

浅析罗尔斯差别原则

王现惠

黑龙江大学哲学学院，黑龙江 哈尔滨

收稿日期：2025年9月30日；录用日期：2025年10月20日；发布日期：2025年11月3日

摘要

本论文旨在对约翰·罗尔斯在其《正义论》中提出的“差别原则”进行分析。文章首先阐述了差别原则的定义、目的及其在罗尔斯正义理论体系中的位置，即作为第二原则的一部分，旨在通过安排社会和经济的不平等，使其最有利于社会中“最不利者”。随后，论文的核心部分转向对该原则面临的三个主要理论困境进行批判性探讨：第一，作为其理论基础的“原初状态”和“无知之幕”是过度理想化的假设，脱离现实；第二，“最不利者”的定义模糊不清，难以在实践中准确界定；第三，该原则可能过分强调社会责任，从而削弱了个体对自己选择和努力应承担的责任。最后，文章总结认为，尽管存在这些争议，差别原则在推动正义理论从形式平等转向实质平等方面具有重要价值，并为思考当代社会公平问题提供了思想资源。

关键词

罗尔斯，《正义论》，差别原则

A Brief Analysis of Rawls's Difference Principle

Xianhui Wang

Faculty of Philosophy, Heilongjiang University, Harbin Heilongjiang

Received: September 30, 2025; accepted: October 20, 2025; published: November 3, 2025

Abstract

This paper aims to analyze the “Difference Principle” proposed by John Rawls in his “*A Theory of Justice*”. The article first elaborates on the definition, purpose, and position of the Difference Principle within Rawls’ justice theory framework, namely, as part of the second principle, it aims to arrange social and economic inequalities in a way that benefits the “least advantaged” in society the most. Subsequently, the core part of the paper turns to a critical discussion of the three main theoretical dilemmas faced by this principle: first, the “original position” and the “veil of ignorance” that serve

as its theoretical foundation are overly idealized assumptions, divorced from reality; second, the definition of the “least advantaged” is ambiguous and difficult to accurately define in practice; third, this principle may overly emphasize social responsibility, thereby weakening the responsibility individuals which should bear for their own choices and efforts. Finally, the article concludes that despite these controversies, the Difference Principle holds significant value in promoting the shift of justice theory from formal equality to substantive equality and provides intellectual resources for thinking about contemporary social fairness issues.

Keywords

Rawls, “A Theory of Justice”, Difference Principle

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

罗尔斯的差别原则主张社会和经济的不平等必须有利于“最不利者”，这一理念体现了对公平和弱势群体的关注。我们首先需要明确的是“差别原则”是什么？在现实中，我们如何确定谁才是真正的“最不利者”？社会差异如此复杂，制度又如何保证差异需求不被忽视？更重要的是，差别原则在强调社会集体责任的同时，是否可能弱化个体对自身选择与努力的道德承担？另一方面，“原初状态”与“无知之幕”设定是否过于理想化？这些争议实际上都指向差别原则在程序性正义框架下，难以充分回应关于个人“应得”(desert)的实质性正义问题。

2. 差别原则的提出

什么是差别原则？罗尔斯在《正义论》中给出了系统性的哲学回应。差别原则并非独立存在，而是从属于罗尔斯正义原则的理论结构中。他提出“正义二原则”：第一原则是“平等的自由原则”，第二原则分两部分，即机会均等原则和差别原则[1]。正义的两条原则分别对应于社会基本结构中的政治和经济部分，目的是优先保障社会成员基本自由问题并调节社会经济不平等问题。其中正义二原则之间有着明确的先后排序关系。第一个原则优先于第二个原则，第二个原则中的机会均等原则又优先于差别原则。差别原则虽然在罗尔斯正义理论中排在词典式序列的最后，但它在整个正义理论中却具有不可替代的地位和作用。

罗尔斯在《正义论》中将差别原则定义为“社会和经济的不平等应该这样安排，使它们在与正义的储蓄原则一致的情况下，适合于最不利者的最大利益”[2]。根据这一观点我们明晰差别原则的对象是不平等的。在社会和经济生活中个体由于自然禀赋和社会出身等原因导致人们获利不同，从而产生社会和经济不平等的现象。针对这一现象，罗尔斯主张用差别原则来进行“平衡”和“选择”。另一方面，差异原则不是无条件适用的。它强调一个重要前提便是“在与正义的储蓄原则一致的情况下”，即在确保自由平等原则和机会均等原则得到充分尊重和落实的基础上，我们才能运用差别原则。此外，罗尔斯差别原则侧重维护最不利者的最大利益来缩小贫富差距，进一步缓解社会和经济的不平等状况[3]。

罗尔斯对差别原则的描述并非经验性的政策建议，而是罗尔斯通过“原初状态”这一思想实验推导出的规范性结论。他假设理性的人们在原初状态的“无知之幕”下选择正义原则，会倾向于考虑最差情境下的自身利益，这也就是“最大最小值规则”。换句话说就是人们在不知道自身天赋、社会出身和

地位等偶然信息的前提背景下,会综合考量各种结果发生的概率,比较每一个选项可能带来的最坏结果,最终会侧重选择那个最坏结果中相对较好的选项。因为在无知之幕的遮蔽下,任何人都有可能成为自然天赋较差或者社会出身不利的“最不利者”。出于规避风险的意识,人们会倾向于选择一种能最大限度地保障弱势群体利益的原则从而保障自身利益。自由平等原则和机会均等原则在一定程度上能够消除社会偶然性因素的影响,但却不能消除自然偶然性因素的影响。这也意味着人的能力和天赋的分配仍有可能决定财富和收入的分配[4]。在罗尔斯看来,人的社会出身和自然天赋等偶然因素不应决定人们的命运,因为这些因素在道德上是任意的、随机的,是个人不应得的[5],因而这样获得的平等是不正当的。所以正义的目标是建立一种分配制度,使得所有人都自愿加入合作体系。无论个体因偶然因素获得何种优势,不管是聪明才智还是富裕的出身,他们利用这些优势的获利行为必须以“改善最少受惠者的最大利益”为前提。反之,如果一个人因为自然资质的偶然性或者社会出身的偶然性而受到损害,个人应该得到社会的补偿。当然,这里提到的“最少受惠者的最大利益”,是在保证其他阶级的利益不受损害或者利益提高的基础上使最少受惠者获得最大利益。差别原则通过这种制度约束,能够将偶然性带来的优势转化为促进共同福祉的工具从而建立互利互惠的社会合作体系,减少不平等的特权。

罗尔斯差别原则从某种程度上是利用个体间的天赋差异来改善社会的共同利益。差别原则并不强求彻底消除个体自然禀赋和社会出身之间的差异,而是通过制度设计让“在自然禀赋和社会起点上具有优势的人”承担起“在自然禀赋和社会起点上处于劣势的人”应有的道德责任。通过制度设定,差别原则使得优势群体充分发挥自身的先天优势来改善弱势群体的处境从而促进整个社会的发展和进步,形成一个良好的互惠互利的发展模式。这种安排不是简单的财富转移,而是通过机会开放和资源倾斜等机制让天赋和出身的偶然优势成为推动整个社会公平的积极力量。从表面上看,罗尔斯差别原则允许存在一定的不平等,但不意味着他默许不平等现象。只有在承认当前社会客观存在的不平等的前提下,致力于将不平等控制在最小必要范围内,从而使每个人都实质性地实现自由和平等的情况下才是合理的。具体来说就是在确保权利平等、机会公平开放这两项基础原则得以落实的前提下,差别原则能够让每个人都能保有尊严地生活,并确保财富分配的不均衡状态不会损害个体所享有的平等自由与公正机会。

3. 差别原则的理论困境

3.1. 原初状态与无知之幕的理想化假设

罗尔斯将“原初状态”(Original Position)与“无知之幕”(Veil of Ignorance)构建为正义原则推导的“代表装置”(representative device)。它的核心功能是排除自然个体天赋、社会出身等偶然性信息的干扰,模拟一个公平的立约环境,确保所有参与者站在无偏倚的立场上选择正义原则,从而推导出能够真正代表所有社会成员利益的正义原则的理论设计。这一设计试图以程序正义的纯粹性来解决社会合作中因偶然优势导致的不公平问题。然而这一“代表装置”在理论功能与现实适用性之间存在着深刻的张力,其理想化假设构成了差别原则的首要困境。

原初状态与无知之幕是罗尔斯在《正义论》中构建的“作为公平的正义”理论的理想化假设[6]。所谓原初状态,并不是指历史上真实存在的“自然状态”,而是被假想的一个初始情境。在这种情境中,被假定为理性且相互冷淡的参与者要共同选择决定一个正义社会的基本结构与原则,但参与者对自己的天赋能力、社会地位、个人特征、教育背景、社会阶级、宗教信仰等信息一无所知。这一构想的逻辑是试图构建一个选择情境来推导出人们在不知道自身特殊身份与利益的情况下,会理性选择出哪些作为社会基本结构的正义原则。显而易见,在罗尔斯的这种设定下,原初状态中的个体最终会选择他所提出的两个正义原则,特别是差别原则。那么为什么理性的个体会倾向于支持差别原则呢?罗尔斯指出,当人们在面对不确定性时,会综合考量利益计算和风险规避,采取所谓的最大最小化策略,也就是优先考虑最坏

情况下的利益最大化,从而避免让自己处于社会最底层。通过构建一个中立、理性的选择情境,罗尔斯试图排除由于个人禀赋和社会出身等原因对正义原则选择过程造成的干扰,从而推导出对所有人都公平、公正且具有普遍适用性的正义原则。

上文提到原初状态与无知之幕的构想起初就是建立在理想化的假设之上的。但是关于这一假设能否在现实中成立或运行,罗尔斯对此并没有给出准确的定论。罗尔斯在正义论中推论道,参与者在原初状态的无知之幕,即“彻底的信息遮蔽”下做出关于正义原则的选择。但是在现实世界中,这种彻底的信息遮蔽是根本不存在的。原初状态与无知之幕的设定本身一定程度上忽视了与现实的联系。现实中人类个体总是生活在特定的社会、文化、历史与家庭背景之中,人们的价值观、情感认同、宗教信仰等因素潜移默化地塑造着他们的正义感与道德判断,并深刻影响着他们对正义与公平的理解。所以原初状态下“彻底的信息遮蔽”这一设定显得过度理想化,甚至在一定程度上脱离了现实实际。

罗尔斯在原初状态中将个体假设为具有同等理性能力且相互冷淡的抽象的存在的人,却忽略了现实中个体之间存在的差异性。现实中的人们在智力、认知、道德等方面存在着各种差异,这种差异会直接或间接地影响人们对正义的理解和选择。罗尔斯假定的“相互冷淡”的设定也不符合现实中人们的交往状态。在现实中人们彼此之间存在着亲情、友情、集体归属感等多种情感和利益关联。这些关联使得人们在思考正义问题时不只是从自身的理性利益出发,还可能会考虑到他人和群体的利益。罗尔斯这种对个体属性的简单设定忽略人们之间的情感和利益关联,使得他构建的原初状态脱离了现实基础。

罗尔斯的正义理论追求的是一种普遍适用且具有道德正当性的正义原则。他假定原初状态下所有参与者基于理性思考会选择同一组正义原则。但这一结论是否具有唯一可能性呢?现实生活中不同的文化传统、宗教信仰和价值观会产生不同的正义观,人们对“正义”的理解便存在多种不同的解读。虽然罗尔斯尝试通过这种程序性的设置来消除所有分歧的想法是好的,但他忽略了价值冲突的必然性和合理性。就算是处于相同的信息约束下,不同个体也会因为各自不同的价值排序和道德信念选择完全不同的正义原则。在当代政治哲学中,功利主义主张最大幸福原则,自由至上主义强调最大自由原则,社群主义推崇共同善原则。不同的价值观和道德基础影响不同个体或群体选择不同的正义原则。由此可见原初状态中的选择结果并不一定是单一导向。罗尔斯只考虑理论上共识的观点可能会阻碍实践中差异化的选择,从而忽视多元共存的必要与价值。

3.2. “最不利者”定义模糊

罗尔斯理论中“最不利者”指的不是某个具体的人,而是指在社会中最不利的群体或阶层。那么什么样的标准能确定谁是“最不利者”呢?罗尔斯将“基本善”的持有份额作为标准,把“最不利者”理解为在社会基本善的分配中处于最低水平的群体。但基本善包括自由、权利、机会、收入、财富和自尊的社会基础等多种因素,以基本善分配中的最低水平来界定最不利者仍然存在争议^[7]。个体之间存在差异,不同个体的劣势根源和需求程度是不同的。一个人的劣势可能来自残疾,也可能来自不良的教育背景。有的人看重收入与财富,有的人可能注重健康。“最大利益”内涵对他们来说是完全不同的。如果我们用收入和财富来界定最不利者,可能会忽视那些像残疾人、少数民族等在非物质层面处于劣势的群体。这些群体的不利处境不一定完全反映在基本善上,但却对他们的生活产生很大的影响。

“最不利者”的处境不是一成不变的,它是动态的。随着政策调整以及经济状况等条件的改变处于不利地位的群体也会发生变化。当下的最不利者可能因教育、政策扶持或偶然机遇在未来境遇中得到改善,而原本处于中间阶层的人也可能因意外、疾病或家庭变故变为新的最不利者。罗尔斯关注社会结构的整体正义,但并未解决如何在不断变化的社会中动态地识别最不利者。在收入水平最低但身体健康的个体和收入水平较高但身患重病的个体之间,哪一方的利益更应当优先得到改善?不同群体间利益冲突

问题应当如何解决？罗尔斯差别原则在这些问题方面依旧面临挑战。

“最不利者”并不是一个需要精确到个体的绝对概念，而是一个相对的范畴。英国政治哲学家布莱恩·巴里在《正义诸理论》中指出罗尔斯的差别原则本质上是一种“比较性正义”(comparative justice)。罗尔斯的差别原则并非追求绝对的“每个人获得完全相同的资源”，而是确保社会中最弱势群体的相对处境得到改善。这种比较性逻辑表明“最不利者”是一个动态的、相对的概念。当社会结构变化时，最不利者的群体构成可能也会随之变化，而差别原则的任务是持续关注这些相对处境最差的群体，而不是固守某个静态的群体标签。虽然在实际操作中难以精确界定最不利者的具体范围，但差别原则提供了一种价值导向，即社会制度应优先考虑改善那些最需要帮助的人的处境。这种相对的界定方式并非缺陷，而是一种基于现实复杂性的务实选择，它允许在具体政策制定中根据不同的社会背景和实际情况进行灵活调整来实现实质意义上的公平。

通过原初状态与无知之幕后这一“公平的程序”，罗尔斯正义原则的推导获得了它的道德正当性。这一程序排除了偶然性因素的干扰，使得所有参与者在“无偏倚的立场”上协商正义规则。虽然原初状态中的个体被抽象为“相互冷淡的理性自利者”，剥离了个体间的情感联结、道德差异和具体社会身份，但这一看似“去差异化”的程序设计却暗含着对差异需求的间接回应。因为它要求选择的正义原则必须能够适用于“所有可能处于任何社会地位的个体”，包括那些因自然或社会因素处于极端劣势的群体。这一程序设计并非忽视差异，而是通过“普遍适用性”的强制性要求，把对差异需求的考量融入正义原则的制定过程之中。当参与者不知道自己未来可能成为残疾人还是低收入者时，他们出于“规避风险”的理性考量也会优先选择那些能够改善潜在弱势群体的正义原则。我们可以设想一个理性参与者，不论他是因出身贫困无法获得优质资源还是在将来因意外而失去劳动能力，差别原则所要求的社会再分配安排也能保障他的基本权益。尽管这种推理逻辑没有明确提及具体差异，却通过“改善所有可能劣势”的程序约束，一定程度满足了个体的差异需求。罗尔斯支持者指出，差别原则虽由理论推导得出，但确定“最不利者”的范围以及调节再分配资源等具体举措仍依赖社会成员的公共讨论和民主决策。在这一过程中，不同群体的差异化需求能够通过政治参与被呈现，并在程序正义的约束下被纳入政策考量。这种程序化的参与机制使得抽象的正义原则能够动态适应个体差异的现实，从而避免差异化的需求被忽视。

3.3. 从“最不利者”到“个体责任”：社会责任和个人能动性的张力

罗尔斯批评者认为，差别原则的设定可能过于偏向社会责任，即社会整体(通过制度)承担调节不平等的主要责任，从而弱化了个人的选择、努力及其相应责任的道德重要性。而罗尔斯差别原则主张社会制度的安排应最大限度地改善最不利者的处境[8]。这一原则将关注的重点放在了社会制度和资源再分配对个体处境的影响上。从罗尔斯的观点中我们可以推论出天赋差异产生的经济收益和社会优势并不完全属于个人，而是具有某种社会共享的正当性。如果个人因为他们较高的天赋、能力或努力程度所获得额外的回报，只有在能够改善最不利者整体利益的前提下才被认为是正当的。这一观点体现了公平分配与对弱势群体的关怀，但在个体责任边界方面依旧存在争议。罗尔斯从制度层面上改善弱势群体处境这一举措可能会逐步产生社会理应对所有结果差异负责的处境。这就隐含着弱化个人对他们自身努力、个体选择和行为所应承担的道德责任的潜在问题。在差别原则的逻辑下，一个因懒惰而贫困的人和一个先天疾病而贫困的人都应该得到社会同等的关注与补偿。但这种处理方式可能会削弱个体对自身行为后果的责任意识，使部分人产生“不用努力，社会也会兜底”的想法，降低个人奋斗的积极性从而影响社会整体发展和进步。差别原则通过补偿因出身贫困、天赋水平较低等非应得的劣势来实现公平这一方案固然具有合理性，但是过度注重“非个人选择因素”，可能间接忽视了个体在公平机会下的主动选择对他们最终境遇的影响。出身相似但选择不同的两个个体最终会取得不同的收入或社会地位，这种差异在多少程

度上应当被再分配机制调节？差别原则的调节逻辑虽然旨在实现结果公平，但却容易将“最不利者”的困境简单归因于外部因素，从而忽视个体自身选择的影响，从而削弱了个体对自身选择所承担的个体责任。

事实上，罗尔斯并没有否认个人责任的存在。自然天赋的差异、家庭出身的优劣以及社会制度的结构缺陷等因素是非个人选择的，这些因素不应该影响人们的生活机会，而正义理论的目标就是调节那些由非个人选择因素导致的不平等[9]。关于个人兴趣不同、努力程度差异、个人决策偏好等个人选择产生的不平等，罗尔斯强调这类后果应当由个体自行承担，因为“人们要为自己的选择承担责任”。罗尔斯差别原则是以“公平的机会平等”为前提条件的。社会制度通过资源分配等方式调节外部环境来消除社会出身和天赋等非个人因素对个体发展的阻碍，而不是直接干预个体努力或选择的结果。在此前提下个体依旧需要在公平的起点上自主决策并承担相应责任。如果两个出身相似但努力程度不同的个体拥有同等的教育资源和就业机会，那么两者最终的收入差异在罗尔斯的理论中仍被视为个体选择的自然结果，而非必须被再分配调节的对象。差别原则对社会责任的强调并不是要取代个体能动性，而是通过弥补非个人选择因素的外部不公平，为所有个体提供一个真正公平的起点，从而使个体责任能够在平等的条件下被充分实现。在罗尔斯的理论中社会责任与个体能动性并非是对立的两极，二者的动态平衡最终是为了实现“作为公平的正义”这一核心目标。

4. 结语

尽管罗尔斯差别原则面临着“原初状态”与“无知之幕”过度理想化、“最不利者”定义模糊以及可能削弱个体责任等理论困境，但这一原则依然具有不可忽视的积极意义。从理论上来看，差别原则主张社会经济的不平等安排必须要改善最不利者的最大利益，将正义的衡量标准从形式的公平转向实质的公平。这一转向不仅体现了罗尔斯对弱势群体关注，更是从社会制度设计层面上改善那些处于社会不利地位且需要基本保障的人群的处境，为构建更加公平正义的社会提供了政策指导。罗尔斯差别原则在保障最不利者的基本权利与发展机会的前提下推动社会资源分配向弱势群体倾斜，一定程度上缩小了贫富差距，缓解了社会矛盾，引导社会朝着更加公平正义的方向发展。

罗尔斯差别原则对弱势群体的注重和对实质公平的追求，为我们思考社会正义问题提供了宝贵的思想资源。虽然这一原则在实践中面临许多争议和挑战，但正是这些争议与反思推动着我们不断深化对“什么是公平”“如何实现正义”这些问题的理解。罗尔斯的正义原则不仅为构建更加公平的社会提供了理论指引，也为我们在复杂现实中寻求正义的实践路径提供了重要启示。

参考文献

- [1] 李义天, 刘畅. 超越差别原则——凯·尼尔森对罗尔斯正义理论的批判[J]. 湖北大学学报(哲学社会科学版), 2021, 48(3): 49-56.
- [2] Rawls, J. (1973) *A Theory of Justice*. The Belknap of Harvard University Press.
- [3] Rawls, J. (2001) *Justice as Fairness—A Restatement*. Harvard University Press.
- [4] [美]约翰·罗尔斯. 正义论[M]. 何怀宏, 何包钢, 廖申白, 译. 北京: 中国社会科学出版社, 2009.
- [5] 齐延平. 论社会基本制度的正义——对罗尔斯正义理论的讨论[J]. 北方法学, 2007(4): 19-38.
- [6] 张怀民, 杨丹. 罗尔斯正义构建的逻辑起点评析: 以博弈论为视角[J]. 学术交流, 2012(8): 34-38.
- [7] [美]约翰·罗尔斯. 正义论[M]. 何怀宏, 何包钢, 廖申白, 译. 北京: 中国社会科学出版社, 2009.
- [8] 龚群. 自由主义与社群主义的比较研究[M]. 北京: 人民出版社, 2014.
- [9] 陈伟. 当代西方政治思潮[M]. 北京: 中国人民大学出版社, 2012.