Constructivism, Strict Compliance, and Realistic Utopianism

By Ben Laurence
University of Chicago

§1: Introduction

The primary concern of political philosophy is justice. Since philosophy is a reflective enterprise, this concern naturally takes the form of constructing a theory of justice. Famously, John Rawls divides this theory into two parts that he calls ideal and nonideal theory. In this essay I argue that Rawls runs together two quite different conceptions of this conceptual dyad. Picking up on his language, I call them “the strict compliance/partial compliance conception” and “the realistic utopia/transitional conception”.¹ I argue that it is confused to treat these conceptions as ways of describing the same distinction. They employ different criteria of classification, are motivated by different concerns, and have different practical upshots.

This raises the question of how, if at all, these heterogeneous dyads are to be related to one another. I draw on aspects of Rawls’ thought to present a rational reconstruction of his view shorn of the confusion. On this amended view, strict compliance is a local assumption motivated by his constructivist account of the principles of justice. Once these principles are in hand, we loosen the assumption of strict compliance to construct a realistic utopia. My proposal on Rawls’ behalf is that we view strict compliance theory as a moment

¹ In this essay, I follow two helpful discussions that each mark some version of this distinction. The first is Laura Valentini, “Ideal vs. Non-ideal Theory: A Conceptual Map” in Philosophy Compass 7/9: (2012): 654-664. The second is Zofia Stemplowska and Adam Swift, “Rawls on Ideal and Nonideal Theory” in A Companion to Rawls, edited by Jon Mandle and David A. Reidy, John Wiley and Sons (Chichester: 2014). Note that these authors do not argue that the two conceptions are in tension in the terms I do, or consider their relation to constructivism, or attribute to Rawls the amended view I propose, and ultimately critique, and so should not be blamed for any shortcomings in my argument.
in a broader project of ideal theorizing along realistic utopian lines that also has partial compliance moments. While this emendation absolves Rawls of the confusion, and seems true to his intentions, I argue that it runs into trouble nonetheless. Although the realistic utopia/transitional conception is well motivated, the strict compliance moment of ideal theory is more problematic even in its more limited role.

One upshot of my discussion is that we can embrace the division of the theory of justice into ideal and nonideal theory, while rejecting the strict compliance understanding of ideal theory. Critics who reject any role for ideal theory by arguing against the strict compliance conception may be right. But they move too quickly, when they urge us on this basis to reject the project of ideal theory. For such objections to strict compliance do not touch the realistic utopia conception, which is the broader and better motivated of the two.

§2: Ideal Theory as Strict Compliance Theory

In _A Theory of Justice_, Rawls first introduces the division between ideal and nonideal theory in the following memorable passage.

[F]or the most part I examine the principles of justice that would regulate a well-ordered society. Everyone is presumed to act justly and to do his part in upholding just institutions. Though justice may be, as Hume remarked, the cautious, jealous virtue, we can still ask what a perfectly just society would be like. Thus I consider what I call strict compliance as opposed to partial compliance theory. The latter studies the principles that govern how we are to deal with injustice. It comprises such topics as the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and conscientious objection to militant resistance and revolution. Also included here are questions of compensatory justice and of weighing one form of institutional injustice against another. Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for a systematic grasp of these more pressing problems…At
least, I shall assume that a deeper understanding can be gained in no other way, and that the nature and aims of a perfectly just society is the fundamental part of the theory of justice.²

Note first that Rawls’ original titles for the two parts into which he divides the theory of justice are “strict compliance theory” and “partial compliance theory”. The language of ideal (and by implication nonideal) theory is only introduced at the end of this passage as an afterthought. Rawls thus originally presents “ideal” as a synonym for “strict compliance”.

The passage begins by asserting that strict compliance theory is the theory of a “well-ordered society”. This is a term of art for Rawls. A society is well-ordered if there is public knowledge among its citizens that (1) they all accept the same principles of justice, and (2) these principles effectively regulate society in that individuals and institutions comply with their requirements.³ So strict compliance theory operates under the assumption that everyone knows that everyone accepts and lives up to the principles of justice that the theory outlines. It asks: suppose they were going to be perfectly realized, what principles of justice should we affirm as a society?

By contrast, partial compliance theory relaxes the assumption of strict-compliance, and asks how we are to respond to injustice when it arises.⁴ The list of subjects Rawls initially includes under this rubric is instructive. In the passage quoted above, he mentions the theories of punishment, compensatory justice, just war, and strategies of protest and resistance to domestic injustice. These topics presuppose departures from the principles of justice, and ask how we are to justly cope with them. For example, the theory of punishment presupposes wrongdoing on the part of the criminal. Similarly, just war theory asks what

---

³ *TOJ*, 453-454.
⁴ *TOJ*, 7-8, 351.
military measures can be taken on the assumption that some nation or rogue group is engaging in unjust military aggression. And so on.

So Rawls originally proposes a two-part theory of justice. The first part operates under the assumption that citizens and institutions comply perfectly with the principles of justice, and asks which principles are the most defensible under that assumption. The second part relaxes the assumption of strict compliance, and asks how we are to deal with departures from the principles when they arise.

Rawls does not explain in this passage why he adopts this assumptions and uses it to divide the theory of justice. But elsewhere he makes clear what he has in mind. When discussing the concept of a well-ordered society and the associated ideal of the citizens as free and equal in the later work *Justice as Fairness: A Restatement*, Rawls says that in order to get a “clear and uncluttered” view of a question, we both abstract, leaving out some of the messy and complicated features of the social world, and substitute idealizations for other features. Such abstraction and idealization, he tells us, allows us to simplify a question, by focusing precisely on the elements that we believe are most significant for developing an answer to our guiding question. In this context, he remarks, “The idea of a well-ordered society is plainly a very considerable idealization.”

The theoretical operations of abstraction and idealization that Rawls mentions are familiar from the social and natural sciences, and are the subject of a burgeoning literature in the philosophy of science. Some authors in this literature helpfully regiment these terms.

---

6 For another reference to “the numerous simplifications of justice as fairness”, see *TOJ* 517.
8 Onora Nell is a critic of ideal theory who quite early distinguished between idealization and abstraction and drew the connection to the philosophy of science. Many later critics build on Nell’s
On their regimentation, abstraction and idealization both involve simplifications employed in the construction of models. (Rawls too explicitly draws a connection between such operations and the construction of models, as we will shortly see.) Abstraction simplifies by omitting information, as when we model the flight of a cannonball using diagrams and physical equations that abstract from the cannonball’s color or the internal composition of its matter. Idealization simplifies in a different way, by attributing features to entities that they do not in fact possess. For example, when modeling the cannonball’s flight, we might represent gravitation force as uniform throughout the arc of its travels, when in fact the gravitational force varies with the changing position of the cannonball in relation to the Earth. Similarly, economists provide models of pricing in competitive markets by attributing to consumers rationally ordered preference and perfect knowledge of the goods on offer, even though consumers always lack these features in reality. These examples could be multiplied indefinitely since the operations of abstraction and idealization are ubiquitous and, although any particular use is open to question, are in general essential to theoretical endeavors.

11 Notice that idealization can be described as omitting features as well. But in abstraction one omits information in the sense that one does not take a stand on the omitted feature (e.g. the color of the cannonball), whereas in idealization one omits a feature in the sense of assuming that it is not present (e.g. varying gravitational force over the arc of the cannonball’s flight). This is why idealization, unlike abstraction, necessarily involves what Godfrey-Smith calls “fictionalizing for the purposes of simplification”. See Godfrey-Smith, “Abstractions, Idealizations, and Evolutionary Biology”, 3-4.
12 Abstraction is essential to all theorizing, since any theory will omit some information from the phenomenon theorized. Idealization is very often useful, although any particular idealization is
Strict compliance is an idealization, or rather family of idealizations, in this regemented sense. It involves attributing to citizens an ensemble of cognitive and motivational properties that they do not currently possess, and that may, like the rational consumers of the micro-economists, be impossible for a community of human beings to achieve. The same could be said as well for the implementation and functioning of institutions that are imagined as seamlessly in accord with the principles of justice. For, we might expect any actual institution to be awash with unintended consequences and uneven functioning.

Now, strict compliance and the concept of a well-ordered society are far from the only idealizations Rawls employs in the construction of his ideal theory. Notable and controversial further idealizations include his stipulation that society is a closed economy that does not interact with the economy of other societies, or that all the citizens enter the community by birth and exit it by death. But these other idealizations are something on the order of auxiliary assumptions that Rawls introduces for specific and discrete theoretical purposes within the broader project of ideal theory. Strict compliance, by contrast, is the idealization that he employs to define ideal theory itself and to demarcate it in a principled fashion from nonideal theory. Among his idealizations, it thus occupies a special place.

Rawls grants strict compliance this central place, because he thinks that it allows us to focus on the elements most relevant for answering the basic question facing a theory of justice: the justification of a set of principles of justice. John Simmons helpfully explains Rawls’ rationale. I should note that Simmons is the most penetrating and lucid expositor of substantial and contestable, and it may be possible to dispense with idealization altogether in some cases. See Godfrey-Smith, “Abstraction, Idealization, and Evolutionary Biology”, 6-7.

13 PL, 12; JAF, 12-14. Both these idealizations are motivated by the decision to bracket questions of international justice while investigating what Rawls takes to be prior questions about domestic political justice. See footnote 15 below for further examples of idealization from the Rawlsian corpus.
the second conception of ideal theory as realistic utopianism. For this reason, his defense of strict compliance is especially interesting. Here is what he says.

[O]ur objective should be to make only realistic assumptions in our ideal theory, in order to avoid idle utopianism. The defense of the ideal theory of justice is, however, to proceed against one straightforwardly unrealistic background assumption: that of “strict compliance”. We are to ask which principles for the basic structure could “well-order” a society on the assumption that most everyone accepts those principles, acts justly, and helps to uphold just institutions… The point of the strict compliance requirement for ideal theory in fact seem fairly plain…if we compare the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles, the different effects we observe can reasonably be taken to be wholly the responsibility of the different ordering principles themselves. So our comparison turns our to be quite strictly a comparison only of the principles of justice.\(^\text{14}\)

First, Simmons revealingly describes the assumption of strict compliance as the one glaring exception to the realistic utopian conception of ideal theory that his essay develops at length. However, he bravely defends this assumption on the grounds that it allows us to achieve an uncluttered view of the point most relevant for the comparison between different principles of justice. For assuming strict compliance allows us to separate two kinds of effect: the effects of realizing principles of justice, on the one hand, and the effects produced by partial failures to realize them, on the other. He thinks that the justification of principles of justice requires comparing the effects these principles would produce in society if fully realized. For the purposes of justification, the failures these principles might engender when joined with actual unruly human subjects and institutions is distracting noise. So the “one straightforwardly unrealistic” assumption of strict compliance allows us to achieve an

---

uncluttered view of the question of the justification of principles of justice, by setting aside the (at least initially) irrelevant complications that arise from noncompliance.\footnote{In Michael Weisberg’s typology of idealizations, strict compliance is a “minimal idealization” that functions to isolate core factors in the explanation of some phenomenon. Weisberg contrasts this with “Gallilean idealization” that involves approximations allowing one to model the phenomenon in a more (mathematically) tractable way. Note that Rawls employs Gallilean idealizations as well, for example, in his division society into three representative positions, or in his discussion of the properties of chain connection and close-knittedness that might hold between these groups. These assumptions simplify the problem of choice in the original position considerably, approximating a more messy reality about the trade offs between different groups. See Weisberg, “Three Kinds of Idealization”, pp. 1-8 and TOJ 70-73.}

It is natural to wonder why separating these effects should be so important for the evaluation of principles of justice. Simmons says that focusing on the effects of compliance allows us to consider the ordering of society for which different candidate principles are “wholly responsible”. He suggests that to consider the effects of partial compliance is to employ as a basis for comparison something other than the content of the principles. Although this seems true, the question is why it would be a mistake. After all, it is a natural thought that the most reasonable or justified principles will precisely be ones that win out in an all-around comparison with other principles. When considering which principles are most justified, why shouldn’t we take into account their full range of effects, insofar as these can be discerned, including those that cannot wholly be accounted for in terms of their content?

Although Simmons is not forthcoming on this point, Rawls has a direct answer.\footnote{Simmons does mention one other rationale for the idealization: that considering the effects of failure to comply is too difficult in the original position, since one would need some way to measure the degree of noncompliance institutions embodying the different candidate principles would elicit. (Note that this would make strict compliance a Gallilean idealization in Michael Weisberg’s typology discussed in Fn. 15 above.) This rationale is, however, undercut by his acknowledgment that institutions will predictably elicit different degrees of support, and that this fact can be used as one criterion to judge between them. See Simmons, “Ideal and Nonideal Theory”, 8-9.}

When Rawls returns to systematically explicate the scattered and terse remarks about the concept of a well-ordered society in \textit{A Theory of Justice}, first in his lectures on “Kantian Constructivism in Moral Theory”, and later in \textit{Political Liberalism}, he relates strict compliance
to his method of constructivism. The leading idea of constructivism is to establish a suitable connection between conceptions of the moral person and society and the first principles of justice. The principles of justice are to be defended by adducing reasons for thinking them to be norms adherence to which express and embody the relevant conceptions. This justification proceeds by means of a procedure of construction that assembles features of these conceptions in such a way that principles of justice may be defended in terms of them.

For these purposes, Rawls employs a series of what he calls “model-conceptions”. One of these model-conceptions is the account of the moral person, understood as the possessor of two moral powers: the sense of justice, and the capacity to form and pursue a conception of the good. The second is the well-ordered society. This is a model-conception of society as a system of social cooperation conducted on fair terms between free and equal citizens. The concept of a well-ordered society functions to provide a formal description of general features that obtain when citizens view and treat one another as free and equal moral persons, engaging with reciprocity in a system of social cooperation.

The third model-conception is that of the original position. It is meant to play a mediating role, helping us to construct from the other model conceptions the fundamental principles of justice. It does this by asking what principles representatives of moral persons living in a well-ordered society would choose. Famously, Rawls constrains this choice

18 “Kantian Constructivism”, 304.
20 “Kantian Constructivism”, 312-313. See also JAF, 18-19.
21 “Kantian Constructivism”, 308-310. See also JAF, 8-9.
through the veil of ignorance, and some other constraints, to lead a fair outcome.\textsuperscript{22} These constraints are precisely designed so that the principles selected will embody relations of reciprocity between free equals.\textsuperscript{23} The original position thus allows us to defend principles of justice as the most reasonable norms for agents conceived of as moral persons, related to one another as free equals in a system of social cooperation.\textsuperscript{24}

Rawls reasons that if we are to justify the principles in this way, then strict compliance and the other idealizations contained in the concept of a well-ordered society appear to be reasonable. For, when strictly complying with publically affirmed principles, agents view themselves and one another through the relevant conceptions and live up to them in their interactions.\textsuperscript{25} But, on the constructivist approach, the fundamental principles of justice simply describe what it is to do this: they tell us what society looks like when people view themselves as moral persons and their social relations as relations of reciprocity between free equals.

The whole purpose of the original position is to approach the question of what principles would be most reasonable for such agents by posing a situation of choice.\textsuperscript{26} So we are to set up the situation of choice so that the representatives will select principles that do embody these conceptions when adhered to. The most direct way to do this is simply to have the representatives choose principles for free and equal moral persons on the assumption that these persons will publically affirm and comply with the selected principles.\textsuperscript{27} Rawls thus has a methodological reason for employing the idealization of strict

\textsuperscript{22} “Kantian Constructivism”, 308-312.

\textsuperscript{23} “Kantian Constructivism”, 310; \textit{PL}, 16-17; for an explicit statement that the original position models the \textit{relation} between citizens see also \textit{LOP}, 30 n.32.

\textsuperscript{24} \textit{JAF}, 16.


\textsuperscript{26} “Kantian Constructivism”, 310-311; \textit{PL}, 22-28.

\textsuperscript{27} \textit{TOJ}, 145.
compliance involved in the model of a well-ordered society. The idealization appears to be licensed, perhaps even required, by his constructivist approach to principles of justice. Or so we might think.

Note that this constructivist approach explains an otherwise puzzling feature of Rawls’ discussion in *A Theory of Justice* of the topics addressed by nonideal theory. We have already seen that he locates in nonideal theory all phenomena that involve an essential reference to injustice. However, in his most extensive discussion in *A Theory of Justice*, he also includes cases involving justified paternalism towards children, the severely cognitively disabled, and the mentally ill. As Rawls understands these cases, they involve agents who blamelessly fail to live up to the moral conceptions of citizen. For example, children do not yet possess the two moral powers to a requisite degree to pursue a conception of the good or relate to their fellow citizens as free and equal social cooperators. Although there is no question of injustice here, Rawls holds that focusing on children and their relations will not serve to specify the fundamental principles of justice. For the fundamental principles of justice describe what it is to express and live up to the relevant conceptions, and children are not positioned to do this. So Rawls sets aside these questions for later nonideal reflection we will be equipped to pursue once the fundamental principles are at hand.

Rawls’ constructivist rationale for strict compliance notwithstanding, many authors have found this use of idealization objectionable, and have urged us to reject the project of ideal theory on this ground. I will allow these critics to speak in their own voices. Here is

---


29 *TOJ*, 249-250

Amartya Sen raising the concern that the simplification in question renders the theory of justice unfit to guide action.

Rawls’s approach, developed with admirable consistency and skill, does involve a formulaic and drastic simplification of a huge and multi-faceted task—that of combining the operation of the principles of justice with the actual behavior of people—which is central to practical reasoning about social justice. This is unfortunate since it can be argued that the relationship between social institutions and actual—as opposed to ideal—individual behavior cannot but be critically important for any theory of justice that is aimed at guiding social choice towards social justice.  

Here is Elizabeth Anderson urging that we must begin from an analysis of people’s actual motivational and cognitive limitations and flaws that lead them to depart from the requirements of justice.

A system of principles that would produce a just world if they regulated the conduct of perfectly rational and just persons will not do so when we ask human beings, with all our limitations and flaws, to follow them. Just institutions must be designed to cancel out our motivational and cognitive deficiencies, to harness our nonmoral motives to moral ends, to make up for each other’s limitations by pooling our knowledge and wills. To craft such designs, we must analyze our motivational and cognitive biases, diagnose how they lead people to mistreat others, and how institutions may redirect them to better conduct.

Finally, here is Charles Mills arguing that this idealization puts up obstacles to understanding the pressing and urgent questions about injustice that it is meant to illuminate.

If we start from what is presumably the uncontroversial premise that the ultimate point of ethics is to guide action and make ourselves better people and the world a

---

better place, then the framework above will not only be unhelpful, but will in certain respects be deeply antithetical to the proper goal of theoretical ethics as an enterprise. In modeling humans, human capacities, human interaction, human institutions, and human society on ideal-as-idealized-models, in never exploring how deeply different this is from ideal-as-descriptive-models, we are abstracting away from realities crucial to our comprehension of the actual workings of injustice in human interactions and social institutions, and thereby guaranteeing that the ideal-as-idealized-model will never be achieved.  

Although I do not evaluate the justice of these criticisms here, later I will raise some related doubts about the claim that we must employ the idealization of strict compliance if we wish to honor constructivist scruples. For now, I wish only to draw attention to the fact that the assumption of strict compliance elicits such objections, and that they are one large factor (certainly there are others) in the reasoning leading critics to reject the division of the theory of justice into ideal and nonideal theory.

§3: Ideal Theory as Realistic Utopianism

In Rawls’ writings, a second conception of ideal theory coexists uneasily alongside the strict compliance conception. In his later work, this conception becomes more prominent. Although it never entirely supplants strict compliance, it tends increasingly to be Rawls’ preferred point of entry for explaining the division between ideal and nonideal theory. This is the conception of ideal theory as presenting a practical good.

This conception is clearly present in A Theory of Justice. There, Rawls writes, “Viewing the theory of justice as a whole, the ideal part presents a conception of a just society that we

33 Charles Mills, “‘Ideal Theory’ as Ideology”, 170.
34 For some representative samples from the later works, see JAF, 4-5; and especially The Law of Peoples, Harvard University Press (Cambridge: 1999), 11-12 and 89-90 (hereafter LOP).
are to achieve if we can.” This reference to achievement sounds a practical note: what ideal theory presents is something to be realized in practice. In the same work, he calls a just society “the end of political action by reference to which practical decisions are to be made.” Ideal theory thus presents an end or telos that is meant to serve as a reference point for making political decisions. Elsewhere he describes this end as providing the “objective” or “long-range goal” towards which we are to strive. On the conception that emerges from remarks like these, the role of ideal theory is to articulate an aim for political practice.

In later works, Rawls expresses this idea with the intentionally paradoxical phrase “realistic utopia.” We could represent each word of this apparently oxymoronic coupling as indicating a requirement on ideal theory. Ideal theory must be utopian, because it uses moral ideals, conceptions, and principles to specify the concept of a reasonable and just society that can serve as an end for our political hope and action. Since neither our society, nor any others with which we are familiar, are just, this involves envisioning a better world that does not exist.

On the other hand, this use of the moral imagination is in the service of articulating a goal for practice and not an idle fantasy or wish. For this reason, it must be realistic. The end it articulates must be, as far as we can tell, possible to achieve, at least under suitably favorable conditions. It must, as Rawls says, “fall under the art of the possible”. The end it determines must be compatible with the facts of human psychology and biology, including our characteristic foibles and vulnerabilities, and with political sociology, economics, and

35 *TOJ* 246.
36 *TOJ* 219
39 *LOP*, 14.
40 *JAF* 185.
social science more generally. It must also be compatible with the material and historical conditions of our social world. Having presented an ideal, one must thus be willing to address the most plausible and empirically grounded objections to its feasibility.

In endorsing this requirement of realism, Rawls would seem to be in familiar territory, for it is widely agreed that something is deeply wrong with a political philosophy if it rests on an unrealistic conception of the political subject, or of our material or historical conditions. A political philosophy that describes a community of angels rather than human beings, or that situates the political community in the Garden of Eden rather than in current material conditions, rightly strikes us as misguided. It’s in this context that Rawls appropriates Rousseau’s methodological dictum in On the Social Contract to take “men as they are, and laws as they might be.” Rawls puts the point elegant in A Theory of Justice, writing, “Conceptions of justice must be justified by the conditions of our life as we know it or not at all.” The theory of justice must be a theory of justice for human beings.

Rawls also connects the requirement of realism with two other features of a theory of justice: workability and stability. A theory of justice is workable if citizens can apply the principles of the theory to their ongoing political affairs. These principles are to provide the terms in which citizens evaluate their institutions, rendering public justification to one

41 LOP, 11-13.
43 For able defenders of the widely accepted requirement of realism, see David Miller, Justice for Earthlings (CUP: 2013), and Joshua Cohen, “Taking People as They Are?” in Philosophy and Public Affairs Vol. 30 No. 4, 363-386. For some strong dissenting opinions, see G.A. Cohen, Rescuing Justice and Equality (HUP: 2008), and Andrew Mason, “Just Constraints” in British Journal of Political Science 34: 251-268. For a more nuanced partial dissent, see David Estlund, “Human Nature and the Limits (If Any) of Political Philosophy” in Philosophy and Public Affairs 39, No. 3, (2011) 207-237. In this paper I say nothing to address the powerful arguments of the dissenters. To make a full case for the realistic utopia/transitional conception one would have to respond to their criticisms.
44 LOP, 7.
45 TOJ, 454.
46 LOP, 13.
another, and hold each other to account for failures of justice. For principles to play this role, they must be formulated with concepts that make it possible to determine whether or not they have been satisfied.\textsuperscript{47} They must also place demands on the citizens that it is possible for them to fulfill, and be known to fulfill, at least under favorable circumstances.

That Rawls includes stability under the rubric of realism is not surprising, since he identifies the stability of a theory of justice at various points with its feasibility.\textsuperscript{48} The concept of stability is a complex one, and I can do no more here than touch on some main points.\textsuperscript{49} A theory of justice is stable if it specifies a basic structure that can maintain its justice in the face of inevitable countervailing forces. The system must be able to set in motion sufficient forces to counteract these injustices when they arise. In that case, “the inevitable deviations from justice are either corrected or held in tolerable bounds by the system”.\textsuperscript{50} Of course, stability can come through many routes, including the barrel of a gun, indoctrination, and moral lethargy. But insofar as this is a requirement on a theory of justice, the stability in question must be what Rawls calls “stability for the right reasons”.\textsuperscript{51} The countervailing forces must consist largely in the sense of justice of the citizen body, and their allegiance to their just institutions. For a theory of justice to be stable, its principles must thus be capable of “generating their own support”.\textsuperscript{52} Since Rawls assumes that the way in which one acquires allegiance to principles of justice is through growing up with an experience of being

\textsuperscript{48} \textit{TOJ}, 144 and 504.
\textsuperscript{49} The most sophisticated discussion is to be found in Paul Weithman, \textit{Why Political Liberalism? John Rawls’ Political Turn} (OUP: 2011), 42-68.
\textsuperscript{50} \textit{TOJ}, 457-458.
\textsuperscript{51} \textit{PL}, xxxvii and 142-43.
\textsuperscript{52} \textit{TOJ}, 456.
benefited by just institutions, stability has an importantly intergenerational form.\textsuperscript{53} The justice of a society must be such that it can be maintained and reproduced through time.

While it might initially seem that realism and utopianism are two unrelated requirements, joined in an ad-hoc fashion, and so limiting one another externally, Rawls clearly intends a deeper unity. The heart of the utopian project consists in the identification and defense of principles of justice, and their use to characterize a just society. The realism constraint is internal to this project, since the theory of justice is intended to be a theory for human beings. The principles of justice are principles \textit{for us}, intended to regulate the public claims we make on one another in light of our shared institutions. The claim that the theory of justice must itself be realistic—taking human beings as they and laws as they might be—is thus not an alien constraint on justice. It rather flows from ideas about what the principles of justice are, and how they must be justified in light of this fact. The utopian aspiration is thus inextricably bound up with an element of realism.

The requirement of realism is, in turn, informed by a utopian element. If we consider the requirement of workability, we can see that it is grounded in a moral idea that the principles of justice serve as public standards through which citizens hold one another to account. The principles of justice are to serve as the public charter of our social world.\textsuperscript{54} They serve as public standards that citizen can employ in their deliberations with one another as free equals.\textsuperscript{55} Justice thus must be something that can be publically appealed to in an open way in our political reasoning. Similarly, when it comes to stability, we are concerned with stability for the right reasons. This is a stability that arises from a firm allegiance to just institutions. The principles of justice must be capable of generating their

\textsuperscript{53} \textit{TOJ}, 453-479.
\textsuperscript{54} \textit{TOJ}, 5.
\textsuperscript{55} The fullest discussion of publicity is found in \textit{PL}, 66-71 and 77-81. See also \textit{TOJ}, 133-136.
own support among the citizen body, and they must be able to do so without the help of illusion or ideology. So the requirement of realism is bound up with utopian elements as well.

Realism and utopianism thus turn out to be two sides of the same coin, united in the concept of a just society, understood as a practical good for us. On this conception, the central task in ideal theory is to combine the cutting edge of utopianism with the hard steel of realism. By realizing this union, political philosophy at its best expands the horizons of our sense of political possibility by showing us a just world that could become a practical reality. This is why Rawls says that ideal theory “probes the limits of practicable political possibility”. If it is successful, it sustains our practical hope that a just world is in fact possible, by showing us how this end might be realistically conceived. Let us call this the “realistic utopia” conception of ideal theory.

The corresponding conception of nonideal theory views it as asking how we are to work from our current condition of injustice towards this realistically utopian goal. As John Simmons puts it in a perceptive discussion, “Where ideal theory dictates the objective, nonideal theory dictates the route to that objective (from whatever imperfectly just condition a society happens to occupy).” If ideal theory discusses the long-range goal, nonideal theory discusses the means that can effectively and justly be taken in pursuit of it in the short and middle range. Simmons aptly describes nonideal theorizing on this conception as transitional in character, since it concerns the movement from situations of injustice to the

---

56 JAF, 184-185.
57 JAF 4.
58 LOP, 89-90.
59 Simmons, “Ideal and Nonideal Theory”, 12.
60 TOJ, 8, 216-217.
end of a just society. I follow Simmons in calling this conception of nonideal theory “transitional theory”.

Since transitional theory is to strategically navigate towards the goal of a just society in this fashion, it must allow us to identify in a systematic way the injustices that are obstacles to forward movement. Having identified them, it must diagnose the underlying causes of these political pathologies, with a view to proposing morally permissible and effective remedies. It must investigate the strategies of opposition, and resistance, and the policies and programs that would be effective in moving us towards conditions of justice. It looks for courses of action and policy that are likely to be effective in moving us towards this goal, and are politically possible, and morally permissible.

Taken together, these ideas about political philosophy express one interpretation of the widely shared idea that political philosophy is practical. On this interpretation, political philosophy is practical insofar as its project is to articulate a realistic conception of a just society that serves as a suitable object of practical hope and political action, and to outline feasible and just steps towards achieving this goal, working out from our currently unjust condition. The practicality of political philosophy thus consists in the fact that it makes explicit and defends a practical good that serves as an end towards which our political action can be directed. Furthermore, it seeks to identify and diagnose the root causes of the injustice of our society in such a way that concrete and pragmatic action can be taken to overcome them.

§4: The Problem with Running the Conceptions Together

61 I take this general metaphor of pathology from Elizabeth Anderson, The Imperative of Integration, Princeton University Press (Princeton: 2010). (Hereafter, Imperative.)
62 LOP, 89-90
Rawls everywhere mingles these two conceptions of ideal theory, and most critics and apologists follow his lead. Here is a representative passage from *A Theory of Justice*.

It will be recalled that strict compliance is one of the stipulations of the original position; the principles of justice are chosen on the supposition that they will be generally complied with. Any failures are discounted as exceptions. By putting these principles in lexical order, the parties are choosing a conception of justice suitable for favorable conditions and assuming that a just society can in due course be achieved. Arranged in this order, the principles define then a perfectly just scheme; they belong to ideal theory and set up an aim to guide the course of social reform.63

In dizzying fashion, Rawls turns back and forth between the strict compliance and realistic utopian conceptions of ideal theory. On the one hand, ideal theory is formulated on the supposition of strict compliance; failures are to be idealized away (“discounted as exceptions”). On the other hand, ideal theory assumes that a just society so described can be achieved “in due course”; it sets up an aim that is to guide the course of social reform. After the quoted passage, Rawls goes on to say that such social reform is treated by a nonideal theory that asks what justice requires when we relax the assumption of strict compliance. In short, ideal theory presents a strictly complying society as a realistic utopia, and nonideal theory considers how we are to transition from a partially complying society to such a realistic utopia.

In this section I argue that running together these two conceptions is problematic. The two conceptions are motivated by different concerns, deploy different criteria for deciding whether something belongs to ideal theory, and set vastly different stakes for locating a phenomenon one side or the other of the ideal/nonideal division. If we slide back and forth the way Rawls does in the quoted passage, we risk the twin errors of pulling

---

63 *TOJ* 215-216. For a very similar passage from the end his career, see *JAF*, 13.
substantive political conclusions out of a hat and building castles in the air. As will become clear shortly, I do not think Rawls commits these errors in practice. His actual practice, on a charitable reconstruction, is considerably less confused than the passages in which he describes his practice. (I exploit this fact in presenting an amended Rawlsian position in §5.) But, for the moment, I wish to take this passage and others like it at face value. For this purpose, let me consider the conceptions in turn.

On the strict compliance conception, as we’ve seen, ideal theory involves the application of the intellectual operation of idealization to the special case of constructing fundamental principles of justice. It does so by working with a model of a well-ordered society, including the assumption of strict compliance. Let me focus on the operation of idealization to bring out two of its general features. The first is that idealizing assumptions are counterfactual, assigning features to the phenomena that they do not possess. Indeed, in many cases of legitimate idealization, a theorist may attribute features to the phenomena for the purposes of the model that are not only false, but impossible given the nature of the phenomena. Even if there can be no fluids without viscosity, or human consumers with full information, it is not necessarily mistaken to consider such counterfactual suppositions when presenting a model of fluid dynamics or price useful for answering some theoretical question.

Note further that the bare operation of idealization does not presuppose that the idealized phenomenon has any particular normative valence. In many theoretical contexts where it is legitimate to employ idealization, e.g. physics, talk of goodness and badness is simply out of place. However, in other cases, we may provide models of phenomena that the theorist judges to be bad. Karl Marx constructed an idealized model of the capitalistic mode of production, while opposing it as an evil to be surmounted. Similarly, sociologists provide
models employing abstraction and idealization that are intended to explain the perpetuation of racial inequality in our society,\textsuperscript{64} or the rate of suicide.\textsuperscript{65}

The strict compliance conception applies the all-purpose theoretical operation of idealization to the special case of constructing fundamental principles of justice. The idealizations employed in the model-conception of a well-ordered society involve attributing counterfactual psychologies to the citizen and properties to the operation of institutions that are both envisioned as in flawless accord with the principles of justice. Some of these assumptions about human motivation and institutional operation almost certainly are impossible given human nature and institutional properties.\textsuperscript{66} Others assumptions are possible for human beings only if certain individuals are excluded from the political community, as when Rawls leaves aside for the purposes of ideal theory citizens with severe mental illness, or cognitive disabilities, or children from the discussion, on the grounds that such agents cannot strictly comply with the principles of justice given their illness, disability, or youth.\textsuperscript{67}

Admittedly, this sounds bad, as though Rawls is saying that society would be better off without children, the severely mentally ill, or some persons with disabilities, or that we should try to achieve a society without them. But in fact there is no such implication. For we have seen that the operation of idealization has no internal connection to practicability or goodness. True, in this case idealization is being employed to isolate a relationship of free

\textsuperscript{64} See, for example, Douglas Massey, \textit{American Apartheid: Segregation and the Making of an Underclass} (HUP: 1998).
\textsuperscript{65} See Emile Durkheim, \textit{On Suicide} (Penguin: 2006).
\textsuperscript{66} Here I diverge from Swift and Stemplowska who interpret Rawls as arguing that strict compliance is realistic, and so practically possible. As I read the relevant passages, Rawls argues instead that in a just society the inclination to injustice would be greatly limited and reduced, so that stabilizing forces would suffice to rectify problems of injustice when they arose. But he does not argue—how could he?—that strict compliance is practically possible for a human society. I discuss this further below in §5. See Swift and Stemplowska, “Ideal and Nonideal Theory”, 115.
\textsuperscript{67} \textit{TOJ}, 216-220; \textit{Political Liberalism}, 16-20.
and equal citizens that is valuable and is related to our ideas about justice. But Rawls is not saying that it would be good if society lacked individuals who lack the features of the citizen, nor is he proposing excluding them in practice. Since the grounds for the idealization in question are theoretical, the only consequences for locating a phenomenon on the nonideal theory side of the divide are intellectual. To say that some phenomenon belongs to nonideal theory is only to declare that it presents complex or untidy difficulties that are best treated in a second moment of theoretical reflection, once clarity on the principles of justice has been achieved. Nothing of political moment or practical consequence seems to be at stake in putting things on one side or the other.

On the realistic utopian conception, by contrast, ideal theory describes the end of a just society. This is a practical good. Nonideal theory discusses how we are to move from conditions of injustice towards this good. This teleological structure of end, and strategic reasoning towards this end, is specific to practical endeavors, and has no correlate in merely theoretical investigations, such as physics, biology or economics.\footnote{This view is not teleological in Rawls’ technical sense of that term. In that sense, a teleological theory is one that defines the good independently of justice (“the Right”) and then defines justice as the maximization of the good. Since the good in question in the realistic utopia conception is a just society, it is obviously not defined independently of justice. Nor is there any appeal to maximization. It is only teleological in that the view involves orientation to an end, or \textit{telos}. See \textit{TOJ}, 21-22.} Since the good in question is \textit{practical}, it follows that the just society treated by ideal theory must be \textit{practicable}, something that could, as far as we can tell, be brought about through our collective action. For this reason, our description of the just society must be compatible with our nature, including facts about human psychology and development, disability and the susceptibility to illness, and the material conditions of our world.

Furthermore, since the end in question is a practical \textit{good}, the classification of something as belonging to ideal theory has immediate normative significance. For, on the
realistic utopian conception, to locate some phenomenon in ideal theory is to say that it is part of a just society that is the proper aim of our political hope and action; to classify some phenomenon as belonging to ideal theory is to say that, as a matter of basic political conviction, we stand for it as a component of a realistic utopia. By contrast, to locate some phenomenon in nonideal theory is to declare it to be in some way bad; it is to say that, as a matter of fundamental political conviction, we stand against it as ultimately incompatible with the just society to which we aspire.

When we put the two features of practicality and goodness together, we see that on the realistic utopian conception, the dividing line between ideal and nonideal has momentous political implications. To say that some phenomenon belongs to nonideal theory is to say that it can and must be overcome. It is at best, a necessary way station, and at worst, an obstacle, on the route to a just society. By contrast, to classify some phenomenon as belonging to ideal theory is to say that it can and must be achieved. It is a component of the just society towards which our action ought to be directed.

The two conceptions of ideal theory I’ve been discussing are radically heterogeneous. The grounds for treating something as belonging to ideal theory on each conception are distinct, and the decision to so treat something has different consequences in each case, in the one case intellectual, in the other case substantive political. Equivocating between the two conceptions is confused. This point is obvious, but is important enough to bear elaboration.

For example, let us follow Simmons and Rawls, and suppose that the assumption of strict compliance is to be justified by the need to separate the effects of instituting principles of justice from the distinct and complicated further effects that deviations from these principles might produce. For this theoretical reason of separating one kind of effect from
another, we assume strict compliance as a way of isolating the feature we see as most
significant for the justification of principles of justice on constructivist lines. Now, as we’ve
seen, the strict compliance conception of ideal theory leads us to classify any phenomenon
that involves deviation from the principles of justice as belonging to nonideal theory. This
list of subjects, if you will remember, includes compensatory justice, the theory of
punishment, just war theory, civil disobedience, and so on. These subjects treated in
nonideal theory in turn justify various institutions including, presumably, criminal law, tort
law, reparations, the maintenance of standing armies, and so on. 69

If we now slide from the strict compliance conception to the realistic utopian
conception, we will view these phenomena not merely as complex and messy phenomena
best treated in a second moment of theoretical reflection, but also as things that can and
must be overcome in the pursuit of a just society. I have no objection to this as a substantive
political commitment: anarchists have thought this about regimes of law backed by a state
monopoly on violence, and peace activists have advocated this concerning the instruments
and practices of war. My point is only that you cannot defend the political conviction, for
example, that justice demands we work towards abolishing standing armies, on the ground
that the existence of war is ruled out by an idealization we had to employ in order to isolate
one effect from another for the purposes of theory construction. This is not the right sort of
reason to support a radical political conviction about the ultimate aims of political action. 70

Here we are trying to generate political convictions about what is bad from a practical point

69 I should here note that Rawls’ remarks about penal sanctions are apparently inconsistent. Although
he original locates them in nonideal theory, in other places he argues that they are justified as an
assurance mechanism, even under conditions of strict compliance in ideal theory. (I find this
justification of punishment thin and implausible, but I cannot argue for this here.) See TOJ, 211-213.
70 Nor is this how Rawls himself argues for this conclusion late in his career, when the realistic utopia
conception has become more central to his thinking. See the discussion of democratic peace in LOP,
44-54.
of view out of theoretical considerations that bear no intrinsic connection to such practical normativity.

Furthermore, as we’ve seen, idealizations can be legitimate even when they involve counterfactuals that are impossible given the nature of the entities in question. So when we slide from the strict compliance to the realistic utopia conception, we may treat some phenomenon that obtains under perhaps impossible counterfactual assumptions as a good to be achieved in practice. Here the danger is that we will be building castles in the air. This error is simply the flip-side of the other: we now are trying to draw conclusions about what is a *practical* good that we can and must achieve from grounds that bear no internal connection to practicability.

So we should not use the realistic utopia and strict compliance conceptions simultaneously as two ways to make what is thought to be the same division between ideal and nonideal theory. In reality, the terms “ideal theory” and “nonideal theory” track different kinds of distinctions when understood in one or the other way. This naturally raises the question how these heterogeneous distinctions are related to one another. Should we employ one but not the other in the theory of justice? Is there a way of finding a place for them both in the theory of justice, without treating them as marking the same distinction?

§5: Combining Strict Compliance and Realistic Utopia: A Rational Reconstruction

It is possible to construct a coherent position from Rawls’ work that is amended to avoid the confusion that I have just documented.\textsuperscript{71} The key to the interpretation is to notice the

\textsuperscript{71} I owe the insight of this section to incisive questions from Arata Hamawaki. He suggested to me that perhaps Rawls was working unclearly with a distinction between justice as a set of principles and justice as an end, and that the two senses of ideal theory I identified might each be associated with
emphasis Rawls puts on stability when explaining the idea of a realistic utopia. In both *A Theory of Justice* and *Justice as Fairness* Rawls divides the argument from the original position—and so ideal theory—into two parts.\(^2\) The first part presents the argument for the principles of justice in the original position, employing the assumption of strict compliance and the other simplifying devices. The second part tests the feasibility of the resulting conception by exploring whether it is stable in light of the psychology of moral learning, and the theory of the human sentiments.

This two-fold division can be further elaborated in connection with the three parts into which Rawls divides *A Theory of Justice*. In Part I, “Principles”, Rawls presents his constructivist defense of the principles of justice using the model conception of the original position. In Part II, “Institutions”, Rawls specifies a set of institutions that would satisfy the two principles defended in Part I. Finally, in Part III, tantalizingly titled “Ends”, Rawls discusses the stability and feasibility of the package of principles and institutions defended in Parts I and II. We might view this as a sequence through which the concept of justice is increasingly determined. Part I takes us from the bare concept of justice to the conception of justice as fairness, consisting of the two principles of justice.\(^3\) Part II further determines the concept of justice, taking us from the conception of justice as fairness to a set of institutions realizing these principles in the basic structure of society.\(^4\) Part III further determines the

\(^{2}\) For the distinction between the concept of justice and its conceptions see *TOJ*, 5-11.

\(^{3}\) In *A Theory of Justice*, the institutions defended include constitutional democracy with production organized by a market. However, Rawls argues that the principles of justice underdetermine whether these market arrangements must involve private ownership of the means of production. He thus offers two sets of possible institutions comparative with the principles of Part I: liberal socialism and property owning democracy. In later work, he develops the idea of property owning democracy further, which he views as a more feasible option for the United States, given facts about our culture
concept of justice, by representing the institutionally realized principles as an end (realistic utopia) capable of being realized and sustained through the activity of a community of (human) citizens.

It is also possible to represent this sequence as involving a progressive relaxing of the assumption of strict compliance. In the first stage, Rawls employs the idealization of strict compliance in order to justify principles of justice through his constructivist methods. Here the assumption is fully in place: the principles are selected in the original position on the supposition that citizens and their institutions will strictly comply with them. In the second stage, Rawls relaxes the assumption of strict compliance sufficiently to consider real institutions that might satisfy the two principles.75 One can see this happening, for example, in his discussion of democracy as a case of imperfect procedural justice, and his discussion of the limitations on majority rule provided by constitution protections along with judicial review.76 Finally, in his discussion of stability, Rawls considers the forces arising from the moral sentiments that can be marshaled by a just society to right the course when injustice arises.77 (Here it is crucial to his argument that in a just society, justice would not be inimical to our good.78) It is also here that Rawls considers the darker moral emotions of spite and envy, and argues that they would not be strong enough to undermine the justice of a just society.79

---

75 This fits with Rawls’ discussion of the four-stage sequence and the relaxing of the restrictions on knowledge through this process. See TOJ, 200.
76 TOJ, 221-234.
77 TOJ, 453-458
78 This is Rawls’ famous argument that the right is congruent with the good. See TOJ, 513-529, and the extensive discussion in Weithman, Why Political Liberalism?, 42-182.
79 TOJ, 530-548.
The picture that emerges from this reading is the following. Ideal theory is a broad and multi-stage process the aim of which is to progressively determine the concept of justice until it acquires the determinacy of an end for practice. This process involves one moment of strict compliance through which the principles of justice are defended. Ideal theory then moves through further moments that introduce elements of partial compliance necessary for further determining the concept of justice, first into an account of a basic structure of institutions satisfying these principles, and then ultimately into a practical good.

On this proposal, what should we then say about nonideal theory? Is it similarly a multi-staged affair? I think that the clearest way to think about nonideal theory on this model is to view it as the complement to the whole package of ideal theory as defined above. Ideal theory terminates with a construction of a realistic utopia; nonideal theory picks up where ideal theory leaves off. It engages with the “pressing and urgent matters” of our everyday experience by reasoning practically from our current conditions of injustice towards the practical good represented by ideal theory. In other words, I think the clearest way to use the term nonideal theory on this reading is simply to mean transitional theory.

Indeed, there is a sense in which the whole of ideal theory on this view could be represented as the multi-staged construction of a realistic utopia, and so there is a sense in which the realistic utopian/transitional conception is the broadest and most complete characterization of the ideal/nonideal theory dyad. But we must hasten to add that ideal theory possesses a strict compliance component corresponding to Rawls’ initial introduction of the concept of ideal theory in *A Theory of Justice*. Thus, both characterizations have their place, although the strict compliance conception is now one component subordinated to the broader enterprise of realistic utopianism.
While I believe this reading is true to Rawls’ intentions, and salvages almost everything of any importance that he wants to say about the division of the theory of justice into ideal and nonideal components, I would be remiss not to mention the many elements of his thinking that are incompatible with this picture. It is, obviously, incompatible with his statements, maintained over the course of his entire career, treating realistic utopia and strict compliance as two interchangeable characterizations. Furthermore, while the reading finds a place for his characterization of ideal theory as strict compliance theory, it finds no place for his characterization of nonideal theory as partial compliance theory. For example, this reading rejects the idea that the discussion of children, or the disabled, belongs to nonideal theory, since both will be present as citizens in a realistic utopia. It also rejects the idea that the mere fact that a phenomenon presupposes injustice settles it that the phenomenon belongs to nonideal theory, since a realistic utopia may have ways of handling inevitable injustice when it arises.

Furthermore, it is at the very least in tension with Rawls’ tendency in talking about institutions in Part II and stability in Part III of a theory of justice to represent himself as limiting himself to “the special case of a well-ordered society”.\(^{80}\) This is the very same language he uses when he signals his focus on strict compliance theory. When first introducing strict compliance, he seems to treats it as a property possessed by a well-ordered society.\(^{81}\) (I followed his lead in this regard in §2 when explicating the strict compliance conception in conjunction with the concept of a well-ordered society.) This is problematic, insofar as on my proposed reading, the later stages of ideal theory relax the idealization of strict compliance, and introduce elements of partial compliance. However, there is a looser sense of the term “well-ordered society” that we might introduce that does not presuppose

---

80 For example at TOJ, 279 and 457.
81 See TOJ, 8.
strict compliance. On this understanding, a society’s being well-ordered is compatible with inevitable infractions by individuals, and temporary or small scale deviations from principles by institutions. This is the sense of being well-ordered on which a realistic utopia might be said to be a well-ordered society. Again, I think this would be true to Rawls’ intentions, and would accord with many of his uses of the term.\textsuperscript{92}

Doubtless, these are costs. My reading belongs to the mode of rational reconstruction, where one tries to stay true to the guiding intentions and central claims of an author, while offering emendations necessary to address problems with the argument as written. The reading is offered with the charitable intention of providing Rawls with a workable alternative consonant with the most important elements of his own thought. However, I am painfully aware of the fact that Rawls is a complex and profound thinker, and there are doubtless ways of understanding him that I have missed. I would be delighted if someone could present an interpretation of his thought that deals satisfactorily with the confusion I identified in §4, while fitting better with the details of his text. For the present, it is the best I can do on his behalf.

\textit{§6: Constructivism without Strict Compliance}

But is rationally reconstructed Rawls right?

Political philosophy is a practical enterprise. Although a highly reflective rational practice, its aim is nonetheless political action. Now, reasoning that aims at action is practical reasoning rather than theoretical reasoning. So we ought to expect political philosophy to

\textsuperscript{92} For example, he frequently speaks of the citizens of a well-ordered society acting justly “for the most part”. These formulations leave it open that some topics dealing with the response to injustice might be features of a well-ordered society dealt with by ideal theory. See, for example, \textit{JAF}, 9 and \textit{PL}, 35.
employ the sorts of concepts and justifications that one finds in practical rather than theoretical reasoning.\(^\text{83}\)

Classically, the practical syllogism—the attempted formal representation of practical reasoning—was conceived as having two premises that worked together to justify a conclusion that was an action. The major premise was a representation of a practical end, or good. The minor premise was a representation of the means towards this end. Together they worked to justify an action in the here and now. The realistic utopia/transitional conception finds this structure present in our thinking about justice. Ideal theory articulates a realistic utopia, a conception of the end of a just society that we are to achieve if we can. Nonideal theory works to identify and diagnose the ways we fall short of this ideal, so that courses of action—practicable remedies—can be proposed as means to this end. Together these components work to identify feasible and just action that will move us towards a more just society. Here political philosophy is being represented as practical reasoning. Indeed, the really striking thing is that the realistic utopia conception employs this structure of practical reasoning as a principle for systematically organizing the discipline of political philosophy. While there are certainly other interpretations of the idea that political philosophy is practical, the realistic utopian/transitional conception is particularly striking and direct.\(^\text{84}\) It is worthy of our attention on this ground alone. I thus find the overarching role that reconstructed Rawls grants this conception attractive.

About the more limited role that reconstructed Rawls gives to strict compliance in defending the principles of justice, I have my doubts. Although I am sympathetic to his constructivism, I find problematic his argument that links constructivism to strict

---

\(^{83}\) Rawls affirms this view at _PL_, 93 ff.

\(^{84}\) I go some way to defending the realistic utopian/transitional conception against rival understandings of the practicality of political philosophy in three papers, “The Priority of Ideal Theory”, “Justice as an End”, and “Scarcity and Realistic Utopia”. 
compliance. To recall, on a constructivist approach, the principles of justice are to be defended as the most reasonable norms for citizens conceived as free and equal moral persons. The principles of justice describe the way the social world looks when people view one another as free equals, and treat one another according. But, Rawls argues, the social world only looks this way when people strictly comply with the principles of justice. To the extent that they depart from the principles of justice they fail to treat one another as free equals, and so their action fails to express the normative point of view articulated by the principles of justice. Such failures are thus irrelevant when characterizing the principles of justice. The entire argument for strict compliance thus hinges on the claim that one can only characterize the relation of free equals if one prescinds from considering the issues raised by deviations from the principles of justice. Or to put the point the other way around, the issues raised by being wronged are irrelevant to characterize the relation of free equals.

But are they really? Here’s an analogy to convey the doubt I have in mind. An understanding of the health of any system of the human organism might plausibly be said to depend on an understanding of the sound anatomy and biological functioning of that system. If we were willing to speak, somewhat loftily, about “the principles of health” for some system, we might say that these principles are grounded in facts about the sound biological functioning of the relevant system, and, ultimately, of the organism as a whole. Such principles, we might say, describe the general functioning of a body insofar as it is healthy. Now imagine a philosopher who tries to infer from this that any mention of illness is out of place in an account of health. Sickness, she reasons, is precisely a departure from

---

85 I found Joshua Kleinfeld and Jorg Schaub, “The Place of Redressive Justice in the Concept of Justice” (Ms.) helpful in thinking about the issues addressed in this paragraph and following.
86 I was helped in thinking through this analogy by Michael Thompson’s “The Representation of Life” in his Life and Action, Harvard University Press (Cambridge: 2009), 25-84. Note that the argument I make here raises complications for the “simple-minded principle of inference” discussed at 80 ff.
health. To the extent that the body is sick, it fails to evince the character of healthy functioning that the principles are intended to capture. So when we describe the principles of health, we must prescind from all mention of the issues raised by illness.

We might point out in reply that part of being healthy is having a well-functioning immune system. One’s immune system functions well if it has effective ways of protecting from illness, and when illness occurs, of responding to it in such a way as to return one to health. The immune system is a crucial part of human anatomy. Given the ubiquity of pathogens in the environment of a terrestrial organism, the health of terrestrial beings depends on, and involves, the defense of the body from illness. However, to represent the sound and healthy functioning of the immune system, one must represent the human body as both under threat from and succumbing to pathogens. That is, to represent a central aspect of human health, one must depart from assumptions about strict compliance and actually represent the body as sick to some degree. The explanation for this is simple and straightforward: part of the health of a healthy body is how it responds to sickness.

No doubt, this analogy has its problems, and I don’t want to rest anything on the claims about health. But it seems to me that an analogous line of thought does hold in the case of justice. If the free and equal citizens of a political community are under standing temptations to commit injustices of various kinds against one another, then part of the account of their sound relations will have to mention justifiable ways they have of preventing injustice and rectifying it when it arises. But, of course, the whole topic of the response to injustice only shows up if one departs from the idealization of strict compliance. Thus, the assumption of strict compliance is out of place, even in the account of at least some aspects of the relationship between free and equal citizens. An essential part of the justice of a just society consists in its response to injustice.
Indeed, whatever the relation of normative principles to the possibility of practical error in general,\(^{87}\) there seem to me to be special reasons for thinking that justice, of all things, must take account of such violations. Rawls is, I think, right to focus his account on the relations between citizens, since justice is an essentially relational phenomenon.\(^{88}\) Justice involves claims that we can justifiably make on one another, often (although not always) to order our shared institutions. This relational character is dramatized in the primal scene of Rawlsian theory, where a person occupying one representative position calls on another occupying a different representative position to justify the inequality between them.\(^{89}\) But the claims of justice, unlike the claims of generosity or charity, are also ones that we can hold one another to fulfilling. To attribute an obligation of justice is already to attribute a right to the object of the obligation. But to attribute a right is to attribute a standing to hold the subject of the obligation to the fulfillment of her obligation. When it is a matter of justice, we do not have to sit idly by while agents wrong us, but are authorized to take an active stance of opposition to the failure to fulfill the obligation.\(^{90}\) There is thus a special “concern that justice involves for responses to injustice, which is internal to our understanding of its relational and demanding character, and sets it apart from the other branches of morality.\(^{91}\) This means that a full characterization of relations of justice—for example between free and

---

87 For some interesting arguments on the more general question, see Douglas Lavin, “Practical Reason and The Possibility of Error” in *Ethics* (114:3), 424-457.


89 Illuminating discussion of relational justification can be found in Elizabeth Anderson, “The Fundamental Disagreement between Luck Egalitarians and Relational Egalitarians”.

90 This is the master thought of Kant’s approach to justice (right) in the *Doctrine of Right*. It is expressed in his famous argument in §D about hindering a hindrance to freedom. I discuss this argument, situating it in the broader context of his relational view of right, in “Kant on Strict Right”.

91 My thinking here has been shaped by Arthur Ripstein, “Authority and Coercion” in *Philosophy and Public Affairs*, issue 32, No. 1, 1-34.
equal citizens—will require reference to the possibility of wrongdoing and justifiable responses to it.

In order to make this concrete, let us focus on the criminal law and the right to a trial. Granted, on some anarchist views these institutions are unjust, since a state with a monopoly on violence is illegitimate and should be overthrown. But suppose that the criminal law is an important manifestation of justice. In that case, isn’t the equal protection provided to all citizens through the criminal law an aspect of the relation of free equals? And similarly, when we defend the right to a fair trial for the accused, isn’t it plausible that we are describing a standing that the accused has as the free equal of her accuser? Or when we discuss the features of punishment, including the idea of commensurability, or the right of appeal, don’t we describe a fragment of the sort of standing that one convicted of wrongdoing has in virtue of her free equality? If the principles of justice describe the relation of free and equal citizens, these would all seem to be candidates for coverage. But they all concern the principled response to injustice.

It would be natural at this point to object to this string of rhetorical questions on rationally reconstructed Rawls’ behalf that they are addressed in Part II of *A Theory of Justice*, when he begins to relax the rigorous assumptions of Part I. This is true. I am not objecting that Rawls doesn’t speak well about these topics, when he does speak about them. The question I am raising is why these topics are ruled out of court for discussion in Part I. The whole argument for excluding them is that they presuppose deviations from the principles of justice and so do not express the relations of free equals. My point is that this argument seems to rest on a strange understanding of justice, and so a strange understanding of the relation Rawls is seeking to characterize. My argument, bolstered by appeal to these examples, is that the response to injustice appears itself to be an essential aspect of the
relation of free and equal citizens. For this reason, strict compliance seems like a misguided
idealization to capture the distinctive relation of free equals, and so a misguided idealization
for the purpose of deriving the principles of justice that are the most defensible norms for
this relationship. It rules out the topic of standing and principled responses to injustice,
when these seem central to understanding the relation in question.  

I won’t pursue this argument further here. If my objection is fair, then Rawls should
not have employed the idealization of strict compliance even for the purposes of defending
the principles of justice along constructivist lines. In a conciliatory spirit, I should add that
this is a limited point, since, on my rational reconstruction, the assumption plays a role only
in the first phase of ideal theory. Furthermore, as far as I have argued, the assumption of
strict compliance functions mainly to exclude certain topics from consideration that perhaps
should be considered, rather than distorting the topics he does discuss. Although this is a
bigger topic that I cannot address here, I do not think that dropping the assumption of strict
compliance far enough to include topics that involve the protection from, and redress of,
wrongs would alter the reasoning of representatives in the original position for, say, the
Difference Principle, or equal liberty. Thus, nothing I have said suggests (much less entails)
that the idealization ruins the arguments that Rawls actually gives in defense of his principles.

---

92 It could be objected that my examples involve concrete institutions—courts and the like. I agree
that the principles of justice should be separate from their institutional realization, and so affirm
the kind of division Rawls draws between Parts I and II. But note that Rawls thinks the principle of
equal liberty protects many proto-institutional rights, including free speech and democratic
participation. The question is why he does not also include a discussion of proto-institutional rights
that pertain to the prevention and redress of injustices that would arise even in a just society. Note
further that Rawls identifies the liberties with constitutionally protected rights. The US Constitution
involves many amendments dealing with rights of the accused and convicted, e.g. the Fourth, Fifth,
Sixth and Eighth Amendments.

93 Many critics have argued that the idealization of strict compliance excludes the topics pertaining to
racial, and other group-based, subordination and discrimination. See, for example, Charles Mills,
“Ideal Theory as Ideology” and also “White Time: The Chronic Injustices of Ideal Theory” in The
But it does suggest that the liberty principle should be construed more broadly, and that there are topics that belong in the reasoning of the party that Rawls does not discuss.

§7: Conclusion

What has been at stake in the arguments I have pursued here? One thing is the interpretation and evaluation of some of John Rawls’ arguments about the structure of the theory of justice. I have isolated two different conceptions of the ideal/nonideal theory dyad. I have argued that it is confused to treat them as two ways of marking the same distinction, as Rawls is prone to do. And I have proposed a rational reconstruction of Rawls that finds a place for each without sliding into confusion. I hope that this work of interpretation contributes to our understanding of Rawls’ rich thought.

But my discussion has also been intended as a contribution to our current systematic thinking about political philosophy. Partly owing to his centrality to the discipline, the confusion in Rawls that I chart in §4 has put up unnecessary obstacles to understanding the rationale for dividing the theory of justice into ideal and nonideal components. In particular, it has made it difficult to appreciate the force, character, and rationale of the realistic utopian/transitional understanding of the theory of justice. Since this conception is an interpretation of the practicality of political philosophy, while the strict compliance/partial compliance conception is not, the conflation has obscured the way in which the division of political philosophy into ideal and nonideal theory can be viewed an expression of the practical character of political philosophy. It has thus contributed unnecessarily to the impression that the project of ideal theory is a possibly misguided and fussy theoretical exercise in idealization.
In addition, strict compliance has attracted the sorts of criticisms I quote in §2. I have made common cause with the critics by contributing my own distinct objections to the use of strict compliance in §6. However, my common cause is limited by the fact that I accept the division between ideal and nonideal theory understood along realistic utopian/transitional lines. Although it is not the only source of their (diverse) objections, I believe that the conflation of the two conceptions has been one major factor leading critics to throw out realistic utopianism on the basis of their perhaps sound objections to the strict compliance conception. Similarly, these critics have rejected the priority of ideal over nonideal theory partly on the grounds that they reject the usefulness of idealizations like strict compliance as guides for the pressing questions of ordinary political life. Although the most plausible construal of the relevant priority is a very open question, this line of criticism fails to engage with the distinct motivation for the priority claim even once we have rejected strict compliance.94

Substantive political questions also depend on the topics I have discussed. If my criticisms of strict compliance are successful, then the prevention and redress of injustice are a topic for the fundamental principles of justice. Furthermore, many of the great questions of political philosophy are, I think, bound up with the idea of realistic utopia. Clearly isolating and explicating this idea is crucial for understanding these debates. For example, the question of Political Liberalism is most naturally interpreted as the question whether a pluralistic society with an overlapping consensus on principles of justice is a realistic utopia. I believe that something similar could be said about the question of socialism, theories of deliberative democracy, anarchism, and much else.

94 I do not mean to say that no critics have engaged with the realistic utopian/transitional rationale for the priority of ideal theory. For example, Elizabeth Anderson engages directly with this rationale. I discuss her arguments, which rely on a genetic epistemological interpretation of the priority claim, in “The Priority of Ideal Theory”.