**南 开 大 学**

本 科 生 学 年 论 文

论罗尔斯对康德自律概念的发展

On Rawls' Development of Kant's Concept of Autonomy

摘 要

自律概念是理解罗尔斯政治哲学理论形而上学基础的关键，而这种自律概念继承自康德。罗尔斯对康德的建构性解读为自身理论开拓了发展空间，起初在构建原初状态时运用并发展了康德的自律概念，继而在完善政治自由理论时区分了合理自律与充分自律，最终使自己的自律概念逐渐独立于康德，并对康德的理论困惑做出了尝试性解答。

关键词：自律；原初状态；理性；合理；建构主义

**Abstract**

The concept of autonomy, inherited from Kant, is the key to understand the metaphysical foundation of Rawls's political philosophy. Rawls makes a constructive interpretation of Kant’s autonomy, which expands the development space for his own theory. Originally, Rawls applies and develops Kant’s autonomy when constructing the original position. Then, in the theory of political freedom, he distinguishes between rational autonomy and full autonomy. Finally, he makes his concept of autonomy gradually independent of Kant’s, and his autonomy makes a tentative solution to Kant's theoretical confusion.

**Keywords:** autonomy, original position, reasonable, rational, constructivism

目 录

[引言 5](#_Toc6445026)

[一、罗尔斯自律概念的源起 5](#_Toc6445027)

[（一）康德关于自律的表述 5](#_Toc6445028)

[（二）罗尔斯对于康德自律概念的理解 8](#_Toc6445029)

[二、罗尔斯对康德自律概念的继承和发展 9](#_Toc6445030)

[（一）“作为公平的正义”的康德式解释的基础 9](#_Toc6445031)

[（二）合理自律与充分自律的区分 14](#_Toc6445032)

[三、罗尔斯自律概念的超越与争议 18](#_Toc6445033)

[（一）罗尔斯自律概念的超越性 18](#_Toc6445034)

[（二）罗尔斯自律概念的理论争议 20](#_Toc6445035)

[结语 21](#_Toc6445036)

[参考文献 23](#_Toc6445037)

# 引言

在《正义论》中，罗尔斯单独开辟了一节对其“作为公平的正义”展开康德式阐释。纵观全书虽然这一节所占的篇幅不长，但却是罗尔斯对其理论的形而上学基础进行阐释的重要环节。康德式阐释为罗尔斯政治哲学理论提供了先验依据。值得注意的是，罗尔斯在本节中明确指出，“作为公平的正义”的康德式阐释是基于康德的自律概念。

自律概念不仅在《正义论》中有所体现，在罗尔斯后续著作中也发挥着重要作用。在《道德理论中的康德式建构主义》、《康德道德哲学诸主题》以及《道德哲学史讲义》的康德讲座中，罗尔斯深入分析了康德式阐释的建构主义特征，加深了对“作为公平的正义”理论的形而上学基础的说明。在《政治自由主义》中罗尔斯进一步将自律概念分为合理自律与充分自律，并对原初状态进行了更深刻地解读，对社会观念与人观念进行分析，使得自身理论更加完善成熟。

罗尔斯的自律概念源起康德。但不同的是，他对康德自律概念做出了经验性的阐释。使其不仅成为自身理论的形而上学支撑，也对康德道德哲学的困惑从契约论角度做出了尝试性解答。

# 一、罗尔斯自律概念的源起

在《正义论》中，罗尔斯指出，自律概念是对“作为公平的正义”进行康德式解释的基础和关键。他在康德的道德形而上学中找到了可供自身理论发展的土壤，即康德道德形而上学具有的先验性和建构性。而这种建构性恰恰是通过康德的自律概念展现的，即意志自律可以从自身出发构建普遍化立法形式。罗尔斯认为这种能从自身建构对象与内容的独创性概念为“作为公平的正义”开辟了理论发展的空间，保证了建构程序的先验性。

## （一）康德关于自律的表述

“自律”在康德的道德哲学中占据很重要的地位，是极为独创性的概念，也是康德为其道德理论寻找到的根基。关于自律比较集中的讨论，体现在康德的两本比较重要的作品中，即《道德形而上学原理》与《实践理性批判》。

《道德形而上学原理》是康德为探求道德形而上学所作的准备。康德在本书中追求一种原则之原则，也就是为不基于经验基础的道德规律及其原则找寻依据。康德在文中总结了观察道德原则的三种方式，并将其在根本上归为同一个规律的三种公式。这种规律或者说原则就是康德寻求的道德律，而三种公式以定言命令的形式对其做出了规定。其一，对于这种在形式上具有普遍性，内容上表现为道德命令的准则，需要进行规定，即“所选择的准则，应该是具有普遍自然规律那样效力的准则。”[[1]](#footnote-1)一方面康德将道德规律原则的有效性与自然规律进行类比。自然规律在一定意义上可以看作因果律的展现。而将二者放在同等的位置，就突出了道德规律的纯粹性，以及不因经验世界的变化而转移。另一方面，康德也强调了这种原则的普遍性。由于这种原则是可以依从的，也就决定了它本身为众人所接受，也就必然是可以普遍化的。总的来说，这种道德原则应该是先验的，即人在道德意义上的善良不能被私欲和贪念影响，而应该表现为超越自身，且对自己和他人同样有效。这也体现了康德对于道德本质的理解。其二，这种原则的选择应该是理性的，而一种理性行为不仅要有原则，也要有目的。不同于主观目的的相对性，对于理性赋予的客观目的而言，“其本性是目的，且是自在目的。”[[2]](#footnote-2)这一规范的作用就是对纯粹相对和任意目的的约束。康德认为，这种自在目的体现了无条件的、绝对的价值。而满足这种绝对价值的只有理性动因或者人自身(person)。这是因为，理性存在者的意志是自由的，它能够自身成为原因，而不需要借助外物作为原因。也就是说，作为理性存在者，人不应该被视为达成目的的工具，这样所体现的价值是相对的。人应当作为自在目的，从而保证善的无条件性，也确定了行动的最高原则的存在。这体现了康德对于人性和目的的理解。其三，所有原则最终指向理性存在者的立法，“通过立法，和可能的目的王国相一致，如同对一个自然王国那样。”[[3]](#footnote-3)这是康德找寻到的对于全部准则的最完整的规定，即自我意志立法。康德认为道德规律之原则是可普遍化的、理性的，并且是和自然规律具有同样效力的原则。而我们必须服从这种原则，正是因为它是自在目的的体现，是人作为理性动因的本性的体现。康德要追寻的道德规律的原则，必定是人自身意志的产物。这体现了康德用意志自律架起了道德本质与人本性目的之桥梁。同时康德对自律和他律也做了区分。自律指的是从自身出发进行普遍性立法，而他律则需要借助外物，即“在某一对象的属性中去寻找规定它的规律。”[[4]](#footnote-4)康德认为意志自律对应着理性世界和本体自我，而若意志走出自身则滑向他律。他律对应的是处于经验世界的自我，即现象自我。经由三个定言命令的层层深入，康德将意志自律总结为最高道德原则的基础。

在《实践理性批判》中，康德也清晰地列出了四个定理，确定意志自律为道德律的依据。他开篇强调了道德律作为一种法则是普遍的、形式的，并通过前两个定理分别指出经验性的实践原则、自爱和自身幸福不能成为法则，经验、自爱和自身的幸福并不能满足普遍化的立法要求。继而在第三个定理中揭示出，想要达成这种普遍的实践法则，理性存在者在思考时，就不能将其视为以质料规定的，而应该视为“只是按照形式包含有意志的规定根据的准则。”[[5]](#footnote-5)一个理性存在者需要将诸主观实践原则思考为普遍化的法则，而只有单纯的形式能做到这一点。意志自律就是这样一种单纯的立法形式，也唯有它符合康德心目中立法的标准。也就是说，康德认为意志的自律性是具有自身决定能力的，能够将自身的实践原则构造为普遍化立法形式。康德在第四个定理中将意志自律与他律进行区别，这种区别体现在与责任义务的联系中。道德律有与之相符的义务，而意志自律则是联通二者的唯一原则。与之相反，他律不会产生任何责任，“而且反倒与责任的原则和意志的德性相对立。”[[6]](#footnote-6)通过康德对于他律的论述，我们可以窥见康德的自律是与责任的原则和意志的德性相关的。但如前所述，康德同样认为意志是自由的，即不为外物影响且是自身原则的根源，这实际上是一种倾向于积极理解的自由，是纯粹的实践的理性自身立法的途径。而当意志与责任相联系时，就表明意志要服从规律、法律，这对意志自由的描述似乎就带上了消极的色彩。康德的解决办法是将自律和自由等同，这样自由要服从的规律就不是外在强加的，而是自身产生的。自律是道德规律，意志自由也从属于道德规律。然而这种论述使得康德的自律概念陷入一种困惑。西季维克抓住了这一点批判康德的自由概念是含混的，他认为康德在双重意义上定义了自由，并且二者是有冲突的。一种是理性的自由，另一种则是在与责任相关的情况下，相对中立的自由，不再与理性相一致。若我们在服从道德法则的条件下做出错误的选择，那么就不是意志自由理性下的立法，则无需承担责任；若意志自律包容了这种错误的选择，即允许错误的选择，那么这就失去了理性的特质。例如恶棍所选择的生活也是他自己自由选择的结果，也体现了因果关系。康德的意志自律在本体自我上的阐述是有一定的问题与困惑的。

总的来说，康德认为道德规律应该是理性的、可普遍化的、形式的法则，而这种形式化的立法须得反映理性存在者的本性，故此，只有理性存在者的意志自律能够做到。康德进而区分了自律与他律，从侧面对自律进行进一步的界定。自律对应着本体自我，而他律对应着经验自我。自律与责任和德性相联系，而他律则与责任和德性相对立。但这种论述也使得康德的理论困惑体现出来，即意志自律的二分或者说悖论。

## （二）罗尔斯对于康德自律概念的理解

在《道德哲学史讲义》中，罗尔斯对于康德的《道德形而上学原理》的解读占了全书很多篇幅。他从康德定言命令的三个公式入手进行具体分析，对于康德在描述道德律时所用的论证方式十分认可。罗尔斯认为“作为纯粹实践理性的道德律从其自身决定（建构）其自身的对象。”[[7]](#footnote-7)换句话说，罗尔斯认为，康德的道德哲学理论本身就包含了建构的力量。他进而由康德的道德律提出道德建构论的观点，并将其与莱布尼茨的合理直觉主义进行对比，突出建构主义的优越性。

自律在康德的道德律中起着拱顶石的作用，是道德律建构性的集中展现，罗尔斯自然没有忽视这一概念。他明确指出，康德自律概念的独特之处在于具有建构性，并以此为出发点做了延展性的解读，大致可以归纳为两点。其一，自律的建构性为建构程序的先验性做出了理论铺设。罗尔斯认为，不是说我们的立法活动要努力去和已经存在的目的王国相协调，而是说，我们在立法时要像我们从理智和良知上遵从实践理性的原则那样，或者说如定言命令程序各式上所显现的那样，“为目的王国构造或建构公共道德律。”[[8]](#footnote-8)罗尔斯认为康德的意志自律可以从自身构建具有普遍性质的道德律，也就是说，不是我们调节自身与外物相适应，而是我们的意志从自身出发去建构和立法。在罗尔斯看来，这种建构性是先验的，这也为建构程序做了理论铺设。其二，理性而合理的自由平等的人观念是自律建构性的基础。“除非这个观念及其蕴含的道德人格力量（我们的人性）以某种方式在人的身上激发了生命活力，道德律在世界上会是无根基的。”[[9]](#footnote-9)在罗尔斯看来，道德律需要自律去触发，也就是说这种道德律要恰好反映人的本性，但这种人观念是有限定的，即理性而合理的自由平等的人观念。没有这种人观念作为载体，道德律将会是无根基的。自律的一个必要条件就体现为这种对人性本质的反映。而莱布尼茨的合理直觉主义则将道德秩序外在于人，所以是他律的。[[10]](#footnote-10)在罗尔斯看来，康德道德律的建构性更多地体现在意志自律上，而这种自律体现出的先验性和对人观念的紧密联系，都证明了康德道德哲学的建构主义解读的优越性。

罗尔斯抓住了康德的自律概念的建构性特征，并强调了这种建构性的自律建立在自由平等的人观念之上，其先验性为建构程序做出了理论铺设，他进而将对自律的建构性解读应用在自己的理论当中，这一点在《正义论》、《政治自由主义》等著作中都有体现。

# 二、罗尔斯对康德自律概念的继承和发展

罗尔斯对康德道德哲学的建构性解读为其创建正义学说拓展了理论空间。他将康德的道德形而上学拉入契约的范畴，为其政治哲学铺设道路。对于康德视为道德律之依据的自律概念，罗尔斯也进行了充分的吸收并运用在自己的理论建构中。总的来说，罗尔斯对康德的自律概念既有继承又有发展。在较早的《正义论》中，罗尔斯将自律看作对“作为公平的正义”进行康德式解释的基础，而在之后的《道德理论中的康德式建构主义》和《政治自由主义》中，又进一步区分了理性自律与充分自律。

## （一）“作为公平的正义”的康德式解释的基础

罗尔斯在《正义论》中明确指出，“作为公平的正义”的康德式解释的基础是康德的自律概念。他认为康德道德哲学的优越性并不在于明确了道德律的普遍化和形式化，其真正的独创性意义在于其中体现出的理性和自律。罗尔斯认为康德对道德律进行界定时所用的理性的条件是十分关键的。这种界定更清晰的表明了，康德的道德哲学面向的是“适当确定的合理抉择及结果。”[[11]](#footnote-11)也就是说，康德的道德律是理性选择的结果，也是自由平等的理性的人合理的选择。这种道德律使得人们在道德世界中的行动是可控的，而当行动的原则恰好反应人的本性时，这种行动就体现出了自律性。在罗尔斯看来，康德的道德哲学的独特之处在于将自律作为道德律的落脚点，而自律概念的独特之处则在于其具有建构性。罗尔斯抓住了这种建构性的特征并将其运用到自己的正义理论当中。更具体地说，罗尔斯在设计原初状态时，也运用了康德道德律中的自律概念。不同的是康德的自律概念侧重的是本体自我（道德哲学领域），而罗尔斯的自律概念则指向集体协议（政治哲学领域）。不过二者虽然处于不同的领域却有共通之处，这种共通在原初状态的主体、条件和特征中都有体现，并且罗尔斯认为，原初状态的种种设计也弥补了康德自律概念的一些缺陷。

1. 主体——自由平等的理性存在者

罗尔斯认为自己对于原初状态的设计既继承了康德的自律概念，又对西季威克提出的自由概念有歧义的问题做出了回应。在罗尔斯看来，康德理论困惑的根源在于，没有阐明意志自由选择反映的本性是理性的，并且这种反映是惟一的。虽然罗尔斯也会承认，康德对一个理性存在者的自律行动是有关于本性的界定的。当人作为自由、平等的理性存在者行动时，表明他是在自律地行动的条件在于，“他的行为原则是作为对他的这一本性的可能是最准确的表现而被他选择。”[[12]](#footnote-12)然而这种界定也存在着问题，道德律既是自由平等的理性存在者的本性的反映，也是意志自由选择的结果且与责任有密切联系。对此，西季维克批判康德在双重意义上定义了自由，比较有反击性的例子就是恶棍理论，即恶棍的选择也是反映本性的自由选择。针对这种批判，罗尔斯认为康德可能会以本体自我和现象自我的区分作为回应。虽然本体自我的选择是理性的，所选出的行动原则也是被一致同意的。但除了理性世界的本体自我，处在经验世界中的现象自我的决定并不能体现自由、平等和理性。也就是说，康德对于这一问题的解答倾向于二分，即在感觉上，本体自我与现象自我是不同的。作为自由平等的理性存在者时，体现的是本体自我。而罗尔斯认为这并不是自律悖论的根源，他认为康德的真正缺陷之处在于，虽然说明了遵从道德律的行为就是表现我们的本质的行为，而且这种道德律是被我们的本性所一致同意的，但却没有说明在自由平等理性的前提下，为什么“按照相反原则的行为则没有表现我们的本质。”[[13]](#footnote-13)也就是说，自由平等的理性人会选择哪一种原则是没有被清晰地限定的。

罗尔斯认为原初状态对于本体自我的设计解答了这了问题。原初状态在程序上避免了这种困局。我们可以把原初状态类比为本体自我理解世界时的状态，对于充当本体自我的各派，无论选择哪一种原则都是自由选择的结果，但虽然是自由选择的状态，各派实际上都有相似的愿望，一方面这种选择要体现出各派在理智王国中是平等的、理性的，另一方面这种选择也要适用于真正的社会生活，也即是说，这种自由的选择可以作为“在他们的社会生活中以原初状态的观点来看待世界、并能够表达这种观点的存在物的本质。”[[14]](#footnote-14)在原初状态的限制下，主体在选择时必然会将自然和社会的种种偶然因素排除在外，而去追求能更好地实现共同体生活的自由的原则。也就是说，在原初状态下，主体本身就站在康德意义上的本体自我的立场。这样作为本体自我展现本性的自由就不会再产生歧义的问题了。罗尔斯在建构原初状态时就在程序上避免了本体自我与现象自我的区分，除此之外，罗尔斯还明确指出了自己的本体自我是与康德不同的，这种本体自我是集体意义上的概念。在原初状态中的各派不是单独个体，而是一种代表设置，是一方利益的代表。可见，罗尔斯在设计原初状态时借鉴了康德的本体自我，但又将这一单个意义上的概念发展到集体意义上，为自己的理论服务。

1. 条件——无知之幕

“无知之幕”是罗尔斯针对原初状态中的主体设的一个先决条件。这个条件对主体进行了限制，排除了选择时的偶然因素。罗尔斯用无知之幕遮蔽了主体自身的天赋气质、所处的阶级地位和所生活的社会的历史经济政治水平，并且对于善观念的具体内容也是空白的。而无知之幕留下的是处于正义环境下的一般事实，所有可能造成争端的特殊事实都被遮蔽了。罗尔斯认为这些特殊的事实会影响主体的选择，如果按照这些事实行事实际上是他律的。在原初状态中，无知之幕排除了“使人能选择他律原则的知识。”[[15]](#footnote-15)各方都处在一种单纯的正义环境中，在做选择时，只有关于正义环境的知识，并且是作为自由平等和理性的存在者去做选择。罗尔斯的无知之幕撇开特殊性，只留下了最一般与普遍的特性。在这一点上，罗尔斯是意图接近康德的。康德在定义道德法则的公式中澄清了选择原则需要具有普遍性。而这种普遍化、一般化的背景也的确为罗尔斯的正义原则的选择提供了便利。

“无知之幕”的设计吸收了康德的观点，但也有自身的独特之处。无知之幕在遮蔽的过程中针对的便是特殊事实，其中自然也包括作为个体自我的偏好。在这种情况下，在无知之幕背后进行选择的就不再是个体，而是具有代表性的一类人。从这一点上看，罗尔斯对于康德的观点是有所补充的。在《实践理性批判》中，康德通过定理将经验原则、自爱和自身幸福排除，并揭示出理性存在者要将诸主观实践原则思考为普遍化的法则，而只有单纯的形式能做到这一点。但对于个体来说，自身的思考要不局限于自身、兼顾他人，从而达成普遍化的法则是很困难的。康德对此也没有十分完美的解答。而罗尔斯的无知之幕作为先决条件，在程序上就已经将这一困难排除在外了。经由无知之幕遮蔽个体的特殊性，主体面对选择时就已经具有可以代表一类人的群体性了。也就是说，无知之幕将能影响主体选择的他律因素遮蔽了，留下了正义的环境下的自律的选择。

1. 特征——相互冷淡的动机假设

罗尔斯不仅为原初状态设计了正义的环境，而且对于各派的动机也做了限定，即相互冷淡的动机假设。这种假设主要是为了避免争议出现时，各派利益纠缠而无法做出理性的选择。所以，罗尔斯假设在原初状态中的各派是相互冷淡的，也就是说各派不愿为他人牺牲自身利益。在这种限定下，各方在进行审慎的思考时不是利他的。这种相互冷淡的特征表明一种不关心他者的利益的倾向。或许有人会质疑，这种相互冷淡的动机假设是否会影响各派的选择自由。罗尔斯从利他主义者入手进行分析，侧面回应了这种质疑。他在《正义论》中指出，当原初状态中各方被假设为利他主义者，即关心他人利益，或者有一些特殊的追求时，这些人的选择的自由会“被限制在与利他主义或享乐主义相容的各种选择中。”[[16]](#footnote-16)也就是说，这种相互冷淡、不关心他人利益的状态恰恰保证了选择的自由，使自身成为目的，而不受其他特殊方式的限制。利他实际上才会造成利益的纠葛，才会成为对自由的一种强制的限制。从这一点上看，罗尔斯相互冷淡的动机假设实际上是与选择的自由密切相关的。他认为在相互冷淡的动机假设下，各方不关心他者的利益，那么就不会牺牲自己的利益去成全他人，同样也不会侵害他者的利益做出自利的行为。这种状态很好的避免了各方利益的牵连和纠葛。主体在选择时是自由的，并不受到各方利益倾轧的影响。

表面上看，这种相互冷淡的动机假设似乎与康德的自律概念是无关的，但罗尔斯认为康德的自律概念实际上是隐含这种相互冷淡的动机假设的，或者说康德对于自律的界定预设了这种不关心他人利益的状态。康德认为，能够完成单纯形式立法的只有意志自律，而这种自律需要恰好反映理性存在者的本性。换句话说，做出选择时，理性存在者的意志是自由的，这体现为它能够自身成为原因，而不需要借助外物作为原因。而这种意志自由不受外在因素限制的特点，正体现着自律性。可见，罗尔斯为原初状态设计的这种各派互不牵扯的状态，是符合康德自律概念的预设的，都排除了外在因素的限制。只不过康德的理论本身含有这种预设。而罗尔斯是以相互冷淡的动机假设进行限定，他指出，在原初状态中，各派只了解基本善，不关心他人利益，这种限定是契约论意义上的限定，“这种契约状态没有对人们可能愿望的东西施加任何先验的限制。”[[17]](#footnote-17)这种契约论意义上的限定，使得原初状态中的正义环境能够更好地维持下去，各派的选择也更加不容易受到干扰。康德认为，意志自由对应着能够充当法则的单纯的立法形式，是不受先验限制的，体现着自律性。而罗尔斯为原初状态设定的相互冷淡的特征是契约意义上的，同样不是先验性的限制。

总的来说，罗尔斯对原初状态的主体、条件和特征都做了设定，而在这些设定中都可以窥见他对康德自律概念的运用与发展。对于主体，罗尔斯很大程度上认同康德对于自由平等的理性存在者的规定，但不同于康德从本体自我出发选择道德规律，罗尔斯在集体意义上重新定义了本体自我，原初状态中的各派是代表设置，对于正义原则的选择属于契约范畴。对于条件，罗尔斯将无知之幕作为主体的先决条件，在程序上排除了偶然因素，使得主体能够从自身出发，而不受他律因素的影响，营造了单纯的正义环境。对于特征，罗尔斯设定相互冷淡的动机假设，保证了各派不关心他人利益，避免了产生利益纠缠，使得各派可以按照意志自由选择。罗尔斯对康德自律概念的继承体现在，他合理的利用了自律所体现出的理性和自身决定的建构性，康德是从本体自我出发选择道德律，而罗尔斯为了选出首要正义原则，设计了原初状态作为过渡，他将原初状态类比为康德意义上的本体自我的立场，以契约论的角度进行限制，避免了康德的本体自我与现象自我的二分。

## （二）合理自律与充分自律的区分

随着罗尔斯不断完善自己的正义理论，他对康德的自律概念也有了进一步的理解和发展。他逐渐确立了自己的自律概念，并明确地与康德的自律概念区别开来，这在《政治自由主义》、《道德理论中的康德式建构主义》等作品中都有体现。在《道德理论中的康德式建构主义》中，罗尔斯阐释了对于复杂的人观念的理解，并将自己的自律概念分为合理自律与充分自律。他认为原初状态中，各派体现出的自律是合理自律，相对来说，这种自律是狭义的，类似于康德的假言命令。而充分自律则与首要正义原则息息相关，这种首要正义原则体现为，社会生活中的公民认可和践行并一致通过的原则，而公民对于这种原则的选择就体现了充分自律。在《政治自由主义》中，罗尔斯也指出，合理自律指的是原初状态下的自律，既是设计的又是非政治性的；充分自律指的则是在罗尔斯两种正义原则下的良序社会中的自律，具有政治力量。

1. 复杂的人观念

在《道德理论中的康德式建构主义》中，罗尔斯提出，康德式建构主义的独特之处在于他对人观念的界定。在罗尔斯看来，康德式建构主义的基本思想与这种人观念息息相关。具体体现为，他认为康德道德律的实质在于，“通过建构程序，在一种独特的人观念与首要正义原则之间建立起一种恰当的联系。”[[18]](#footnote-18)

在康德的自律概念中，恰好表现了理性存在者的本性的普遍化理性立法形式才是道德律，并且这种本性的展现是本体自我自由选择的结果。可见，康德的人观念是独立于经验世界的，具有先验性。而罗尔斯论述的康德式建构主义以及他所建设的作为公平的正义中，人观念则是复杂的，是与社会生活紧密相连的。他指出，以康德论述道德律的方式看待正义学说时，是离不开特殊的人观念的。这种人观念不仅要满足在特性上是自由平等的，在能力上是能够理性且合理地行动，还要依据这种特质与能力去参与“如此构想的人们之间的社会合作中去。”[[19]](#footnote-19)可见，罗尔斯的人观念并不是完全先验的，而是潜在地与社会公共文化相关，并且要经得起反思检验，具有真实内容。这种人观念与人的本性理论是相区别的。人观念是一种道德理性，与组织有序社会的观念相伴相生，如此便要求这种道德理想是能够遵循的、可行的。而人的本性理论出现在原初状态中，具体体现在可为各派知晓的一般性事实中，这种本性理论不同于人观念体现出的理性，只是在原初状态中体现的一种合理性，是一种有待添加解释的要素，并没有明确其在真正社会生活中的可行性。

罗尔斯在构建“作为公平的正义”的理论时，作为基本模型的观念有两个，即良序社会观念(a well-ordered society)与道德人观念。良序社会观念是为了探讨作为自由平等的公民的社会关系概念的本质，而道德人观念的设定则是为了探究我们作为道德人自我概念的本质。罗尔斯的正义原则体现的是他对良序社会中的公民关系的理想描绘，他认为只有正义原则能够为公民们实现良序社会的理想状态。为了论证正义原则的有效性，以及选择这种正义原则的自明性与必然性。他提出了原初状态，作为这种道德人观念与正义原则之间的过渡性设计。原初状态是一种理想化的状态，通过种种设计和限制，公民以道德人的立场在良序社会中为他们的社会挑选首要正义原则。而在秩序良好的社会中，作为道德人的公民在人格上有两方面的特质，那就是合理的(rational)与理性的(reasonable)。而原初状态的不同方面的设计分别体现着合理与理性的特质，这也为接下来对自律概念的区分做了铺设。

1. 纯粹程序正义——合理自律

罗尔斯在阐述合理自律与充分自律的区别之前，在建构主义的框架下再次解读了原初状态理论。罗尔斯认为原初状态的无知之幕排除了社会运气和自然偶性的信息，从而保证各派在审慎思考时对社会基本正义原则的选择。这样的原初状态所设定的思考方式在一个最高层次上体现了纯粹程序正义，其结果定义着恰当的正义原则，这使得我们可以很好地解释原初状态中的各派是自律的。

罗尔斯认为原初状态理论体现了一种纯粹的程序正义，而恰恰是这种纯粹的程序正义的特质塑造了合理自律，也就是说，合理自律体现在原初状态中作为代表的各派的慎思方式里。具体表现为“个人的理智能力和道德能力。”[[20]](#footnote-20)理智能力指的是能够与其他人促成一致性的契约，既体现出个人与他人都能运用理性去选择，又体现为个人和他人都服从理性约束。道德能力体现为对善观念的追求与运用，具体来说，个人对善观念应该是一种不断追求的状态，在形成善观念后，不断地进行修正，并能够按照善观念进行思考，不仅追求善观念，还要将其贯穿在自身行为与选择中。也就是说，这种合理自律一方面表现为各派在审慎思考时，虽然是有理智能力的，但并不要求应用任何先在的正义原则，另一方面表现为各派仅受最高利益的驱动，并没有一种独立于原初状态各派最高利益的外在视角。原初状态中的各派在基本善观念的背景下，运用理智进行选择，体现的是合理性。在这种情况下，各派只是原初状态设计下的个人，也就是仅仅作为原初状态的代表设置，并没有参与到社交世界或者说社会合作中。各派在原初状态中审慎地思考选择时表现出的自律，只体现了在秩序良好的社会中的公民的道德人格的一个方面，也就是合理性。而另一方面的理性则是通过原初状态中作为限制的理性条件来表现的，对于身处其中的各派来说，这些理性条件显然是外在的。故而，罗尔斯将原初状态中各派审慎思考表现的自律定义为合理自律。

1. 良序社会——充分自律

相比于合理自律体现的是原初状态中的各派审慎思考下的合理选择，充分自律则对应着理性因素。罗尔斯对充分自律的讨论不再局限于原初状态，而是延展到社会生活中的公民。合理自律体现在各派的选择，而充分自律则更多地在公民身上体现，倾向于社会生活。但罗尔斯在论述充分自律的塑造时，依然是从原初状态开始分析的。他指出，这种充分自律是经由“原初状态的结构性方面来塑造的”[[21]](#footnote-21)也就是说，不同于合理自律指向的是各派在原初状态下的思考，充分自律直接指向了各派之间关系的形成，以及各派在审慎思考时受到的信息限制。如果说合理自律的塑造依赖于已有原初状态下的思考，那么充分自律则依赖于这些思考背后的理性条件，是结构层面的。在无知之幕的重重限制下，原初状态处于单纯的正义的环境，这些限制条件作为背景约束着各派，这是理性因素的体现。而罗尔斯认为充分自律就来自于这些理性因素。

在《政治自由主义》中，罗尔斯对于这种充分自律体现的理性的论述更多地偏向了社会生活领域，充分自律体现为公民遵循正义原则而行动，这种正义原则是原初状态中各派站在自由平等的人的立场做出的一致决定，是一种公平合作条款。这种公平合作条款意味着参与合作的人要共担风险，共享受益。体现着互惠互利这种相互性的观念。罗尔斯认为社会合作这一要素体现着理性，而当中每个参与者的利益对应着合理性。理性和合理性在原始状态下是统一的。“理性预设着合理性，而合理性从属于理性。”[[22]](#footnote-22)而在社会合作中同样如此，公民的行为要符合正义原则，进一步来说公民需要按照正义原则来行动，充分自律表明，公民不仅可以认识到原始状态下采用的原则，而且可以明智地使用它们。根据罗尔斯的观点，这种充分自律更多的体现的是政治价值，而不是道德价值，这也一定程度上为他的理论建设服务。充分自律可以适用于包括社会生活和个体生活在内的整个生活，是一种政治自律。在良序社会中，公民受到公共要求的约束，因此不会因政治正义而受到指责。在罗尔斯看来，每个人都拥有政治正义感，并且是同等效力的，这种政治正义感也是充分自律能够实现的必要因素。也就是说，公民拥有道德人格的两个方面，即合理与理性。在政治正义感的指导下，两者在公民理解和运用正义原则的过程中统一起来。

罗尔斯的合理自律与充分自律都偏离了康德的自律概念，康德的自律是归属于个体的一种理性的意志，排除了外在影响和先验限制，是在道德领域的一种前后一贯的完备性学说。而罗尔斯将这种自律运用到自己的学说中，就面临着困难，即如何将这种完备性的观念赋予经验性的解读，从而更好地适用于政治哲学。故而，罗尔斯逐渐将自己的自律概念与康德区分开，是必然并且可以预见的。他并没有放弃康德的自律观念，而是对其进行了发展。在《正义论》中，罗尔斯对康德自律概念的运用较多的体现在原初状态的设计中，一方面借鉴其建构性来论证自己的理论结构，另一方面从契约的范畴对本体自我进行限制，避免二分。在此处，罗尔斯对自律的运用就已经有别于康德，是一种集体意义上的复数概念。而合理自律与充分自律的区分，更体现了罗尔斯的自律概念已经渐趋独立，虽然源起康德，但更多是服务于正义理论。罗尔斯提出这种区分也是为了完善原初状态和良序社会的理念，他明确了人的两种道德能力，即合理与理性，并以此作为合理自律与充分自律区分的基础。不同于康德自律概念归属于定言命令，合理自律似乎接近于康德的假言命令，指的是各派在原初状态中的慎重选择。而充分自律则是公民能认识原初状态中达成一致的正义原则，并在社会生活中遵循并运用这种原则，从而完成良序社会的理想。这两种自律若是在康德看来，似乎都带上了他律的色彩。实际上，在罗尔斯理论逐渐发展的过程中，自律成了更倾向于政治意义的理论，而非道德完备理论。

# 三、罗尔斯自律概念的超越与争议

罗尔斯的自律概念源起于康德，随着自身理论的深入发展，他对自律概念的界定也有了新的变化，进一步区分了合理自律与理性自律。一方面，罗尔斯的区分从契约论的角度对康德的自律概念做了新的理解。另一方面，他也进一步解释了“作为公平的正义”理论的自我一致性。虽然罗尔斯为正义理论开创了新的思路，但他对康德自律概念的解读偏重建构性，并在此基础上发展了倾向于政治力量的合理自律与充分自律，针对于此，也出现了许多批评与争议。罗尔斯的自律概念既有超越性，也相对存在着理论争议。

## （一）罗尔斯自律概念的超越性

罗尔斯的自律概念的超越性在于他从康德到康德式的跨越，罗尔斯在吸收和继承康德的观念后，剥离出根本性的建构模式发展了自己的理论。康德的道德哲学是先验的，其中作为依据的意志自律对本体自我的选择具有规范作用。罗尔斯以经验性对其进行解释，并将自律概念引入政治领域。在《正义论》中，虽然他声称自己的理论此时是接近康德的，但在他将自律概念当成“作为公平的正义”的基础时，他所理解的本体自我与康德意义上的本体自我是不同的。罗尔斯的本体自我是一个群体性的概念，代表的并不是作为个体的“人”，而是作为代表的“人们”。换句话说，当罗尔斯将自律概念的建构性作为突破点建设自己的理论时，他的关注就从康德的人与他人更多的迁移到社会关系中了。这是一种从“我”到“我们”的转变。在政治自由主义等时期，罗尔斯的理论更加成熟，他对自律概念在正义理论的应用更加细化了，合理自律与充分自律都和康德意义上的自律相背离了，合理自律成为原初状态审慎思考的体现，而罗尔斯更重视的充分自律则更贴向政治价值。罗尔斯的自律概念与公民的社会生活息息相关，已经不再是个体与他者关系的体现，更多的是作为公民面对政治生活时的自主性。为重叠共识和多元民主社会的理念作铺设。

罗尔斯的自律概念的另一个超越之处体现在，对道德秩序做出了发展性阐释。罗尔斯认为，康德的道德哲学是区别于合理直觉主义的，这一点体现在道德秩序上。合理直觉主义认为存在一种先在的道德的价值秩序。而以罗尔斯的理解，对于康德而言，这种合理直觉主义是他律的。合理直觉主义认为，从概念的角度看，基本道德观念与自然观念是相互独立的，并且道德价值秩序不仅是先天综合的，而且也可以把握。康德的自律概念则要求，不能有先于和独立于这些观念的道德秩序，正是这些观念规定着界定自由和平等的人之间正当和正义原则的程序形式。罗尔斯将康德的自律视为自我建构的，或者说构成性的。道德价值秩序和政治价值的秩序想要被构成或者说建立，“必须通过实践理性的原则。”[[23]](#footnote-23)在此基础上，罗尔斯认为，对于政治自由主义，判断一种政治观点的是否具有自律性，关键在于其能否表现出秩序化的政治价值。这种政治价值是基于实践理性原则的，并且这种原则与合适的、政治的社会观念和个人观念相联系。在这一点上，罗尔斯认为应该撇开康德的构成性自律。作为理性的理念，社会观念和个人观念不是被建构出来的，而是相似的和相互联系的。可以试图通过反思的方式，来使这些想法在实际思考中显现出来，并建立一个相互关联的秩序。也就是说，罗尔斯认为价值秩序的建构的关键在于人，只有人参与其中才能促成价值秩序。这也体现了罗尔斯的自律向人观念与社会观念的迁移。在这种倾向下，自律更好地为罗尔斯的正义理论服务，各派从原初状态自然过渡到社会生活，成为其中的公民，不仅可以认识正义原则，而且可以遵循并运用这种原则，以期实现良序社会的理想。

总的来说，罗尔斯的自律概念的超越性体现为两个方面，其一是对康德自律概念的建构性解读，并将其拉入政治领域，从契约论角度对其进行了补充；其二是在此基础上，发展了自律概念，使其倾向于人观念与社会观念，更具政治力量，在服务于自身正义理论时逐渐转向自主性。罗尔斯的自律之独特，在于其继承了康德的论证方式，但却将其运用于政治领域，为正义理论开辟了新的视角。对于自律，完成了从康德的到康德式，最后逐渐独立的发展过程。

## （二）罗尔斯自律概念的理论争议

罗尔斯对于康德道德学说的经验式解读的确给予人们新的理解方向，但这种康德式的理论也同样充满争议。一种观点认为，罗尔斯对于康德的建构主义解读实际上是对康德的一种误解，更多的应该是罗尔斯自己的理论发挥。另一种观点认为，罗尔斯的建构式解读并没有解决康德哲学的困惑，罗尔斯的理论体现着一种妥协式的无奈。

巴德在他的论文中批判了罗尔斯的康德式观点，他认为所谓的康德式解释是罗尔斯将自己的观点强加在康德的观点上。[[24]](#footnote-24)巴德认为罗尔斯对康德的解释清楚地表明，他确信康德的理论需要他自己理论的元素才能奏效。这是由于罗尔斯不能或不愿相信康德理论的某些要素(先天的和形式的)确实有效。然而，在忽略了这些因素之后，他又为这些因素在康德理论中所扮演的角色找到了新的因素。在这里，罗尔斯介绍了与他自己的理论方法和工具相似的特征，也就是建构性，他认为这些特征对于处理哲学问题是必要的，比如正义原则的确定。巴德认为罗尔斯实际上改变了康德的观点，具体例如无知之幕，理性选择等理念。而这些变化改变了康德理论的整个性质。它使经验的和假设的推理成为特定范畴规定的基础，并使道德原则的权威依赖于(经验的)欲望或者说以欲望为条件。巴德的质疑点在于我们能否说这样的变化之后的理论仍然是康德式的。因此，巴德认为罗尔斯应该宣称，他的理论不是康德式的，而是经过适当调整的康德才是罗尔斯式的。

面对类似巴德的这种比较普遍的质疑，奥诺拉·奥尼尔给出了另一种思路，她认为康德的道德正当性概念比罗尔斯提出的更具根本性的建设性。[[25]](#footnote-25)但她并没有十分明确地阐明这种根本性，而是指出康德和罗尔斯最终会遇到类似的困难，他们都提出了一个建设性的程序来确定具体的实践原则，但都不能令人信服地表明这个程序是建立在或表达(一个充分的概念)合理性的基础上的。他们的立场可能是建设性的，因为他们所提出的程序有限，作为代表设置的各派可以用这些程序来确立指导行动的原则，但不能从更充分的意义上为这些原则辩护。虽然面对质疑，但罗尔给出的康德式解释以及道德建构主义也是一次非常有意义的尝试。

除了巴德的这种质疑，还有对于罗尔斯的理论是否真的解决了康德哲学的困惑的质疑，例如奥利弗·约翰逊。他认为原初状态中的自律选择不是因为特定事项或突发事件,同样也不是因为他们想要某种东西，真正的原因在于由于无知之幕，他们被排除考虑他律的因素。[[26]](#footnote-26)虽然无知之幕一定程度上避免了他律因素的干扰，但各派的动机问题仍然存在困扰。约翰逊特别指出我们必须研究这些人的动机，他们的动机不会因为无知之幕而改变。只有他们作出决定的条件发生了变化。他们仍然选择他们所做的原则，动机是希望以最好的方式促进他们自己的利益。换句话说，每个人的决定仍然有它的基础，仍然需要欲望，也就是说这种决定实际上是各派利益的妥协。而在康德看来，自律与他律的区别不在于行为发生的环境，而在于行为发生的动机。一个原本是他律的行为，如果其动机的本质不变，即使是在无知之幕下进行，也不会被赋予自律性。在他看来无知之幕下的自律并没有解决康德的悖论。康德的确在一定意义上将自律与纯粹的理性存在者相连，这点上来看约翰逊的质疑是有道理的。这样看来，在《正义论》中，罗尔斯所阐述的自律似乎只能存在于种种限制下，是一种较弱的理论，也就是说，因为我们有这种自律的能力，使得我们能够参与立法。虽然不是康德意义上的绝对的纯粹的立法者，但却以道德能力的形式体现在我们身上。而随着自身理论的发展，罗尔斯则致力于将自己的自律概念与康德独立开来，更倾向于政治领域的阐释。

# 结语

罗尔斯的自律概念虽然源起康德，但却十分不同，是对康德自律概念的发展。他对康德的自律概念的理解更多的偏向对其建构性的特征的把握，并以这种建构性为基础开拓自身的理论空间。罗尔斯将自律概念视为“作为公平的正义”的康德解释的基础，在原初状态的设计中贯穿了对于自律概念的理解与发展。而随着理论的不断深入，罗尔斯区分了合理自律与充分自律，更重视自律的政治力量以及自身理论的一脉相承。

罗尔斯对康德自律概念的这种发展既有争议，也有独特意义。一方面，对于罗尔斯自身的理论来说，这种建构性的把握为之后的重叠共识等概念做了铺设。另一方面，这种经验性阐释规避了康德的道德哲学在本体自我中的困难，将视野转向社会合作或者说共同体，这种全新的尝试虽然具有争议但也开拓了路径，适用于更广泛的领域。总的来说，罗尔斯将道德律中的建构因素应用于社会正义，用集体协议来解决自律在个体自我上出现的困惑，也从契约论的角度为康德的自律概念产生的困局提供了不同的解决尝试。

# 参考文献

[1] [德]伊曼努尔·康德著，苗力田译，《道德形而上学原理》，上海人民出版社，2005年。

[2] [德]康德著，邓晓芒译，杨祖陶校，《实践理性批判》，北京：人民出版社，2003年。

[3] [美]约翰·罗尔斯著，顾肃、刘雪梅译，《道德哲学史讲义》，中国社会科学出版社，2012年。

[4] [美]约翰·罗尔斯著，何怀宏、何包钢、廖申白译，《正义论》，北京：中国社会科学出版社，1988年。

[5] [美]约翰·罗尔斯著，陈肖生等译，《罗尔斯论文全集》，吉林出版集团有限责任公司，2013年。

[6] [美]约翰·罗尔斯著，万俊人译，《政治自由主义》（增订版），南京：译林出版社，2011年。

[7] 丛占修.康德道德哲学对罗尔斯正义理论的影响[J].道德与文明,2011(02):54-57.

[8] 乔洪武,曾召国.罗尔斯对康德伦理学的“建构主义诠释”——兼论《正义论》证成结构的思想“原型”[J].哲学研究,2011(12):98-104+125.

[9] 王佳鑫. 对“作为公平正义的康德式解释”的再解释[D].吉林大学,2012.

[10] 孙小玲.自律的悖论与罗尔斯的契约论解决[J].哲学研究,2016(01):89-96.

[11] 李爱龙.从自律到自主:罗尔斯对康德的经验性阐释[J].广西社会科学,2018(07):81-85.

[12] Budde, K. *Rawls on Kant: Is Rawls a Kantian or Kant a Rawlsian?* *European Journal of Political Theory*, *6*(3), 2007， 339–358.

[13] Oliver A. Johnson, *The Kantian Interpretation,* Ethics 85, no. 1 ，Oct., 1974， 58-66.

[14] Oliver A. Johnson, *Autonomy in Kant and Rawls: A Reply,* Ethics 87, no. 3 ，Apr., 1977，251-254.

[15] Onora O’Neill, *Constructivism in Kant and Rawls*, *The Companion to John Rawls*, Samuel Freeman, Cambridge University Press,2003,347-367

[16] Stephen L. Darwall, *A Defense of the Kantian Interpretation,* Ethics 86, no. 2 ，Jan., 1976，164-170.

[17] Thomas E. Hill, *Kantian Constructivism in Ethics,* Ethics 99, no. 4 ，Jul., 1989， 752-770.

1. [德]伊曼努尔·康德著，苗力田译，《道德形而上学原理》，上海人民出版社，2005年，57页 [↑](#footnote-ref-1)
2. [德]伊曼努尔·康德著，苗力田译，《道德形而上学原理》，上海人民出版社，2005年，57页 [↑](#footnote-ref-2)
3. [德]伊曼努尔·康德著，苗力田译，《道德形而上学原理》，上海人民出版社，2005年，57页 [↑](#footnote-ref-3)
4. [德]伊曼努尔·康德著，苗力田译，《道德形而上学原理》，上海人民出版社，2005年，63页 [↑](#footnote-ref-4)
5. [德]康德著，邓晓芒译，杨祖陶校，《实践理性批判》，北京：人民出版社，2003年，33页 [↑](#footnote-ref-5)
6. [德]康德著，邓晓芒译，杨祖陶校，《实践理性批判》，北京：人民出版社，2003年，43页 [↑](#footnote-ref-6)
7. [美]约翰·罗尔斯著，顾肃、刘雪梅译，《道德哲学史讲义》，中国社会科学出版社，2012年，207页 [↑](#footnote-ref-7)
8. [美]约翰·罗尔斯著，顾肃、刘雪梅译，《道德哲学史讲义》，中国社会科学出版社，2012年，179页 [↑](#footnote-ref-8)
9. [美]约翰·罗尔斯著，顾肃、刘雪梅译，《道德哲学史讲义》，中国社会科学出版社，2012年，212页 [↑](#footnote-ref-9)
10. 这主要是罗尔斯的观点，康德并没有明确表明合理直觉主义是他律的。 [↑](#footnote-ref-10)
11. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，250页 [↑](#footnote-ref-11)
12. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，251页 [↑](#footnote-ref-12)
13. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，254页 [↑](#footnote-ref-13)
14. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，255页 [↑](#footnote-ref-14)
15. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，251页 [↑](#footnote-ref-15)
16. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，253页 [↑](#footnote-ref-16)
17. [美]约翰·罗尔斯著，何怀宏、何包钢等译，《正义论》，北京：中国社会科学出版社，1988年，253页 [↑](#footnote-ref-17)
18. [美]约翰·罗尔斯著，陈肖生等译，《罗尔斯论文全集》，吉林出版集团有限责任公司，2013年，348页 [↑](#footnote-ref-18)
19. [美]约翰·罗尔斯著，陈肖生等译，《罗尔斯论文全集》，吉林出版集团有限责任公司，2013年，345页 [↑](#footnote-ref-19)
20. [美]约翰·罗尔斯著，万俊人译，《政治自由主义》（增订版），南京：译林出版社，2011年，67页 [↑](#footnote-ref-20)
21. [美]约翰·罗尔斯著，万俊人译，《政治自由主义》（增订版），南京：译林出版社，2011年，71页 [↑](#footnote-ref-21)
22. [美]约翰·罗尔斯著，陈肖生等译，《罗尔斯论文全集》，吉林出版集团有限责任公司，2013年，358页 [↑](#footnote-ref-22)
23. [美]约翰·罗尔斯著，万俊人译，《政治自由主义》（增订版），南京：译林出版社，2011年，91页 [↑](#footnote-ref-23)
24. Budde, K. *Rawls on Kant: Is Rawls a Kantian or Kant a Rawlsian?* *European Journal of Political Theory*, *6*(3), 2007， 339–358. [↑](#footnote-ref-24)
25. Onora O’Neill, *Constructivism in Kant and Rawls*, *The Companion to John Rawls*, Samuel Freeman, Cambridge University Press,2003,347-367 [↑](#footnote-ref-25)
26. Oliver A. Johnson, *The Kantian Interpretation,* Ethics 85, no. 1 ，Oct., 1974， 58-66. [↑](#footnote-ref-26)