

# 从预期最大化到移情偏好

## ——数理学派公平与正义理论透视

乔洪武 沈昊驹\*

**摘要:**自从罗尔斯的《正义论》问世以来,公平或者说正义成为伦理学、法学和政治哲学聚集的一个问题。哈萨尼、高德和宾默尔等学者采用现代数学方法尤其是博弈论的最新成果,对公平正义理论进行研究,提出了自己的公平正义理论。哈萨尼认为“无知之幕”之后的“理想观察者”根据贝叶斯决策原理订立公平正义的契约,以追求个人效用贝叶斯期望值最大化;高德则认为新古典经济世界的经济人会根据“最小最大相对让步原理”公平地分配“合作剩余”;宾默尔的公平理论要解决的是一个根据移情偏好而非个人偏好来计算效用的新问题,他认为公平的契约是生存博弈的均衡。笔者对数理学派中公平理论从预期最大化到移情偏好这一发展逻辑进行梳理比较,分析其对中国社会主义公平正义伦理体系建设的借鉴意义。

**关键词:**理想观察者 协定道德 合作剩余 生存博弈 道德博弈

在当代西方学术界,公平(Fairness)或者说正义(Justice)是伦理学、法学和政治哲学三大学科聚集的一个中心问题。哈佛大学的罗尔斯(John Rawls)从当代社会发展果实中汲取崭新的理论灵感和营养,建立了他的《正义论》的伦理学——政治哲学体系。罗尔斯的“正义论”以一种崭新的社会契约主义和自由主义传统的理论方式力图为现代社会、尤其是美国民主社会和市场秩序建立“公平正义”的道德基础,重新开拓西方规范伦理学传统之路。照罗尔斯看来,一个正义的社会应该能被视为一个建立在“互惠”和“平等地追求相互利益的人们之间合作”基础上的“自愿企划组合”。罗尔斯的政治建构主义正义论,除了遭到政治哲学中的自由主义“内部”例如诺齐克的严厉批评和挑战外,还遭到以麦金太尔(A lasdair Mac Intyre)等为代表的社群主义的猛烈攻击。当代西方学界的这场大论战,引起了包括哈萨尼(John C. Harsanyi)、帕菲特(Derek Parfit)、哈耶克(Friedrich A. Hayek)、萨金(Robert Sugden)、高德(David Gauthier)、宾默尔(Ken Binmore)等当代伦理学家、经济学家和政治哲学家在同一个层面上和同一个聚集点上进行对话。在这一对话中,哈萨尼、高德和宾默尔等学者采用现代数学方法尤其是博弈论的最新成果,对公平正义理论进行研究,提出了自己的公平正义理论,形成了经济伦理中的数理伦理学派。本文从比较的视角,主要介绍经由宾默尔模型化了的罗尔斯正义理论、哈萨尼的功利主义正义理论、高德公平理论、宾默尔的公平理论,并进行比较,以从中汲取对中国社会主义公平正义伦理体系建设的借鉴意义。

### 一、哈萨尼对罗尔斯正义理论的修正

当代美国影响最大的政治哲学家罗尔斯是从他的“无知之幕”(veil of ignorance)开始对正义问题的研究

\* 乔洪武,武汉大学经济与管理学院,邮政编码:430072,电子信箱:qiaohongwu@126.com;沈昊驹,武汉大学经济与管理学院,邮政编码:430072。

韦森:《经济学与伦理学:探寻市场经济的伦理维度与道德基础》,中文版,37页,上海,上海人民出版社,2002。

韦森称之为“数理伦理学”,并称著名哲学家德雷克·帕菲特为“数理伦理学家”(参见韦森:《经济学与伦理学:探寻市场经济的伦理维度与道德基础》,中文版,38页,上海,上海人民出版社,2002)。据作者了解,国内使用“数理伦理学”这一概念的还有清华大学的赵南元教授,赵南元教授认为:“伦理学为了科学化,应该而且可能向经济学靠拢,像数理经济学那样建立‘数理伦理学’。这种理论的伦理学主要解决我们需要什么道德的问题。”(参见赵南元:《伦理学的若干基本问题》,载 [http://szpx.zjnu.net.cn/webedu/llx/llx\\_8.htm](http://szpx.zjnu.net.cn/webedu/llx/llx_8.htm)) 本文将利用博弈论等数理工具分析经济伦理问题的这一新倾向,归纳在经济伦理学数理学派这一宽泛的旗帜之下。

的。罗尔斯认为：“自由和平等的人们要想就结构的首要正义原则达成公平的协议，就必须消除交易优势”。为此，就必须设计出“无知之幕”，只有“无知之幕”才能保障这一点。在“无知之幕”之后，所有人都不知道自己是什么人，不知道自己在社会中的地位，不知道自己属于哪个阶层，不知道自己的天赋和才能，甚至不知道自己喜欢什么追求什么，因而所有人都不带偏见。“相互冷淡的各方除了有关社会理论的一般知识，不知道任何有关个人和所处社会的特殊信息”。正因为所有的人都处在这样一种“无知之幕”背后的“原始状态”，每个人所做出的决策便毫无偏见；当他们做出决策时他们所达到的一致公认的社会契约，就是正义的。

罗尔斯认为，躲在“无知之幕”背后的人，最关心自己在现实中最好的境况，社会契约的制定者会最大化自己可能的最差境况，因而公平正义的社会契约应满足“适合于最少受惠者的最大利益”这一最大最小标准。假如亚当和夏娃试图就某一分配问题达成一项社会契约（为了简化，我们以两人为例，在原始状态，亚当和夏娃并不知道自己是亚当还是夏娃），这一过程就是亚当和夏娃的博弈过程。最大最小标准就是尽量使得其中获得最小支付的博弈者获得的支付尽可能大，即福利状况最差的人排在最优先的位置。社会福利由福利状况最差的人决定。若用  $(x_A, x_E)$  分别表示亚当和夏娃在这一社会契约下可能获得的支付组合，罗尔斯认为这种支付是可以进行人际比较的“基本善”，则罗尔斯主义的最大化的社会福利函数  $W_R$  为：

$$W_R(x) = \min\{x_A, x_E\}$$

用  $x = (x_A, x_E)$  的集合  $X$  的边界来表示亚当和夏娃博弈的效率均衡，可行集  $X$  是凸和完备的。罗尔斯认为因为存在着需要进行矫正的“出身与天赋的不平等”，因此，该集合非对称。 $W_R(x) = \min\{x_A, x_E\} = d$  为不同水平  $d$  下亚当和夏娃获得的支付组合，这个组合形成一条无差异曲线。显然，可行集  $X$  与无差异曲线的切点是满足最大最小标准的最大效用点。罗尔斯认为，在这一点上，既满足博弈的可能支付和最大最小标准，又使得社会福利最大化，因而是人人可以接受的，所以亚当和夏娃的这种支付分配是公平的。由无差异曲线  $W_R(x) = \min\{x_A, x_E\} = d$  的特性可知在该点  $x_A = x_E$ ，这就是自由主义者攻击的罗尔斯的平均主义。

“无知之幕”的目的是要去掉人们之间的差异以便在一个完全平等的基础上进行讨论。哈萨尼的功利主义支持用“无知之幕”使个人进入原始状态重构他们的个人偏好，但是与罗尔斯根据拉普拉斯理由非充分原理把最大最小标准作为在原始状态的理性决策原理不同的是，他提出了一个“理想观察者”，“理想观察者”关心的不是自己将来最坏的状态，而是根据主观概率来决定人们的社会角色定位和根据一定的期望来计算作为理性决策的原理。哈萨尼认为：“……理性人会努力把社会福利函数建立在个人效用单位的‘真实’的交换比率的基础上（交换比率可能是由对每个人的个人特征和控制人们行为的心理规则有充分信息的观察者来使用的）。但是如果他没有充分的信息来计算这些‘真实’的交换比率，他将尽最大努力来对后者进行估计。”

哈萨尼根据萨维奇 (L. Savage) 在冯·诺依曼和摩根斯坦 (J. Von Neumann and O. Morgenstern) 工作基础上发展起来的贝叶斯决策原理，通过冯·诺依曼-摩根斯坦效用确定支付的方法来处理不确定性条件下的决策问题，以定义冯·诺依曼-摩根斯坦形式的期望效用函数：

$$W_R(x) = Ux_A + Vx_E$$

此处  $U, V$  分别为成为亚当和夏娃的主观概率。当估计成为亚当和夏娃的主观概率相等时，这个函数有

约翰·罗尔斯：《作为公平的正义：正义新论》，中文版，26页，上海，上海三联书店，2002。

约翰·罗尔斯：《正义论》，中文版，6页，北京，中国社会科学出版社，1988。

约翰·罗尔斯：《正义论》，中文版，83~84页，北京，中国社会科学出版社，1988。

因为在无知之幕之后参与人不知道自己将是亚当还是夏娃，其只关心自己可能的最坏情况，因此亚当和夏娃获得基本善的组合 (5, 3) 和 (3, 6) 对其来讲是无差别的，因为最坏的情况都是 3，组合 (5, 3) 和 (3, 6) 是同一条无差异曲线上的点。显然，组合 (6, 4) 和 (4, 5) 构成的无差异曲线比前者福利水平高。

此处包含一个重要假设：即“想当然地认为对亚当的支付的测度单位可以直接与夏娃支付的测度单位进行比较。确切地讲，一个隐含的假设是每一单位的亚当的支付可以与夏娃的一个单位的支付进行交换。”（参见肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，61页，上海，上海财经大学出版社，2003）。假如亚当和夏娃获得支付的测度单位不同，而其比为  $V:U$ ，则社会福利函数为  $W_R(x) = \min\{Ux_A, Vx_E\}$ ，公平博弈则沿  $Ux_A = Vx_E$  推进。罗尔斯定义这条曲线为贡献曲线，由于不同时期  $U$  和  $V$  的比值是变化的，公平不再是按直线推进，这条贡献曲线假定了由社会基本结构确定的社会合作是互相有利的（参见约翰·罗尔斯：《正义论》，中文版，77页，北京，中国社会科学出版社，1988）。

Harsanyi, 1977. Rational Behavior and Bargaining Equilibrium in Games and Social Situations Cambridge University Press, Cambridge, p. 60.

种简化形式： $W_R(x) = x_A + x_E$ 。因此，在上例亚当和夏娃的博弈支付可行集  $X$  中，哈萨尼的期望社会福利函数  $W_R(x) = x_A + x_E = c$  形成了不同水平  $c$  上的功利主义的无差异曲线。显然，该无差异曲线与约束集  $X$  的切点为哈萨尼认为的现代功利主义公平点，即哈萨尼认为这点使得在支付可能的约束内，人人的期望效用均达到了最大化，若把亚当和夏娃放在“无知之幕后”，他们就会同意在这个点上达成的社会契约，哈萨尼认为这种契约是公平的。

罗尔斯认为，“无知之幕后”要厚到无法估计自己将来角色的概率，因此其与哈萨尼的理想观察者的区别在于，前者使用最大最小规则而后者追求根据主观概率计算的期望最大化。由于可行集  $X$  具有罗尔斯描述的需要进行矫正的“出身与天赋的不平等”，因此显然功利主义公平点与罗尔斯公平点不同。罗尔斯认为，“在其他情况相同的条件下，……功利主义允许较大的不平等。”哈萨尼在追求个人效用贝叶斯期望值最大化时并不关注这个总和在个体之间的分配状况，而认为在支付集合约束下社会效用期望达到最大的结果就是一种公平的结果，这是“理想观察者”们能够达成的一种公正的契约结构。因此，森认为，“功利主义方法不仅被广泛应用于判断分配判断，甚至还成为了平等主义的一个标准，这多少有些令人匪夷所思。”

## 二、高德德的协定道德理论

自从 1971 年出版《正义论》以来，罗尔斯的伦理政治哲学受到诺齐克、麦金太尔和桑德尔以及哈耶克等学者的批评与挑战；另一方面，也有一些西方思想家沿着罗尔斯的分析思路进一步探索。道德哲学家高德以一种契约论的方法来确定道德的理性基础，其 1986 年出版的《协定道德》(Morals by Agreement) 一书就是这一领域的又一学术力作。

首先，高德德的“协定道德”的理论起点不是罗尔斯的“无知之幕后”的原始状态，而是经济学中的新古典世界。其理论建构的基本假设中，协定道德的主体就是新古典经济学世界中的理性经济人。在这个新古典世界中，“一个人被构想为不受约束的行为中心，致力于以他的能力和资源去实现他的利益。”第二，高德定义了一个重要的概念——合作剩余。合作剩余就是指社会合作生产产生非合作生产所不能达到的惠利，用经济学的话来说就是“贸易的好处”。高德认为，理性人实际上并不是就社会分配中的个人份额多少进行讨价还价，而是对这种“合作剩余”各自所得的份额进行讨价还价，在讨价还价的境遇中，人们必须协定规则，来决定参与者的行动和合作剩余中的分配份额，并达成一项可避免效率损失而陷入“囚徒困境”的有约束力的协议。高德认为，这一协议能得到所有人的同意，就是公平、道德的。

根据高德的理解，假设亚当和夏娃在规划某种合作项目，以产生一定量的合作剩余，亚当和夏娃可获得的支付组合  $(x_A, x_E)$  形成可行集  $X$ 。其要解决的就是从“原始讨价还价状态”  $(x_A, x_E)$  到效率均衡这部分合作剩余如何公平分配的讨价还价问题  $(X, \cdot)$ 。高德认为，每个参与到合作体系中的人对合作剩余的最初要求，应当等于他对这种合作剩余的边际贡献价值。

高德设计了一个二阶段的理性讨价还价的过程。为了达到一致同意，每个人都对初始要价做出一定的让步。他把“参与人的妥协的相对重要性”即“相对让步”定义为  $C_i = (X_i - x_i) / (x_i - x_i)$ ， $x_i$  为参与人  $i$  可获得最高支付，此处  $i = A, E$ 。在此定义下，高德提出，假如参与合作的讨价还价者同样是理性的，那么“每个人作为效用最大化者，在努力最小化他的让步，没有人会期望其他理性人将会愿意做出他自己不愿做出的相同让步来让步”。这里的“相同让步”可用“相对让步”的量值来表示。他认为理性的讨价还价者选择的是使  $C_A$  和  $C_E$  中较大的一个最小化的让步，这就导致这样一个结论：一致同意将会在一个最小化的最大相对让步值这一点上取得。高德称之为“最小最大相对让步原理”(the principle of minmax relative concessions)。这一原理意味着每一个人和其他人一样一般会做出相等的相对让步。换句话说，合作剩余将按每个人要价的同一比例进行分割。因此社会福利函数为：

$$w_{ks}(x) = \min\{(x_A - x_A) / (x_A - x_A), (x_E - x_E) / (x_E - x_E)\}$$

社会福利函数与支付可行集  $X$  切于  $k$  点 (该解称之为卡莱-斯莫尔定斯基讨价还价解，因此福利函数为  $w_{ks}$ )，即高德认为的公平点。照高德看来，每个人都同意做出相等的相对让步，这既是理性的，也是公平正义的。这种人人受益、人人接受的协定契约，是合乎道德的。

显然，高德所竭力证明的公平的道德，“既不是在我们的同情心中，……也不是在任何据称是独立于我

约翰·罗尔斯：《正义论》，中文版，78页，北京，中国社会科学出版社，1988。

阿马蒂亚·森：《论经济不平等——不平等之再考察》，中文版，14页，北京，社会科学文献出版社，2006。

Gauthier, D., 1986 *Morals by Agreement* Oxford: Oxford University Press, p. 9.

Gauthier, D., 1986 *Morals by Agreement* Oxford: Oxford University Press, p. 143.

们个体利害关系的客观义务中，……而是以受益于每一个人，并且也为每个人理性接受的方式，在我们互惠事务的明智排序中。”

### 三、宾默尔的公平理论

作为博弈论的大师，宾默尔不仅成功地把罗尔斯、哈萨尼和高德等人的观点博弈模型化，而且在对前人公平观念模型比较的基础上，提出了自己的公平观念。

首先，宾默尔对人性的假设是新古典经济学的假设，即假设人们在他们自己的文明中由自利动机引导行为，但宾默尔又引入了“移情”这一概念，以试图来解决人际比较问题。宾默尔指出：“‘移情’指的是一个过程，通过这一过程我们想象自己从他人的立场和角度在看问题……在亚当对夏娃产生移情时，可以理解是他在设身处地地为她着想并且根据她的观点进行推理，但这并未对他的个人偏好产生任何影响。”“一个放高利贷的人也可能对电影里一位老夫人产生了同情而陷入悲痛。但是没有什么可以阻止他在离开剧场后擦去眼泪继续以电影中学来的方式敲诈另一位年迈的寡妇。”宾默尔认为经济人必须在一定程度上具有移情能力。正是借助“移情”这一概念，宾默尔展开了对公平理论的分析。

在其公平博弈理论的起点问题上，宾默尔认为“与其追随哈萨尼支持进入原始状态重构他们的移情偏好，重新使他们认为自己具有同样的理想观察者的观点，不如简单地估计个人是根据他们在真实生活中的移情偏好来设想自己在原始状态的角色的。”因此，宾默尔公平理论的起点，并不是罗尔斯“无知之幕”之后的原始状态，而是一种根据真实生活设想的原始状态。这种原始状态前的“无知之幕”是薄的无知之幕，“它与罗尔斯的无知之幕的不同之处在于它不要求亚当和夏娃忘掉现实社会的组织方式。在薄的无知之幕后面，他们保留了有关当前的自然状态的知识，他们也没有忘记类似哈萨尼故事中的移情偏好。”

那么，移情偏好又有何特点呢？在宾默尔看来，“在一个社会中占主导地位的移情偏好在很大程度上是一个文化现象。……每一个构筑了真实社会的内部之间相互连接的亚社会都具有自己的文化而且还可能有自己内部的移情模式。”移情偏好的特点是：不同的人对移情的度量是有差别的，这种差别就如同用两支分别用华氏和摄氏表示的温度计来测量一样，移情偏好具有可塑性，它会随着社会的演化而不断地发生变化。这两个特点也正是移情偏好与新古典经济学中的理性经济人的不同之所在。

其次，在上述分析基础上，宾默尔借用企业讨论生产成本的经济理论术语，区分了三个时间段：短期、中期和长期。用“固定偏好”来表示每个人的个人偏好，这种个人偏好的典型特征是稳定性；用“可变性偏好”用来表示个人的移情偏好。“我们对待生活方式的偏好的可塑性就是一个具有普遍性的移情偏好模型。”个人偏好的均衡是生存博弈的均衡；而移情偏好的均衡是道德博弈的均衡。

在生存博弈中，宾默尔所面临的是同高德一样的讨价还价问题  $(X, v)$ 。宾默尔定义了社会福利函数：

$$W_N(x) = (x_A - A)(x_E - E)$$

显然， $W_N(x)$  在与可行集  $X$  的切点达到最大化，宾默尔称这一生存博弈的均衡为纳什讨价还价解  $v$ ，宾默尔认为，纳什讨价还价解与高德的卡莱-斯莫尔定斯基解一样，不具有任何道德上的含义。

在短期，亚当和夏娃的个人偏好与移情偏好都是固定的，虽然他们的个人偏好可能相去甚远，但是他们的移情偏好的比是固定的，例如为  $v:U$ 。在无知之幕后面，他们并不知道自己的个人偏好，他们根据移情偏好来讨价还价，得到一条罗尔斯主义的公平推进曲线  $U_{x_A} = V_{x_E}$  或修正的罗尔斯公平推进曲线  $U(x_A - A) = V(x_E - E)$ 。宾默尔称这一解为比例讨价还价解，比例讨价还价解是道德博弈的均衡。

在中期，个人偏好是固定的，但作为社会演化结果的移情偏好可能会变化。宾默尔认为通过模仿和教育，“在中期，社会演化的力量为个人的移情偏好定了型”，这种“作为移情偏好集合的移情均衡是演化过程的最终产品而不是参与人有意计算的结果”，而且“一旦到达均衡位置，演化过程就会停止。”宾默尔假设，若将亚当和夏娃从他们的文化和道德博弈中分离出来，“他们将直接就社会契约问题进行谈判而不必去想除生存博弈之外的任何其他博弈的情形。此时，他们仅进行生存博弈，而不进行任何道德博弈，因而必然

Gauthier, D., 1990 *Moral Dealing: Contract, Ethics and Reason* Cornell University Press, p. 9.

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，72页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，39页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，79页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，405~406页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，79页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，80页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，82页，上海，上海财经大学出版社，2003。

会得到一个不具有任何道德内涵的理性纳什讨价还价解。在社会演化有时间使他们的移情偏好适应于这个讨价还价问题后,宾默尔再将亚当和夏娃重置于他们的文化中,但是由于中期移情偏好已经发生了变化,即“让  $U$  和  $V$  具有这样的值,比例讨价还价解 就与纳什讨价还价解  $v$  重合。”“当这些移情偏好被嵌入到道德博弈中时,由此产生的社会契约正是我们认为不具任何道德内涵的纳什讨价还价解。在中期,生存博弈导致的社会演化过滤掉了道德博弈中的所有道德内涵,但“*Adam* 和 *Eve*看上去并不是这样,他们仍根据从文化中继承而来的移情偏好在进行道德博弈。”

也就是说,在中期,由于社会演化的力量使得道德博弈产生了与生存博弈重合的均衡解  $v$ ,假设在长期中出现了产生新技术的机会,可行集不再是  $X$ ,而是发生了向右推移的  $Y$ 。在长期里需要解决的是一个以中期均衡  $v$  为原初讨价还价状态的新讨价还价问题  $(Y, v)$ 。在一个较长的时期内将会产生一个新的道德博弈与生存博弈重合的讨价还价解。在长期中,个人偏好容易受到演化压力的影响而发生变化,就像在中期移情偏好容易受到影响发生变化一样。因此,显然长期里的从  $v$  到 的公平契约推进曲线不再是中期里的从 到  $v$  的公平契约推进曲线的自然延伸,其斜率发生了变化。如果时期足够长,连接不同时期讨价还价解得到的是一条斜率不断变化的弧线的长期公平推进曲线。

宾默尔的公平理论要解决的是“一个根据移情偏好而非个人偏好来计算效用的新问题。”由道德博弈达到的公平观念,是一种移情均衡,移情均衡受生存博弈演化的影响。“一项公平的契约不过是一个生存博弈中的均衡,条件是如果道德博弈中使用这一均衡的战略不会对道德博弈的参与人产生求助于原始状态装置的激励。……事实上,道德博弈只不过是一个在生存博弈中进行均衡选择的协调装置。……对于规则有约束力的道德博弈产生的是一个生存博弈的均衡。”因而宾默尔认为,“并不存在先验于社会的道德的基石”;由此得出了与高德一样的“市场是一个道德无涉区”的结论。

#### 四、比较与借鉴

综观数理学派中代表人物哈萨尼、高德和宾默尔的公平正义理论,他们与罗尔斯一样,均采用契约论方法作为分析理路,即把公平的观念都建立在一种人人均接受的社会契约上。总体上讲,他们的公平正义理论在以下三个方面存在着异同:

第一,罗尔斯理论中博弈的起点,是处在“无知之幕”之后的原始状态。罗尔斯认为,由于原始状态的人最终可能处于收入分配的任何一个位置上,所以,他采用最大最小规则制定社会契约。哈萨尼所代表的现代功利主义的公平理论,也考虑公平的博弈起点,理想观察者站在原始状态,用主观概率预计期望效用最大化。罗尔斯与哈萨尼的公平正义理论都有平均主义的倾向,虽然这种平均的程度存在着差别。他们的理论与诺齐克有着重大区别。以诺齐克为代表的自由意志主义 (*Libertarianism*) 者认为,机会平等比收入平等更重要。诺齐克将意志自由放在第一位,强调过程公平的重要性,他认为,“政府应该强调个人的权利,以确保每个人有同样发挥自己才能并获得成功的机会。一旦建立了这些游戏规则,政府就没有理由改变由此引起的收入分配。”

高德公平理论的起点是新古典经济学世界,而不是原始状态。当然,高德的公平理论并不在于起点是否公平的强调,而着重于过程中按最小最大相对让步原理对“合作剩余”分配的公平原则上,这点显然与自由意志主义者相同。宾默尔公平理论中的起点,既不是躲在无知之幕之后的原始状态,也不是新古典经济学世界,而是一种个人根据他们在真实生活中的移情偏好来设想的原始状态。

第二,罗尔斯认为,处于原始状态的人会特别关注处于收入分配最底层的可能性。因此在设计公共政策时,其目标应该是提高社会中状况最差的人的福利。“哈萨尼与罗尔斯的理想观察者的区别在于前者追求期望最大化而后者使用最大最小规则。”高德则借用新古典经济学中边际贡献的概念,提出最小最大相对让步原理,强调按贡献分配的公平性。宾默尔则跳出了这一框架,将社会博弈分为生存博弈和道德博弈,认为“一项公平的社会契约因而也成为道德博弈的均衡,但绝对不能忘记的是,它还必须是生存博弈的均衡——否则,它将不具有可行性”,而且“那些与我们人类一同发育的道德博弈的规则只不过是包含在我们

肯·宾默尔:《博弈论与社会契约(第1卷:公平博弈)》,中文版,108页,上海,上海财经大学出版社,2003。

肯·宾默尔:《博弈论与社会契约(第1卷:公平博弈)》,中文版,105页,上海,上海财经大学出版社,2003。

肯·宾默尔:《博弈论与社会契约(第1卷:公平博弈)》,中文版,55页,上海,上海财经大学出版社,2003。

曼昆:《经济学原理(微观经济学分册)》,中文版,425页,北京,北京大学出版社,2006。

肯·宾默尔:《博弈论与社会契约(第1卷:公平博弈)》,中文版,98页,上海,上海财经大学出版社,2003。

文化中虚构的内容”，道德博弈的均衡最终要与生存博弈均衡重合。由强调道德博弈的均衡最终要与生存博弈均衡重合，宾默尔走向了道德虚无主义。

第三，从道德哲学角度来看，哈萨尼同罗尔斯一样，属于道德先验主义者，这点同高德、宾默尔有本质的区别。众所周知，罗尔斯在康德哲学基础上建构了正义理论，认为道德观念是先验的。哈萨尼借用布兰德(Brunt)行为功利主义和规则功利主义的概念，认为，“即使违约并不会受到惩罚，道德人也谨守契约，因为他们的道德指令告诉他们要谨守契约。”高德却从新古典经济学中的理性人假设出发，放弃了行为主体的道德前提，认为通过理性讨价还价得到的结果是公平的，市场是一个道德无涉区。宾默尔认为任何社会契约，首先是符合生存博弈均衡的解，其次才考虑道德含义。并且宾默尔认为在社会博弈中，道德博弈中的移情偏好最终要受生存博弈的影响，因而得出了与高德一样的市场道德无涉的结论。

与罗尔斯的正义论一样，数理学派提出的公平正义理论受到了来自各方的挑战。例如哈萨尼提出的理想观察者的看法是基于一种基数效用，难以成功解决人际比较的问题。“哈萨尼公理在目前的情况下解决不了任何问题。”而“高德并没有赋予卡莱-斯莫尔定斯基解任何伦理方面的美德，他的故事中的主人为了讨价还价并没有穿过无知之幕，由此得出在他的讨价还价故事中包括人际效用比较在内的任何伦理内容都一定聚集于他对状态的选择。”而宾默尔试图从生存博弈中推导出道德来，“游走于道德与理性、实然与应然之间，纯粹理论理性的推理自然会有超越自己有限性的冲动，并不时僭越地坐在道德推理的法官席上”的道德虚无论，更是饱受萨金等人的批评。

数理学派的公平正义理论尽管存在这样那样的缺陷，但从理论和实际两方面来看，仍有可资借鉴之处。

首先，从哈萨尼、高德与宾默尔对罗尔斯正义论的发展来看，一方面，他们的理论假设的起点从“无知之幕”之后的原始状态到新古典经济学世界再到个人根据他们在真实生活中的移情偏好来设想的原始状态，表明数理学派的公平正义理论的研究假设力图一步步地更接近现实社会，更真实地揭示出社会公平正义的科学内涵。尽管目前尚未达到这个目标，但这一理论探索方向无疑是正确的。另一方面，他们所得出的公平正义的结果各不相同，这既为进一步深化研究开辟了不同的路径，又为我们提供了这样一个重要启示：多样性的公平正义判断标志本身就是与多样性的现实社会相吻合的，而单一的公平正义价值观反而会存在片面性。

其次，从我国当前实际来看，党的十七大明确提出实现社会公平正义是发展中国特色社会主义的重大任务。通过对数理学派经济伦理中的公平正义理论进行比较分析，我们以为它对于我国社会主义公平正义伦理体系建设有着重要借鉴意义。

第一，发展市场经济要建立互利互惠的道德法则以为其提供经济伦理支撑。罗尔斯与数理学派虽然对公平的观念各异，高德和宾默尔甚至得出了市场无涉道德的结论，但有一点是共同的，即公平正义的分配必须是博弈各方的帕累托改进，是对合作剩余的分配。一个阶层和群体的利益增进不能以损害另一个阶层和群体的利益为前提条件；换言之，富裕群体的发展和困难群体的生活改善应当同步推进；同样地，在强调保障劳动群体的利益快速增长的同时，也应当依法维护非劳动群体的合法权益。这一点，恰恰是从市场经济体制本身引申出的最基本的道德法则。也只有坚持这一点，我们才能真正保障社会公平正义，不断促进社会和谐，形成全体人民各尽其能、各得其所而又和谐相处的局面。

第二，应明确按贡献分配所得也是社会主义公平正义的重要内容和标志之一。在高德的协定道德中，公平是取得合作剩余的必要条件，不公平的让步原则只会让成员放弃合作，整个社会也不可能取得合作剩余。公平正义原则要求在对合作项目收益进行分配时，遵循按照贡献进行分配的原则，使合作成员得到自己所应

---

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，55页，上海，上海财经大学出版社，2003。

行为功利主义作为一种理论，是指社会效用最大化的功利主义原则应该直接应用于每个人的行为：对于任何给定的情况，道德上正确的行为应该是使社会效用最大化的某种特定的行为。相反，规则功利主义的理论认为，功利主义的选择标准应首先应用于可供选择的可能的道德规则，而不是直接应用于可供选择的可能的行为：对于任何给定的情况，道德上正确的行为是与适用于该情况的“正确的道德规则”相符合的行为；而“正确的道德规则”在这种情况下，被定义为如果每个人都遵守的话，将使社会效用最大化的特定的行为规则。

Harsanyi, 1987. "Review of Gauthier's 'Morals by Agreement'." *Economics and Philosophy*, Vol 3, pp. 339 - 343.

盛庆崃：《统合效用主义与公平分配》，108页，杭州，浙江大学出版社，2005。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，78页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，104页，上海，上海财经大学出版社，2003。

肯·宾默尔：《博弈论与社会契约（第1卷：公平博弈）》，中文版，6页，上海，上海财经大学出版社，2003。

乔洪武：《互利——市场经济最基本的道德法则》，载《光明日报》，1993-09-01。

得到的那一份。这不仅可以消除平均主义的影响,还具有至关重要的激励作用。十七大报告强调,必须坚持把发展作为党执政兴国的第一要务。发展,对于全面建设小康社会、加快推进社会主义现代化,具有决定性意义。而要坚持发展,就不能否认在市场经济的初次分配中,必须始终明确按贡献分配所得就是社会公平正义的题中应有之义。而在再分配中,更加注重分配结果的公平,这也是符合现代文明社会公平正义原则的一项重要举措。

第三,要积极推进机会平等和分配公平正义的社会基本结构和制度建设,使社会博弈不仅要达到生存博弈的均衡,而且还要达到道德博弈的均衡。宾默尔的公平理论指出,个人既不是躲在无知之幕后原始状态,也不是新古典经济学世界,而是一种个人根据他们在真实生活中的移情偏好来设想的原始状态。由道德博弈达到的公平观念,是一种移情均衡,它最终要受生存博弈演化的影响。也就是说,生存的利益博弈对道德博弈起支配作用。这一理论可用原捷克著名经济学家奥塔·锡克更通俗的话来说明——经济利益是人的最基本利益,“是一种一般劝说和鼓动等等所不能改变的,而直接由人们在社会中的基本地位引起的。……是一种有时会使认识失去作用,并且只要认识与之相对立便会否认这种认识的利益。”虽然市场本身与道德无涉,但市场所依存的、规范和调节人与人利益关系的社会基本结构和制度是道德有涉的。不道德、非正义的社会基本结构和制度所规定的利害关系,既会诱使人们产生不道德的“移情偏好”,又会毒化人们生存竞争的环境,导致优败劣胜、尔虞我诈,人与人像狼一样的生存博弈结局。因此,公平正义的社会基本结构和制度建设既是一个社会是否实现公平正义的标志,又是保障社会公平正义实现的基本条件。诚如罗尔斯所说,“社会基本结构之所以是正义的主要问题,是因为它的影响十分深刻并自始至终。”而公平正义的社会基本结构和制度首先应做到,在参与财富等社会资源分配之前,根据机会平等的规则摒弃先赋性(如特权、身份等级)等不公正因素的影响,保证每一位社会成员能够有一个平等竞争的条件,能够得到公正的对待,从而拓展自由创造的空间,最大限度地发挥自己的能力和才干。其次应在财富生产的过程中,健全劳动、资本、技术、管理等生产要素按贡献参与分配的制度,确保按贡献分配所得的规则得到遵守。最后,在财富再分配的过程中,还有必要建立健全利益补偿和协调机制,着力提高低收入者收入,逐步提高扶贫标准和社会救助标准,以扭转收入分配差距扩大趋势。惟其如此,我国社会主义市场经济中的社会博弈将不仅能达到生存博弈的均衡,而且还可以达到道德博弈的均衡。

#### 参考文献:

1. Binmore, K., 1994. *Game Theory and Social Contract, Vol. I: Playing Fair*. Cambridge, Mass.: The MIT Press.
2. Binmore, K., 1998. "The Evolution of Fairness Norms." *Rationality and Society*, Vol. 3, pp. 275 - 301.
3. Gauthier, D., 1986. *Morals by Agreement*. Oxford: Oxford University Press.
4. Gauthier, D., 1990. *Moral Dealing: Contract, Ethics and Reason*. Cornell University Press.
5. Harsanyi, J., 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy*, Vol. 63, No. 4 (Aug., 1955), pp. 309 - 321.
6. Harsanyi, J., 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, Cambridge.
7. Harsanyi, J., 1978. "Bayesian Decision Theory and Utilitarian Ethics." *American Economic Review*, Vol. 68, No. 2, (May, 1978), pp. 223 - 228.
8. Harsanyi, J., 1987. "Review of Gauthier's 'Morals by Agreement'." *Economics and Philosophy*, Vol. 3, pp. 339 - 343.
9. Sugden, R., 1989. "Spontaneous Order." *Journal of Economic Perspectives*, Vol. 3, No. 4, pp. 85 - 9.
10. 阿马蒂亚·森:《论经济不平等——不平等之再考察》,中文版,北京,社会科学文献出版社,2006。
11. 奥塔·锡克:《经济-利益-政治》,中文版,北京,中国社会科学出版社,1984。
12. 费尚军:《道德的博弈何以可能——对高塞尔协定道德的一种解读》,载《华中科技大学学报(社会科学版)》,2007(6)。
13. 肯·宾默尔:《博弈论与社会契约(第1卷:公平博弈)》,中文版,上海,上海财经大学出版社,2003。
14. 曼昆:《经济学原理(微观经济学分册)》,中文版,北京,北京大学出版社,2006。
15. 乔洪武:《互利——市场经济最基本的道德法则》,载《光明日报》,1993-09-01。
16. 盛庆峰:《统合效用主义与公平分配》,杭州,浙江大学出版社,2005。
17. 韦森:《经济学与伦理学:探寻市场经济的伦理维度与道德基础》,中文版,上海,上海人民出版社,2002。
18. 约翰·罗尔斯:《正义论》,中文版,北京,中国社会科学出版社,1988。
19. 约翰·罗尔斯:《作为公平的正义:正义新论》,中文版,上海,上海三联书店,2002。

(责任编辑:刘成奎、彭爽)

奥塔·锡克:《经济-利益-政治》,中文版,303页,北京,中国社会科学出版社,1980。

约翰·罗尔斯:《正义论》,中文版,7页,北京,中国社会科学出版社,1988。