

罗尔斯公平与效率道德基础的建构逻辑

马丹阳

郑州大学哲学学院, 河南 郑州

收稿日期: 2026年7月27日; 录用日期: 2026年8月18日; 发布日期: 2026年8月28日

摘要

在现代社会正义理论的论争中, 公平与效率的关系重构始终是核心议题。罗尔斯在《正义论》中通过“原初状态”与“无知之幕”的思想实验, 突破了古典功利主义将公平依附于效率的理论困境, 构建了“公平的正义”理论体系。该理论以“人是目的”为终极道德依据, 依托“原初状态”与“无知之幕”的伦理论证, 建构起了公平与效率的道德基础, 并通过与功利主义、诺齐克主义等流派的对比, 确立了“公平优先于效率”的价值排序, 为效率划定了明确的道德边界。不过, 在诺齐克等当代思想家的攻击中, 公平优先于效率的当代合法性意义, 需要持续深入体悟。总之, 罗尔斯《正义论》的建构逻辑是当代自由主义对时代问题的最有力回应, 当代的所有公平正义秩序的建立, 都必须尊重这点。

关键词

罗尔斯, 原初状态, 无知之幕, 公平的正义, 效率, 道德基础

The Constructive Logic of Rawls's Moral Foundation for Fairness and Efficiency

Danyang Ma

School of Philosophy, Zhengzhou University, Zhengzhou Henan

Received: July 27, 2026; accepted: August 18, 2026; published: August 28, 2026

Abstract

Within the ongoing debates on justice theory in modern society, the reconfiguration of the relationship between fairness and efficiency has consistently remained a central issue. In *A Theory of Justice*, Rawls employs thought experiments based on the “original position” and the “veil of ignorance” to overcome the theoretical dilemma inherent in classical utilitarianism, where fairness was subordinated to efficiency, and thereby constructs a theoretical framework of “justice as fairness.” This theory takes the premise that “the individual is an end” as its ultimate moral foundation; relying on the

文章引用: 马丹阳. 罗尔斯公平与效率道德基础的建构逻辑[J]. 哲学进展, 2026, 15(9): 75-81.

DOI: 10.12677/acpp.2026.159422

ethical reasoning derived from the “original position” and the “veil of ignorance,” it establishes a moral basis for both fairness and efficiency. Furthermore, by contrasting this approach with schools such as utilitarianism and libertarianism, it establishes a value hierarchy wherein “fairness takes precedence over efficiency,” thereby delineating clear moral boundaries for efficiency. However, in response to critiques from contemporary thinkers such as Rawls’ opponents within the liberal tradition, the contemporary legitimacy of prioritizing fairness over efficiency requires continued and profound reflection. In summary, the constructive logic underlying Rawls’s *A Theory of Justice* represents the most robust response by the mainstream contemporary liberalism to the pressing questions of our era; the establishment of any contemporary order of fairness and justice must necessarily respect this principle.

Keywords

Rawls, Original Position, Veil of Ignorance, Justice as Fairness, Efficiency, Moral Foundation

Copyright © 2026 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：理论缘起与核心问题

1.1. 功利主义正义观的理论困境

自边沁、密尔创立古典功利主义以来，“社会整体福利最大化”始终是其核心正义标准。功利主义主张，“(就整个共同体而言)当一项行动增大共同体幸福的倾向大于它减小这一幸福的倾向时，它就可以说是符合功利原理，或简言之，符合功利。”[1]因此，判断一项社会制度或政策是否正义，关键在于其能否实现“最大多数人的最大幸福”，即通过效用计算实现社会总福利的最优状态。在这种理论逻辑中，“公平”并非具有独立道德地位的价值，而是作为实现“效率”(总福利最大化)的工具性存在。然而，这种将“效率”置于绝对优先地位的理论范式，不可避免地陷入了个体权利与集体福利的冲突困境，即为了实现社会总福利的提升，少数人的基本权利与合法利益可能被正当化地牺牲。学者姚大志指出，“在原则上，功利主义会允许为了得到更大的功利总额而牺牲某些人的利益，甚至会为了更大的利益而剥夺少数人的自由和权利。”[2]例如，在功利主义框架下，若牺牲少数弱势群体的财产权能换来多数人的普遍福利提升，这种牺牲将被视为正义之举。罗尔斯以其康德式的义务论伦理学立场，否定功利主义的目的论伦理学主张。在罗尔斯看来，“自由与权利的要求”必须要优先于“对社会福利总量之增加的欲求”。罗尔斯认为，“社会的每一成员都被认为是具有一种基于正义、或者说基于自然权利的不可侵犯性……正义否认为使一些人享受较大利益而剥夺另一些人的自由是正当的。”[3]

表面上看，功利主义似乎是个人主义的，但罗尔斯深刻地洞察到，功利主义并不是个人主义的，因为功利主义只是把适合于个人的选择原则扩展应用于社会。学者王立指出，“功利主义将社会总体欲望体系视为个人欲望的集合，因而，个人选择原则和社会原则是同一个原则。”[4]基于此，罗尔斯尖锐地指出，功利主义的根本缺陷在于其并不认真对待人与人之间的差别性与分离性，由于将社会视为一个整体的具有自身生命的有机体，功利主义忽视了每个个体都是独立的、不可替代的道德主体。这种理论逻辑本质上是将人视为实现集体福利的手段，而非目的本身，严重背离了对个体尊严的平等尊重。因此，从“个人主义”的逻辑起点出发，功利主义走向的却并非“个人主义”的目的地，并从根本上，将“自由与权利的要求”置于“对社会福利总量之增加的欲求”之后。

1971年罗尔斯《正义论》出版，终结了功利主义在现代政治伦理学领域百余年的话语主导权，重新复兴了社会契约论传统，将“正义是社会制度首要价值”的义务论立场确立为现代政治哲学的核心基准。随着现代社会对个体权利重视程度的提升，功利主义的理论局限性日益凸显，亟需一种新的正义理论来重构公平与效率的关系，为社会基本制度提供更具正当性的道德基础。

1.2. 核心问题：社会基本制度的正义性根基重构

现代性的基本价值“自由”与“平等”经过几百年的发展，前者在法政层面已经基本落地，而后者还是时代的焦虑。“罗尔斯把正义视为现代政治哲学的主题，并提出‘正义总是意味着平等’，这意味着他将政治哲学的主题由自由变为平等。罗尔斯的历史地位和重要意义就在于他完成了西方政治哲学主题的转换。”^[5]为实现“平等”的落地，罗尔斯将研究视野聚焦于“社会基本结构”，认为社会基本制度的正义性是整个社会正义的基础。在罗尔斯看来，现代社会的核心正义问题在于：如何设计社会基本制度，才能既保证社会合作的效率，又维护个体的平等权利与尊严，实现公平与效率的协调统一？这一问题的本质是为公平与效率确立正当性的道德根基。传统理论要么陷入“效率至上”的功利主义误区，要么走向“绝对平等”的平均主义极端，均无法为现代社会提供可行的正义方案。罗尔斯的核心任务便是跳出这种二元对立，通过理论创新构建一套既能彰显公平的独立道德价值，又能为效率提供合理空间的正义体系。

所以，罗尔斯的贡献是在“自由”已经取得历史性进步的同时，对于“平等”的正义观念取得的时代突破，就是可以干预部分人的自由(往往是强者)而实现比较普遍的“平等”，而这种“平等”虽是对少数精英的损害，却对整体“社会基本结构”是更有利的，成就一种整体性的最大的善。“所谓正义，也就是让每一个人的权利都得到保障的一种社会价值。”^[6]所以当代伦理学也是政治学，论证上就是政治哲学。

2. 核心概念界定：原初状态与无知之幕的逻辑建构

2.1. 原初状态：理想的契约环境设定

为解决上述核心问题，罗尔斯回归并创新了古典契约论传统，提出了“原初状态”和“无知之幕”的思想实验方法。与霍布斯、洛克等古典契约论者将契约视为历史真实存在的状态不同，罗尔斯明确指出，“原初状态”纯粹是一种假设的状态，其价值不在于描述历史事实，而在于为我们提供一种反思性的视角，让我们能从无偏私的立场出发选择正义原则。这种思想实验的核心优势在于，原初状态提供了一种“客观的视角”，它能够通过设定特定的约束条件，排除那些在道德上无关的偶然因素(如出身、天赋、财富等)对正义原则选择的干扰，确保“立约者”在平等、理性的环境中达成共识。通过这一方法，罗尔斯成功将正义原则的选择转化为一个理性选择问题，为公平与效率的道德基础建构提供了严谨的逻辑支撑。

为了给正义原则的选择设定一个理想的契约环境，罗尔斯给“原初状态”做出如下定义，“它是一种其间所达到的任何契约都是公平的状态，是一种各方在其中都是作为道德人的平等代表、选择的结果不受任意的偶然因素或社会力量的相对平衡所决定的状态。这样，公平的正义从一开始就能使用纯粹程序正义的观念。”^[3]“原初状态”的核心特征在于为“立约者”提供了平等、理性的选择前提，在这样一个“正义环境”中，既要求平等的客观环境，例如共有一个确定的地理环境、各人的身体和精神状况等能力相似、以及各个领域存在中等程度的匮乏；同时也要求平等的主观环境，涉及合作的主体，也就是各个“立约者”。“立约者”的设定具体包含三个关键要素：

其一，平等的立约者。在原初状态中，所有参与正义原则选择的立约者都是“自由、平等、理性且合理的道德人”。这里的“自由”意味着立约者能够自主选择并追求自身的生活目标；“平等”则体现为立

约者拥有平等的选择权利与议价能力，不存在身份、地位、权力的差异；“理性且合理”则要求立约者既能理性地追求自身利益最大化，又能尊重他人的合理诉求，具备达成共识的道德能力。这种平等的立约者设定，排除了现实社会中因权力不平等导致的规则偏向性，为公平选择奠定了主体基础。

其二，相互冷淡的理性动机。为了保证正义问题出现时人们的行为和动机公平平等，罗尔斯将“立约者”的动机假定为“相互冷淡”，这意味着“立约者”既不追求利他，也不刻意损人，而是专注于自身“基本善”的最大化。所谓“基本善”，是指所有理性人都需要的、对实现任何生活目标都不可或缺的资源与权利，包括自由、权利、机会、财富、自尊等。这种动机设定既避免了功利主义的利他假设过于理想化，又克服了极端利己主义无法达成共识的缺陷，确保“立约者”能够在理性权衡的基础上选择普遍适用的正义原则。

其三，共识性的目标追求。所有“立约者”的共同目标是选择一套能够规范社会基本结构的正义原则，这套原则将成为社会合作的基本准则，具有终极性与不可违背性。“立约者”清楚地知道，他们选择的原则将直接决定自己未来在社会中的生活前景，因此会以审慎的态度进行选择，确保所选原则能够兼顾自身与他人的利益，实现社会合作的稳定与持久。

2.2. 无知之幕：无偏私选择的核心约束机制

如果说“原初状态”为正义原则选择提供了理想的环境框架，那么“无知之幕”则是确保选择无偏私性的核心约束机制，是罗尔斯理论中最具创新性的设计。

为了实现纯粹的程序正义，排除各种特定偶然性的影响，罗尔斯假定各方处于“无知之幕”之后。具体来讲，“无知之幕”的核心逻辑是：在立约者选择正义原则时，将被一层“幕布”遮蔽掉所有关于自身的、偶然的、非道德的个人信息，从而避免“立约者”因自身特殊处境而偏向于选择有利于自己的原则。

被“无知之幕”遮蔽的信息包括四个层面：一是个人的先天禀赋，如智商、体能、相貌、健康状况等；二是个人的社会处境，如家庭出身、阶级地位、财富水平、职业身份等；三是个人的价值偏好，如宗教信仰、人生理想、生活方式等；四是社会的具体客观条件，如经济发展水平、文化传统、政治体制等。这些信息的共同特征是“在道德上具有任意性”，一个人拥有高智商、出身富裕家庭，并非源于自身的努力，而仅仅是运气；反之，出身贫寒、天生残疾也并非个人过错。罗尔斯认为，这些偶然因素不应该成为决定社会规则的依据，因此必须通过“无知之幕”予以排除。

需要强调的是，“无知之幕”并非将“立约者”变成“没有理性的白痴”，它保留了“立约者”做出理性选择所必需的核心信息：一是“立约者”拥有普遍的理性与道德能力，能够进行逻辑推理与价值判断；二是“立约者”了解“基本善”的重要性，清楚自身对基本善的需求；三是“立约者”掌握人类社会的基本事实与一般规律，如资源的稀缺性、合作的必要性、人性的基本特征等。这些保留的信息确保立约者能够在无偏私的前提下，做出符合自身与社会整体利益的理性选择。

“原初状态”与“无知之幕”的组合，其核心价值在于构建了一个无偏私的公平选择语境，排除了现实社会中普遍存在的背景依赖。在现实生活中，人们的利益诉求往往与自身的特殊处境密切相关：富人会反对累进税，强者会拒绝弱者保护，特权阶层会维护现有不平等的制度。这种基于自身特殊利益的选择，必然导致所制定的规则缺乏普遍性与公平性。

而在“无知之幕”的遮蔽下，“立约者”无法预知自己未来在社会中的具体位置，既可能成为社会顶层的富豪，也可能沦为底层的贫民；既可能是健康的强者，也可能是残疾的弱者。在这种“未知自身祸福”的状态下，理性的“立约者”会意识到，任何偏向某一群体的原则都可能让自己陷入不利境地，因此必然会选择一套对所有群体都公平的、具有普遍性的正义原则。“原初状态”与“无知之幕”的组合能够

让“立约者”摆脱因社会地位与背景产生的偏见与歧视，为正义原则的制定创造良好环境。这种通过排除偶然特权实现的选择公平，是罗尔斯构建公平与效率道德基础的逻辑起点。

2.3. 正义原则：公平与效率的分层价值排序

在“原初状态”与“无知之幕”的约束下，“立约者”如何推导出具体的正义原则？罗尔斯认为，此时理性的“立约者”会遵循“最大最小值规则”(maximin rule)进行选择。这一选择逻辑的合理性源于“无知之幕”带来的不确定性。由于“立约者”无法预知自己未来在社会中的具体位置，他们无法通过概率计算来权衡不同原则的收益与风险，因此会采取一种“审慎的保守策略”，避免选择那些可能让自己陷入极端不利境地的原则。例如，若选择功利主义原则，虽然可能获得较高的收益，但也存在被牺牲的风险；而选择一套能够保障底层群体利益的原则，则能确保即使自己沦为最少受惠者，也能获得基本的权利与福利保障。因此，遵循最大最小值规则，是“立约者”在“无知之幕”下的理性选择。

基于最大最小值规则，罗尔斯推导出了“公平的正义”的两大原则，第一原则是“平等的自由原则”，第二原则是“机会公平平等原则与差别原则”；并明确了它们之间的优先次序，第一原则优先于第二原则，在第二原则内部，机会公平平等原则优先于差别原则。这两大原则共同构成了罗尔斯正义理论的核心，为公平与效率的关系设定了清晰的价值排序。

罗尔斯的两大正义原则，本质上构建了一套“公平优先、兼顾效率”的价值架构，明确了效率的合法性来源，即只有在符合公平原则的框架内，效率才具有道德正当性。罗尔斯并非反对效率，而是反对“为效率而效率”的功利主义逻辑，认为脱离公平约束的效率是赤裸的，不具备正义性。

3. 道德基础的三重建构：从终极依据到具体支撑

3.1. 终极根基：继承康德“人是目的”的道德内核

罗尔斯公平与效率道德基础的终极依据，源于对康德道德哲学“人是目的，而非手段”核心原则的继承与转化。“因为理性存在者全都服从这条法则：每一个理性存在者都应当决不把自己和其他一切理性存在者仅仅当作手段，而是在任何时候都同时当作目的自身来对待。”^[7]也就是说，在康德看来，人作为理性的存在者，具有内在的价值与尊严，必须被当作目的本身来对待，而不能仅仅被当作实现其他目的的手段。罗尔斯将这一道德原则从个体道德领域拓展到社会制度领域，认为社会基本制度的设计必须以尊重每个个体的平等尊严为底线，尤其是罗尔斯的“差别原则”，规定任何经济不平等都要能够最大化弱势群体的利益。

在罗尔斯看来，功利主义的根本错误在于违背了康德“人是目的”的这一原则，将个体视为实现社会总福利的工具。而他的“原初状态”与“无知之幕”，本质上是将“人是目的”的原则转化为具体的制度设计：通过无知之幕排除偶然因素的干扰，确保每个立约者的利益都能被平等考量，没有人会被当作“牺牲者”来换取他人的利益。这种对个体平等尊重的强调，构成了罗尔斯公平与效率道德基础的终极内核，即无论效率提升的幅度有多大，只要违背了对个体平等尊严的尊重，就不具备道德正当性。不过，也有学者对罗尔斯的“差别原则”提出质疑，“道德直觉上，我们应该给予穷人利益更大的道德重要性，优先提高穷人的利益才对。差别原则依循了这一直觉的方向，优先考量弱势群体的利益。但它走向了另一个极端，把对穷人利益的‘优先考量’变成了‘唯一考量’。”^[8]这确实是一个尖锐的批判角度，其间张力需要学术界继续探讨。不过，罗尔斯是用“最大最小值”原则来替代功利主义的“最大多数人的最大幸福”原则，因此，当强调穷人利益之时，罗尔斯的本意是包含穷人的所有人，而非仅仅考虑大部分人。因此，当其考量穷人利益之时，全部人的利益已尽在考量之中了。

总而言之，罗尔斯的理论不仅具有重要的学术价值，还为现代社会的制度设计提供了坚实的道德依

据,对现实政策产生了深远影响:在税收制度方面,差别原则为累进税制度提供了道德依据,对富人征收更高比例的税收,并非对富人的“惩罚”,而是为了让效率的红利能够共享,保障最少受惠者的基本利益。在教育公平方面,机会公平平等原则推动了义务教育的普及、贫困地区教育资源的均衡配置等政策,为弱势群体提供了平等的发展机会。在社会保障方面,罗尔斯的理论支持建立普惠性的社保体系,保障公民的基本生活权利,减少社会不平等。在政治制度方面,平等的自由原则为公民的基本政治权利提供了保障,推动了民主制度的完善与发展。

3.2. 公平根基:纯粹程序正义的自证性支撑

为了论证公平的独立道德价值,罗尔斯区分了三种正义类型:完善的程序正义、不完善的程序正义与纯粹的程序正义。完善的程序正义具有明确的正义结果标准与可实现该结果的程序;不完善的程序正义有明确的结果标准,但没有能保证实现该结果的程序;而纯粹的程序正义则没有预设的正义结果标准,只要程序本身是公平的,通过该程序产生的结果就是正义的。

罗尔斯认为,“原初状态”与“无知之幕”所构建的,正是一种纯粹的程序正义。这一程序没有预设任何“理想的分配结果”(如绝对平均或按贡献分配),但它满足了所有的公平条件:平等的立约者、无偏私的选择环境、理性的商议过程。因此,通过这一程序推导出的两大正义原则,以及依据这些原则制定的社会制度,其公平性是“自证的”,不需要依赖外在的标准来评判,程序的公平性直接赋予了结果的公平性。

这种纯粹程序正义的设计,解决了公平道德基础的核心问题:“为什么公平是正义的?”罗尔斯的答案是,公平的道德价值并非源于其工具性(如公平能带来效率),而是源于其内在的程序正当性。公平本身就是一种值得追求的道德价值,是社会基本制度不可或缺的德性。

3.3. 效率根基:“正义的效率”的双重道德约束

公平与效率之间存在张力与博弈,“公平意味着平等,意味着社会的稳定。而稳定是一个社会发展的前提,没有稳定,就无从说起效率……一个只有效率而没有公平的社会也是不可能长久地存在,这样的社会必定充满着暴力和革命,或迟或早它将会被颠覆。”^[9]罗尔斯认为效率构建的道德基础,核心在于将效率纳入公平的框架,使其成为“正义的效率”,并为其设定了双重道德约束:

第一重约束是“优先性约束”:效率必须从属于公平,不能凌驾于基本自由与机会平等之上。罗尔斯明确指出,帕累托最优等效率标准只是经济学标准,而非正义标准。一个实现了帕累托最优的社会,可能是极度不平等的(如1%的人占有99%的财富),这种效率是“不正义的效率”,不具备道德正当性。只有当效率的提升符合两大正义原则,不损害公民的基本自由与机会平等时,这种效率才是可接受的。

第二重约束是“目的约束”:效率的终极目的是实现“所有人的善”,尤其是最少受惠者的利益。罗尔斯认为,效率提升的本质是增加社会整体的“基本善”总量,从而让所有公民都能获得更多的自由、机会与福利。“差别原则”的核心就是确保效率的红利能够向下流动,而非向上集中。这种目的约束使得效率不再是单纯的经济指标,而是服务于社会整体正义的工具,其道德价值源于对公平目标的促进。

这双重约束共同构成了效率的道德基础,使得效率从属于公平的终极价值,同时也为效率的合理追求提供了正当性依据。在罗尔斯的理论框架中,效率与公平并非对立关系,而是一种“目的与手段”的关系。效率是实现更高水平公平的手段,公平是效率的终极目的与约束边界。

罗尔斯通过原初状态与无知之幕的思想实验,构建了“公平的正义”理论体系,其核心理论贡献在于突破了功利主义与绝对平等主义的二元对立,重构了公平与效率的关系框架。罗尔斯明确了公平的独立道德地位,将公平视为社会制度的首要德性,同时为效率划定了明确的道德边界,实现了公平与效率

的协调统一。与功利主义相比，罗尔斯的理论坚守了个体权利的不可侵犯性，避免了为效率牺牲公平的误区；与诺齐克主义相比，罗尔斯的理论兼顾了结果公平的兜底价值，保障了弱势群体的基本利益；与保守主义的效率优先论相比，罗尔斯的理论强调了效率的终极目的是服务于公平，确保了社会发展的正义性。这一理论体系为现代政治哲学提供了新的研究范式，推动了正义理论的发展与深化。

总之，罗尔斯的原初状态与无知之幕，不仅为公平与效率构建了坚实的道德基础，也为现代社会的正义追求提供了清晰的方向。在当代社会，公平与效率的冲突依然存在，罗尔斯的理论为我们解决这些冲突提供了重要的思想资源，推动我们不断构建更加公平、正义、高效的社会制度。

参考文献

- [1] 杰里米·边沁. 道德与立法原理导论[M]. 时殷弘, 译. 北京: 商务印书馆, 2009: 60.
- [2] 姚大志. 罗尔斯与功利主义[J]. 社会科学战线, 2008(7): 13-19.
- [3] 罗尔斯. 正义论[M]. 何怀宏, 等, 译. 北京: 中国社会科学出版社, 2009: 22.
- [4] 王立. 罗尔斯对功利主义的三重批判[J]. 社会科学辑刊, 2024(1): 5-13+237+2.
- [5] 姚大志, 罗翔. 罗尔斯是如何证明两个正义原则的? [J]. 浙江社会科学, 2024(6): 104-112+159-160.
- [6] 李佃来. 现代政治哲学的两个正义概念与马克思的正义观[J]. 武汉大学学报(哲学社会科学版), 2023, 76(6): 112-121.
- [7] 康德. 康德著作全集(第四卷) [M]. 李秋零, 译注. 北京: 中国人民大学出版社, 2013: 441.
- [8] 汪志坚. 康德“人是目的”命题的分配正义意蕴[J]. 哲学动态, 2025(11): 122-134+174-175.
- [9] 陈建平. 公平与效率: 平等与自由的博弈[J]. 理论月刊, 2015(10): 99-103.