

利他行为及其经济学意义^{*}

——兼与叶航等探讨

齐良书

摘要：本文指出了叶航(2005)和叶航、汪丁丁、罗卫东(2005)建立的“利他行为演化均衡”模型的错误,重新梳理了利他行为与经济学的关系。现代经济学的分析框架并不排斥利他偏好,利他行为难以进入经济学实证分析,主要是由于利他行为难以量化。人类利他行为是生物进化和社会发展的综合产物,道德情感和道德规范在其中起了关键性的作用,但这并不妨碍经济学把利他偏好当作人性的一部分。经济学可以通过解释道德情感的进化根源、道德规范的形成机制以及道德规范对个人行为方式的影响,来解释人为什么有利他行为。

关键词：利他 经济学 进化稳定均衡 道德

一、引言

对利他行为的研究关系到对人性的正确认识,历来受到社会科学各分支以及生物学的重视。相比之下,经济学对利他行为持比较淡漠的态度。这是人们批评经济学缺乏伦理关怀的主要理由。近年来,随着行为经济学的发展,越来越多的实验证据表明人有利他倾向。这促使经济学家重新思考利他行为在经济学研究框架中的位置,并与其他学科的学者共同探讨利他行为的起源问题。

叶航、汪丁丁、罗卫东发表于《经济研究》杂志2005年第8期的文章《作为内生偏好的利他行为及其经济学意义》以及叶航发表于《经济学家》杂志2005年第3期的文章《利他行为的经济学解释》探讨了利他行为的经济学意义,并试图从进化角度直接解释利他行为(这两篇文章的观点和主要内容相同,为行文方便起见,以下简称“叶文”,只在必要时注明引文具体出处)。然而,叶文的模型不能成立,内在逻辑也有问题。受叶文启发,本文将从两个方面重新梳理利他行为与经济学的关系:第一,经济学怎样看待人的利他行为?第二,经济学应该如何解释利他行为?

二、叶文的理论背景、内在逻辑和基本结论

20世纪60年代以来,新达尔文主义生物学在个

体选择理论上解释了自然界的利他现象,即生物个体牺牲自身的生存适应性以增进其他个体适应性的行为。生存适应性指期望后代数目。Hamilton以亲缘选择理论解释了生物个体对亲属的利他行为:由于受惠者与施惠者具有一部分相同的基因,后者的自我牺牲在一定条件下可以增加自己的遗传基因在以后世代中的份额。为了解释非亲属之间的利他行为,Trivers提出了互惠利他理论:如果生物个体的利他行为可望在将来得到回报,使该行为的成本得到弥补,那么利他行为就可以在自然选择过程中保存下来。Axelrod和Hamilton证明了在满足一定条件的囚徒困境中,如果博弈是多次且局数不定的,从长远看,利他/合作就有可能比利己/不合作对个体更有利,并通过计算机模拟,发现在各种可能的合作策略中,“针锋相对”的互惠利他行为是最好的策略。在此基础上,Stephens证明即使适当放宽条件,互惠利他行为仍然可以存在。

显然,亲缘利他和互惠利他最终还是有利于增进生物个体的适应性,实际上是迂回的自利。主流生物学家认为,生物界不存在亲缘利他和互惠利他以外的“真正的”利他,或曰纯粹利他(Dawkins,1976; Sober and Wilson,1998)。原因很简单,得不到回报的纯粹利他行为必将减少生物个体的期望后代数目,因而利他基因将在进化中被淘汰。即使是一个完全

* 本文作者感谢李子奈教授的评论和意见。当然,文责自负。

由利他者组成的群体,一旦有利己者侵入,经过若干代自然选择之后,利他者也会被利己者所取代。也就是说,得不到回报的利他不是进化稳定策略(evolutionarily steady strategy, ESS)。生物进化过程只能产生直接或间接的自利行为。

叶文认为,“要解释利他行为,仅仅满足于亲缘利他和互惠利他是不够的。我们必须超越这两者,对纯粹利他作出令人信服的解释”(叶航,2005)。这对经济学尤其具有重要意义:“现代主流经济学把自利作为人类行为的基本前提,从而本质上排斥了利他行为对经济研究的意义……但经济学对偏好的定义事实上不依赖于偏好的伦理取向。换言之,经济学所谓的偏好,既可以包括利己偏好也可以包括利他偏好。……正是在这种理论背景下,利他行为逐步进入主流经济学的研究视野,但这些研究主要是在外生给定利他偏好的前提下进行的。……可是,一旦经济学家试图将偏好内生,需要解释利他偏好的形成原因时,经济学家就会发现他们将面临和生物学家同样的问题:‘减少个人适应性的利他行为如何能够通过自然选择而得以进化?’于是,很多经济学家试图从伦理学、社会学、心理学或者人类学中寻求答案,把利他偏好的原因归结为道德规范、文化教育甚至宗教信仰,然而……即便经济学家可以从这些途径找到答案,对经济学来说,这可能仍然于事无补。利他偏好的内生,需要经济学家在经济学的逻辑体系和框架中寻求这一问题的解释”(叶航、汪丁丁、罗卫东,2005)。怎样解释呢?叶文认为,必须找到“利他行为演化均衡的微观基础”,回答“导致自身适应性降低的利他行为怎样才能在严酷的生存竞争中对利己行为保持相对的遗传优势,从而使自己得到进化”(叶航、汪丁丁、罗卫东,2005)。为此,叶文建立了一个“利他行为演化均衡模型”,试图证明:有可能出现一种进化均衡,群体中一定比例的利他者和利己者共存;或精确表述为“在生物种群及个体行为模式中,利己行为与利他行为将以某种大致固定的比例同时存在”。最后,叶文得出结论:(纯粹)利他是一种进化稳定策略;经过自然选择和进化产生的人类心智与人类行为,不仅与自利心相容,而且与利他心相容;主流经济学家长期以来对利他行为的偏见与忽视应该得到纠正,利他行为应该纳入经济学实证分析的范畴。

叶文的内在逻辑可以总结为以下几点:

- (1) 要解释人类的利他倾向,必须对亲缘利他和互惠利他以外的(纯粹)利他作出解释。
- (2) 要解释(纯粹)利他,必须证明(纯粹)利他是进化稳定均衡。

(3) 经济学需要解释利他偏好的原因,即将利他偏好内生。

(4) 只有证明(纯粹)利他是进化稳定均衡,才算是经济学的逻辑体系和框架内解释了利他偏好的原因。

(5) 证明了(纯粹)利他是进化稳定均衡,就说明人类的利他行为是一种“实然”,应该将其纳入经济学实证分析。

此外,叶文还介绍了桑塔费学派(Santa Fe Institute)经济学家关于“强互惠”(strong reciprocity)行为的研究成果。本文认为,强互惠行为虽然与主流生物学家的互惠利他理论中描述的互惠合作行为不同,但仍不是生物学意义上的纯粹利他。而且,强互惠行为只有在群体周期性地面临灭绝威胁的条件下才能成为进化稳定均衡,这一点被叶文忽略了。

叶文的核心部分是“利他行为演化均衡模型”。本文将在第三部分指出这个模型的主要错误,在第四、第五、第六部分逐步阐明本文关于上述各问题的观点。

三、叶文的“利他行为演化均衡模型”不能成立

为了便于分析,现将叶文的“利他行为演化均衡模型”照录如下:

博弈支付矩阵为:

	利他者	利己者
利他者	$V/2, V/2$	$0, V$
利己者	$V, 0$	$(V - C)/2, (V - C)/2$

其中, V 代表利他合作带来的全部收益, $V/2$ 代表每个利他者平均享有的合作收益; 0 与 V 表示利他者遭受背叛而利己者独占全部合作收益; C 代表合作剩余, $(V - C)/2$ 表示丧失合作剩余情况下利己者的平均收益。设利己者数量为 X , 利他者数量为 Y , 则两者的期望适应性分别为:

$$EU_x = VY + (V - C)X/2 \dots\dots\dots (1)$$

$$EU_y = (V/2)Y \dots\dots\dots (2)$$

(以上两式表达有误。由于 X 、 Y 分别表示利己者和利他者的数量,而不是在群体中所占的比例,故两式等号右边均应除以 $X + Y$ 。不过这个错误并不影响以下推导过程和结论)。

演化均衡 * 为:

$$* = EU_x = EU_y \dots\dots\dots (3)$$

此时利他者与利己者的预期适应性相等(应该指出,式(3)的表达方式是错误的。在满足演化均衡*的条件下, $EU_x = EU_y$, 但 * 并不等于 EU_x 或 EU_y)。

将式(1)、(2)代入式(3),有:

$$\theta^* = Y/X = (C - V)/V \dots\dots\dots (4)$$

当 $C > V$ 时有非负解。

叶文将上述演化均衡表示为图 1。该图表明,在 $0 < Y/X < \theta^*$ 区间, $EU_y > EU_x$; 在 $0 > Y/X > \theta^*$ 区间(两篇文章原文皆如此,有误,应该是 $Y/X < 0$ 或者 $Y/X > \theta^*$), $EU_x > EU_y$ 。 Y/X 收敛于 θ^* , 故 θ^* 是稳定均衡, 利他行为与利己行为都是进化稳定策略。 θ_0 表明, 在任何一个生物种群中, 由突变、迁徙或其他原因产生的利他者必须超越一个阈值才能得到进化。而“由于亲缘利他的存在, 其他利他行为可以从学习和摹仿中滋生。因此, 超越这个阈值, 在生物长期进化中并非一个不可能的事件。”

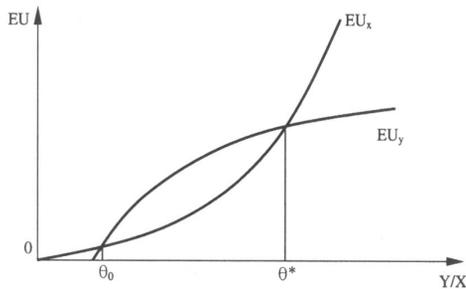


图 1

以上模型的主要错误在于：

1. 该模型错误地否定了主流生物学家的理论前提, 而实际上, 该模型的理论前提与主流生物学并无不同。

叶文认为：“道金斯等主流生物学家的观点存在着明显的疏漏。个体选择理论断言利己行为比利他行为具有更大的适应性, 也许产生于一个长期的偏见和误导, 因为这一结论得以成立的前提, 是孤立地考察利他行为与利己行为对适应性的贡献。我们认为, 这个前提是不正确的。生物适应性是一个全面、综合的评价指标, 它不应该而且也不可能被某个单一的事件或关系所决定。具体地说, 一个利他者的生存适应性不仅取决于他与自私者的个别交往, 而且取决于他与其他利他者的交往, ……同样道理, 一个自私者的生存适应性不仅取决于他与利他者的个别交往, 而且取决于他与其他自私者的交往 ……” (叶航、汪丁丁、罗卫东, 2005)。

叶文误解了主流生物学家的理论前提。主流生物学家并不是“孤立地”考察利他行为与利己行为对适应性的贡献。在他们的模型中, 每个生物个体都既与利他者交往, 也与利己者交往。他们假设群体成员一对一地打交道, 每个个体每次以一定的概率遇到利他/合作者或利己者, 在此基础上计算每类个体生存适应性的期望值。实际上, 叶文的模型与此毫无二致, 之所以得到不同的结论, 是由于模型参数

设定有误。

与此形成对照的是, 桑塔费学派建立的强互惠模型假设群体成员不是一对一地打交道; 合作行为的成果是一种公共品, 在群体所有成员之间分享。但这也只是一种可能性, 不能由此断言主流生物学家的假设不正确。

2. 该模型有非负解时, 不是囚徒困境模型; 囚徒困境何以消失, 叶文没有给予必要的说明。

解释非亲属之间利他/合作行为的难点在于解释博弈双方如何摆脱囚徒困境: 虽然双方都利他/合作对双方最有利, 但由于利己/不合作是占优策略, 也就是说, 不管对方采取什么策略, 自己采取利己/不合作战略总是能占便宜, 至少不会吃亏, 所以双方都会采取利己行为, 导致彼此陷于困境。在这种情况下, 利他/合作行为如何能够出现并持续, 是研究者需要解决的关键问题。主流生物学家和桑塔费学派都没有回避这个问题, 而是证明在一定条件下囚徒困境有合作解, 而这些条件在生物进化过程中是有可能出现的。

叶文又是如何解决这个问题的呢? 叶文没有对模型中的参数加以初始设定, 而是直接认定模型有混合策略均衡解。而事实上, 如果 $C < V$, 该模型是一个典型的囚徒困境模型, 利己行为是唯一的均衡解。只有当 $C > V$ 时, 该模型有混合策略均衡解。这从式(4)也可以看得很清楚: $\theta^* > 0$ 的充要条件是 $C > V$ 。此时该模型不再是囚徒困境模型, 而变成了一个斗鸡博弈 (chicken game) 模型。这种模型不存在占优策略, 而是有两个纳什均衡: 一方进, 另一方的最优战略就是退; 反之亦然。具体到这个模型, 一方利己, 另一方的最优战略就是利他; 反之亦然。这样就避开了囚徒困境问题。

叶文这样解释 $C > V$ 的意义: “在生存条件极其恶劣、选择压力特别巨大的环境中, 个体间的合作是一个生死攸关的选择。囚徒困境下的纳什均衡解对背叛者不会有任何意义, 因为我们面临的是莎士比亚的问题 ‘To be or not to be, that is the question. ‘合作意味着生存, 不合作意味着死亡! 这正是模型中 $C > V$ 的深刻涵义: 巨大的合作剩余孕育出具有道德意蕴的纯粹利他, 引领人类超越死亡的囚徒困境” (叶航, 2005)。然而, 用“合作意味着生存, 不合作意味着死亡”并不能摆脱囚徒困境。困境之“困”就在于: 背叛别人总比被别人背叛好, 即使面临死亡威胁也是如此 (根据桑塔费学派的研究成果, 在面临死亡威胁时更是如此! 见本文第六部分)。不管“合作剩余”多么巨大, 如果互不合作的利己者得不到“合作剩余”, 那么被利己者背叛的利他者不仅也得不到

“合作剩余”，而且应该比利己者遭受更大的损失。这才是包括人类在内的生物逃不开的囚徒困境的真正含义。而在该模型设定的支付矩阵中， $C > V$ 将导致这样一种情形：被利己者背叛的利他者的处境（收益为 0）优于互不合作的利己者的处境（收益为 $(V - C)/2 < 0$ ）。囚徒困境消失了。究竟什么样的选择压力能使这种情形在生物进化过程中成为可能呢？对此，叶文没有给予必要的说明。

3. 阈值问题

关于“参数之间可能具有的非线性性质”，叶文语焉不详，并申明“如上基于 ESS 的模型不是为了描述利他行为演化均衡的具体过程以及这一过程的动态特征，所以我们忽略了一些技术细节”。然而，正是这些“技术细节”决定着模型能否成立。很容易看出，在目前给定的条件下，图 1 是错误的。首先， EU_x 和 EU_y 曲线只可能有一个交点。其次，由式(1)、式(2)可见，若 $C > V$ 成立，则当且仅当 $Y/X = (C - V)/2V$ 时， $EU_x = 0$ ；而 EU_y 总是大于等于 0。故正确画法应如图 2。其中， $\theta_0 = (C - V)/2V$ ， $\theta^* = (C - V)/V$ 。

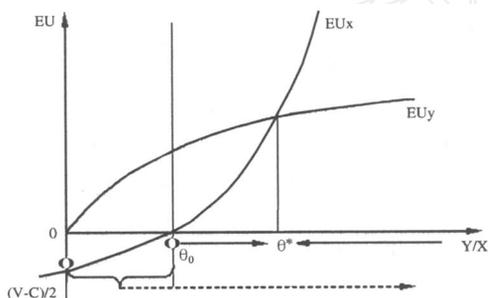


图 2

图 2 表明，全部由利己者组成的群体将灭绝；不过利他者只要一出现就可以生存。这就导致非常奇怪的结果。若初始状态是 $Y/X > \theta_0$ ， Y/X 将收敛到 θ^* 。若初始状态是 $0 < Y/X < \theta_0$ ，利己者将死光，群体中只剩下利他者， Y/X 变成无穷大。此时如果由于基因突变出现利己者， Y/X 又将收敛到 θ^* 。所以在这个模型中，阈值 θ_0 根本没有意义！

退一步讲，即使图 1 正确，叶文关于阈值的观点也不对。叶文认为，当突变、迁徙等小概率事件无法满足超越阈值的要求时，“由于亲缘利他的存在，其他利他行为可以从学习和模仿中滋生”。然而，这是一个进化模型，描述的是通过遗传而延续或扩散的行为方式。如果利他是一种后天获得的特性，是不可能通过基因遗传下去的。而且，如果利他行为一开始就能通过学习和模仿传播，此后的利他行为更应该可以，那么也就没有必要证明利他行为是通过进化产生的了。

4. 博弈策略的多样性和稳定性问题。

该模型中只有两种人，确切地说，只有两种策略：利他/合作和利己/不合作。最后的均衡策略是两者的混合：个体以一定的概率（随机地）采取利他/合作和利己/不合作行为。然而，在重复博弈中，特别是在生物进化这种漫长的重复博弈中，各种策略都可能出现。要证明一种策略是进化稳定策略，必须证明它能够“抵御”其他策略的“入侵”，就是说，当一种新的策略由于突变而产生时，不会战胜现有的策略。该模型中的随机混合策略是相当脆弱的，许多策略都有可能战胜它，比如“小气鬼”（miser）策略：与合作过或没有打过交道的对手合作；而一旦遭到背叛，就永不与背叛过自己的对手合作。又如“针锋相对”策略：一开始合作，以后每次都采取对手上次的策略（Dawkins, 1976）。仅从这一点看，该随机混合策略也不是进化稳定策略。

综上所述，本文认为，叶文的“利他行为演化均衡模型”不能成立。

四、人类利他行为与生物学意义上的利他行为

叶文认为要解释人类的利他倾向，就必须从进化角度对亲缘利他和互惠利他以外的（纯粹）利他作出解释。这里存在逻辑上的问题。产生问题的深层原因在于，混淆了作为社会科学研究对象的人类利他行为与生物学意义上的利他行为的概念。

在社会学领域，利他并没有一个统一的定义，但各科学者对利他的界定无不包含两个基本要素：第一，行为主体预期该行为能够增进他人的福利，包括生命、健康、金钱、实物、服务、名誉等；第二，行为主体不期望该行为增进自己的福利，即行为动机必须是非自私的。人类利他行为的判断标准主要是行为动机，而不是行为结果。通常人们会认为，“好人有好报”（怀着利他动机去做某事，无意中增加了自身的福利）或“好心办坏事”（怀着利他动机去做某事，但最终损害了他人的福利）是利他行为，而“无心插柳”（无利他动机，但客观上增进了他人的福利）或“偷鸡不成蚀把米”（怀着自利动机去做某事，最终却损害了自身的福利、增进了他人的福利）不是。当然，在多数情况下，行为动机与行为结果是统一的。

而生物学判断利他行为的标准是行为的客观结果。只有当施惠者的一种行为客观上导致自身适应性降低、受惠者适应性提高时，该行为才是生物学意义上的利他行为。

亲缘利他和互惠利他在表面上或短期内看是利他，但归根结底还是自利，似乎不符合人类利他行为的标准。故而叶文认为，只有亲缘利他和互惠利他

以外的(纯粹)利他才能与人类利他相对应。这是不对的。首先,绝大多数人类利他行为对本人的适应性毫无影响,根本不属于生物学意义上的利他,例如做好事不留名(没有危险的情况下)、为团体利益牺牲个人利益(不危及自己的生命和后代)等。其次,有些行为以人类标准衡量是无私利他,但从生物学角度看是亲缘利他,例如向近亲捐赠人体器官的行为,父母在生死关头不顾自身安危保护儿女的行为等。再次,有些人类行为从生物学角度看既非亲缘利他,也非互惠利他,其最终效果是减少了自身的适应性,增进了他人的适应性,可以说是生物学意义上的纯粹利他,但按人类标准衡量却不是利他行为,例如一些人(通常是高收入者)出于自利目的而减少生育数目的行为。此外,也有一些人类利他行为同时也属于生物学意义上的纯粹利他,如出于高尚目的而收养没有血缘关系的儿童,舍己救人(无血缘关系)的英雄行为等。但这类行为只是人类全部利他行为的一小部分,而且没有理由认为这是人类利他行为中起源最早的部分。简言之,人类利他实际上是一种行为动机,生物学意义上的纯粹利他是一种行为结果,二者是两个不可通约的概念。因此,前者在人类社会中的存在不必以后者在生物进化过程中的存在为前提。

事实上,从生物学意义上的利他定义本身即可推出,纯粹利他行为不可能是生物进化的产物。一种利他行为必须要通过某种途径得到回报,最终增进生物个体的适应性,否则必将被自然选择过程所淘汰。亲缘利他和互惠利他就是通过亲缘关系或互惠合作关系迂回地增进了生物个体自身的适应性,才得以在进化中延续并扩散。那有没有可能出现一种既非亲缘利他,又非互惠利他的非纯粹利他行为呢?有可能。强互惠行为可能就是这样一种利他行为。本文将在第六部分对此进行讨论。

五、利他行为与经济学

1. 利他行为与经济学基本假设

经济学必须对人的行为方式作出必要假设,以此为基础展开研究。早期经济学家一般认为人拥有双重特性:既是对自身福利感兴趣的自利者,又是具有公益精神、能关心他人福利的利他者。经济学之父亚当·斯密一生留下两部著作:《道德情操论》和《国富论》,分别研究了以利他和自利动机为基础的两种人类行为及其对社会的影响,这已为人们所熟知。不那么为人所知的是,以在经济学研究中使用数学方法而闻名的埃奇沃思也认为人不是完全自利的。在他的合约和交换理论中,埃奇沃思探讨了所

有参与者力图最大化共同利益的情形。即使在个人最大化经济利益的行为中,埃奇沃思也引入了一个利他系数,因为他认为个人行为可以是出于对他人利益的同情。

“边际革命”之后,随着经济学与伦理学之间的界限日益明晰,经济学对利他的重视渐渐淡化。多数经济学家认为自利假设基本符合现实。如斯蒂格勒所说:“在我们生活的世界上,有理性,并掌握充分信息的人们,在机智地追求着他们的自利”;“当自利与伦理价值的口头忠诚发生冲突时,如果让我来预测关于行为的系统与广泛的试验结果,那么,在很多情况下,事实上,是在绝大多数情况下,自利理论将会获胜”(Stigler,1981)。虽然如此,但效用理论“并不包含任何有关利己主义在人类行为中的作用的假说”(约瑟夫·熊彼特,1952)。在现代经济学中,以数学语言表述的“理性行为”假设只包含两个条件:偏好的完备性和传递性。其含义是,只要有一束能够定义并排序的偏好,并能在给定的约束条件下最大化偏好,人的行为就被认为是理性的。偏好中可以包含他人的福利,也就是说,理性人可以有利他倾向。现代经济学分析框架并不排除在利他行为假设下讨论制度安排和个人的权衡取舍选择问题。

2. 经济学视野中的人性

经济学对人的行为方式的基本假设是以对人性的认识为基础的。何谓人性?人性是否只限于直接通过生物进化产生的特性?

人性有多种定义。在经济学家眼中,人性是指在长时期内、不同文化背景下,人类行为所具有的不变特性。一种特性,如果对人的行为有明显的、系统性的影响,不论它是不是通过进化得来的,都被经济学视为人性的一部分。

我们怎么知道他人是理性的?米塞斯回答:观察(Kirzner,2000)。同样,我们怎么知道人是自利的还是利他的?也是通过观察。90年代以来,人性中有利他倾向已为越来越多的经济学家所认同,就是因为实验观察为人的利他倾向提供了越来越多的证据。

近年来,行为经济学通过实验发现,人的行为与传统的自利假设有着显著的、系统性的差异。在实验条件下,人们的初始禀赋被标准化为筹码或现金,人们的行为由于必须遵从实验规则而大大简化,因而人们的偏好可以清晰地显示在实验数据中。使用多种博弈规则和实验框架,在数十个国家所做的数百个实验表明,人们关心公平和互利,并愿意牺牲自己的所得来改进物质资源的分配状况,有时还会采取行动去奖励亲社会(pro-social)行为,或惩罚不愿

做出亲社会行为的人。而且,人们的利他行为是理性的,符合一般性显示偏好公理(generalized axiom of revealed preference),包含了利他行为的效用函数是连续、单调的凸函数。

“经济学家所研究的是实际存在的人”(阿尔弗雷德·马歇尔,1920)。观察表明实际存在的人具有哪些特性,经济学就应该以哪些特性作为对人的行为的基本假设,而不论这些特性是来自生物进化过程,还是源于社会因素。人的利他倾向是不是“天然”,与利他策略是不是进化稳定均衡无关。

3. 经济学需要将利他偏好内生化的吗?

一切科学的理论归根结底都是建立在若干公理假设基础上的逻辑体系,而这些公理假设是本学科无法证明的,只能通过归纳法(观察)得出。经济学也不例外。

经济学一旦假设人有某种偏好,便将之视为公理。标准的个人效用函数中包含消费品和闲暇,经济学从来不试图将人们对消费品或闲暇的偏好内生化的。同理,如果假设人有利他偏好,那么效用函数中就应引入他人的福利,但毋需将其内生化的。

当然,一个学科的基本假设越少越好。爱因斯坦指出:“科学的目的是,一方面是尽可能完备地理解全部感觉经验之间的关系,另一方面是通过最少个数的原始概念和原始关系的使用来达到这个目的(在世界图像中尽可能地寻求逻辑的统一,即逻辑元素最少)”。经济学家希望对人性的假设尽可能简单。如果能在更少的公理基础上解释人的自利-利他偏好,对经济学理论来说是件好事。Becker(1976)就是朝这个方向所作的努力。但,把利他偏好视为外生给定的,完全不会妨碍经济学研究。

4. 把利他纳入经济学研究框架所面临的困难

既然现代经济学的分析框架并不排斥利他偏好,实验证据又显示人有利他倾向,那为什么对利他行为的经济学实证研究迄今仍如此少见呢?

理论与实证并重,是经济学研究方法的根本特征。经济学家并非不愿对利他行为作实证研究。然而要把利他行为纳入经济学研究框架,就必须从数量上对利他动机的作用加以预测和检验。不幸的是,利他行为很难加以量化。这是因为,直接或间接的自利行为归根结底是一种交换,交换能够形成价格(不管这种价格是否采取货币形式)并显示出该行为的成本。成本是通过外在效果衡量主观动机的标尺,是经济学分析人的行为的必要条件。相反,因为利他行为与交换无缘,所以无法显示出成本,也就难以量化并被纳入经济学的研究范围。马歇尔在《经济学原理》中指出:“责任感和对邻人之爱所激起的

那些活动,大多是不能分类、被归纳为规律和加以衡量的;因为这个理由——并非因为这些活动不是基于利己心的缘故——经济学的方法才不能用之于这些活动”(阿尔弗雷德·马歇尔,1920)。

虽然近年来行为经济学有了不少进展,但目前对利他行为的实证研究还仅限于以大学生为对象的实验,对现实生活中各种各样的利他行为尚无法进行系统的测量和记录。或许今后行为经济学的进一步发展以及社会生活的网络化、数字化能够逐渐解决这一难题,但可以预料的是,在相当一段时期内,自利假设必然仍将主宰经济学研究,自利-利他假设只能应用于少数特殊场合。目前以自利-利他假设为前提的经济学研究主要集中在:家庭内部或亲属之间的利他行为;利他性遗赠或慈善行为;以及利他动机对公共政策的可能影响等狭窄领域。

六、利他行为的经济学解释

1. 利他行为与道德

人是生物进化和社会发展的综合产物。生物学家、社会心理学家普遍认为,人类的利他倾向——确切地说是利他动机——源于道德。而道德又分为两个方面:个人道德情感和社会道德规范。

道德情感是道德规范在个人意识中的对应物,是利他行为的微观基础。按照常识,人类的利他行为总是在同情、内疚、正义感等道德情感驱使下做出的。不同学派的哲学家都认为,道德与情感密不可分。斯密在《道德情操论》中强调了道德情感,特别是怜悯或同情在美德中的重要地位。现代心理学也认为,同情心是对困境中的他人实施利他行为的重要条件。从经济学角度看,利他之所以成为人的一种偏好或者说行为倾向,正是由于利他行为能够通过道德情感给个人带来心理上的满足。

道德情感显然有其进化根源。按照人类行为学家的观点,与人类几百万年的进化史相比,人类的大规模聚居是非常晚近的事情,以至于现代人从生物意义上讲仍然是原始部落成员,内心深处本能地认为与自己交往的人都是自己的亲属(Morris,1969)。社会生物学家也指出,在人类历史上的大部分时间里,主要的社会单位都是家庭和密切的亲属关系网(Wilson,1975)。最早的城市出现在距今约8000年前。此前人类一直生活在非常狭小的交往范围内,交往对象主要是亲属,即使不是亲属,彼此也必定非常熟悉,可以建立长期合作关系。各种道德情感就是在长期的亲缘选择和互惠合作过程中进化而来的。正因如此,有时人们的利他行为显得非常情绪化。

道德情感与其他动机一起支配着个人的行为,通过人与人之间行为的相互影响,外化为一种对所有群体成员都具有约束力的行为规则,即道德规范。随着人类群体范围的扩大,道德规范的作用范围也逐步扩大到家庭、亲属、熟人圈以外。生物学意义上的纯粹利他行为开始在人类社会中出现。一些新的道德规范产生了,例如爱国主义。道德规范对个人的道德情感和行为方式具有反作用。心理学研究显示,儿童和青少年随着年龄增加、个人对道德规范敏感性增加、道德准则逐渐内化,利他行为逐渐增加。那些道德判断水平较高、或认为道德规范对自己具有强制力的人更倾向于无私帮助他人。

2. 经济学如何解释利他行为

如前所述,仅仅把人的利他倾向当作一个外生假设丝毫无碍于经济学的科学性。但是,为了简化经济学的基本假设,同时也为了加深对人性的理解,经济学可以把解释利他行为作为一个课题。什么样的解释算是经济学解释呢?经济学从权衡取舍角度考察人的行为。所谓权衡取舍,不应理解成有意识的计算和比较,而应理解成某种行为的结果使得行为主体的效用最大化,从而显得好像是经过了权衡取舍一样。正是从这个意义上讲,现代生物学对亲缘利他和互惠利他的解释也可以看作是经济学解释。

人类利他行为的判断标准是行为动机。经济学上的利他偏好也是一种行为动机。基于上一节的分析,利他动机可归结为道德情感。而道德情感一方面植根于亲缘选择和互惠合作,另一方面又是道德规范内化的结果。因此,要解释利他行为,必须回答以下几个问题:第一,道德情感是如何在进化过程中产生并延续下来的;第二,道德情感如何通过人与人之间行为的相互影响外化为道德规范;第三,道德规范如何影响个人的行为方式。

第一个问题已经在经济学逻辑体系内得到了解释。根据认知心理学家的研究,情感属性是那些有利于生物生存的功能上的涌现特性(Johnson, 1999),即:一种使生物个体倾向于采取某种行为的基因也会作用于神经系统,使个体产生“必须如此行动,才能安心、满足”的感觉。促使人们采取利他行为的情感可视为“利他/合作”基因的涌现特性。如果亲缘利他和互惠利他有利于生物的生存,那么作为这些行为的涌现特性的情感,包括同情、对亲人的爱、对团体的忠诚、得不到回报的愤怒、自己疏于回报的歉疚等,也会与亲缘利他和互惠利他一起被自然选择筛选并保存下来。因此,证明了亲缘利他和互惠利他能够成为进化稳定均衡,也就从经济学角度解释

了上述道德情感(包括其神经基础)的进化根源,表明经过自然选择和进化产生的人类心智与人类行为,不仅与自利心相容,而且与利他心相容。

第二个问题实际上是给定偏好下的制度形成问题。由于进化稳定策略不是唯一的,所以不是每个个体都具有相同的道德情感,也就是利他偏好。但在一定条件下,通过人与人之间行为的相互影响,某种利他行为有可能在群体中扩散开来,成为对所有群体成员都有约束力的行为规则。Becker (1976) 论证,在与关心他人福利的利他者交往时,利己者为了最大化个人效用,最优策略是像利他者那样行动,也就是假装利他。那么从效果上看,关心他人福利就外化成了一种道德规范。关于这个问题,经济学家还有许多工作可做,因为人类社会中存在多种多样的道德规范,其形成机制以及对社会成员的约束方式也是多样的,有待于进一步解释。

对于第三个问题,目前最完备的解释是 Simon 提出的“顺从性”理论。该理论认为,人必须通过社会学习,获取关于生活技能和“正当行为”(即社会公认的价值观、行为规范)的知识。擅长社会学习的人乐于学习并相信社会中的其他人希望他们学习并相信的东西,这种人被称为“顺从的”。由于人的理性的有限性和生存环境的复杂性,顺从能够极大地增加人的适应性。顺从者在接受社会教育的过程中,有可能不考虑或者不能充分考虑是否对自己有利,而接受利他观念,并采取生物学意义上的纯粹利他行为。这样,顺从-利他者以牺牲自己的适应性为代价,为他人和社会做贡献,提高了整个社会的平均适应性。同时,由于利他观念只是社会灌输给顺从者的生活技能和正当行为中的一小部分内容,顺从-利他者从社会学习中得到的全部好处远远大于利他行为的成本。所以,顺从-利他者的适应性将超过不顺从-利己者的适应性,生物学意义上的纯粹利他行为因而能够在人类社会中存在并扩散。也就是说,生物学意义上的纯粹利他行为只是顺从-利他者通过社会学习得来的全部生活技能和正当行为的一个子集。这种行为本身降低了顺从-利他者的适应性,而顺从-利他者的全部行为却增进了自身的适应性。这就解开了生物学意义上的纯粹利他行为何以能在人类社会中存在之谜,同时从宏观上解释了道德规范如何影响个人道德情感。不过,人与人之间毕竟有着偏好差异,即便是顺从者的行为方式也不可能完全相同。作为一种非正式制度的道德规范如何对个人的利他行为产生影响,还需要经济学作进一步研究。

3. 道德情感与道德规范在合作过程中的共同

发展——一种可能机制

上述第一个和第二个问题不是截然分开的。也就是说,自然选择对道德情感的筛选过程与道德规范的形成过程可以相伴进行。桑塔费学派关于强互惠行为的研究成果揭示了某些道德情感和道德规范在进化过程中共同发展起来的可能机制。

强互惠是指与他人合作,并以个人利益为代价去惩罚那些破坏合作规则的人。强互惠行为在现实生活中有广泛的经验基础。人们普遍憎恶欺骗、失信、背叛,并倾向于对背信弃义的行为实施惩罚,即使个人需要为此付出成本也在所不惜。行为经济学研究为此提供了规范化的实验证据(Gintis, et al., 2003)。强互惠行为与主流生物学家定义的互惠利他行为不同。后者(以下称“一般互惠利他”)是一种能够直接增进其他个体的生存适应性的合作行为,并能通过其他个体的相同的合作行为得到回报。而前者的关键部分是一种“利他惩罚”(altruistic punishment),该行为本身不能直接增进其他个体的生存适应性,而是通过维护合作秩序增进整个群体的平均适应性;该行为也不是通过其他个体的相同行为得到回报,而是通过群体中合作行为的增加而得到回报。

Gintis(2000)建立了一个理论模型,证明在群体周期性地面临灭绝威胁的条件下,一般互惠利他将瓦解,而强互惠行为可以在进化过程中延续。因为当群体灭绝的可能性很大时,一般互惠利他所获回报的期望值非常低,采取一般互惠利他行为的个体将被利己者所淘汰。换言之,互惠合作只在不太坏的生存环境中才可能对个体有利;大难将临时,只顾自己是最优策略。而强互惠者虽然不能直接从自己的利他惩罚行为中得到回报,但至少可以在群体中维持一个由部分成员组成的小合作团体,彼此从合作行为中受益。与此同时,破坏合作秩序者(利己者)因受到强互惠者的惩罚,生存适应性降低。这样,强互惠行为就得以在群体中扩散开来。

Bowles 和 Gintis(2004)用计算机仿真实验模拟了距今 10 - 20 万年前渔猎 - 采集部落的人类生活,表明在一定条件下,少数强互惠者可以侵入全部由利己者组成的人群,并导致合作者人数增加。经过千百代的进化,强互惠者、合作者和利己者的比例将大致稳定下来。该结果不需要假设群体面临灭绝威胁,但需要对群体的生存状态和生存环境,包括群体规模、群体个数、突变率、迁移率、生产方式、交往方式、惩罚方式等作出十分细致的假设。由于假设条件过多,该结果无法用理论模型加以证明。而且,该仿真实验没有引入任何一般互惠利他策略,因而不

能表明强互惠在一般情况下是进化稳定策略。“我们不能断言人类的强互惠倾向是以我们描述的方式进化来的。但我们的仿真实验表明有这种可能”(Bowles and Gintis, 2004)。

“强互惠者自身承担代价,而使团体成员受益,从这个意义上讲,强互惠行为是利他的一种形式”(Gintis, 2000)。但从生物学意义上讲,强互惠行为仍然不是纯粹利他。虽然强互惠者并未期望得到回报,也不是从其他个体的相同行为中直接得到回报,因而不是一般意义上的“互惠”,但事实上强互惠行为还是通过团体中其他成员的合作行为得到了回报,该回报弥补了强互惠者实施惩罚行为的成本。可见,以行为效果衡量,强互惠行为有利于行为者自身,只不过与一般互惠利他相比,强互惠行为获得回报的途径更加曲折,要求的条件更加严格。

强互惠理论显示了道德情感和道德规范共同演进的一种可能性。最初的强互惠者由基因突变产生,可将其视为先天具有一种嫉恶如仇的道德情感,见到破坏合作规则者便欲惩罚之而后快(不妨将强互惠基因的这种涌现特性称为正义感)。Fehr 等(2004)初步证实了这种道德情感的神经基础。客观上,强互惠行为维持了合作秩序,改善了合作者的生存条件,降低了不合作者的生存机会,使群体在恶劣生存条件下不致灭绝,并最终有利于自身。随着强互惠基因在群体中的扩散,具有正义感的个体渐渐增多,同时合作者也渐渐增多,不合作者通常会受到惩罚,合作成为一种道德规范。

七、结论

人类利他行为与生物学意义上的纯粹利他行为是两个不同的概念,前者在人类社会中的存在不必以后者在生物进化过程中的存在为前提。证明(纯粹)利他是进化稳定均衡一个不可能完成的任务,这是由生物学对利他的定义所决定的。人类利他行为是生物进化和社会发展的综合结果。伴随着长期的亲缘选择和互惠合作进化而来的道德情感是人类利他倾向的根源,而人们在社会交往过程中形成的道德规范巩固和扩展了这一倾向。经过自然选择和进化产生的人类心智与人类行为,不仅与自利心相容,而且与利他心相容。

现代经济学的分析框架并不排斥利他偏好。随着行为经济学发现越来越多的证据表明人有利他倾向,经济学应当把利他行为纳入研究范围。但由于利他行为很难量化,故在相当长的一段时期内,对利他行为的实证研究仍将极其有限。这不是任何人的主观愿望所能改变的。(下转第 70 页)

人力资本的高回报将吸引人力资本从不发达地区流向发达地区,加速发达地区的经济增长,从而导致更高的人均收入,继续吸引人力资本流入,地区间的经济增长将呈现出俱乐部收敛现象,富的地区越富,穷的地区越穷。要使区域经济协调发展,加快不发达地区的增长速度,可以采取各种合适的经济手段,来提高不发达地区的人力资本水平,或者是扭转人力资本流动方向,使其从发达地区向不发达地区回流,促进不发达地区的经济增长,达到整个经济的趋同。

参考文献:

1. Arrow, Kenneth J., 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies*, Vol. 29, June, pp. 155 - 173.
2. Barro, Robert J., 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, Vol. 106, No. 2, May, pp. 407 - 443.
3. Barro, Robert J. and Xavier Sala - I - Martin, 1997. "Technological Diffusion, Convergence, and Growth." *Journal of Economic Growth*, March, pp. 1 - 26.
4. Baumol, William J., 1986. "Productivity Growth, Convergence, and Welfare: What the Long - Run Data Show." *American Economic Review*, Vol. 76, No. 5, December, pp. 1 072 - 1 085.
5. Ben - David, Dan, 1998. "Convergence Clubs and Subsistence Economics," *Journal of Developmental Economics*, Vol.

55, February, pp. 153 - 169.

6. Deardoff, Alan V., 2001. "Rich and Poor Countries in Neoclassical Trade and Growth." *The Economic Journal*, Vol. 111, pp. 277 - 294.

7. De Long, J. Bradford, 1988. "Productivity Growth, Convergence, and Welfare: Comment." *American Economic Review*, Vol. 78, No. 5, December, pp. 1 138 - 1 154.

8. Lucas, Robert E. Jr., 1988. "On the Mechanics of Development Planning." *Journal of Monetary Economics*, Vol. 22, No. 1, July, pp. 3 - 42.

9. Romer, Paul M., 1986. "Increasing Returns and Long - Run Growth." *Journal of Political Economy*, Vol. 94, No. 5, October, pp. 1 002 - 1 037.

10. 蔡昉等:《制度、趋同与人文发展——区域发展和西部开发战略思考》,北京,中国人民大学出版社,2001。

11. 蔡昉、都阳:《中国地区经济增长的趋同与差异——对西部开发战略的启示》,载《经济研究》,2000(10)。

12. 刘强:《中国经济增长的收敛性分析》,载《经济研究》,2001(6)。

13. 魏后凯:《中国地区经济增长及其收敛性》,载《中国工业经济》,1997(3)。

14. 王绍光、胡鞍钢:《中国:不平衡发展的政治经济学》,北京,中国计划出版社,1999。

(作者单位:东北财经大学产业组织与企业组织研究中心 大连 116025
东北财经大学研究生部 大连 116025)
(责任编辑:N)

(上接第 48 页)

经济学没有必要非把利他偏好内生生化不可。将利他偏好当作一个外生假设丝毫无碍于经济学的科学性。不过,经济学可以通过解释道德情感的进化根源、道德规范的形成机制、道德规范对个人行为方式的影响,来解释人为什么有利他行为。经济学家在这些方面已经做了一些工作。其中,桑塔费学派关于强互惠行为的研究成果揭示了某些道德情感和道德规范在进化过程中共同发展起来的可能机制。

参考文献:

1. 阿尔弗雷德·马歇尔:《经济学原理》,中文版,北京,商务印书馆,1964。
2. 叶航:《利他行为的经济解释》,载《经济学家》,2005(3)。
3. 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)。
4. 约瑟夫·熊彼特:《经济分析史》,中文版,第3卷,北京,商务印书馆,1996。
5. Becker, G. S., 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *Journal of Economic Literature*, Vol. 14, No. 3, Sep., pp. 817 - 826.
6. Bowles, S. and Gintis, H., 2004. "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations." *Theoretical Population Biology*, Vol. 65, pp. 17 - 28.

7. Dawkins, R., 1976. *The Selfish Gene*. Oxford Press.

8. Edgeworth, F. Y., 1881. *Mathematical Physics, An Essay on the Application of Mathematics to the Moral Sciences*. Reprints of Economic Classics, Augustus M. Kelley Publishers, 1967.

9. Fehr, E. et al., 2004. "The Neural Basis of Altruistic Punishment." *Science*, Vol. 305, 27 August.

10. Gintis, H., 2000. "Strong Reciprocity and Human Society." *Journal of Theoretical Biology*, Vol. 206, pp. 169 - 179.

11. Gintis, H., et al., 2003. "Explaining Altruistic Behavior in Humans." *Evolution and Human Behavior*, Vol. 24, pp. 153 - 192.

12. Johnson, V. S., 1999. *Why We Feel: The Science of Human Emotion*. Cambridge, MA: Perseus Books.

13. Kirzner, I. M., 2000. "Human Nature and the Character of Economic Science: The Historical Background of the Misesian Perspective." *The Harvard Review of Philosophy*, Vol. 8, pp. 14 - 23.

14. Morris, D., 1969. *The Human Zoo*. Jonathan Cape.

15. Sober, E. and Wilson, D. S., 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge MA: Harvard University Press.

16. Stigler, G. J., 1981. "Economics or Ethics?" in S. McMurrin, ed., *Tanner Lectures on Human Values*, Vol. II. Cambridge: University Press.

17. Wilson, E. O., 1975. *Sociobiology. the New Synthesis*, Harvard, Belknap Press.

(作者单位:清华大学经济管理学院 北京 100084)
(责任编辑:N)