

后果主义道德评价的社会合作理论

——以“囚徒困境”为视角

刘 宇*

【摘 要】 后果主义的分析路径是研究道德评价的合理途径。在社会合作的理念下,通过比较分析囚徒困境与道德评价两种模型,可以发现,对囚徒困境解决方式的进一步深入研究,可以确立社会合作的基本原则,即最大收益原则与共同受益原则。最大收益原则与共同受益原则分别表现在效益式的后果主义理论与公平式的后果主义理论之中。通过普遍模仿验证的后果主义道德评价理论,有可能将二者结合起来,更好地实现社会合作理念。

【关键词】 社会合作;囚徒困境;效益式后果主义;公平式后果主义;普遍模仿验证

道德评价,即依照一定的标准对自己或他人的行为进行善恶、好坏、荣辱等道德价值的评论和判断。任何一种道德评价理论,都必须选择某种分析路径,即某种评价的效力来源和最终标准。近代以来,有关道德哲学的分析路径大体可以分为三种:后果主义道德理论,义务论伦理学 and 美德论伦理学。本文选取后果主义的分析路径,但非简单沿用其传统观点,而是意在通过对社会合作根本原则的研究,重新建立一个能够更好实现社会合作理念的后果主义道德评价理论。本文的思路是:第一部分,在社会合作理念下比较囚徒困境与道德评价,并论证囚徒困境能够成为研究后果主义道德评价理论的合理途径;第二部分,分析囚徒困境的各种解决方案,得出只有通过添加限制条件来改变博弈收益才能走出囚徒困境的结论;第三部分,从囚徒困境的解决方案推导社会合作的两个根本原则,并论证两原则分别为效益式后果主义与公平式后果主义所追求的目标,最终通过普遍模仿验证的后果主义道德评价理论将二者结合在一起,更好地实现社会合作理念。

一、囚徒困境:研究后果主义道德评价理论的合理途径

囚徒困境的基本预设如下:A和B两人共谋犯罪被关入监狱,两人不能互相沟通且都想尽快出狱。警方要对二人进行审讯,规则如下:如果两人都保持沉默,则因证据不足,两人都将被关押一年

* 中国政法大学博士研究生,100088。

后释放;如果一人招供而另一人沉默,则揭发者因立功而立即获释,沉默者入狱十年;如果两人都主动招供,则因证据确凿,两人同时被判入狱五年。此时,虽然两人同时选择沉默是最优方案,但最终两人却都会选择招供,因为无论对方选择沉默还是选择招供,自己选择招供都比选择沉默被判刑更少。

“囚徒困境”之所以能够成为研究后果主义道德评价理论的合理途径,其原因有二:

首先,在后果主义视角下,囚徒困境与道德评价同时具有社会合作的本质属性。社会合作意味着在一定的环境之下,社会中的人,为了共同的目标,以一定的方式互相配合行动的过程。囚徒困境与道德评价的社会合作属性体现在以下四个方面:其一,社会合作必然发生在一定的环境之下:囚徒困境发生在假定的博弈环境之下,而道德评价发生在现实的博弈环境之下。其二,社会合作需有相应的合作参与者:囚徒困境的合作参与者是A、B两人,而道德评价的合作参与者是全体社会成员。其三,社会合作的参与者应为了实现共同目标而合作:囚徒困境的参与者为了共同的目标——减少判刑而合作,而道德评价的参与者为了共同的目标——幸福生活而合作;社会合作的参与者为实现共同目标需以一定的方式相互配合而行动:囚徒困境的参与者以何种方式协调其行动才能实现减少判刑的共同目标正是囚徒困境中需要解决的问题,而道德评价的参与者以何种方式评判行动才能实现幸福生活的共同目标也是道德评价中需要分析的问题。因此,由于性质上的共通,囚徒困境问题的解决与道德评价问题的分析有着本质上的联系。

其次,在后果主义视角下,囚徒困境是社会合作的简单模式而道德评价是社会合作的复杂模式,从简单模式向复杂模式的推理是分析和解决问题的有效途径。囚徒困境是一种假想状态下的合作问题,即最“小”的社会合作模型,参与人数最少——两人,而道德评价是一种现实状态下的合作问题,即最“大”的社会合作模型,参与人数为最多——全体社会成员;囚徒困境所涉及的是最“简单”的合作模型,因为在假想的社会合作中只有一个限制条件,两个参与者只受到自我利益的驱使而行动,而道德评价所涉及的是最“复杂”的合作模型,因为在现实的社会合作中存在多种限制条件,无法单纯通过理性博弈来确定人们的选择。对于囚徒困境这种假想的合作问题,唯一解决方式就是施加限制条件来改变博弈环境实现最优的合作选择,而对于道德评价这种现实的合作问题,限制条件的作用同样在于通过规制人们的行为来建立最佳的合作模式。如果能够通过施加限制条件让参与者顺利走出囚徒困境这种假想的合作模型,那么此条件所引申出的合作原则,同样应该能够对道德评价这种现实的合作问题具有重要启发作用。

由此可见,囚徒困境与道德评价的共同点体现在博弈环境的价值追求上,而不同点体现在博弈环境的复杂程度上。正是这样的特点决定了利用囚徒困境研究道德评价可能成为一种可行的方案。在囚徒困境中存在一个已知的最优结果,而在道德评价中却不存在这个已知的最优结果。囚徒困境是一个已知最优结果,探索何种方法能够达到此结果的问题;相反,道德评价是一个未知最优结果,需要以一种方法来帮助其寻找并达到最优结果的问题。在已知最优结果且博弈环境更为简单的囚徒困境之中,寻找方法的可能性较高;相反,在未知最优结果且博弈环境更为复杂的道德评价之中,寻找方法的可能性较低。最后,由于双方同时具有社会合作的根本性质,所以囚徒困境中的解决方案与道德评价中的分析方法应该是相通的。然而,在囚徒困境中所添加的限制条件却无法直接应用于道德困境之中,需要引入一个第三者作为两者之间的桥梁,这个第三者就是社会合作的根本原则。社会合作,作为囚徒困境和道德评价的共同追求,其本质有相通之处。所以,囚徒困境的解决与道德评价的分析均可上升并统一于社会合作的根本原则之中。于是,从解决囚徒困境所添加的限制条件中应该能够推导出社会合作的根本原则,而社会合作的根本原则也应该能够应用于道德评价的分析之中。这便是文章研究的根本逻辑所在。

二、囚徒困境的解决方案

囚徒困境的解决就是要在其基本模型上添加一定的限制条件,促使博弈双方能自发地选择同时招供,以达到利益最优。如何解决囚徒困境的问题就是如何添加限制条件的问题。限制条件的添加,也就是博弈条件的改变,可分为三个方面:改变博弈者、改变博弈环境以及改变博弈收益。下文将分别对这三个方面进行分析,并认为:改变博弈者与改变博弈环境并不能真正走出囚徒困境,只有改变博弈收益才能真正走出囚徒困境。

(一) 改变博弈者

所谓改变博弈者,就是变利己理性人的预设为利他理性人的预设。众所周知,在利己理性人的初始状态下,博弈双方最终会选择“同时招供”;相反,在利他理性人的预设模型下,博弈双方最终会选择“同时沉默”。但笔者认为,作为博弈模型的初始预设,利己的理性人要比利他的理性人更为合适。这是因为:其一,作为一个有效的理论起点,仅仅能够处理“最佳可能性”是不够的,必须同时能够容纳“最差可能性”。在理性人的模型构建中,利他理性人只能处理“最佳可能性”,只有利己理性人才能容纳“最差可能性”。其二,任何策略,甚至包括道德策略,都必须具有生存优势,否则无法成为真正的优势策略。在理性人的模型构建中,利己理性人比利他理性人更能获得生存上的优势,能够在成为道德优势策略的同时为生存负责。其三,利己的理性人模型在现实中出现的可能性更高且更加稳定。因此,可以作如下两种假设:

假设1:所有人均为利他的理性人,而仅存在一个利己的理性人。此利己理性人在竞争中必定会比其他利他理性人获得更多的利益,从而成为优势策略。那么此时,此利己理性人由于占据博弈优势而必然被其他利他理性人所模仿。最终,社会终将蜕变为由利己理性人所构成的社会。

假设2:所有人均为利己的理性人,而仅存在一个利他的理性人。此利他理性人在竞争中必定会比其他利己理性人获得更少的利益,从而成为劣势策略。那么此时,此利他理性人由于没有博弈优势而必然被其他利己理性人所淘汰。最终,社会终将稳定在由利己理性人所构成的社会。由此,利己理性人是比利他理性人更为稳定的初始模型,所以也是更合理的初始模型。

(二) 改变博弈环境

所谓改变博弈环境,主要指变单次博弈为重复博弈以及变沉默博弈为对话博弈。重复博弈和对话博弈分别以艾克斯罗德(R·Axelrod)与哈贝马斯(J·Habermas)各自的理论为代表。艾克斯罗德通过计算“重复博弈回合”模拟中“一报还一报”策略的胜出来证明在重复博弈中善意合作策略均要优于恶意背叛策略^①;哈贝马斯通过设置“理想的对话环境”和“有效的沟通标准”来证明如果语言活动足够“正确”就有可能促成由背叛到合作的进化^②。但笔者认为,无论是重复博弈还是对话博弈,最终都无法真正地走出囚徒困境。

重复博弈理论的缺陷在于:每个人在试验中都不可能杀死或淘汰。“在艾克斯罗德试验之中,每个博弈者可能失败即得分低,但不可能被消灭即杀死。这个与真实世界不符的规定不能正确表达人生命运……采用了‘杀不死’假定生死游戏弱化成输赢游戏,命运不在严肃。‘杀不死’等于永远有

^①R. Axelrod, “Effective Choice in the Prisoner’s Dilemma”, *Journal of Conflict Resolution*, vol. 24, 1980, pp. 3—25.

^②J. Habermas, *Moral Consciousness and Communicative Action*, Cambridge, UK: Polity Press, 1990, p. 89.

机会卷土重来,这就限制了不择手段竞争的威力,所谓背叛就只能占到小便宜,从而造成良善合作策略具有博弈优势的假象。”^①在理论的囚徒困境中虽然不存在被杀死或淘汰的可能,但在现实的囚徒困境中一定会出现被杀死或淘汰的状况。更真实的状态是在重复博弈的囚徒困境中设置一个遭受N次背叛就被判处无期徒刑或死刑的条件。在这样一个更真实的博弈中,虽然仍有善良的合作者最终会胜利,但也有一些善良的合作者最终会被淘汰出局。于是,一些本来的善良合作者就会动摇自己的立场,甚至选择背叛。

对话博弈理论的缺陷在于:对话只能保证互相理解却不能保证互相接受。“哈贝马斯的成功对话需要这样的过程:理性沟通→互相理解→一致同意……问题出在互相理解与一致同意之间漏掉了一个重要的环节——互相接受。互相理解无法保证一致同意,至多是必要条件而绝非充分条件。在许多情况下,对话双方完全互相理解,但绝不接受对方主张,甚至于水火不容……完美的成功对话过程应是这样的:理性沟通→互相理解→互相接受→一致同意。”^②这里的一致同意在囚徒困境中可以理解为一致行为,因为由心而至的一致同意才会导致一致行为,而虚假敷衍的一致同意不会导致一致行为,这就需要通过互相接受而不仅仅是互相理解。这就造成了在现实的囚徒困境中,即使双方能够进行理性沟通,因为自身利益存在根本冲突,最终也无法达至真正的一致同意,即使达成表面上的共同协议,也不会真正遵守。

(三) 改变博弈收益

综上,无论是改变博弈者——“理性人”的初始状态,还是改变博弈环境——变单次博弈为重复博弈或是变沉默博弈为对话博弈,都无法有效走出囚徒困境。笔者认为,只能是通过添加限制条件来改变博弈收益,只有博弈收益改变之时博弈者才会自发地改变博弈策略,实现最优的合作选择。

囚徒困境的参与者之所以通过理性选择会最终选择招供而不是沉默,根源在于两个基本心态:其一是规避最大风险;其二是争取最大收益。规避最大风险体现在当博弈的对方选择招供时,本方可能面临博弈中的最差结果,即选择沉默被判10年徒刑,于是为了规避最差结果则必然选择招供;争取最大收益体现在当博弈的对方选择沉默时,本方可能面临的是博弈中的最优结果,即选择招供获得无罪释放,于是为了争取最优结果则也会选择招供。由此,同一种博弈策略既包含了一方的最优选择也包含了另一方的最差选择,双方都希望这种博弈策略发生,同时尽一切可能争取成为优势方,规避成为劣势方。争取成为优势方需要选择招供,规避成为劣势方也需要选择招供,招供就成为了唯一可能的选择。

于是,问题的关键在于一方招供而另一方沉默的策略之中,真正需要改变的博弈收益,也正是此种策略的博弈收益。假如改变本方招供对方沉默之时的博弈收益,使其不再是最优选择;或是改变本方沉默对方招供之时的博弈收益,使其不再是最差选择;再或是同时改变这两者;是否就能够改变囚徒困境的博弈状况呢?

为了便于分析,利用收益不被改变的两种博弈策略(同时选择招供和同时选择沉默)将本方收益由多到少的利益区间分为三个部分:A区域、B区域及C区域,如图1所示:

A区域为本方收益最多,其获利大于双方同时选择沉默(各判1年)的情况;C区域为本方收益最少,其获利小于双方同时选择招供(各判5年)的情况;B区域为本方收益的中间值,其获利大于双方同时选择招供但小于双方同时选择沉默。由此,在囚徒困境的初始模型中,本方招供对方沉默的策略(设为:策略E)收益最多,即位于利益区域A之中;本方沉默对方招供的策略(设为:策略F)收

^①赵汀阳:《冲突、合作与和谐的博弈哲学》,《世界经济与政治》2007年第6期。

^②赵汀阳:《冲突、合作与和谐的博弈哲学》,《世界经济与政治》2007年第6期。

益最少,即位于利益区域 C 之中。
 (“×”表示为选择“沉默”;“√”表示为选择“招供”。)



图 1 利益区间的三个区域

1. 当策略 F 不再是最差选择,策略 E 收益不变之时,情况有四:①将策略 F 提升至利益区域 B;②将策略 F 提升至利益区域 A,但收益劣于策略 E;③将策略 F 提升至利益区域 A,且收益优于策略 E(见表 1)。④将策略 F 提升至利益区域 A,与策略 E 收益相同(见表 2)。

表 1 囚徒困境模型的收益分配

本方获益	本方选择			对方选择		
	①	②	③	①	②	③
最大	√	√	×	×	×	√
其次	×	×	√	×	√	×
再次	×	×	×	√	×	×
最少	√	√	√	√	√	√

情况①:当对方选择沉默(×)之时,本方选择招供(√)的收益更大;而当对方选择招供(√)之时,本方选择招供(×)的收益更大。由此无法构成博弈的优势策略。
 情况②:当对方选择沉默(×)之时,本方选择招供(√)的收益更大;而当对方选择招供(√)之时,本方选择招供(×)的收益更大。由此仍无法构成博弈的优势策略。
 情况③:当对方选择沉默(×)之时,本方选择招供(√)的收益更大;而当对方选择招供(√)之时,本方选择招供(×)的收益更大。由此也无法构成博弈的优势策略。

情况④:当本方选择招供(√)时,依据对方选择,既有可能获得最优利益(对方沉默×),也有可能获得最差利益(对方招供√);而当本方选择沉默(×)时,依据对方选择,仍有可能获得最优利益(对方招供√),也有可能获得中间利益(对方沉默×)。由此,在无论本方做何种选择都有可能获得最优利益的前提下,出于规避最差可能性的考虑,本方应该选择沉默(×),因为选择招供(√)有可能获得最差利益,但是选择沉默(×)则不会。

表 2 囚徒困境模型的收益分配

本方获益	本方选择	对方选择
最大	√ = ×	× = √
其次	×	×
最少	√	√

2. 当策略 F 不再是最差选择,策略 E 收益不变之时,情况有四:①将策略 E 下降至利益区域 B;②将策略 E 下降至利益区域 C,但收益优于策略 F;③将策略 E 下降至利益区域 C,且收益劣于策略 F(见表 3)。④将策略 E 下降至利益区域 C,与策略 F 收益相同(见表 4)。

表 3 囚徒困境模型的收益分配

本方获益	本方选择			对方选择		
	①	②	③	①	②	③
最大	×	×	×	×	×	×
其次	√	√	√	×	√	√
再次	√	√	×	√	×	√
最少	×	×	√	√	√	×

情况①:当对方选择沉默(×)之时,本方选择招供(×)的收益更大,而当对方选择招供(√)之时,本方选择招供(√)的收益更大。由此无法构成博弈的优势策略。

情况②:当对方选择沉默(×)之时,本方选择招供(×)的收益更大,而当对方选择招供(√)之时,本方选择招供(√)的收益更大。由此仍无法构成博弈的优势策略。

情况③:当对方选择沉默(×)之时,本方选择招供(×)的收益更大,而当对方选择招供(√)之时,本方选择招供(√)的收益更大。由此也无法构成博弈的优势策略。

情况④:当本方选择招供(√)时,依据对方选择,既有可能获得最差利益(对方沉默×),也有可能获得中间利益(对方招供√);而当本方选择沉默(×)时,依据对方选择,仍有可能获得最差利益(对方招供√),也有可能获得最优利益(对方沉默×)。由此,在无论本方做何种选择都有可能获得最差利益的前提下,出于追求最优可能性的考虑,本方应该选择沉默(×),因为选择沉默(×)有可能获得最优利益,但是选择招供(√)则不会。

表 4 囚徒困境模型的收益分配

本方获益	本方选择	对方选择
最大	×	×
其次	√	√
最少	√ = ×	× = √

3. 当策略 F 不再是最差选择,同时策略 E 不再是最优选择之时:既然单独改变策略 E 或策略 F 一方既可以走出囚徒困境;则同时改变策略 E 与策略 F 双方则必然在某些利益组合下也可以走出囚徒困境。如果以上两种方法都能够达到效果,则应更倾向于选择尽量少的改变,尽量简洁的设计。也就是说,在效果相同的情况下,选择添加一个限制条件要优于选择添加两个限制条件。改变策略 E 或策略 F 其中的一者意味着添加一个限制条件;同时改变策略 E 与策略 F 二者则意味着添加两个限制条件。如此,改变策略 E 与策略 F 其中一者的方案要优于同时改变二者的方案。

两种可行策略的比较:虽然将策略 E 的收益降低为与策略 F 的收益相等的方案,及将策略 F 的收益上升为与策略 E 的收益相等的方案,均可以促使博弈双方同时选择沉默,但前者必然优于后者。因为,现实之中,达到收益下降的手段是惩罚,而达到收益上升的手段则是奖励。作为一个普遍的制度设计,显然利益上的惩罚要比奖励更容易实现。另外,也是最重要的,当策略 F 的收益上升为与策略 E 相同时,集体的最优选择已经从同时沉默——各判 1 年,变为一方沉默一方招供——招供者本身已无罪释放,而将沉默者的收益上升为与招供者相同则同样无罪释放。将后一种情况视为最优选择显然是不合理的,因为只有存在背叛才可能同时无罪释放,这无异于鼓励了背叛。

结论:“不准通过他人受损而使自己获利”的限制条件

由此,为走出囚徒困境所添加的限制条件需要以实现“将策略 E 的收益下降为与策略 F 的收益相等”为条件。笔者认为,该条件可以表述为“不准通过他人受损而使自己获利”。如果一旦有人因为他人受损而使自己获利,则必须放弃所得利益而与他人面临相同的处境。只要有一方因为他人沉默而自己招供获得无罪释放,则必须放弃所获利益与另一方共同面对 10 年徒刑。如此,经过博弈双方的理性分析,一定会同时选择沉默,从而走出囚徒困境。

事实上存在很多通过添加条件而解决囚徒困境的范例,例如设置有仇人追杀的限制条件,如果有人单独出狱则会因为势单力薄而被仇人杀害,如果两人同时出狱则会合二人之力震慑仇人使其不敢施以报复。但归其根本都反映了“不准通过他人受损而使自己获利”的原则,可以说都是该原则的具体应用,例如仇人追杀的限制条件致使博弈者必须获得相同年限的判决,否则就会被杀害以致被淘汰出局。

三、社会合作的两原则及其统一：普遍模仿的后果主义

（一）囚徒困境解决方案的启示：社会合作的“共同获益”与“最大收益”原则

社会合作的基本原则必须满足两方面的要求：其一，该原则必须促使社会中的人进行合作而不是各自为战；其二，该原则必须促进社会中的人选择更优的合作方式。前者是如何保障合作产生的问题，后者是如何选择合作方式的问题；前者是合作产生的前提，后者是合作追求的目标。

从囚徒困境解决方式的分析可得到能够满足这两方面要求的社会合作基本原则：在囚徒困境中，双方都知道同时沉默是最佳选择，那么如何促使博弈双方选择沉默的问题就是如何保障合作发生的问题；另外，双方同时沉默作为最佳选择的原因则可能成为解答如何选择合作方式问题之关键。

对于第一个问题的解答，在囚徒困境中保证双方能自发地进行合作的关键在于“不准通过他人受损而使自己获利”的原则。该原则可以被概括为“共同获益”原则，即“自己获利当且仅当他人也从中获利”。通过上述对于囚徒困境的分析可以看出，如果一旦抛弃了“共同获益”的前提，则博弈双方通过个人理性一定会重新回到囚徒困境的不合作状态——明知同时沉默为最佳方案却最终会选择同时招供。

对于第二个问题的解答，在囚徒困境之中双方选择沉默作为最佳合作方式，其原因有二：首先，双方沉默则双方同时被判一年徒刑，任何人都没有从他人的利益受损中获利，是典型的共同获益；其次，在共同获益的前提下，双方都选择沉默同判一年显然比双方都选择招供同判五年所获利益更大。这就是“最大收益”原则，即在共同获利的基础之上，尽可能地追求所获得利益的最大化和所损失利益的最小化。

这里需要明确的是：最大收益原则一定是建立在共同获益原则基础之上的，因为如何选择合作方式只有在保证合作确实发生的前提下才有意义。如果逻辑顺序颠倒，则更加无益于社会合作。这个社会合作的根本原则，即在保证共同受益基础上追求最大收益的原则，有一个更为流行且简练的表述，那就是孔子说的：己欲立而立人，己欲达而达人。

（二）关注“最大收益”与“共同受益”的后果主义理论

在社会合作的理念之下，一个完整的后果主义理论，既要实现社会合作的“最大收益”原则，也要实现社会合作的“共同受益”原则。传统的后果主义最大的缺陷就在于不能兼顾二者。能够体现“最大收益”原则的后果主义理论，乃是基于功利原则的效益式后果主义；能够体现“共同受益”原则的后果主义理论，乃是基于正义原则的公平式后果主义。任何后果主义理论，都至少要具备两个要素：首先要定义所要追求的后果，即何为根本的“好”；其次要定义如何追求这个后果，即以何种方式实现这个根本的“好”。下文证明“效益式”后果主义与“公平式”后果主义都具备这两个基本要素。“效益式”后果主义的缺陷是由忽略公平造成的；“公平式”后果主义的缺陷是由忽略效益造成的。

1. 关注“最大收益”的后果主义理论：效益式的后果主义

“效益式”后果主义在何为“好”的问题上选择了“幸福”作为所要求追求的后果；而在以何种方式追求“好”的问题上则选择了“效益”理念，即如何实现最大化的幸福。

“效益式”后果主义对于效益的计算并不依赖于每个单独的个体获得了多少效益，而只关注对每个单独的个体所获得的效益进行加总所得到的整体效益。产生整体效益最大的行为就是道德上的最佳选择，产生整体效益最小的行为则是道德上的最差选择。根据“效益式”后果主义的计算方法，如果有两个公共政策，它们的实施可以用不同的方式影响五个人，其中具备“效益式”后果主义特点

的影响结果可以设定为:政策1: $A=1;B=1;C=1;D=1;E=1$;政策2: $A=5;B=2;C=-5;D=15;E=-10$ 。 $A-E$ 代表五个不同的个体,数字则代表政策对这些个体所造成的效益影响。通过效益加总,政策1的总体效益为5;政策2的总体效益为7,所以政策2比政策1在道德上更有优势。但是,“效益式”后果主义没有考虑的是,政策1对五个人的影响都是均等的,没有人因为政策的实施而受到任何的利益损害;政策2对五个人的影响是不等的,其中有两人因为政策的实施而受到严重的利益损害。

在理论缺陷上,对“效益式”后果主义的批评主要集中于三个方面:其一,幸福是否可以作为唯一的重要价值;其二,效益的计算是否对人们提出了过分严苛的要求;其三,“效益式”后果主义有可能带来对公平分配的忽视。第一个问题可以通过变单一价值追求的后果主义为多元价值追求的后果主义来解决;第二个问题可以通过变行为后果主义为规则后果主义来解决;只有第三个问题才是其理论中无法调和的根本缺陷。长久以来“效益式”后果主义的反对者们所提出的理由,包括无视个人权利、无视社会正义、甚至在最大幸福理念庇护下的暴政最终都可以归因于对公平分配的忽视。由此可见,“效益式”后果主义是一个关注结果“最大收益”原则的后果主义理论,忽略了“公平”原则导致其理论存在重要缺陷。

2. 关注“共同受益”的后果主义理论:公平式的后果主义

“公平式”的后果主义在何为“好”的问题上选择了“基本善”作为所要追求的后果;而在以何种方式追求“好”的问题上则选择了“公平”理念,即如何平等地分配基本善。

“公平式”后果主义在权利与自由之间画上了一道不可逾越的界限,并且认为,基本权利之内的利益都必须受到保护,基本权利之外的自由都不应受到侵犯。权利为公平规定了底线,自由为公平提供了空间。即使不能保证人们所获利益完全相同,至少可以保证利益分配的差异仅源于自由意志的选择差异,如此利益分配便确定是公平的。但是,权利和自由就能够保证公平的实现么?人们所获得利益的差异还有一部分源于先天条件的差异,有的人天生聪明,有的人天生强健,更有的人天生就拥有财富和权力,那么天生条件的分配是应得的么?如果天生条件的分配就是不公平的,那么如何又能说明在天生条件基础上通过自由意志行为所导致的利益分配就是公平的呢?由此,自由主义通过改变契约订立的初始状态,塑造了一个真正公平的缔约环境,提出了在尊重自由与权利的基础之上,一定程度地向弱势群体提供政策上的倾斜,实现“有利于弱势者的不公平才是真正的公平”之理念。

在理论缺陷上,无论是由“自由”与“权利”保障的“公平”,还是加入了“差异原则”的有利于弱势者的“公平”,都没有考虑权利与自由界限的严格划分及偏向弱势方的政策倾斜在多大程度上会延缓社会整体福利的提升。从经济学的观点看,当资金集中在能够利用它创造更多财富的那部分人手中时,社会财富才能被更快速的积累,继而也会有助于弱势者利益的提升。虽然社会资源分配并不均衡,但弱势者在“效益优先”的社会模式下所获得的利益却未必少于在“公平优先”的社会模式下所获得的利益,继而也就不能认为“公平式”的后果主义一定会造成有利于弱势者的结果。由此可见,“公平式”的后果主义是一个关注“共同受益”原则的后果主义理论,而忽略了“效益”原则导致其理论仍存在重要缺陷。

(三) 统一社会合作两原则的方法:普遍模仿验证的后果主义道德评价理论

综上,作为追求“最大收益”的后果主义理论,“效益式”后果主义存在缺陷的根本原因在于对“公平”的忽视;作为追求“共同受益”的后果主义理论,“公平式”后果主义存在缺陷的根本原因在于对“效益”的忽视。无论是“效益式”后果主义还是“公平式”后果主义最终都只关注了结果的某一方

面价值而忽略了另一方面,只体现社会合作两原则中的一个而忽略了另一个。由此,完整的后果主义需要同时关注“效益”与“公平”,也必然同时体现出“最大收益”原则与“共同受益”原则。

如何达至二者的统一呢?笔者认为,可以通过将某特定局部的后果分析推广至最大空间域的后果分析,将某特定时刻的后果分析推广至最大时间域的后果分析,来统一两者。这种推广需要借助“普遍模仿验证”的理论:“假定存在一个通用的博弈语境,在其中每个人都是理性的,都追求利益最大化,而且每个人都有足够的学习能力。在这样的条件下,人人都能在博弈过程中互相学习,并且在博弈中模仿习得优势策略。能力更强的人不断推出高明的策略以便获得优势,但优势总是暂时的,高明的策略很快会变成公开的知识而被大家所模仿,直到各种优势策略都出现并且被普遍模仿。此时大家拥有足够饱和的共同知识和对称知识(知己知彼),将出‘集体黔驴技穷’现象,大家都一致模仿被证明为最具优势的策略,于是达到普遍的策略均衡。”^①

道德评价的分析对象,某种行为(或价值)在进行“普遍模仿”验证时,情况如下:在特定情形S中,对行为A进行道德评价。首先假定行为A在特定情形S中是道德上的最优选择,那么普遍模仿则意味着在通用博弈语境中(全社会的空间维度或全历史的时间维度)的每个人,在类似的情形S'中采取类似的行为A',此时A'被假定为优势策略而必然被普遍模仿。那么,对行为A进行道德评价的标准则在于其被普遍模仿之后回馈给行为人什么样的后果。如果被众人模仿的行为A'对每个人(回馈基本相同)都造成正向的回馈,则可以证明该行为是一个道德上好的行为;同理,如果被众人模仿的行为A'对每个人(回馈基本相同)都造成负向的回馈,则可以证明该行为是一个道德上坏的行为。依据此方法,自私自利、忘恩负义、乘人之危、落井下石、姑息养奸等行为一旦被普遍模仿,则会对社会中的每个人都造成负向的回馈,即道德上坏的行为;自由、平等、公正、法治、和谐、民主等价值一旦被普遍模仿,则会对社会中的每个人都造成正向的回馈,即道德上好的价值。^②依此理论进行道德评价,可以统一后果主义的两条路径,形成一个完整的后果主义评价方法。

首先,此理论关注社会合作的“最大收益”原则。此方法最终判断道德上正确行为的标准在于是否对个人有正向的回馈。以社会中每个人所获得的回馈基本相同为前提(此为“其次”的论证对象),每个人都获得正向的回馈将导致社会全体的后果总量增加,从而社会所创造的后果总量越大,此行为所具有的道德价值(行为被提倡的程度)也就越大;每个人都获得负向的回馈将导致社会全体的后果总量减少,从而社会所损失的价后果总量越大,此行为所具有的非道德价值(行为被压制的程度)也就越大。由此,每个人获得回馈的正负与大小就决定了社会后果总和的正负与大小,这与“最大收益”原则的理念相吻合。

其次,此理论关注社会合作的“共同受益”原则。利用普遍模仿将特定环境下的行为推广到最大空间域和最大时间域后,每个人在类似情形下都将采取同样的行为策略,所获得的回馈也应该是基本相同的。即使存在必须牺牲部分人利益的状况,普遍模仿的应用也会导致在最大空间域和最大时间域的范围,在博弈状态下,即在可能性上,也就是进入真实状态之前,形成共同受益。假设初始的后果分析应为:是否应该以牺牲一个人(利益)为代价来拯救五个人(利益)。而普遍模仿的后果分析之所以能够形成共同受益其原因有二:1. 普遍模仿的后果分析注重多次行为所造成的后果,而初始的后果分析只重视单次行为所造成的后果。在多次行为所造成的后果中,行为A中的牺牲者未必再是类似行为A'中的牺牲者,行为A中的被拯救者也未必再是类似行为A'中的被拯救者。虽然对每个相同的个体而言,成为牺牲者的可能性不同于成为被拯救者的可能性,但是对于每个不同的个

^①赵汀阳:《共在存在论:人际与心际》,《哲学研究》2009年第8期。

^②赵汀阳:《共在存在论:人际与心际》,《哲学研究》2009年第8期。

体而言,成为牺牲者的可能性相同,成为被拯救者的可能性也相同。所以,当行为无限次地被重复,初始行为中牺牲者的收益与被拯救者的收益也就趋于相同了。2. 普遍模仿的后果分析对每个人而言都是复合可能性的(成为牺牲者同时成为被拯救者),而初始的后果分析对每个人而言都是单一可能性的(只可能成为牺牲者或被拯救者)。如此,行为被普遍模仿的后果分析则变为:个人在同时成为牺牲者和被拯救者时是否能够获得正向的回馈。虽然对每个相同的个体而言,成为牺牲者的收益不同于成为被拯救者的收益,但是对于每个不同的个体而言,成为牺牲者的收益相同,成为被拯救者的收益也相同。所以,同时成为牺牲者和被拯救者时的收益也是基本相同的。

依据普遍模仿,分析每个收益相同个体的利益得失,在方法上需要借助“无知之幕”^①的博弈模型,在标准上要依赖于特定社会环境下的文化基础。普遍模仿的后果分析并不是类似于数学公式的单变量带入,而是多元化多层次的分析理念。在特定的社会文化环境下,要以特定的社会文化所决定的价值作为所要追求的后果来评价行为;当评价所处的文化环境改变时,则要以改变后的社会文化所决定的价值作为所要追求的后果来评价行为。这样的后果主义道德评价理论既能适应当今社会文化多元和文化宽容的大趋势,又能给道德评价建立一个通用且灵活的标准。而且,此方法的评价对象绝不仅限于评价行为,也可以对权利、权力、价值等进行评价。如果以某种价值作为标准,即后果可以评价行为,那么以人类更基本的需求作为标准,即后果则一样可以评价价值。笔者相信,如此不断深化对于后果的理解,最终必然超越道德相对主义的界限,实现某种由人的共性而阐发的,不依赖于任何社会文化环境所存在的基本需求。笔者认为,这个终极的“后果”应该是人的生存需求,从生存出发对行为、价值、权利、权力等进行的后果主义道德评价,才是最终的道德评价。

(责任编辑:杨嵘均)

Consequentialism as Moral Evaluation Theory in Social Cooperation: From the Perspective of Prisoner's Dilemma

LIU Yu

Abstract: The path adopted by consequentialism is an appropriate approach to moral evaluation. A comparison between the two models of the prisoner's dilemma and the moral evaluation in the light of the concept of social cooperation shows that a deepened investigation of the solutions to prisoner's dilemma can help establish the basic principles of social cooperation, namely the principle of maximum benefit and principle of common benefit. These two principles can respectively be found in utility consequentialism and justice consequentialism. The consequentialism after undergoing a widespread imitation verification can combine these two principles, thus leading to a better realization of the concept of social cooperation.

Key words: social cooperation; prisoner's dilemma; utility consequentialism; justice consequentialism; widespread imitation verification

^①J. Rawls, *A Theory of Justice*, Cambridge, MA: Harvard University Press, 1971, p. 137.