

互惠伦理的内涵、依据及其践行

龚天平 胡莹

摘要 互惠伦理是人类伦理价值规范系统的重要构成。互惠意指个体基于自我意愿或利益帮助他人以增进他人利益并期待未来可以获得某种回报的动机或行为;互惠伦理意指交往各方以互惠方式,把自己的自利或利他偏好付诸行为,且对方应当自觉负有回报义务的伦理。互惠伦理主要包括积极互惠伦理和消极互惠伦理。互惠伦理之所以可能,是建立在行为能力、理性能力、同感能力等一系列具体条件的基础上的。只有得到人们的恰当践行,互惠伦理才能走向稳定。互惠伦理践行机制主要包括社会奖惩制度、他人评价与监督等外在约束和个体内心道德自律即内在约束,实践中还需要合理对待欺骗者,互惠行为也要禁害第三方以趋善避恶。

关键词 互惠伦理的含义;互惠伦理的可能性;互惠伦理的践行

中图分类号 B82 **文献标识码** A **文章编号** 1672-7320(2025)04-0065-11

基金项目 国家社会科学基金项目(21BZX023)

人是关系性存在,关系意味着互惠,因而也可以说人是互惠性存在。互惠意味着协作行动,即人与人在情感、行为等方面相互顾及、相互考虑、相互成全。德国行为经济学家阿明·福尔克说:“人类行动的基本原则之一是互惠:当别人合作时,我们才会合作;当别人对我们好时,我们才更愿意公平地对待他。”^[1](前言Pxi)这意味着,互惠是一个伦理规范,它表明人们都是以合作响应合作、以公平回报善举的,或者说都是既考虑自身获得的物质利益和其他私利,也考虑他人利益的。因此,与幸福、义务、良心、自由、平等、公正等一样,互惠也是一种具有典型意义的伦理,即互惠伦理,它是人类伦理价值规范系统的一个重要构成。学界关于互惠的研究成果比较丰富,但专门论述互惠伦理的研究成果比较少见。那么,互惠伦理到底是何种含义,其何以可能,又如何践行并走向稳定?

一、何谓互惠伦理

行为和实验经济学把互惠称为互惠偏好,认为其是社会偏好中与利他偏好、不平等厌恶偏好具有同等典型意义的一种类型,最先由马修·拉宾给予界定,意指“当别人对你友善时你也对别人友善,当别人对你不善时你也对别人不善”,而且他还从经济学角度把其明确规定为“如果你在损失自己效用(收入、利益等)的情况下去损害别人的效用,就被定义为你对别人不善;如果你在损失自己效用的情况下去增进别人的效用,就被定义为你对别人友善”^[2](P62)。其后许多学者大都是在此种意义上理解互惠。

(一) 互惠的若干理解

对于互惠,学术史上许多思想家比如克鲁泡特金、亚当·斯密等都作过深入研究,但特别具有代表性并被广泛引用的大致有如下几种。

第一,马林诺夫斯基的互惠范畴。人类学家马林诺夫斯基曾提出这样一个疑问,为什么在原始社会中,野蛮人(savage)会遵守一些行为规范?他提出,野蛮人愿意遵守苛刻的行为规范,可能并非仅源于对

社会习俗的敬畏或对命令的顺从,也可能源于人与人之间相互关联的责任、义务,这可以在互惠原则中得到答案。他将互惠定义为“发生在一个长期稳定的合伙关系中,或与明确的社会关系有关,或与非经济事务中的相互关系相结合”^[31](P67),是一种相互满足的交换商品和服务的模式,或是对劳动分工的补充和实现。他认为,即使不是所有的经济行为,大多数经济行为也属于某种互惠赠礼和反赠礼的链条,从长远来看,这种链条是平衡的,双方都能平等受益^[31](P67-68)。即使是在原始社会,野蛮人也会因为想要维护社会秩序平衡而遵守相应的行为规范。由此看来,互惠是稳定社会秩序的有效标准。

第二,古尔德纳对互惠的理解。在阿尔文·古尔德纳看来,互惠是一种相互交换满足的模式(a pattern of mutually contingent exchange of gratifications),即一方给予帮助,另一方具有回报的义务^[4]。这种理解的独特性在于互惠被定义为一种道德规范,维护社会秩序、社会制度是其重要功能。在他看来,互惠是一种以增加对方利益为目的的双向互动,人们有义务回报曾经给予过自己帮助的人。其中的原理为,如果你想要得到别人的帮助,你就要帮助别人(这种帮助既要符合一般性规范,又要符合他人的特殊性要求,既是恰当的又是有效的)。詹姆斯·C.斯科特根据古尔德纳的理解将此表述为,作为一条渗透整个社会生活的道德原则,互惠“根植于这一简单观念:一个人应当帮助那些帮助过自己的人”,或者“至少不损害他们”,“它意味着被接受下来的礼品或服务为接受者带来了相应的义务——有朝一日要以相当的价值给以回报”^[5](P215-216)。

第三,特里弗斯的互惠理论。完整的互惠理论不能不提到罗伯特·特里弗斯,他基于理性视角提出了清洁共生生物的行为(behavior involved in cleaning symbioses)、鸟类的警告叫声(warning cries in birds)、人类的互惠(human reciprocal altruism)^[6]三个互惠模型。他认为,非人类物种只是出于本能经验而活动,非人类之间也会出现因付出未得到回报或被欺骗而产生的报复行为,但根据目前的生物学研究成果,只能暂将其归为动物本能的进化。与非人类不同,人类有自我意识,会认知、思考、辨识等,因而人类互惠行为具有更复杂的心理因素,如友情、讨厌、道德攻击(moralistic aggression)、感激、同情、信任、怀疑、值得信赖、内疚、不诚实、虚伪等^[6]。从近亲之间的互惠到非近亲但有感情基础的人之间的互惠最后到陌生人之间的互惠,付出一回报机制逐渐凸显。因为预期有回报所以会付出,因为付出所以产生回报意愿,这是互惠最突出的潜在规范。

在特里弗斯看来,互惠的生存进化可能与这些因素有关:彼此的寿命长短(可预估双方的互惠频率)、彼此的可替代性(可预估双方的互惠黏性)、彼此的信任指数(可预估双方的背叛概率)、彼此的亲密程度(可预估双方对付出—回报模型的容差程度)、社会公正(可预估背叛成本)。其一,互惠双方寿命越长,他们之间的关系越稳定,寿命不仅是对个体较有生存意义的数值,同时对彼此而言,对方寿命越长还意味着自我付出或被回报的频率越多,想到这一点,人们通常会与寿命更长者(身体更健康者)保持互惠关系。其二,互惠双方的不可替代性越强,这一互惠关系就越具有黏性;相反,彼此可替代性越强,人们就会优先选择与能够为自己带来最大收益的对象(如更值得信赖、更高效率、更低期待回报率等)构建互惠关系。其三,互惠双方会优先选择值得自己信赖的对象构建互惠关系,值得信赖就意味着被欺骗率低,即相信自己的付出能够得到回报的概率,互惠对象的可信任指数越高,付出能够得到的回报率也就越高。其四,互惠者之间的亲密程度越高,彼此感情越深厚,他们对付出一回报差额的包容度也就越高,人们普遍会对自己的亲人、朋友、爱人具有更深的包容心和共情心,他们的情感会抵消因付出与回报不相符而造成的部分差额,这就是人们所说的人情味。其五,社会的公平正义对维护互惠氛围提供了外在保障,建立在社会规范之上的社会奖惩制度具有的社会效益会使互惠关系背叛者受到应有的惩罚,当背叛成本大于背叛收益时,社会中的背叛行为就会有效减少。

第四,阿克塞尔罗德的互惠理论。互惠即意味着合作或联合行动,对合作作出有说服力的解释且被人们广泛借鉴的成果是罗伯特·阿克塞尔罗德的《合作的进化》。该书从“在每一个人都有自私动机的情况下,怎样才能产生合作”出发,通过三届“重复囚徒困境博弈计算机竞赛”,发现在每轮竞赛中获胜的都

是最简单的“一报还一报”或“针锋相对”策略。这些竞赛结果表明,“在适当的条件下,合作确实能够在没有集权的自私自利者的世界中产生”^[7](P14)。那么这些条件是什么?阿克塞尔罗德在书中提出了一系列命题,这些命题不仅说明了合作产生的条件,也说明了合作演化的进程。这些命题或条件可以总结为如下说法:“合作的进化要求个体有足够大的机会再次相遇,使得他们能形成在未来打交道的利害关系”^[7](P14),具体地说,就是个体之间的“交往”“相互回报”“基于回报的策略”和对“其他不太合作的策略的侵入”的防止^[7](P14)。当这些条件具备时,即便“没有友谊和预见,合作也可以产生”^[7](P47),因而“合作的基础不是真正的信任,而是关系的持续性”^[7](P126)。即是说,互惠是合作的基础。显然,阿克塞尔罗德的“一报还一报”重复策略立基于个体的自利,往往是“理性的行动者”习惯采用的策略。正如他本人所言,如果“对策者能通过对双方有利的可能性的试错学习、通过对其他成功者的模仿或通过选择成功的策略剔除不成功的策略的盲目过程”^[7](P126),即通过群体生活、交往即互惠的经验学会如何合作,那些常规性的合作方法也会随着时间的流逝和互惠经验的积累从而演变为文化的产物。

(二)互惠的几种类型

阿明·福尔克说:“互惠,既强大又普遍,几乎没有别的人类动机能与之相比。它表现在谈判中、市场上等各种场合……它存在于所有文化中,只是程度不同而已。”^[1](P117)这意味着,互惠是普遍的,但并不是所有互惠都是同质的,而是可以分为多种类型。

第一,互惠可以分为普泛互惠、对等互惠和负面互惠。马夏尔·萨林斯从三个方面来解释互惠:“回报的平等性(equivalence)、即时性(immediacy)和利益性(interest)”;斯帕罗和林登在萨林斯研究的基础上总结出三种互惠类型:普泛互惠(generalized reciprocity)、对等(平衡)互惠(balanced reciprocity)、负面互惠(negative reciprocity);乌尔-比恩和马舍林参照林登、斯帕罗和韦因的观点提出了两种类型的互惠规范:负面互惠规范(negative norm)和正面互惠规范(positive norm),这两者之间包含了负面互惠、对等互惠和普泛互惠三种类型。普泛互惠是指“交换各方不会明确说明对方给予报答的时间、回报资源的数量和质量”,“是给予者对他人利益行为的关注”的互惠;负面互惠“是交换各方明确规定回报的时间并强调回报资源的等价性”,“个体的目的是维护自己的利益并尽可能使自己的利益最大化,有时为了达到这种目的甚至会损害他人利益”的互惠;对等互惠“是交换各方同时给予对方等价的资源”,“接受者要在短时间内偿还赠予者价值相近的资源”的互惠^[8]。

第二,互惠可分为积极互惠和消极互惠。阿明·福尔克把互惠分为两种:一是积极互惠,意指“别人对我公平、友好,而我也对他们和善、友好、积极配合”,二是消极互惠,意指“当别人对我表现出不友好或不公平的行为时,我就对他们进行制裁和惩罚”^[1](P115)。

第三,互惠可分直接互惠和间接互惠。马丁·诺瓦克也针对互惠提出有直接互惠和间接互惠两种机制,“直接互惠,就是简单的平等交换原则。我给你挠了背,也指望着你能为了回报我而给我挠挠背”^[9](P36);间接互惠是“如果我给你挠了背,我就树立起了一个好榜样,可以鼓励他人效仿这个做法,如果足够幸运的话,也会有人给我挠背”^[9](P72)。

第四,互惠可分为强互惠和弱互惠。桑塔费学派鲍尔斯、金迪斯等将互惠分为强互惠和弱互惠两种。强互惠包括两方面:一是关心他人福利、维护社会规范的行为;二是“即使付出个人成本”“也会惩罚那些盗用他人合作行为成果的人”,从而为群体利益或联合项目的成功做贡献的行为,这种行为能产生如此道德后果,即或者“满足、骄傲甚至欢欣”的积极后果,或者因不如此行为而成为“羞耻和内疚的源泉”^[10](P1-2)的消极后果。强互惠的“强”是指即便行动者需要承担成本甚至担负成本后不太可能有净收益也仍会采取的利他行动,此行为又称利他性惩罚(Altruistic Punishment)。对此,叶航形象地说道:“所谓强互惠行为,就是我首先和别人合作,如果对方背叛合作,哪怕这种背叛不是针对我,我也要进行惩罚,甚至不惜花费个人成本……事实上,这就是我们人类所特有的‘正义感’。”^[11]弱互惠是指以自利为动机的互惠,主要包括基于自利的合作、市场交换、礼尚往来等^[12](P271)。

综合以上理解,我们可将互惠定义为个体基于自我意愿或利益帮助他人以增进他人利益并期待未来可以获得某种回报的动机或行为。阿明·福尔克说:“互惠,指的是一个人回报友好行为和惩罚不公平行为的动机,即使这种回报或惩罚不会给他带来任何战略意义上的好处。”^[11](P117)用伦理学的语言来说,这其实就是善以善报,恶以恶惩。当人们运用互惠来调节人际关系时就会形成互惠伦理,其意是指交往各方以互惠方式,把自己的自利或利他偏好付诸行为,且对方应当自觉负有回报义务的伦理。这样看来,互惠就是一种道德规范。“互惠义务是一条典型的道德原则……它既适用于地位相同的主体之间的关系,也适用于地位不同的主体之间的关系。”^[15](P217)

互惠与道德在内涵上是一致的。道德在于“利益的相互性、对等性和平衡性”,它既为自利辩护又反对极端自利,既鼓励利他又反对极端利他。在道德领域,自利并不是完全不道德的,适度自利恰恰是道德应该为之提供理由的行为,不道德的是那些不顾及他人利益的极端自利;利他是道德的,但也并非没有边界,道德并不无条件地提倡自我牺牲式的利他,虽然这种行为令人感佩,但以此要求普通人,就与道德渐行渐远了。道德的实质是“某个人在某个时候为另一人付出精力和注意力,并期待着某个时候会有一种对此的回报”^[13](P63),“‘道德行为最好的论证是相互性原则’,道德对于所有的人都应是物有所得”^[13](P63-64)。互惠也同样是这种含义。它意味着“某人暂时把自己的精力与注意力放在另一个人的身上,是指望某个时候能得到‘回报’”^[14](P256),“道德层面的互惠法则,即‘你希望别人怎样对待你,你就该怎样对待别人’”^[15](P56)。因此,无论是道德还是互惠都不过是自利与利他、付出与回报对等、平衡的浓缩化、精炼化表达。道德就意味着互惠,互惠也意味着道德。

至于互惠伦理的类型,我们赞成阿明·福尔克从性质上对互惠的类型划分,即积极互惠伦理和消极互惠伦理。前者指我对人好,人对我好;后者指人对我不善,我必惩人。对互惠伦理作这种性质上的划分,有助于恰当厘定互惠伦理在道德规范体系中的地位。如果我们把道德规范分为提倡性的高线层次规范(比如奉献)、允许性的中线层次规范(比如道德权利)、禁止性的底线层次规范(比如禁害),那么积极互惠伦理就属于中线层次,消极互惠伦理因较为复杂而不能简单化处理,一方面,消极互惠伦理包括第三方的利他性惩罚,这是允许的;另一方面,消极互惠伦理中的“人对我不善,我必惩人”实质上意味着互害因而属于底线层次。

二、互惠伦理何以可能

互惠伦理一般以两种形式存在,一是动机,即互惠意愿;一是行为活动,即互惠行为。当人们在动机驱动下把互惠意愿转化为互惠行为,就形成互惠伦理关系。从互惠意愿到互惠伦理关系,是借助行为得以完成的。但是,互惠从意愿转化为行为并非自然而然的,而是建立在一系列具体条件上的,当互惠意愿不能转化为互惠行为,互惠伦理关系就难以形成。那么,互惠伦理需要一些什么条件才能成为可能?

(一) 行为能力

从伦理学角度看,互惠实际上就是自利与利他的平衡、协调。互惠“不是单纯由他者所决定,也不是完全由自我所决定的,而是在自我与他者的交互关系中所决定的”^[16],只有在这种交互关系中自利和利他都能完成,互惠才能达成。根据互惠行为的难度系数,我们可以将其大致分为能够轻易完成的、难以完成但又有意愿完成的、难以完成而放弃的三种。对第一种,行为者不需要付出过高成本就可以为他人带来利益增长。因为行为者并未付出过高成本,因对回报的期待也不会过强,甚至不要求任何回报。对方的轻微回报都可能良好地维持这段关系,使互惠伦理关系得以呈现。对第二种,虽然行为者需要付出高额的成本,但是过强的意愿依旧促使其将互惠意愿转化为行为。过强意愿中包括过强的利他情感,这种情感可能产生于对方曾经为自己提供过的援助,也有可能是对方的处境激起了自己的恻隐之心,或是因为与对方关系亲密等。此时,行为者在完成互惠行为时付出了高额的成本,难免会产生对回报的预期,因而需要对方在当下或未来回报这一行为以维持彼此的互惠伦理关系。对第三种,即使行为

者具有互惠意愿,但因现实状况超出自己的能力,导致意愿并未付诸行为。如果强迫人们完成自己无法完成的行为,就可能会给行为者造成巨大的行为负担和心理负担,既不利于互惠的产生,也可能使其作出危害他人、社会的行为。是否具有与需求信号相符的行为能力对互惠意愿是否能够现实化至关重要。人们只有不断提升自己的行为能力,关注自利同时也超越自利,才有助于社会形成良好的互惠伦理氛围。

行为能力有限就可能会产生一个问题:对方无回报能力而拒绝互惠邀请,那么此时不接受他人的互惠邀请是不道德的吗?即使我们可以假设互惠者能够预判对方是否具有回报能力,但在现实中,这一预判能力很难准确发挥作用。当不知情的互惠邀请者发出互惠信号,被无力回报者拒绝时,互惠邀请者可能会觉得对方不友善;当对方接受互惠邀请者的付出而无回报时,互惠邀请者也可能觉得自己受到了欺骗。但事实并非如此,假若对方不具有回报能力,却强行将对方拉入付出一回报式的交往模式,会给对方造成过重的心理负担。因而在践行互惠伦理时,人们应提升自我预判能力,避免强人所难,学会理解他人、关怀他人、尊重他人,既要分辨搭便车者,又要避免因自我苛刻而使善良但无回报能力者受到伤害。可见,仅具有行为能力还不足以使互惠意愿以道德的方式付诸实践,行为者还需基于自我理性能力、感性能力对真实情景作出具体情况具体分析。

(二) 理性能力

理性是人特有的能力,有助于人们对互惠伦理关系的可行性、合理性、公正性、平等性等展开分析、评判,也有助于人们预估这一关系的后续影响。

第一,理性使人们关注利益所能达到的最大程度。在福利经济学领域,涉及生产资源最优配置时,最有公平和效率的为帕累托最优,即付出成本不变而实现利益最大化,帕累托改进则是达到帕累托最优的途径和方法,即通过调整资源的分配,实现最优化生产,此时的分配方式与生产结果均为最优,无需任何改进。在构建互惠伦理关系时,人们会基于对方的可信度、寿命(涉及互惠伦理关系保持的长久程度,避免互惠中途中断)、不可替代性、处境、能力、道德水平、知识体系等,预估对方的行为能力、互惠意愿等,也会判断互惠伦理关系中自我正当利益是否得到了相应的最优化提升。

第二,理性使人们关注利益分配的合理程度。互惠伦理关系是互惠各方基于付出一回报模式构建的关系。理性行为者会考虑这一关系的合理程度,如稳定性、平等性、公正性等。其中,稳定性取决于行为双方对彼此的善意、双方的寿命、行为能力、道德水平以及付出与回报的差额;平等性与公正性也可以参照互惠双方各自付出与回报差额的差额。在讨论利益分配时,我们把重点放在付出与回报的差额上。

假设一种情景,A、B之间已经达成互惠意愿,A做有利于B的事,付出成本为M,为B带来的效益增加为M',B回报A有益于A的事,付出成本为N,为A带来的效益增加为N'(其中,M为A的主观付出,M'为因A的付出为B带来的主观效益,N为B主观付出即回报,N'为因B的回报为A带来的主观效益)。因为A、B的身份地位、心理状态、认知水平、道德能力等不同,且A、B对自我付出与效益的衡量具有主观性,难以达成 $M=M'$, $N=N'$, $M=N$ 。因此,A、B之间的互惠平衡可以依据 $M=N'$, $M'=N$ 来判断,此时,A、B的付出符合他们的预期收益,实现了建立在合理性之上的互惠伦理关系的动态平衡。对A来说,付出一回报的差额为 $M-N'=Pa$,对B来说付出一回报的差额为 $N-M'=Pb$ 。当 Pa 与 Pb 均小于或等于0时,意味着A、B双方所得回报大于付出,有助于提高这一互惠关系的稳定性。其中,当 Pa 、 Pb 越接近0时,越能体现这段互惠伦理关系的平等性与公正性,可能会促进下一轮互惠行为中M和N的增加,并以此循环。

第三,理性使人们关注对背叛行为的惩罚。这就涉及博弈论的相关知识,在合作博弈中有一种非合作博弈类型——纳什均衡,指当其他方策略不变时,任何一方单方面改变自己的策略,都不会提高自身效益。在博弈过程中,当个体能够确定其他人的策略时,会选择最优策略并且不会改变这种策略。在这种非合作博弈过程中,往往会出现背叛,因为无论其他人选择合作还是背叛,背叛都会是个体的最优选择,然而,人的理性却不会容忍背叛。因此,在次数不确定的博弈中,对背叛的惩罚会影响人们的纳什均

衡策略。我们以A的视角作策略分析:

假若在单次博弈中,B利他时,A回报的收益为3,背叛的收益为5,那么A会选择背叛;假若在次数不确定,且存在背叛惩罚的博弈中,B利他时,A回报的收益为3,背叛的收益为5,背叛的惩罚为4,那么A会选择回报。作为理性的行为者,人们不仅会避免自己成为欺骗者,还会配合揭露欺骗事件及欺骗者,支持对欺骗行为的惩罚。可见,对背叛的惩罚有助于减弱人们的背叛倾向,对自我与他人互惠意愿的现实转化有保护作用。

但是在现实生活中,互惠的进化过程是次数未定的非单次博弈,倘若有人存在侥幸心理试图以欺骗手段占有互惠行为带来的效益提升,即使这一欺骗行为在当下未被揭穿,但这一行为从此埋下恶果,一旦被发现,依旧会使欺骗者受到惩罚。在无限时空中,道德朝向正义、真理与自由,任何无视道德权威、道德真理的侥幸群体,最终都会受到来自正义的惩罚。其次,以道德的方式有效减少互惠欺骗者、破坏者,增加互惠伦理关系的守护者、捍卫者,均有助于互惠伦理的形成和稳固。

互惠伦理作为一种社会交往模式(我们可以称为非强制的付出一回报式交往),大致由两个环节构成:意愿付出、自觉回报。前面也分析了付出一回报式的数学模型,尽管可以抽象地概括出互惠各方的付出一回报差额,但是其具体数值确实很难判定。正因这一数值难以精确化,互惠各方(前提是各方均为互惠者,且均自愿维护这一互惠关系)可能会为了避免自己回报较少便采取超额回报的方式维护互惠关系,超额的回报给予对方后,对方有可能也会超额回报。只要存在这一心理,即避免自己处于付出一回报模式的负债地位,就会有一双“看不见的手”发挥协调功能,使互惠伦理得以形成并维持。

这一过程类似“和面”过程中“面多了加水,水多了加面”原理,以使面团达到软硬适中的平衡状态。在互惠关系中,一方的付出多了,另一方的回报就会多,如果后者的回报超过了前者的付出,就需要前者再回报,以此循环,最终目的是达到互惠关系的平衡。但因付出与回报值的难以精准衡量,究竟各方的付出一回报是否达到平衡呢?最终只能基于主观的方式判定,这就要求我们进一步考察互惠伦理何以可能的另外方面的依据,即感性能力。

(三) 感性能力

人并非纯粹的理性机器,而是具有感性情感的存在。假如按照康德对理性人的假设,过于追求纯粹理性之下的利益提升,在现实生活中会遇到很多情理困境。在互惠伦理关系呈现时,理性使我们关注外在利益,感性使我们更在意内在利益。互惠意愿转化为互惠伦理关系的过程,体现了行为者理性与感性的内在平衡,尤其不可忽视感性如亲密关系因同感而产生的同情关怀、宽容仁慈、愤怒、羞愧内疚等对理性的影响。

第一,人们在实践中会在意关系的亲密性。交往实践构建了人际交往关系,其中包括亲密关系。亲密关系不仅包括亲缘关系,还包括因某种善意情感而缔结的友爱关系,内含彼此间的信任、真诚、包容、善与爱。这种关系并非由陌生人随意构建,而是因情感的作用拉近了彼此的生活空间;是因情而结并具有巨大的力量,可以使人具有超越理性的能力,激发人的内在潜能。人们践行互惠意愿时,并非单纯基于理性,而是还会考虑自身已构建的亲密关系,亲密性可以部分或全部抵消理性要求的利益回报。比如,处于亲密关系中的人们会更偏向于利于他人,甚至无所谓对方是否给予回报;但不具有亲密关系时,人们的理性可能会更占上风。互惠行为也影响自我与他人关系的亲密度,有助于关系亲密度的提升,如此良性循环,互惠行为的内在价值能得到更深的挖掘,有助于互惠伦理的进化。

第二,人们具有作为人类成员的同感能力。在非亲密关系的人群中,人们依旧能够产生一种平衡理性的情感,这一情感源于人们的同感能力,同感能力使人产生同情关怀、宽容仁慈、愤怒、羞愧内疚等。这些因同感而产生的情绪能够配合人的理性,协助互惠伦理关系的构建。

首先,同感使人关照他人,产生同情与关怀。同情(sympathy)一方面体现了自我通过主观方式对他人情感的带入而产生的共同感受(fellowing-feeling),另一方面体现了当他人遭遇苦难与不幸时自我对

他人的怜悯情绪。同情建立在自我的想象力之上,人们的道德认知体系越完善,其对他人处境中道德内涵的想象就越丰富。同情他人会产生关怀他人的情感,当这种关怀情感足够强时,会推动利他动机转化为行为。关怀体现了自我对他人利益的关心,使人在力所能及的范围内产生有助于减弱他人苦难而增加他人利益的行为。

其次,同感使人产生强大的理解能力和宽厚心胸。个体心胸的宽厚程度体现了个体对他人的宽容程度、理解程度、尊重程度等,尤其体现在对他人曾经错误的理解予以包容,以及对他人改过自新的宽容。这种情绪强化了人们的互惠意愿,即使理性告诉我们欺骗者不可信,但这种对他人的包容却让我们愿意一试,理解他人的欺骗动机,以自己宽厚的心胸感化他人,给欺骗者走上正轨的机会。理解、宽容不仅是一种超越自我理性的强大力量,还是对欺骗者人性的尊重,理解互惠中某一方未能履行回报义务的处境,对互惠伦理关系的构建具有促进作用。

再次,同感会使人产生愤怒的情绪。愤怒是个体对与自我认可价值不符的事件产生的情绪。同感使人具有归纳类似事件的能力。当他在践行互惠意愿时遭遇了欺骗,人们会吸取类似被欺骗经验防止自己被欺骗。即使被欺骗的人不是自己,个体也会将自我情绪带入他人处境中即移情,会对这种欺骗行为感到愤怒。愤怒的情绪(无论来自当事人还是具有同感能力的第三方)会产生对欺骗者的惩罚,从而对被欺骗者有利。当产生互惠意愿时,发现对方具有被欺骗的经历,这种对欺骗的愤怒情绪也会促使其意愿转化为行为,以安慰被欺骗的心灵,避免欺骗者因被骗而产生不良行为,缓解对方的弱势地位。

最后,同感会使人产生羞愧、内疚的情绪。比如,具有互惠意愿的一方A,做有利于提升另一方B的利益的事,但B并未意识到(可能因为缺乏同感能力、道德能力等)这一行为产生于A。因B未意识到A的善举而导致A的付出未得到回报。事情已经发生而无法改变,但可以弥补,弥补意愿就源于同感。当B将自我情感带入A的处境,产生羞愧内疚的情绪,这种情绪就促使B产生对A利他行为的弥补。不仅如此,曾经产生羞愧与内疚情绪会成为未来的实践经验,这种情绪存在于人心之中,激励人们将互惠意愿转化为互惠行为,甚至会产生额外的亲社会善行。

三、互惠伦理如何践行

研究互惠伦理不能仅基于模型的分析,还要与现实生活紧密结合,分析其如何践行。作为一种伦理,互惠伦理只有得到人们的恰当践行才能走向稳定,其践行机制主要包括外在约束和内在约束。当外在约束转化为内在自我约束时,就意味着个体道德能力的增强、道德水平的提升。同时,互惠伦理的践行还需要合理对待欺骗者,互惠各方的行为要无害于第三方。

(一) 社会奖惩制度和他人评价、监督机制

社会奖惩制度和他人的评价、监督是互惠伦理的外在约束机制。

第一,社会奖惩制度。鲍尔斯、金迪斯认为,强互惠是能够避免人类背叛行为的方式。在强互惠中,互惠关系任何一方有背叛行为或倾向的,另一方或第三方会采取自己付出也得不到补偿的惩罚措施,即需要付出额外成本的惩罚。当一个群体存在强互惠者时,这个群体中的自私者就会受到相应的“强互惠”惩罚。但有一个问题,谁来担任这个强互惠角色?未承担强互惠角色者是否属于搭便车者?这同互惠搭便车者有什么区别?总之,即使是存在强互惠去维持群体中的互惠平衡,也难免会有人多付出而少回报。由于个体在承担强互惠角色时,不仅会付出超额的成本,损害自身利益,还可能会违反社会规定,甚至会掉入强互惠的陷阱而付出巨额代价。这就表明,让社会奖惩制度担任强互惠角色,社会群体中的个体负责监管,比个体直接承担强互惠角色更具有优势。

与人为的主观惩罚相比,社会奖惩制度受法律保护、民众认可,具有客观性、强制性、公开性、透明性。社会这双“看得见的手”解决了“谁来承担这个强互惠者角色”的问题,也避免了新的搭便车者的产生。但是,社会规范的制定依旧离不开人的参与,人在其中主要担任的是商讨制定、监督管理、修补完善

规范的工作,也坚持以人为本、与时俱进原则的角色。人们通过协商制订出能够维护互惠伦理关系的奖惩制度,并且社会群体能够自发监督管理,检举违反或试图破坏这一良好秩序的行为,在新问题出现时,能够重新商讨、完善奖惩制度体系。在此过程中,奖惩制度不仅要合乎尊重人性、人格完整性和自由的原则,还要顺应时代发展,不断更新。

社会奖惩制度应当合理有效,避免奖励过重而适得其反,或惩罚过重而造成不可挽回的结果。作为一种亲社会行为,互惠行为体现出人们的社会偏好,社会应当用心呵护这些行为,为这些行为的实现提供良好的实践氛围。社会应当树立互惠行为的道德模范,以这些模范为榜样激励群体内形成良好的交往氛围。对道德模范的奖励应当适中,不然就会出现“人前一套,人后一套”的虚假道德模范。对于互惠行为的欺骗者,根据欺骗行为给社会带来的负面影响程度,社会应当给予相应惩罚。采取恰当的方式对欺骗者群体中有意愿接受道德改造的人们以道德教育,尽量给予他们改过自新的机会。但是,对情节严重者,或改造意愿不强烈、依旧对社会具有危害性的欺骗者群体,社会应当采取更重的惩罚。

第二,他人评价、监督。特里弗斯在互惠模型第二个例子中提到欺骗者,鸟类是否存在欺骗者、如何对待欺骗者、欺骗者是否会受到惩罚,这些都无法得到客观上的验证。相对于其他动物,人类还有一个特长就是凭记忆来识别、揭发欺骗者。在交往过程中,人们不仅基于经验判断对方是否会满足自我需求,还会基于对方的声誉、信誉来决定是否互惠。涉及声誉评估的互惠属于间接互惠,生物学家亚历山大认为,间接互惠在人类社会(可能在一些灵长类动物、社会性犬科动物和其他群体中也是如此)中很重要,它涉及声誉和地位,并导致群体中的每个人不断地被评估和重新评估(assessed and reassessed)^[17](P205)。社会中存在第三方力量奖励具有良好声誉的个体,人们也会对具有良好声誉的个体更加和善。马丁·诺瓦克和卡尔·西格蒙德也通过建立数学模型分析印象分^[18]的评估及其对互惠的影响。在制定互惠行为策略时,人们会以对方的印象分为参照,此时,印象分等同于互惠的信誉度,它的评估源于人们累积的互惠经验。在构建互惠伦理关系的过程中,人们会判断对方的互惠意愿以及维护互惠伦理关系的程度,还会分享这种经验,一旦社会出现一个欺骗者,人们就会凭借经验以防止自己与他人后续被欺骗。正是因为声誉评价和监督的存在,人们为了避免自我声誉受影响,提升自我的互惠印象分,会尽其所能维护已构建或待构建的互惠伦理关系。但其中有一个问题,声誉的评估具有主观性,并非具体的可视化数值,是以经验感知的计量方式评估人们的行为。这种评估存在误差,即可能高估或低估人们的信誉。这表明,他人的评价和监督对维护互惠伦理关系有一定作用,但也要考虑到监管误差,至少尽可能准确评估对方是否有互惠声誉。

分享欺骗者的经验可能会使自己受到欺骗者的报复,但为何人们依旧自愿揭发检举欺骗者呢?其一,人们内心具有正义感,这种情感使人们遇到欺骗者时会向周围人分享欺骗者的外貌、行为特征,避免其他人再次受骗。如同强互惠中的第三方,自愿付出额外成本去惩罚互惠伦理的背叛者,这一惩罚情绪的来源之一就是内心的正义。正是这份正义使他们自愿维护互惠伦理。他们并非因为自私,担心自己的利益受到侵害,仅仅为了捍卫这份正义。其二,即使是自私的人也会为维护互惠伦理而去揭发欺骗者。他们更在意自己的利益,但欺骗者的存在对他们的利益具有潜在威胁,为了避免自己的利益再受损失,他们也会主动监督,并分享自己的受骗经验,同大家一起协同清除社会中的欺骗者。其三,人们对内在利益的追求。麦金泰尔在《追寻美德》中分析了内在利益和外在利益,认为美德“使我们能够获得那些内在于实践的利益”^[19](P242)。对互惠伦理关系中欺骗者的监管,可能并不会给自己带来任何外在利益,然而他们这种监督行为的本质,使他们的内心得到了满足。

但是,人们主观评估声誉时依据的是行为者的言行特征等,根据这些因素评估行为的互惠性,会造成欺骗者模仿互惠者的现象产生。前者通过经验学习伪装成后者,进而展开互惠中的欺骗行为,这是难以避免的。因此,提升行为者内心自律非常重要。在互惠伦理的进化中,任何外在制约,无论源于社会还是他人,其效果都可能低于人们的内心自律。

(二) 内心道德自律

外在约束毕竟源于他者,在外在约束转化为内在自律之前,这些约束可能会使人产生不情愿的情绪。要减少这些情绪,就需要依靠道德熏陶,使外在约束转化为行为者的内心自律,使行为者自觉遵守互惠伦理,维持互惠关系的动态平衡。这就要求人们在互惠伦理关系中做到如下几点。

第一,表达自我真实需求,拒绝伪善。如克鲁顿—布洛克所述,人们之所以与其他动物相区别,是因为人类能够表达自我意图和期望,这也意味着存在真实的或虚假的意图和期望^[20]。互惠伦理的践行,不仅需要人们具有利他之心,还要有表达自我真实愿望之意。比如,一方的真实愿望为X,虚假愿望为X',另一方的真实愿望为Y,虚假愿望为Y'。假若两者之间构建了互惠伦理关系,但因两者在构建关系时均未表达自己的真实愿望,最终导致虚假愿望的实现。这就会造成心理预期满足偏差(X-X')和(Y-Y')的出现,但这一偏差出现的原因应该归于两者本身。可能是因为他们对对方保持不信任的态度,也可能是他们的真实愿望难以启齿,不管怎样,表达自己的虚假愿望,属于欺骗行为。既是对对方劳动的不尊重,又造成自我真实愿望的遮蔽,是一种不利人也不利己的行为。

第二,保持互惠之情。互惠之情是人们意愿互惠的感情,利他之心是人们自愿为他人的利益考虑。互惠是自利与利他这两个范畴的结合,它为自然界存在的众多利他行为提供了恰当辩护。自然界的互惠意愿和行为究竟是源于先天还是后天,在学界是有争论的。在非人类生物中,有大量实验数据结果为它们之间存在的本能互惠行为提供了辩护,无论这些生物是否具有亲缘关系,是否属于同一物种;在人类社会,也有大量数据支撑人类幼崽具有互惠倾向。故笔者认为,互惠意愿是先天与后天综合作用的结果。人们具有的先天之善很重要,但在后天实践中,如何合理运用这些善也很重要。利他意味着关心他人的利益,保持互惠之情,心存利他之意,能够有效协调德性与人性,促进互惠伦理的合理践行。

第三,不断提升道德认知水平、道德判断能力以及道德敏感度。当互惠中一方未意识到利他者为互惠中的另一方时,就说明前者的道德判断能力与道德敏感度尚待提高,是对后者的不公平。为了避免这种情景,需要人们不断积累现实经验,提升自我道德认知水平,与人多交流、多沟通,提升自己的道德认知、评价与判断能力,积累道德模范、道德榜样的标志印象,识别互惠信号,提升自我道德敏感度。

第四,捍卫自我正当权益,但不侵害他人权益。在现实生活中,大多数人往往因为担心自己破坏了社会秩序而得到报复,因而小心翼翼地坚守社会规范,甚至甘愿多付出,也就是人们常说的“吃亏是福”。但正是存在这些甘愿吃亏的人,才让另一些人占了便宜而少付出。基于正义达成的资源分配平衡,是互惠付出一回报模式平衡的基础。其实,大部分情况下的互惠伦理难以维持,不是因为人们过于自私,而是因为人们对自我正当权益的维护失当,也可以说是对追求正义的害怕、畏缩;对自我正当权益的放弃就是对不义的纵容。因此,我们应该宽容待人,也要多多关注自己的应得利益。

(三) 合理对待欺骗者

如何对待欺骗者是谈论互惠伦理践行问题时需要解决的问题。通常情况下,这些欺骗者激起人们愤怒的情绪,使人们无法原谅,甚至希望他们受到巨大的惩罚。对那些并没有对社会、群体、个体造成巨大伤害的欺骗行为,人们应当采取理解、宽容的态度,感化对方。因为惩罚过重会使人麻木,对未来无望可能使这些被惩罚者对道德更加冷漠,甚至对生活失去感知,对人生失去信心。给予那些愿意主动承认错误的人以新的机会,重新以道德的方式踏入社会,就是对成全的践行。使受骗者原谅欺骗者可能显得有些强人所难,但“人非圣贤,孰能无过”,理解他人的欺骗动机,宽以待人,给他人改过自新的机会,成全他人,也是在成全自己的善心。至于那些对社会造成不可逆转危害的人群,就另当别论了。

社会各方应当采取恰当的方式对欺骗者进行道德改造,使其产生羞愧与内疚的心理。道德改造尽量通过道德诱导的方式,因为恰当的内心诱导比强制性的外在权威更能使人产生懊悔和羞愧的心理。懊悔与羞愧的心理会使人产生弥补过失的情绪,进而增强自我的利他意愿。对曾经的欺骗者,假如在未来产生互惠意愿,这种内疚的情绪也会促使这一意愿快速转化为实践,以消解这种情绪。

(四) 禁害第三方以趋善避恶

互惠伦理的践行会遇到一些不能排除的情况,其中一种情况就是是否会对与互惠关系无关的第三方构成影响。互惠各方在这一过程中是否做到趋善避恶,一般说来,有以下几种可能:

第一,互惠各方在践行互惠伦理时未影响到第三方(甚至有益于第三方),这一互惠行为就符合不伤害他人的道德要求,是道德的。这样的互惠伦理实践,不仅能得到互惠各方的认可,还能得到这一行为的所有波及者的认可,甚至会得到社会的鼓励、倡导。

第二,互惠各方在践行互惠伦理时无意识地影响到第三方,造成了不伤害他人的动机与第三方受到伤害的结果不符,互惠各方可能并没有伤害第三方的意图,但伤害了第三方,互惠各方的行为可能得到义务论者的辩护,但得不到后果论者的原谅。因此,互惠各方在践行互惠伦理时,应该以“不伤害”为行为原则,充分考虑各种因素、可能性结果以及波及范围和人群,不断提升自己的道德认知水平、道德预判可信度等。

第三,互惠各方在践行互惠伦理时有意识地影响到第三方,当行为人为自我利益而有意损害他人利益时,由此产生的行为后果很难说是道德的,难以得到第三方的原谅及社会的倡导。

互惠伦理的践行不仅需要互惠各方具有互惠伦理意愿,还要使这一意愿转化为行为。在行为中,互惠各方要平衡自我理性与感性,发挥自我意识能动性,使自我互惠意愿在合理、可行范围内转化为行为。不仅如此,人们处于社会共同体之中,任何关系都并非独立存在,而是会牵扯到其他关系,互惠伦理的践行不仅要考虑各方的利益,最重要的是不能有意损害第三方的利益。

因此,互惠伦理要长久发挥作用,需要社会、他人、自我的协同运作:社会奖惩机制合理,并坚持以人为本、与时俱进、不断更新完善,他人评价真实、监督管理有效;个体基于自我互惠意愿践行利他,表达自我真实需求而拒绝伪善,不断提升自我道德水平,正当维权、为自我与他人负责;欺骗者意愿接受道德改造,改过自新,践行互惠;互惠各方坚持不伤害第三方。互惠伦理秩序如没有外力的施加,是无法形成的,个体就相当于互惠伦理的开关,他人就相当于互惠伦理的监管者,欺骗者就相当于互惠伦理中的错误,社会则为互惠伦理秩序的形成提供了源源不断的能量。

四、结 语

互惠作为人类行动的一个基本原则,意味着人们都是既考虑自身物质利益和其他私利,也考虑他人利益的。它与道德在内涵上具有一致性,因而是一种道德规范。当人们以互惠来从事人际交往就会形成互惠伦理。互惠伦理表现为付出一回报模式。互惠伦理从性质上可分为积极的和消极的两种类型,前者指我对人好,人对我好;后者指人对我不好,我必惩人。互惠伦理的形成要求个体具有良好的行为能力、理性能力、感性能力等一系列条件。只有得到人们的恰当践行,互惠伦理才能走向稳定,其践行机制主要包括外在约束和内在约束。当外在约束转化为内在自我约束时,就意味着个体道德能力的增强、道德水平的提升。同时,互惠伦理的践行还需要合理对待欺骗者,且无害于第三方。

总之,作为人类伦理价值规范系统的一个重要构成,互惠伦理具有极为广泛的适用性,对人类生活具有值得重视的积极价值。它可以促进社会团结,使个人与社会协调共存,人与人之间相互担责、履行义务;也可以促进社会公平,使社会充满仁爱、同情、温馨;还可以促进社会信任,信任又有助于合作,因为如果我们认为他人值得相信,我们会生活得更好。正如有人所说的,互惠伦理是一种几乎没有别的人类动机能与之比肩的强大的、普遍的人际交往原则,它既存在于现实中的各种场合,也程度不同地存在于所有文化系统。因此,在促进人类实现美好生活上,互惠伦理是不可忽视的精神力量。

参考文献

- [1] 阿明·福尔克.人性的博弈——为什么做个好人这么难? 杨亚庆译.北京:中信出版社,2024.

- [2] 叶航,陈叶烽,贾拥民.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,2013.
- [3] Bronislaw Malinowski. *Crime and Custom in Savage Society*. London, New York: Routledge, 2017.
- [4] A. W. Gouldner. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 1960, 25(2).
- [5] 詹姆斯·C.斯科特.农民的道义经济学——东南亚的反叛与生存.程立显、刘建等译.南京:译林出版社,2013.
- [6] R. L. Trivers. The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 1971, 46, (1).
- [7] 罗伯特·阿克塞尔罗德.合作的进化.吴坚忠译.上海:上海人民出版社,2017.
- [8] 邹文箴,田青,刘佳.“投桃报李”——互惠理论的组织行为学研究述评.心理科学进展,2012,(11).
- [9] 马丁·诺瓦克,罗杰·海菲尔德.超级合作者.龙志勇、魏薇译.杭州:浙江人民出版社,2013.
- [10] 塞缪尔·鲍尔斯,赫伯特·金蒂斯.合作的物种——人类的互惠性及其演化.张弘译.杭州:浙江大学出版社,2015.
- [11] 汪丁丁,罗卫东,叶航.人类合作秩序的起源与演化.社会科学战线,2005,(4).
- [12] 克里斯蒂娜·M.方,塞缪尔·鲍尔斯,赫尔伯特·金蒂斯.互惠与福利国家//赫尔伯特·金蒂斯,塞缪尔·鲍尔斯,罗伯特·博伊德等.道德情操与物质利益——经济生活中合作的基础.李风华、彭正德、孙毅译.北京:中国人民大学出版社,2015.
- [13] 甘绍平.伦理学的当代建构.北京:中国发展出版社,2015.
- [14] 克里斯·布斯克斯.进化思维——达尔文对我们世界观的影响.徐纪贵译.成都:四川人民出版社,2022.
- [15] 科林·迈耶,布鲁诺·罗奇.互惠经济——企业互惠、长期主义与共同富裕.彭相珍译.北京:中译出版社,2022.
- [16] 吴先伍.从“不安”到“不忍”——孟子对孔子人己思想的发展.云梦学刊,2024,(4).
- [17] R. D. Alexander. *The Biology of Moral Systems*. London, New York: Routledge, 2017.
- [18] M.A. Nowak, K. Sigmund. Evolution of Indirect Reciprocity by Image Scoring. *Nature*, 1998, 393(11).
- [19] 阿拉斯戴尔·麦金泰尔.追寻美德.宋继杰译.南京:译林出版社,2024.
- [20] Tim Clutton-Brock. Cooperation Between Non-kin in Animal Societies. *Nature*, 2009, 462(5).

Reciprocal Ethics: Connotations, Foundations and Practice

Gong Tianping, Hu Ying (Zhongnan University of Economics and Law)

Abstract Reciprocal ethics is an important component of the human ethical values and normative system. Reciprocity refers to the motivation or behavior of an individual to help others, based on their own will or interest, in order to promote the interests of others with the expectation of receiving some form of return in the future. Reciprocal ethics denotes an ethical framework where parties in interaction put their selfish or altruistic preferences into action in a reciprocal manner, and each party should consciously assume an obligation to reciprocate. Reciprocal ethics mainly encompasses positive reciprocal ethics and negative reciprocal ethics. The possibility of reciprocal ethics is based on a series of specific conditions, including behavioral capability, rational capability and empathy ability, etc. Reciprocal ethics can be stable only when properly practiced by people. Its practice mechanism mainly includes external constraints such as the social reward and punishment system, and peer evaluation and supervision, along with internal constraints represented by individuals' inner moral self-discipline. It also requires reasonable treatment of deceivers, and reciprocal behaviors should not pursue goodness and evade evil at the cost of third parties.

Key words the connotations of reciprocal ethics; the possibility of reciprocal ethics; the practice of reciprocal ethics

-
- 作者简介 龚天平,中南财经政法大学哲学院教授,湖北 武汉 430073;
胡莹,中南财经政法大学哲学院博士研究生。
- 责任编辑 涂文迁