 36, p. 91.

参见 Friedman, M., 2007. *Price Theory*. New Brunswick, N. J.: Aldine Transaction Publishers, p. 13.

Bowles 这样定义偏好的情境依存性。“假设 s_i 是一个表示状态 i 的向量,它是可能状态集合 S 的一个元素, $U_i(s_j)$ 是状态 s_j 对于一个目前正经历状态 s_i 的个人而言的效用。令 $U_i(s)$ 表示当一个人处于状态 s_i 时对于所有可能状态的偏好排序。于是,如果存在一定的 i 和 k , 对于相同的个人而言,他在另一个不同状态下的排序 $U_k(s)$ 和由 $U_i(s)$ 给出的排序不同,那么这个偏好就是情境依存的。参见 Bowles, S., 1998. "Endogenous Preference: The Cultural Consequences of Markets and other Economic Institutions" *Journal of Economic Literature*, Vol 36, p. 76.

出现这种报价的原因在于在几内亚社会中,竞争性的礼物赠送行为被作为建立地位与隶属关系的一种方式 (Henrich, et al., 2004)。

和繁衍后代。然而,当我们把眼光放在真实世界中的个体行为上,偏好稳定仅仅体现了一种特殊性,广泛存在的偏好改变与演化才具有一般性。因此,对于偏好的拓展,必然将偏好演化纳入其内涵。

偏好的演化观表明了偏好的情景依存性,表 2 列出的博弈实验证实了这种情景依存的偏好。情景依存的偏好一般可以通过主客观两方面进行改变与演化。一方面,通过个体自身的主观心理等因素而演化。行为经济学大量的理论诸如损失厌恶、禀赋效应、前景理论等可以很好地予以验证。Kahneman 和 Tversky (1979) 研究表明个体的行为决策往往会出现相似性偏差、可利用性偏差、表征性启发、后视偏差、锚链效应与调整、认知分歧与群体影响。该理论认为,效用函数的自变量是状态或事件的变化而非状态本身,因此个体给状态赋予的价值取决于状态与现状之间的关系,因而价值函数如图 1 所示。利用价值函数可以解释很多主流理论不能解释的现象。

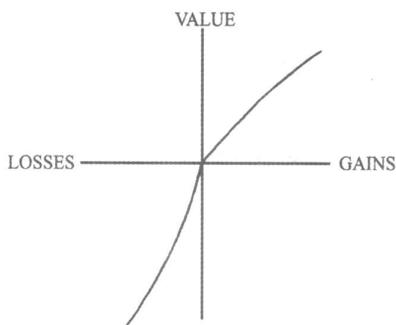


图 1 价值函数 (Kahneman and Tversky, 1979)

另一方面,通过个体外界客观环境的变动而相应改变,如时间的差异、地区的变动、社会制度的变迁等等。我们通过图 2 表示这一情景依存的偏好:

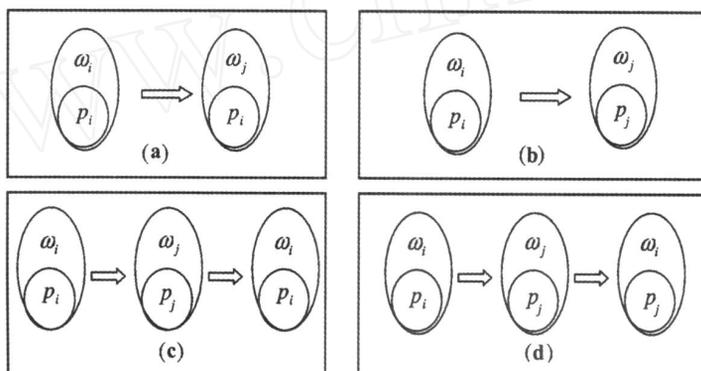


图 2 情景依存的偏好

- (a) 在情景 ω_i 中,个体有偏好 p_i ;当情景变为 ω_j 时,个体的偏好依旧是 p_i ;
- (b) 在情景 ω_i 中,个体有偏好 p_i ;当情景变为 ω_j 时,个体的偏好变为 p_j ;
- (c) 在情景 ω_i 中,个体有偏好 p_i ;当情景变为 ω_j 时,个体的偏好变为 p_j ;当情景再变为 ω_i 时,个体的偏好再变回 p_i ;
- (d) 在情景 ω_i 中,个体有偏好 p_i ;当情景变为 ω_j 时,个体的偏好变为 p_j ;当情景再变为 ω_i 时,个体的偏好依然为 p_j 。

情景依存的偏好为偏好演化的一般性理论奠定了基础。先前的研究表明,个体偏好是不断形成与演化的。偏好是一个社会的、历史的、动态的范畴,是经济社会中的制度、资产、价格等的复合函数;偏好的形成与塑造离不开文化因素、社会经济制度、社会政治制度的制度因素和资产、价格、利率等其他因素的参与和渗透。

Robson, A., 2001. "The Biological Basis of Economic Behavior" *Journal of Economic Literature*, Vol XXXIX (March 2001), pp. 11 - 33; Robson, A., 2002 "Evolution and Human Nature" *Journal of Economic Perspectives*, Vol 16, No 2, Spring, pp. 89 - 106

此价值函数具有三个特征:价值是定义在财富的变化而非财富水平上、价值函数在现状点是折弯的、损失厌恶系数大约为 2 或稍大。参见 Kahneman, D. and Tversky, A., 1979. "Prospect Theory: An Analysis of Decision under Risk" *Econometrica*, Vol 47, pp. 263 - 291.

透;偏好影响制度,制度也影响偏好,偏好和制度共生演化,他们各自影响着对方的发展。

四、偏好融合的理论假说与意义

通过上述偏好拓展的分析研究,本文尝试将这些偏好理论进行综合,初步给出一般意义上的个体偏好,为偏好的融合奠定基础。因此,我们提出偏好融合的理论假说:个体偏好是自涉偏好与他涉偏好的统一;偏好是情景依存的,内生于社会制度之中;偏好是异质的;偏好是演化的。因而,我们构建出融合后的个体偏好 $P = F[U(B_i), C(B_i), \dots, I, \dots]$,这个简洁的偏好函数基本上可以表达本文的主要观点:

1 个体偏好可以利用效用函数来表征,不同于主流经济学的工具性的效用函数理论,本文的偏好与效用是相互依存相互联系相互影响的。 B_i 为广义的行为,可以是体现自涉偏好的利己行为,也可以是体现他涉偏好的利他行为等等,可以是一些实际经济行为如生产、消费,甚至还可以是一些精神行为、心理行为如羡慕与嫉妒等等。 $U(B_i)$ 表示个体从行为 B_i 中获取的效用,如果 B_1 、 B_2 分别代表利己与利他行为,则 $U(B_1)$ 、 $U(B_2)$ 为个体从利己行为与利他行为中获取的效用。因此,偏好函数体现了自涉偏好与他涉偏好的统一,体现了偏好的异质性。

2 特定情景下个体的偏好取决于偏好结构。 $C(B_i)$ 为成本函数,表示个体实施行为 B_i 所付出的成本,这个成本是实施行为 B_i 所有成本的加总。因此,沿用主流经济学分析思路,当 $\frac{\partial U(B_1)}{\partial B_1} / C(B_1) > \frac{\partial U(B_2)}{\partial B_2} / C(B_2)$ 时,个体的自利 (B_1) 偏好大于利他 (B_2) 偏好,个体表现出自利行为;反之成立。特别地,个体的效用最大化行为必然是 $\frac{\partial U(B_1)}{\partial B_1} / C(B_1) = \frac{\partial U(B_2)}{\partial B_2} / C(B_2) = \dots = \frac{\partial U(B_n)}{\partial B_n} / C(B_n)$ 。

3 偏好函数表达了情景依存的偏好,是对 2 的深化与拓展。正如前述图 1 情景依存的偏好所示,主流经济理论忽略了情景对偏好的影响,而状态 S_i 是影响偏好的一个重要因素。偏好的传递性在独立场景如 S_1 中可以满足,但从 S_1 过渡到 S_2 则不一定能满足。在 S_2 的分析基础上,在 S_1 中个体表现出强烈的自利偏好,而在 S_2 中则可能利他偏好更为强烈。这个思路的启发是很大的,它警示我们个体具有的公平、利他偏好并不是一定会导致公平、利他行为,因为个体决策不仅依赖于其偏好,还依赖于情景或制度。公平、利他偏好一定条件下也可能导致自利行为;同样,纯粹自利偏好在另一条件下也可能导致公平或利他行为。这便为社会政策的制定提供了理论参考。因此,个体偏好的这种情景依存性,是偏好融合的重要部分。

4 情景依存的偏好必然内生于更为广泛意义上的制度之中。偏好函数的这种处理并不割裂状态 S_i 与制度 I 之间的关系,而是强调制度与构建偏好与演化偏好的互动关系。表 3 给出了主流偏好理论与融合后的偏好理论比较的一个简要的框架。

表 3 主流偏好理论与融合后的偏好理论比较

主流范式	本文观点
自涉偏好	自涉与他涉偏好的统一
外生给定	情景依存的偏好;偏好内生
偏好同质	偏好异质;稳定与改变、演化
偏好稳定	偏好稳定的特殊性与演化的一般性
工具性的效用函数理论	偏好与效用相互影响相互制约
偏好与选择范式	偏好、制度与选择的互动
合作悖论	合作的产生与解释

Bowles利用复制者动态方程与计算机仿真模型化与模拟化了偏好与制度共生演化理论,详见其著作第 13 章;关于偏好与制度的关系,拙作 (2009) 也做了初步探讨。

有趣的是,对利他行为的考察,生物学家做了大量的研究,取得了很多富有启发性的成果,如 Edwards (1962)、Wilson (1975) 认为利他行为可以伴随着种群的胜利而成功演化;Hamilton (1964) 从个体角度解释了生物世界普遍存在的亲缘利他行为;Trivers (1971) 借助博弈论解释了生物个体之间的互惠利他行为;Dawkins (1976) 认为基于基因研究角度,利他主义存在但其体现了一种自私性或利己性;Sober 和 Wilson (1998) 探索了人类中利他主义动机的演化。这些理论为他涉偏好提供了生物学基础。

效用函数 $U(B_i)$ 可以不是凹性,它包含凸性的效用函数;它满足可加性,因为本文的效用函数是建立在扩展与融合后的偏好理论之上,偏好是可以计量的。最新的福利经济学以及神经经济学等的发展可以为这种度量提供方法,本文不讨论偏好的计量问题。

融合后的偏好理论是我们解释、预测与指导真实世界中个体决策、制度设计等问题的核心,具有重大的理论意义与现实意义。一方面,陌生个体间的一次交往并不会陷入囚徒困境的纳什均衡,得益于他涉偏好的贡献,双方的效用函数只要涉及了对方的利益,合作便成为可能。这也是真实世界中普遍存在的情形。因此,偏好的融合将解决主流理论困境的合作的产生问题,甚至助推经济研究从“自涉-竞争-一般均衡”向“他涉-合作-演化”范式的转变。偏好的融合,为经济学的个体决策即理性选择模型的重构提供了一个新的视角:不再是个体在有限的约束下求解最大化偏好的程序算法问题,而是立足于偏好内涵的扩充、偏好的演化来重构一般化的理性行为者模型。另一方面,以纯粹自利的偏好为前提的经济政策出现了偏差,如 Fehr(2000, 2004)、Gneezy和 Rustichini(2000)及 Cardenas(2000)的研究表明,建立在自利基础之上如使用罚金等惩罚机制来增强员工工作的努力程度的激励政策会适得其反,这给我国政策的制定提供了经验。

而且,偏好的融合是对人类价值多元化的认可,与中国的传统文化产生共鸣。中华文化强调中庸与仁和,强调助人与平等等,均与公平偏好、互惠偏好与利他偏好不谋而合。汶川地震后社会各界的慷慨解囊、奥运志愿者无偿的热心服务均诠释了他涉偏好的普遍性。中华五千年历史的文化沉淀,使得我国的人文社会观念与制度环境独具中国特色。在这一特色下,个体的性格特征、价值观念、道德基础以及我们的历史传统、文化底蕴等都与西方有着很大的差异。因此,在对中国经济现象与社会问题进行研究与解决时,需要体现自己的本土性,从而建立符合我国具体社会制度条件的理论模型,促进我国经济理论的研究和现实经济政策的制定。作为贫富差距最大的国家之一,公平分配问题已经成为和谐发展中的突出问题。从“效率优先,兼顾公平”到“初次分配注重效率,再分配注重公平”到“更加注重社会公平”,体现了公平偏好在经济决策从而在经济发展中的地位;另外,建立和谐社会这一时代课题要解决的是市场自由的效率与体现社会公平的社会均衡结合的问题,其最终目标是经济增长成果的共享和增长过程的社会和解(杨春学, 2009)。偏好的拓展与融合正为我们完成这一历史使命提供了基础。

参考文献:

1. [美] 萨缪·鲍尔斯:《微观经济学:行为、制度和演化》,中文版,北京,中国人民大学出版社,2006。
2. 黄有光:《社会福祉与经济政策》,北京,北京大学出版社,2005。
3. 汪丁丁等主编:《神经经济学:实证与挑战》,上海,上海世纪出版集团,2007。
4. 汪丁丁等主编:《走向统一的社会科学:来自桑塔费学派的想法》,上海,上海世纪出版集团,2005。
5. 杨春学:《和谐社会的政治经济学基础》,载《经济研究》,2009(1)。
6. 叶航、汪丁丁、罗卫东:《作为内生偏好的利他行为及其经济学意义》,载《经济研究》,2005(8)。
7. 周小亮、笪贤流:《效用、偏好与制度关系的理论探讨——反思消费者选择理论偏好稳定之假设》,载《学术月刊》,2009(1)。
8. 朱宪辰、黄凯南:《基于生物学基础的行为假设与共同知识演化分析》,载《制度经济学研究》,2004(4)。
9. Andreoni, J. and Miller, J., 1993. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence" *Economic Journal*, Vol 103, pp. 570 - 585.
10. Berg, J.; Dickhaut, J. and McCabe, K., 1995. "Trust, Reciprocity and Social History." *Games and Economic Behavior*, Vol 10, pp. 122 - 142.
11. Blount, S., 1995. "When Social Outcomes Aren't Fair: The Effect of Casual Attributions on Preferences" *Organizational Behavior & Human Decision Processes*, Vol 7(3), pp. 346 - 380.
12. Bowles, S. and Gintis, H., 2004. "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations" *The American Economic Review*, Vol 65(1), pp. 17 - 28.
13. Camerer, C. and Fehr, E., 2001. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists" *Working Paper Series, ISSN 1424 - 0459, No 97*.
14. Camerer, C. and Thaler, R., 1995. "Ultimatums, Dictators and Manners" *Journal of Economic Perspectives*, Vol 9, pp. 209 - 219.
15. Cardenas, J.; Stranlund, J. and Willis, C., 2000. "Local Environmental Control and Institutional Crowding-out" *World Development*, Vol 28(10), pp. 1719 - 1733.
16. Charness, G. and Rabin, M., 2002. "Understanding Social Preferences with Simple Tests" *Quarterly Journal of Economics*, Vol 117, pp. 817 - 869.
17. Cox, J., 2000. *Trust and Reciprocity: Implications of Game Triads and Social Contexts* Mimeo, University of Arizona at Tucson.
18. Dawes, R., 1980. "Social Dilemmas" *Annual Review of Psychology*, Vol 31, pp. 169 - 193.
19. Fahr, R. and Irlenbusch, B., 2000. "Fairness as a Constraint on Trust in Reciprocity: Eamed Property Rights in a Reciprocal Exchange Experiment" *Economics Letters*, Vol 66, pp. 275 - 282.

(下转第 152页)

45. Shakotko, R. A. ; Edwards, L. N. and Grossman, M. , 1980. "An Exploration of the Dynamic Relationship Between Health and Cognitive Development in Adolescence." NBER Working Paper, No. 454, Cambridge, MA: NBER.
46. Smith, J. P. , 1999. "Healthy Bodies and Thick Wallets: the Dual Relation between Health and Economic Status." The Journal of Economic Perspectives, Vol 13 (2) , pp. 145 - 166.
47. Spiess, C. K. and Schneider, T. , 2004. "Midlife Caregiving and Employment: An Analysis of Adjustments in Work Hours and Informal Care for Female Employees in Europe." in Mortensen J. ; Spiess C. K. ; Schneider T. ; Costa - Font J. and Patxot C. , eds, Health Care and Female Employment: A Potential Conflict? Occasional Paper, No. 6.
48. Thomas, Duncan and Strauss, John, 1997. "Health and Wages: Evidence on Men and Women in Urban Brazil." Journal of Econometrics, Vol 77, pp. 159 - 185.
49. Weil, David N. , 2007. "Accounting for the Effect of Health on Economic Growth." Quarterly Journal of Economics, Vol 122 (3) , pp. 1265 - 1306.
50. Young, Alwyn, 2005. "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations." Quarterly Journal of Economics, Vol 120 (2) , pp. 423 - 466.
51. Young, Alwyn, 2007. "In Sorrow to Bring Forth Children: Fertility Amidst the Plague of HIV." Journal of Economic Growth, Vol 12 (4) , pp. 283 - 327.

(责任编辑:彭爽)

(上接第 130 页)

20. Falk, A. ; Fehr, E. and Fischbacher, U. , 2003. "On the Nature of Fair Behavior." Economic Inquiry, Vol 41 (1) , pp. 20 - 26.
21. Falk, A. ; Fehr, E. and Fischbacher, U. , 2000. "Testing Theories of Fairness - intentions Matter." Institute for Empirical Research in Economics, University of Zurich, Working Paper, No. 63.
22. Falk, A. and Fischbacher, U. , 2006. "A Theory of Reciprocity." Games and Economic Behavior, Vol 54, pp. 293 - 315.
23. Fehr, E. and Gaechter, S. , 2000. "Fairness and Retaliation: the Economics of Reciprocity." The Journal of Economic Perspectives, Vol 14, pp. 159 - 181.
24. Fehr, E. and Götte, L. , 2004. "Do Workers Work More When Wages are High." IZA Discussion Paper, January, No. 1002.
25. Fehr, E. ; Kirchsteiger, G. and Riedl, A. , 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." Quarterly Journal of Economics, Vol 108, No. 2, pp. 437 - 460.
26. Fehr, E. and Schmidt, K. , 1999. "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics, Vol 114, pp. 817 - 868.
27. Forsythe, R. ; Horowitz, J. ; Savin, N. and Sefton, M. , 1994. "Fairness in Simple Bargaining Games." Games and Economic Behavior, Vol 6, pp. 347 - 369.
28. Gintis, H. ; Bowles, S. ; Boyd, R. and Fehr, E. , 2003. "Explaining Altruistic Behavior in Humans." Evolution and Human Behavior, Vol 24, pp. 153 - 172.
29. Gneezy, U. and Rustichini, A. , 2000. "A Fine is a Price." The Journal of Legal Studies, Vol 29, pp. 1 - 17.
30. Güth, W. ; Schmittberger, R. and Schwartz, B. , 1982. "An Experimental Analysis of Ultimatum Bargaining." Journal of Economic Behavior and Organization, Vol III, pp. 367 - 388.
31. Harsanyi, J. , 1997. "Utilities, Preferences, and Substantive goods." Soc Choice Welfare, Vol 14, pp. 129 - 145.
32. Henrich, S. ; Boyd, J. ; Bowles, R. ; Camerer, C. ; Fehr, E. and McElreath, R. , 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small - Scale Societies." The American Economic Review, Vol 91 (2) , pp. 73 - 78.
33. Hirschleifer, J. , 1977. "Shakespeare vs. Becker on Altruism: the Importance of Having the Last Words." Journal of Economic Literature, Vol 15, pp. 500 - 502.
34. Hoffman, Elizabeth; McCabe, K. ; Shachat, K. and Smith, V. , 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior, Vol 63 (2) , pp. 131 - 144.
35. Kahneman, D. ; Knetsch, J. and Thaler, R. , 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." American Economic Review, Vol LXXVI, pp. 728 - 741.
36. Köhler, S. , 2003. Difference Aversion and Surplus Concern - An Integrated Approach. Mimeo, European University Institute, Florence.
37. Ledyard, J. , 1995. "Public Goods: A Survey of Experimental Research." in A. Roth and J. Kagel, eds, Handbook of Experimental Economics. Princeton: Princeton University Press, Chap. 2.
38. Lindbeck, A. and Weibull, J. , 1988. "Strategic Interaction with Altruism: the Economics of Fair Accomplish." Journal of Political Economy, Vol 96, pp. 1165 - 1182.
39. Offerman, T. , 1999. Hurting Hurts More than Helping Helps: the Role of the Self - serving Bias. Mimeo, University of Amsterdam.

(责任编辑:彭爽)