One of the many distinctions introduced in John Rawls's monumental work, *A Theory of Justice*, was that between (what he called) the *ideal* and the *nonideal* theories of justice. Like many of Rawls's dichotomies, this one has come to frame much subsequent discussion in moral and political philosophy, and use of the Rawlsian terminology has become perfectly commonplace. Unlike many of Rawls's other distinctions, however, relatively little serious critical attention has actually been focused on the distinction itself—and where the ideal-nonideal distinction has received such attention, I will argue, this attention has been largely either superficial or misguided. Within the world of "Rawls scholarship," the distinction seems to be treated as straightforward and uncontroversial. Neither of the two largest recent books about Rawls (weighing in, respectively, at 585 and 550 pages),¹ for instance, contains more than a passing reference to the ideal-nonideal distinction, nor do either of the two latest large-scale contributions to the vast literature on Rawls and his theories.² Where the divide between ideal and nonideal theory is discussed at all in that world, Rawls's distinction is usually characterized as "clear"³ or

For their helpful suggestions concerning earlier drafts of this article, I would like to thank Jody Kraus, John Arras, Simon May, an audience at the University of Toronto Law School, and the Editors of *Philosophy & Public Affairs*.


© 2010 Wiley Periodicals, Inc. *Philosophy & Public Affairs* 38, no. 1
“unproblematic.” 4 Indeed, even the two recent philosophical books with “nonideal theory” in their titles (whose authors clearly take themselves to be engaged in nonideal theorizing) contain no substantial discussion of the distinction. 5

This lack of attention is surprising. Rawls’s account of the ideal-nonideal distinction, of course, constitutes his proposed solution to the venerable problem of characterizing the relationship between philosophical theory and political practice. Given the importance of that problem and the apparent widespread acceptance of Rawls’s approach to it, one might reasonably have expected his model of ideal and nonideal theory to have received some sustained attention. Instead, those who have adopted the Rawlsian language of ideal and nonideal theory seem mostly to have taken Rawls’s version of the distinction to be obviously correct. And those who criticize the Rawlsian approach (in the remarkably thin literature that actually addresses the distinction) seem mostly to dismiss it as flawed in relatively simpleminded ways. In neither case has much effort been taken to be clear about exactly what Rawls’s position amounts to and what the alternatives to it might be.

My aims in this article are three. First, in Sections I through III, I will lay out a “rational reconstruction” of Rawls’s position on the relationship between and the content of ideal and nonideal theory. I hope that the reading I offer will be more thorough and more helpful than is common in the existing literature on the ideal-nonideal distinction. Second, in Section IV, I will examine some possible alternative approaches to the ideal-nonideal distinction and suggest some reasons for favoring the Rawlsian approach. Third, in Section V, I argue that the few recent criticisms of the Rawlsian position are in fact either unpersuasive or based on confusions about Rawls’s understanding of nonideal theory. I conclude that it is not Rawls’s employment of the distinction itself or the version of it that he employs that should be the target of those who find unconvincing Rawls’s (or Rawlsian) political philosophy.

I. RAWLS’S IDEAL THEORY

In its barest outline, of course, the idea behind the distinction is simple and appealing, no doubt accounting for the widespread acceptance of (or indifference toward) Rawls’s version of it: “the intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances.”6 This “ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception” (TJ, p. 246). “Nonideal theory asks how this long-term goal might be achieved, or worked toward, usually in gradual steps. It looks for courses of action that are morally permissible and politically possible as well as likely to be effective.”7

Political philosophy should begin, Rawls tells us, with the defense of those (moral) principles that ought to constrain the design and operation of the basic structure of each society.8 When the basic institutions fully conform to the requirements of these principles, the society’s institutional framework is then perfectly just. Rawls understands this ideal theory of justice as giving an account of what he comes to call a “realistic utopia,” that is, the best we can realistically hope for, “taking men as they are and laws as they might be” (in Rousseau’s words; LOP, p. 7). We ask what could come into existence as a result of our choices, given the limits set by our moral and psychological natures and by facts about social institutions and how humans can live under them. Ideal theory, then, “probes the limits of practicable political possibility” (JAF, pp. 4, 13). We will undoubtedly “have to rely on conjecture and speculation” in our determination of what is in fact “practically possible” (LOP, p. 12); and

8. The basic structure of a society is its “major social institutions,” i.e., its “political constitution and the principal economic and social arrangements” (TJ, p. 7). As examples of such “institutions,” Rawls mentions basic constitutional liberties, competitive markets, “private property in the means of production,” and the monogamous family (ibid.). Rawls later characterizes the basic structure as “the background social framework within which the activities of associations and individuals take place” (Justice as Fairness: A Restatement, ed. Erin Kelly [Cambridge, Mass.: Harvard University Press, 2001] [hereafter JAF], p. 10).
what counts as practically possible may be in certain respects historically relative. But our 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 (that is, acts justly, etcetera, according to the principles of justice being defended or considered). Although some critics attack this assumption of strict compliance as dammingly unrealistic—given the obvious fact in real societies of widespread unwillingness to behave justly—the point of the strict compliance requirement for ideal theory in fact seems fairly plain.

First, of course, 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 out to be quite strictly a comparison only of the principles of justice. If instead we try to evaluate principles in terms of how societies governed by them would operate with a “normal” amount (or a certain percentage) of noncompliance with them (supposing we can even make sense of that hypothesis), we will likely find both that our evaluations yield quite indeterminate results and that the results

9. Rawls at one point suggests that a realistic utopia concerns the “practically possible here and now,” meaning what is practically possible in the “historical conditions of our age.” The conditions of our specific age include “reasonable pluralism” (within liberal democracies) and diversity of culture and tradition between nations (LOP, pp. 11–12), although neither of these “conditions” seems exactly to concern either human nature or the nature of political and legal institutions. This, of course, reflects Rawls’s concern that his account of social stability in A Theory of Justice failed to take adequate note of the ineliminable pluralism of our age—pluralism both within liberal democratic societies and between the nations of the world (Political Liberalism [New York: Columbia University Press, 1993] [hereafter PL], pp. xv–xviii). And an institutional scheme that is unstable is not “practically possible” (in Rawls’s sense of that idea). As a result, Rawls’s conception of domestic justice is recast in Political Liberalism as a conception that could be accepted by proponents of all of the reasonable “comprehensive conceptions of the good” that are likely to flourish in a liberal society, while his “law of peoples” is similarly fashioned in order to be acceptable to all reasonable nations (or peoples) in the world. Oddly, however, Rawls appears to suggest as well (in the middle of The Law of Peoples) that although “the specific conditions of our world at any time” affect the appropriate conclusions of nonideal theory, they should not influence ideal theory (LOP, p. 90).
depend on more than simply the different ordering effects of the principles being compared. Exactly what form does the assumed noncompliance take, for instance, and how does that form of noncompliance itself influence the behavior and the expectations (etcetera) of others? What will we understand to be the “normal” forms of noncompliance, given that the behaviors or the institutional structures that count as noncompliant with different, competing principles of justice must be themselves quite different?

Further, the assumption of strict compliance should not be understood to amount to an assumption that all possible principles for ordering society’s basic structure are equally good at motivating compliance. We must, of course, assume that persons (and their representatives in the original position) are “capable of a sense of justice,” so that they “can rely on each other to understand and to act in accordance with whatever principles are finally agreed to” (*TJ*, p. 145). But institutions that embody different conceptions of justice will predictably affect differently the degrees of support (and, ultimately, compliance) that those institutions can expect from their subject populations. And a conception of justice that “is unlikely to generate its own support, or lacks stability” (ibid.) will for that reason be unacceptable—and will certainly be inferior to competing conceptions whose requirements on social institutions can, in light of the basic principles of moral psychology, predictably generate widespread and multigenerational support for and compliance with those conceptions (*TJ*, pp. 496–99).

So the strict compliance assumption is designed only to allow us to imagine the results of getting “up and running” the institutions embodying different conceptions of justice, which requires imagining that those subject to those institutions support and comply with them, at least initially. But it may turn out that some conceptions—such as utilitarianism—will predictably involve “strains of commitment” that “exceed the capacity of human nature” (*TJ*, p. 176), such that long-term support and compliance are unlikely or impossible. The strict compliance assumption, then, “still permits the consideration of men’s capacity to act on the various conceptions of justice” (*TJ*, p. 145).10

---

10. For further discussion of the point of the strict compliance assumption, see Zofia Stemplowska, “What’s Ideal about Ideal Theory?” *Social Theory and Practice* 34 (2008): 331–40.
It is the ideal theory of justice, understood in this way, on which Rawls focused his principal attention in his writings. *A Theory of Justice* does contain (in chapter VI) the sketch of a nonideal theory of civil disobedience and conscientious refusal, and *The Law of Peoples* lays out the form of a nonideal theory of international duties. But these brief forays into nonideal theory are plainly seen by Rawls as secondary to what should be the political philosopher’s first concern. Anticipating the many later complaints that it is instead nonideal theory that should be our primary focus, Rawls responds that he first defends his ideal theory of justice in order “to formulate a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies . . .”; “it is natural to conjecture that once we have a sound theory for this case, the remaining problems of justice will prove more tractable in the light of it” (*TJ*, p. 8; see also *TJ*, p. 391). So this “ideal theory” of “the nature and aims of a perfectly just society is the fundamental part of the theory of justice” (*TJ*, p. 9), with nonideal theory being secondary and dependent upon it. Rawls later says that this conjecture, “that it is ideal theory that is fundamental,” has been “confirmed” by the book’s main arguments (*TJ*, p. 241).

The specific content of Rawls’s ideal theory of justice is, of course, given by the results of—just as the reasoning supporting that theory is modeled in—the original position choice problem, in which rational, mutually disinterested representatives of the members of the society are required to select, under a veil of ignorance, principles of justice to guide the institutional structure of their social lives together. Once Rawls turned his mind seriously to questions of international justice, however, it was necessary for him to characterize his ideal theory as having “two parts” (*LOP*, pp. 4–5). The first part is an account of justice appropriate to liberal democratic societies, that is, principles for a perfectly just regime that can “come about and be made stable under the circumstances of justice” (*JAF*, p. 13). The second part of ideal theory extends the same ideas to the international community, deriving principles for a “society of peoples”—using a similar original position model, with the contractors now representing societies or peoples11—that can bring about and make stable (that can “well-order”) a just international “society.”

11. Rawls first suggests the idea of an international original position, so conceived, at *TJ*, p. 378. The idea is fully fleshed out in *The Law of Peoples*. 
 Needless to say, these two parts of ideal theory will have two corresponding nonideal theories, one dealing with failures of domestic institutions to live up to the requirements of domestic ideal theory, the other dealing with failures of societies to live up to the requirements of international ideal theory. There is, however, a third part of ideal theory, plainly requiring a corresponding third kind of nonideal theory, that Rawls does not mention in his division of ideal theory into “two parts” (although this third part got equal billing in Rawls’s earlier division of “the concept of right” into three parts \(TJ\), p. 109)). That third part is the set of “principles for individuals,” derived separately from both the principles for just domestic institutions and those for peoples in a just international order.  

The principles for individuals mentioned in *A Theory of Justice* are the “principle of fairness,” which “[accounts] for all requirements that are obligations” (that is, requirements that are voluntarily assumed and that arise from special relationships \(TJ\), p. 111), and the various principles of “natural duty,” including the duties “not to injure,” “not to harm the innocent,” to give “mutual aid” and “mutual respect” \(TJ\), p. 109), and to “support and comply with just institutions that exist and apply to us” \(TJ\), p. 115). Precisely which of these principles should be taken to survive Rawls’s transition to a purely political conception of justice is less clear. However it is filled out, though, nonideal theory for

12. By the time Rawls wrote *Justice as Fairness*, he appears to have conceived of the three parts of ideal theory differently, identifying “three levels of justice”: “first, local justice (principles applying directly to institutions and associations); second, domestic justice (principles applying to the basic structure of society); and finally, global justice (principles applying to international law)” \(JAF\, p. 11\). Domestic justice has priority over “local justice” in the sense that the former’s principles “constrain (or limit), but do not determine uniquely” the latter’s \(JAF\, pp. 11–12\). Since the principles governing the ways in which individuals may associate locally (and the subsequent duties imposed on individuals by the rules of the voluntary associations they create) are only some of the moral constraints suitable to individuals in societies, I take the principles of “local justice” to be just one subpart of the more general category of principles for individuals. I suspect that Rawls’s reluctance to continue directly discussing the more general principles for individuals stemmed from his concern that such language would encourage the (by this point rejected) idea that political philosophy was just one application of a more general, comprehensive moral theory, one that included the principles governing interpersonal life. But even after Rawls’s insistence on the autonomy of political philosophy, he was, of course, still free to articulate a set of ideal principles for individuals. The only requirement was that these principles be conceived of as principles governing the behavior of citizens of just societies (rather than the behavior of persons in all settings), principles deriving from the needs of the just institutions defined by his autonomous political philosophy.
individuals—as opposed to that for domestic institutions or for whole societies in the international order—must presumably explain how we should deal with failures of individual persons to discharge the obligations and duties defined by these ideal principles for individuals, all with an eye to bringing about full compliance with those principles.

II. THE STRUCTURE OF RAWLS’S NONIDEAL THEORY

As we have seen, nonideal theory (or what Rawls at first also calls “partial compliance theory”) “studies the principles that govern how we are to deal with injustice” (*TJ*, p. 8). Where ideal theory dictates the objective, nonideal theory dictates the route to that objective (from whatever imperfectly just condition a society happens to occupy). We have, Rawls says, a natural duty (of justice) “to remove any injustices, beginning with the most grievous as identified by the extent of the deviation from perfect justice.” While “the measure of departures from the ideal is left importantly to intuition,” “the lexical ranking of the principles [of the ideal theory] specifies which elements of the ideal are relatively more urgent” (*TJ*, p. 246). So—to focus first on domestic nonideal theory for the basic structure of society—violations of Rawls’s First Principle of justice (which requires extensive and equal basic liberties) are relatively more “grievous” than violations of part (b) of the Second Principle (the “equal opportunity principle”), which are in turn more grievous than violations of the Difference Principle (part [a] of the Second Principle). The “idea of a well-ordered society” thus can “help to clarify the goal of reform and to identify which wrongs are more grievous and hence more urgent to correct” (*JAF*, p. 13). The “special conception of (ideal) justice” (in which we have two—or three—principles in “lexical order”) provides nonideal theory with at least a very general priority rule in determining the required manner of addressing injustice.

Unfortunately, Rawls’s only remarks (in *A Theory of Justice*) on the overall structure of nonideal theory are more confusing than helpful. Nonideal theory, he tells us, has “two rather different subparts. One consists of the principles for governing adjustments to natural limitations and historical contingencies, and the other of principles for meeting injustice” (*TJ*, p. 246). The second of these “subparts” he at first calls “the partial compliance part of nonideal theory” (*TJ*, p. 247), later calling it simply “noncompliance theory” (e.g., *LOP*, p. 90). But since the
whole of nonideal theory was originally described in terms of principles for dealing with injustice—and thus with “partial compliance” (or full noncompliance) with the principles of ideal theory—it is surprising to here find Rawls characterizing only the second part of nonideal theory in terms of remedying injustice and responding to partial compliance.

There appear to be two distinct ideas at work when Rawls discusses the subject of the puzzling first part of nonideal theory: “natural limitations and historical contingencies.” On the one hand, Rawls discusses “reasonable restrictions” of liberty that “derive from the natural limitations and accidents of human life, or from historical and social contingencies,” where he has in mind cases like the reasonable regulation of liberty of thought and conscience and “the lesser liberty of children” (TJ, p. 244). Such restrictions, of course, involve no obvious injustice at all (“the question of the justice of these constraints does not arise” [ibid.]). What is hard to see, of course, is why such cases require any special principles or any version (subpart) of nonideal theory for guidance. The restrictions Rawls mentions appear to involve no more than a straightforward derivation from ideal principles of institutional arrangements that are sensitive to basic, uncontroversial facts about, e.g., human psychology, as, presumably, all just institutions must be. Similarly, the “principles of paternalism” that Rawls goes on to discuss a few pages later appear to be no more than obvious qualifications of the ideal principles that would be accepted in the first place by the original position contractors. Indeed, Rawls himself at one point appears to suggest precisely that in the case of making “adjustments to natural limitations” (as opposed to the case in which “we are confronted with injustice”), it is “the ideal conception of justice” that “applies” (TJ, p. 351).

There is, however, a quite different kind of “natural limitation” or “historical contingency” that Rawls also mentions, although this reference is not so prominently placed in his discussion of nonideal theory. In explaining the lexical ordering of his ideal principles (and the priority of liberty that this ordering entails), Rawls discusses the case where “social conditions do not allow the effective establishment of [the] rights” required by the ideal principles in lexical order. In this kind of

13. “The principles of paternalism are those that the parties would acknowledge in the original position to protect themselves against the weakness and infirmities of their reason and will in society” (TJ, p. 249).
case—Rawls appears to be thinking primarily, at least, of cases involving severe societal poverty—"one can concede their [i.e., the rights'] limitation," but only where this limitation is "necessary to prepare the way for a free society," by "raising the level of civilization" (IJ, p. 152). In this case, Rawls suggests, society should be governed primarily\(^\text{14}\) by the general conception of justice (IJ, p. 62), which lacks the separate, lexically ordered principles of ideal theory, until it becomes reasonable, as a result of "improvement in economic well-being," to be guided by the two separate principles of ideal theory in their serial order (ibid.). The general conception dictates that over the long run lesser liberties for all must benefit all, while unequal liberties must benefit those with lesser liberty.\(^\text{15}\) We can think of the general conception of justice as a kind of intermediate conception, lying between the full, lexically ordered ideal principles of "justice as fairness" and the more narrowly pragmatic, transitional principles that govern the second, noncompliance part of nonideal theory.

What is striking about this second notion of "natural limitations and historical contingencies," of course, is that it does seem to involve injustice (in the less extensive liberty mandated by the society), hence it involves only "partial compliance" with ideal principles; and it does appear to be the kind of case that calls for some special, nonideal principles of guidance: in this case, perhaps as Rawls suggests, the guidance of the more general principle of justice of which Rawls’s ideal theory is a special case. What this second kind of case may not involve is deliberate or easily avoidable noncompliance and injustice. This second kind of

\(^{14}\) I say "primarily" here, because the general conception, by itself, does not identify any departures from the ideal as more or less "grievous" than any other. And, as we have seen, the "grievousness" of injustices, as measured by the lexical order of the principle the injustice violates, is to serve as a priority rule in nonideal theory (see below). So Rawls writes, "When we come to nonideal theory we do not fall back straightway upon the general conception of justice. The lexical ordering of the two principles . . . suggest[s] priority rules . . ." (IJJ, p. 303).

\(^{15}\) In Colin Farrelly’s recent attack on Rawlsian ideal theory, he (oddly) supposes that its inclusion of "the absolute priority of liberty" indicates that Rawls "presumes rights are costless" and ignores "the realities of non-compliance and scarcity" ("Justice in Ideal Theory," Political Studies 55 [2007]: 844–64, at p. 853). Rawls’s ideal theory plainly involves presuming only that the “cost of rights” is significantly less than their value in decent economic circumstances, and that in less happy circumstances where this is not the case, principles of nonideal theory are to be followed instead of the lexically ordered principles of ideal theory.
“contingency” thus makes considerably more sense as the subject of a special “subpart” of nonideal theory than does the (more prominently located) first notion of contingency. And we can in consequence imagine a nonideal theory that distinguishes between two different kinds of noncompliance and injustice: those (in the first part) that are, due to societal misfortune or special history, largely inevitable and nondeliberate; and those injustices (in the second part) that are more clearly avoidable or deliberate.

Dividing Rawlsian nonideal theory in this fashion makes even more sense in light of Rawls’s own later division of international nonideal theory along precisely these lines. The Law of Peoples, like A Theory of Justice, also divides nonideal theory into two parts, described again as the parts concerned, respectively, with “unfavorable conditions” and with “noncompliance.” In international nonideal theory, however, the “cases” covered by the two parts are two different kinds of societies, and nonideal theory now dictates how other, liberal and “decent” societies are required and permitted to act toward them. The “unfavorable conditions” cases are what Rawls calls “burdened societies,” where “historical, social, and economic circumstances make their achieving a well-ordered regime . . . difficult if not impossible” (LOP, p. 5). Burdened societies “lack the political and cultural traditions, the human capital and know-how, and, often, the natural and technological resources needed to be well-ordered” (LOP, p. 106). The “noncompliance” cases are the “outlaw states” that simply refuse to comply with the requirements of justice, either internally (in violating the human rights of their own subjects) or externally (in aggressing against their neighbors) (LOP, p. 90, note 1). So here we have the structure for nonideal theory suggested above, with its parts devoted, respectively, to unfortunate and deliberate noncompliance.

What, though, are we to make of Rawls’s insistence that a part of nonideal theory concerns cases that involve no injustice or “partial compliance” at all? What seems to unite all of the cases we’ve considered as candidates to be covered by Rawls’s “first subpart” of nonideal theory is this: in each case restrictions on liberty are acceptable only if they are part of a process aimed at achieving the fullest possible societal justice. Rawls appears to be claiming that in such cases there is no actual injustice (or partial compliance), despite the fact that the ideal principles are not fully complied with. Thus, it is not unjust to give lesser liberty to
children, the severely retarded, or the insane if this is necessary for and aimed at their overall good. Impoverished societies may justly restrict liberty or make liberties unequal if necessary to and aimed at the economic prosperity or stability required for just institutions. And it is not unjust to regulate liberty of thought, conscience, or participation in order to preserve public order or guarantee that these very liberties won’t be voted away. The salient feature of such cases is not so much the fact that noncompliance with ideal principles can be traced to “natural limitations” or “historical contingencies.” Rather, the idea seems to be that temporary institutional “adjustments” that are necessary to respond to misfortune and to eventually bring about full compliance—and that are motivated by and aimed at that end—do not really count as injustices or as cases of partial compliance.

We can, then, think of Rawls’s first part of nonideal theory as intended to cover a substantial variety of cases: those involving societies that, due to (say) unfortunate cultural or economic histories, have never been able to achieve just institutions; those involving societies that, due to new “contingencies” or crises, must temporarily “adjust” their just institutions in order to preserve them for the long term; and those involving unfortunate individuals (within otherwise just societies) whose liberties must be curtailed for their own good. All seem to involve departures from ideal principles that are either less blameworthy than avoidable, deliberate injustice, or that are not blameworthy at all. As such, they require, Rawls believes, different treatment than the cases (in his second part of nonideal theory) involving deliberate or avoidable noncompliance with ideal principles of justice.

If all of this is correct, then we would do better to think of Rawls’s nonideal theory as having six parts, rather than two. Each of the three parts of ideal theory will have a corresponding nonideal theory, with each of these branches of nonideal theory in turn divided into two subparts, one dealing with deliberate noncompliance and the other with unfortunate noncompliance. So nonideal theory for the basic structure will deal both with deliberate or avoidable institutional injustice (including unjust conduct by public officials in their application of institutional rules) and with unfortunate noncompliance (due to societal poverty or cultural obstacles to justice), recommending civil disobedience and conscientious refusal (and, in more extreme cases, violent resistance) in the former case and temporary institutional adjustments guided by the
general conception of justice in the latter. In both cases, the object of the required nonideal policies or actions is eventual full compliance with ideal principles for the basic structure. Similarly, in the case of nonideal theory for individuals, deliberate wrongs will require reparation and punishment (aimed at correcting injustice and at disabling and deterring criminal activity), while the principle of paternalism will govern those cases of unfortunate inability to comply (due to, e.g., immaturity or insanity). As before, the ultimate goal of the policies guided by these nonideal principles is full compliance with the ideal principles for individuals. Finally, in the case of nonideal theory for nations, deliberate noncompliance may require international intervention or just war aimed at bringing about compliance, while unfortunate noncompliance may require international assistance in order to allow noncompliant nations to become well ordered.

The chart below lays out this way of organizing the various kinds of noncompliance discussed by Rawls and his suggestions concerning the subject matter of the corresponding parts of nonideal theory:

<table>
<thead>
<tr>
<th>Ideal Theory</th>
<th>Nonideal Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Basic structures</td>
<td>a. deliberate noncompliance (institutional injustice): civil disobedience, etcetera.</td>
</tr>
<tr>
<td></td>
<td>b. unfortunate noncompliance (poverty, culture): the general conception</td>
</tr>
<tr>
<td>II. Individuals</td>
<td>a. deliberate noncompliance (wrongs, crimes): punishment/reparation</td>
</tr>
<tr>
<td></td>
<td>b. unfortunate noncompliance (insanity, immaturity): paternalism</td>
</tr>
<tr>
<td>III. Nations</td>
<td>a. deliberate noncompliance (outlaw states): just war, intervention</td>
</tr>
<tr>
<td></td>
<td>b. unfortunate noncompliance (burdened societies): international assistance</td>
</tr>
</tbody>
</table>

There is a final complication, never directly addressed by Rawls. To precisely which agents, we might ask, are the various rules of nonideal

16. This organizational scheme seems not entirely adequate to characterizing noncompliance that is neither deliberate nor merely unfortunate—such as (individual) negligent injury of others or the institutional or national analogues of negligent noncompliance. It might, then, be better to divide noncompliance into the “blameworthy” and the “non-blameworthy.”
theory addressed? Rawls’s comments on nonideal theory might be taken to suggest that the classes of agents who are directed by nonideal theory to respond in certain ways to unjust noncompliance divide as neatly as do the agents responsible for the injustice. Thus, we might take Rawls to be suggesting that where domestic institutions (or the officials occupying institutional positions) are unjust, the required response is from the citizens of that society. Where private individuals act unjustly, it is their domestic institutions that should respond according to the applicable nonideal principles. Where nations are unjust, it is other nations that should respond. But that picture is plainly too simple and just as plainly cannot be what Rawls intends. For some kinds of domestic injustice—for instance, serious human rights violations—seem to call for action both by the citizens of the offending society (in the form of disobedience, say) and by other nations (in the form of intervention, say). Indeed, the non-ideal theory governing such a case would presumably include as well special rules for the conduct of public officials in the unjust society, and possibly even rules for the conduct of foreign private citizens. So while we have identified six kinds of noncompliance and six corresponding branches of nonideal theory, each of those branches may contain complex sets of principles separately governing the actions of a variety of different classes of agents, including private citizens (and their voluntary associations), public officials, and nations.

III. THE CONTENT OF RAWLS’S NONIDEAL THEORY

Nonideal theory requires us to work toward achieving the perfectly just basic structure that would result from strict compliance with the ideal theory of justice. The specific “policies and courses of action” it mandates must be (i) “morally permissible,” (ii) “politically possible,” and (iii) “likely to be effective” in moving society toward the ideal of perfect justice (LOP, p. 89). Rawls explains neither how these three requirements are to be weighed against one another (e.g., is a politically popular but likely less effective policy to be preferred to one that has questionable political support but is more likely to be effective?) nor whether nonideal theory requires the optimal policy (i.e., the one that best satisfies these three requirements together, however that might be calculated) over others that merely satisfy each of the three requirements. Beyond those worries, though, something else is plainly missing from Rawls’s list of
three nonideal requirements on policies. As we’ve seen already, we also need a fourth requirement: namely, that relatively more “grievous” injustices are to be dealt with before less severe ones.\footnote{17} Again, however, it is unclear whether judgments of “grievousness” are to trump concerns about, say, political possibility, leaving us obligated to ignore lesser injustices until more serious ones have been successfully remedied, or whether such judgments of grievousness are simply another factor to be weighed in determining the required course of action.

Although much is obviously left vague here, we can at least infer this much from Rawls’s remarks: nonideal theory will require judgments of both philosophical and social-scientific sorts. Determinations of a policy’s “moral permissibility” obviously lie in the proper domain of moral and political philosophy, as do judgments of grievousness, which depend on prior ideal theorizing. Determinations of “political possibility” and “likely effectiveness,” on the other hand, seem more naturally to require the expertise of, e.g., political scientists, economists, and psychologists. Further, since the general principles of nonideal theory are intended to have implications for the policies of particular societies, all three kinds of determinations will require reasonably specialized knowledge of the structure and workings of those particular societies.

It seems correct, of course, that, other things equal, policies are better that are more likely candidates for gaining widespread acceptance and for actually bringing about the required institutional changes. When other things are \textit{not} equal, as when the better policy in terms of likely political success is also weaker than alternatives in terms of likely effectiveness, we are left to simply weigh intuitively the various desiderata as best we can (with such weighing \textit{also} having to include consideration of the relative grievousness of the injustices to be addressed and any relative moral permissibility of the policies). Since the four requirements of nonideal theory (may) all admit of degrees of satisfaction, the conclusions of ideal theory, applied to particular injustices in particular societies, are likely to be somewhat speculative (and certainly nothing like simple deductions from those requirements conjoined with societal data).

\footnote{17}{The lexical ranking of the principle of justice that is violated by an injustice is to determine the relative seriousness of the injustice and the relative urgency of our response, thus setting a priority rule (\textit{TJ}, p. 303) to function along with the three other requirements for policies in nonideal conditions.}
There is at least a plausible case to be made for the view that we really should not expect anything less sloppy than this in our quest for moral guidance in the nonideal realm. After all, short of successfully defending some one-value consequentialist moral theory (of the sort that utilitarianism aspires to be), we have been given no reason yet to expect conclusive moral guidance even in individual or small-scale “moral dilemmas”; even less, it seems, should we expect conclusive guidance in the case of large-scale social injustices, especially where the various desiderata for remedial policies point us in different directions. Perhaps it is enough to be able simply to articulate the various considerations—or “moral vectors”—that must be taken seriously in any reasonable defense of policy in nonideal circumstances.

Rawls does not expand on the simple claim that policies in nonideal conditions must be morally permissible, so we are left to guess about his meaning. Perhaps he has in mind policies condoning acts by public officials (in their pursuit of injustice-remediying policies) that would be plainly impermissible, such as murdering the most stubborn opponents of the policies. Or perhaps the “priority rule” for nonideal theory that was just discussed suggests that it would be morally impermissible to remedy any less grievous injustice by instituting an injustice that is more grievous (according to the lexical order of the principles violated). Or, since individual citizens are not bound to “further just arrangements” in ways that involve “too much cost” for themselves (TJ, pp. 115, 334), perhaps Rawls was thinking that morally permissible policies must not impose excessive or unreasonable costs on individuals.

Such cost restrictions suggest a more general kind of “moral impermissibility”: namely, morally impermissible kinds of transitional unfairness. In changing even unjust institutional rules, we are generally sensitive to the moral dangers of “rug-pulling” (famously identified by Sidgwick in his discussion of “conservative justice”),18 that is, cases where people base life plans or important activities on the reasonable expectation that the rules will remain unchanged (or, worse, where they are simply required to structure their activities around existing rules) and then have the rug pulled from beneath them by sudden institutional

change. The loss and suffering that such changes in institutional rules can bring about may often both be considerable and seem unfair to those who had little choice but to rely on the future being like the past. This fact seems to many to argue for a requirement that institutional changes proceed gradually, with ample prior warning, or that such changes be accompanied by policies providing compensation for those who have innocently relied (to their detriment) on even unjust rules. And perhaps such requirements are a part of what Rawls has in mind in insisting that policies in nonideal conditions be morally permissible.

As we have seen, nonideal theory requires that we pursue policies that are not only morally permissible, but also politically possible and likely to be effective. But there are, of course, (at least) two ways in which we might assess a policy's political possibility and likely effectiveness. We might assess the policy, first, as it relates to the elimination of the specific injustice addressed by the policy; or we might assess it, second, as it relates to the overall achievement of perfect (ideal) justice in the society, that is, relative to the elimination of all institutional injustice. These two approaches to the requirements could presumably yield quite different results. A particular policy might, for instance, be a good bet for remedying a particular injustice (or kind of injustice), while at the same time being a policy that retarded, stalled, or set back efforts to achieve overall justice. Or a policy that would completely eliminate some injustice at a stroke might in fact be a worse policy than some more modest one, if we focus not on the particular injustice in question but instead on the goal of eventually achieving perfect justice. Compare, for instance, the small, gradual legal steps that might be necessary to actually one day achieve the ideal of racial equality in a racist society with much more dramatic possible legal steps—steps that might at once make our institutions more closely approximate those of an ideally just society, but that would be likely to precipitate a conservative backlash that would leave us stranded forever or for a much longer period of time in a deeply unjust society.

Because the object of Rawls's nonideal theory is the eventual achievement of the ideal of perfect justice, not simply the elimination of particular or salient injustices, I take nonideal theory's requirements of political possibility and likely effectiveness to be best understood in our second way, that is, as requirements that policies be politically possible and likely to be effective as parts of a strategy for the complete
elimination of all societal injustices. This means that we must understand Rawlsian nonideal theory as both strongly transitional (as opposed to simply comparative) in character and as offering us an integrated, not a piecemeal, goal as our target (for assessing policies’ possibility and effectiveness). A good policy in nonideal theory is good only as transitionally just\(^\text{19}\)—that is, only as a morally permissible part of a feasible overall program to achieve perfect justice, as a policy that puts us in an improved position to reach that ultimate goal. And good policies are good not relative to the elimination of any particular, targeted injustices, but only relative to the integrated goal of eliminating all injustice. Although the policy of completely eliminating some particular targeted injustice may often be the best way to approach the elimination of all injustice, especially where the particular injustice is among the basic structure’s most grievous, there is no guarantee that this will be the case.

If this is correct, of course, some familiar claims about nonideal theory may turn out not to be true, at least with respect to Rawls’s version of it. For instance, Ingrid Robeyns has argued both (a) that one of the essential tasks of nonideal theory is “to enable us to make comparisons between different social states and evaluate which one is more just than the other,” a task which is “presupposed” by nonideal theory’s other task of “guiding our actions” toward the ideal; and (b) that it is possible to develop a “partial ideal theory” that concerns justice only in some particular domain (in her case, the domain of “gender justice”).\(^\text{20}\) But (to first consider claim [b]) if the goal for nonideal theory (that is presented to it by ideal theory) is an integrated goal of overall perfect justice, then neither ideal nor nonideal theory can be pursued “partially” or in piecemeal fashion. There is no reason to suppose in advance that justice in one domain is independent of justice in other domains. So ideal theory cannot set “partial” targets until it first determines that hitting those targets will be consistent with all other aspects of overall societal justice, which implies, of course, the need to first determine at least most of the content of that integrated ideal.

\(^{19}\) The term “transitional justice” will thus be used here rather differently than its more common use to refer to the ways in which states address past human rights violations (during their transitions to social stability).

Consider next ideal theory's supposedly essential task of making comparative judgments about the relative justice of various nonideal conditions (claim [a] above). If we understand nonideal theory as genuinely transitional, relative to ideal theory's integrated goal, it simply is not necessary for nonideal theory always to yield such judgments before endorsing actions or policies. There might, for instance, be only one action or policy (P) that is morally permissible, politically possible, and likely to be successful in reaching a social condition (C) from which the ultimate achievement of perfect justice is more likely than it is from the status quo. But if so, then in order to endorse P, nonideal theory need not first determine that C is a “more just” social state than the status quo (or than any other imaginable social state). C might well be less just than the status quo in terms of their relative resemblances to a perfectly just basic structure and still be more just in transitional terms. If it is necessary to take one step backward in order to take two steps forward, Rawlsian nonideal theory will endorse that step “away from” resemblance to the ideal.

What Rawls’s nonideal theory does not appear to do is simply recommend policies that aim us toward the “more just” of two possible social conditions (at least if “more just” means only “more closely resembling a perfectly just institutional scheme”). We are all confident, for instance, that the institution of human slavery is deeply unjust. But, says Rawls, “there may be transition cases where enslavement is better than current practice” [such as where states that had previously put prisoners of war to the sword agree to hold them as slaves instead!(TJ, p. 248) Nonideal theory might recommend in such a case that a nonslaveholding society institute legalized human slavery, as part of the only feasible, transitionally just route to a condition in which the First Principle rights of persons can one day be fully institutionalized. Worse, of course, comparative judgments may simply be very difficult or impossible to make. While we can confidently say that, for Rawls, an institutional arrangement that includes a more grievous injustice is “less just” than an otherwise identical one that includes a less grievous injustice, it is far less obvious how we are to compare arrangements involving different injustices of the same “rank,” or to compare multiple lesser injustices with a single greater injustice (and so on).

Rawls’s position, then, appears to be that nonideal theory consists in transitional principles relative to an integrated ideal. And there is, I
think, a great deal to be said for this position. While our attacks on particular, especially offensive injustices may be for us understandably compelling, few devotees of “partial justice” would be able to sustain their single-minded commitments in the face of clear evidence that their efforts were setting back or permanently blocking movement toward overall societal justice. The temptation to deny this, I think, is based on the same kind of concern as is the inclination to deny that nonideal theory should be strongly transitional: namely, doubts about our ability to reliably predict how current efforts in the cause of justice will actually affect society over the long run. Some may believe that we simply cannot confidently make the social-scientific (or moral?) judgments on which must rest such a transitional theory with an integrated target. Given such uncertainty, why should we not simply seize whatever local and comparative gains we can get?

Rawls’s answer, I think, should be this. Where we are genuinely unable to make the kinds of well-justified empirical (or moral) judgments which Rawlsian nonideal theory requires, nonideal theory simply offers no conclusive guidance to our political practice. Then, presumably, we must simply muddle through the best we can, keeping the target ideal of justice firmly in mind. Then it may seem acceptable to cross our fingers and just accept whatever comparative gains in justice we can get or single-mindedly attack some particular, salient injustice. But it is important to see that committing ourselves to such practices as a general rule would in fact amount to an abandonment of the goal of any systematic theoretical guidance of political practice. Where “comparative gains” or targeted attacks in fact set back the cause of overall social justice, it is hard to see why anyone who is committed to that cause would regard this as nonetheless a positive development.\footnote{As an Editor of Philosophy \& Public Affairs pointed out to me, one might also object to Rawls’s approach on grounds of intergenerational justice, arguing that the current generation should not have to endure greater burdens so that future generations might enjoy full justice. The Rawlsian response here should be, I think, to reemphasize the “moral permissibility” and “cost” restrictions stated above. An important part of the theoretical work to be done will presumably involve specifying the correct distribution across generations of the costs of doing justice.} Rawlsian nonideal theory, like any theory purporting to provide moral guidance for political practice in an unjust world, must concede that its guidance can reach only as far as our ability to apply it. But that fact constitutes no reason for skepticism about the theory itself.
It is occasionally observed that Rawls's nonideal theory is closely related to the more familiar, nontechnical notion of "the second best," as that idea is sometimes used by political and legal theorists (with "the best" corresponding to Rawls's ideal theory). We can now see, however, that this comparison is misleading. The idea of "the second best" is not normally understood to include (and arguably does not include in its more formal use) the transitional aspects of Rawlsian nonideal theory. A state of affairs can be "second best" because, say, it is the arrangement that most closely resembles "the best" in the set of those arrangements that are politically feasible (i.e., in cases where the ideal arrangement is simply not feasible), without this second-best arrangement constituting genuine progress toward actually achieving the ideal. Second-best arrangements may well involve departing from the feasible path to the ideal in such a way as to actually preclude ever achieving the ideal. Yet the arrangement might still be the second best, in the sense of being "nearer to" (that is, more closely resembling) the best nonideal policy than any feasible rivals. Nonideal theory, according to my reading of Rawls, may dictate the pursuit of "third-" or "fourth-best" options instead, which may constitute less direct, less dramatic, less rapid progress toward the ideal, if this is what is necessary if we are ever in the future to actually reach the ideal.

IV. ALTERNATIVE VERSIONS OF IDEAL AND NONIDEAL THEORY

Rawls's theory is, however, plainly only one version of the ideal-nonideal distinction as this can be found in recent philosophical literature. Here I will distinguish from it three other versions of the distinction (without intending to suggest that no further variations are possible). The first of these is the version utilized by Liam B. Murphy in his book Moral Demands in Nonideal Theory. Murphy appears to believe that his principal disagreement with Rawls (on the use of the ideal-nonideal distinction) is in their difference on the autonomy of political philosophy (or the "continuity" of moral philosophy across the domains of moral principles)

22. R. G. Lipsey and Kelvin Lancaster famously argued (in "The General Theory of Second Best," The Review of Economic Studies 24 [1956–57]: 11–31) that if one of the conditions for Pareto optimality cannot be satisfied, the next best solution may involve moves away from what would otherwise be optimality conditions in other areas. But because there is no analogue of fixed, perfect justice in this efficiency model, second-best theory cannot here be transitional in exactly the fashion of Rawlsian nonideal theory.
for institutions and for individuals). There is no suggestion that Murphy takes himself to be departing in any substantial way from the Rawlsian conception of nonideal theory. It is only the proper scope of that theory that he disputes (favoring a more expansive scope for it than did Rawls).

In fact, however, the differences between Murphy and Rawls seem considerably broader than Murphy’s remarks suggest. In particular, it seems clear that Murphy’s version of nonideal theory lacks the strongly transitional character of Rawls’s version. The principles on which Murphy focuses are primarily moral principles for individual conduct. Presumably, though, given his rejection of Rawls’s “discontinuity” or “dualist” view, Murphy would have to say similar things about the moral principles that should govern institutional change in nonideal conditions. And Murphy’s solution to the problem of partial compliance, of course, is not that persons should then act in whatever way will be most likely to bring about full (or fuller) compliance with ideal principles. Rather, his solution is that persons are obligated to do no more than they would have been obligated to do had others in fact fully complied—or, less formally, that persons are obligated only to do their “fair shares” of what morality demands of us, not to take on the shares of others simply because others refuse to do their parts.

Now, however, Murphy seems obliged either (a) to reject—implausibly, in my view—the transitional objectives of nonideal theory for institutions, or (b) to explain why nonideal theory for individuals should not have the transitional objectives that should guide nonideal theory for institutions, and this without simply reintroducing the “dualism” or discontinuity to which he objects in Rawls’s view. Remember, further, that nonideal theory governing international requirements can also be structured with either transitional objectives or concerns about “fair play” in mind. Rawlsian nonideal theory requires that peoples adopt policies for bringing about full compliance with ideal theory’s requirement that the nations of the world form a “society of peoples.” The international analogue of Murphy’s nonideal theory for individuals would require, instead, that peoples/states do no more to support the institution of a society of peoples than they would be required to do were all other states fully complying with the rules of such

an international society. In the end, since it seems clear (to me) that at least some aspects of nonideal theory should have transitional objectives, Murphy owes us a better explanation of why nonideal political philosophy should not be characterized in the transitional fashion that Rawls has embraced. None of this is to say that considerations of cost and the fair distribution of burdens should have no role in nonideal political philosophy. It is to say, however, that such considerations could only very implausibly be taken to trump the transitional considerations by which Rawlsian nonideal theory is oriented.

There is yet another, third, version of the ideal-nonideal distinction that can be drawn from Joel Feinberg’s well-known discussion of what he calls the “ideal directives” that are disguised as rights-claims. Feinberg argued that many apparent claims of right (or justice) are in fact only applications of “right names” to what are better understood as “ideal directives, addressed to those in appropriate positions to do their best for a particular kind of human value. . . .”24 As such, the ideal at issue (in this version of “ideal theory”) has a more aspirational status, functioning in certain ways like a Kantian “regulative ideal.” The ideal is not so much a condition to be achieved (like the Rawlsian ideal of perfect institutional justice), but rather a personal or institutional value to be always taken seriously in any practical deliberations. This model of ideal theory, then, would be concerned with establishing the ideal directives that are to function in this regulative fashion. The corresponding non-ideal theory would presumably specify requirements on actions, policies, or institutions in circumstances where one or more of these ideal directives was not fully or adequately reflected in practical deliberation. The requirements of nonideal theory might or might not have transitional aims on such an approach.25

Rawls, of course, would reject such a version of the ideal-nonideal distinction as intuitionistic26 through and through. Although Rawls allows

26. Theories of justice count as intuitionistic in Rawls’s sense if “first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another” (TJ, p. 34).
that intuitionism of this sort is not “intrinsically irrational” and even “may be true” (TJ, p. 39), he rejects every intuitionist conception of justice as “unconstructive” (TJ, p. 41), lacking as they do clear priority rules for settling conflicts between the values that make up the ideal. Feinberg, in turn, regarded claims to be able to establish such priority rules as simply “astonishing,” adding that Rawls’s own nonideal theory itself relies heavily on intuition. We have seen already, of course, that this latter charge seems correct, that the requirements of nonideal theory that are noted by Rawls do seem to leave considerable space to be filled by intuitive weighing of the various considerations at issue. Rawls would concede, I think, Feinberg’s claim that for actors “on the ground” in the real, unjust world, priority rules cannot be expected to dictate solutions in nonideal theory, at best rather just identifying those injustices that are most grievous. But however sloppy and intuitionistic we might think nonideal theory must inevitably be, the proper characterization of ideal theory will depend on whether Rawls’s (or others’) arguments for priority rules in political philosophy are adequate to overcome intuitionistic skepticism. And so long as it is not clear that such a defense of priority rules must fail—and, in my view, this is not yet clear—we have not been given any strong reason to abandon Rawls’s characterizations of either the nature of ideal and nonideal theory or the relationship between them.

I consider, finally, a familiar complaint about Rawlsian ideal theory that could be understood to be pointing toward still another version of the ideal-nonideal distinction. The complaint in question is that Rawlsian ideal theory is too ideal, setting the target for our endeavors unrealistically or impractically high. In one of the more restrained versions of this kind of criticism, Allen Buchanan has argued that a defensible ideal moral theory should be “feasible,” “accessible,” and “morally accessible.” A theory is feasible “if and only if the effective implementation of its principles is compatible with human psychology, human capacities generally, the laws of nature, and the natural resources available to human beings”; it is “accessible if it is not only feasible, but if in addition there is a practicable route from where we are now to at least


28. Although I cannot discuss it here, there is also a familiar (contrary) view that Rawlsian ideal theory is not ideal enough, that it is in fact too sensitive to facts about “the human condition,” about economics, human psychology, etcetera. See, e.g., G. A. Cohen, Rescuing Justice and Equality (Cambridge, Mass.: Harvard University Press, 2008).
a reasonable approximation of the state of affairs that satisfies its principles”; and a theory is “morally accessible” if “the transition from where we are to the ideal state of affairs should be achievable without unacceptable moral costs.”29

Rawls, as we have seen, accepts the requirement that Buchanan calls “feasibility”; theories premised on dramatic changes in human nature or in the world in which humans live must be left to the utopian branch of political philosophy. But it seems clear that in Buchanan’s insistence that ideal theory also be “accessible” and “morally accessible” he has drawn into ideal theory much that was for Rawls the substance of non-ideal theory, making the boundaries between ideal and nonideal theory quite obscure. As a result, it seems clear as well that the kind of ideal theory that Buchanan is prepared to accept will be both a *weaker* (or more modest) theory and a more *variable* theory. The requirements that there be a practicable and a morally acceptable route to whatever ideal condition our ideal theory identifies suggests, first, that the ideal may have to be made more modest (and possibly far more modest) in order to satisfy those requirements. And those requirements suggest as well that ideal theory can be expected to consist of quite different principles, thus requiring quite different things of the basic social institutions, both in different societies30 and at different times within the same society. As Buchanan says, there might be some, otherwise defensible, ideal principles that the “contingencies of our history or culture or the inertia of our severely defective social institutions might bar us from” realizing.31 And “whether the moral costs of transition [to the otherwise defensible ideal] are acceptable will depend in part upon how defective the current state of affairs is.”32

It is, of course, a little puzzling that in an account like Buchanan’s, which characterizes the principles of ideal theory as setting the “ultimate standards” or “ultimate targets,” ideal theory turns out to be so potentially variable (both between and within societies). And the

30. I leave to one side the fact that Rawls’s own ideal theory, during the period between *A Theory of Justice* and *Political Liberalism*, “evolved” from a theory apparently intended to have universal application to a theory only for the basic institutions of liberal democracies.
32. Ibid., p. 62.
33. Ibid., pp. 60, 61 (my emphasis).
requirement of theoretical “modesty” under which Buchanan’s preferred version of ideal theory operates seems to me to concede far too much to the exercises of physical power and emotional manipulation that are so often employed in rendering “impracticable” the reform of unjust social institutions. While in the year 1920 there might well have been no “practicable route” to the reform of Jim Crow laws (at least in Buchanan’s sense of “practicable”), it would surely be unconvincing to argue that on that ground alone our ultimate standards of justice did not, even in 1920, condemn social institutions that included such laws. It seems to me as a result desirable to set for ourselves, with Rawls, a genuinely “ultimate” ideal at which to aim, an ideal set as high as the “practically possible” permits, leaving theorizing about more transitory practical requirements—those concerning the most “accessible,” efficient, and morally permissible routes to that ideal—squarely in the domain of nonideal theory.

V. THE PRIORITY OBJECTION

In addition to the intuitionist’s objections to Rawlsian priority rules for ordering the principles of ideal theory, there is another kind of “priority objection” that is both familiar in literature critical of Rawls’s project and relevant to the ideal-nonideal distinction. Since the first critical discussions of Rawls’s first book, it has been common for critics to complain that our principal focus in theorizing about justice should be on the part of a theory of justice where, as it were, the rubber meets the road, namely, on nonideal theory. We live in societies that are not even approximately just according to Rawlsian ideal principles; so Rawlsian ideal theory, consisting as it does of principles for perfectly just institutions, can provide no direct guidance for those seeking to counter injustice. What we clearly should care about, as possible agents of political change, is the principles of justice that ought to guide our efforts to address and remedy injustice in the real world. Yet Rawls, as we have seen, expended very little effort on nonideal theory. As a result of Rawls’s fixation on ideal theory, it is claimed, subsequent arguments about justice have also become bogged down in the abstract realm of “armchair theorizing.” Instead, the argument goes, we can and should

address first the real issues of justice in the domain of nonideal theory before we worry over the details of some ideal theory or the grand metaphysical questions of abstract moral and political philosophy. Some of Rawls's critics have even suggested, further, that we can do without the ideal theory of justice altogether. The assumptions of ideal theory, they argue, are often so far removed from the actual circumstances in which we find ourselves that they are more or less useless in generating guides for practical reasoning: “the institutional prescriptions [Rawls] advocates are of little use when applied to real societies.”

If we want to know what we ought to do here and now, we have to begin by assuming not fictions like “strict compliance,” but instead more facts that actually resemble that here and now. In other words, many significant facts about the nonideal world we actually inhabit—facts such as scarcity of resources, noncompliance, the various histories of injustice, the moral apathy of typical persons, and the lack of communal solidarity in an increasingly atomistic world—need to be taken account of before we can hope to generate sound practical guidelines. And such facts seem to simply be ignored in the generation of Rawls’s ideal theory. Of what relevance, then, could such ideal theory possibly be to the concerns of those who care about justice? Indeed, the mere fact that ideal theory, by its very nature, is derived through “the exclusion, or at least marginalization, of the actual” has seemed to some critics to altogether disqualify such ideal theory as respectable political philosophy. “The abstractions of ideal theory are not innocent.”

To this latter complaint about counterfactual assumptions and “significant falsehoods,” the appropriate Rawlsian response seems simple and straightforward. Rawls’s ultimate target—the ideal of his ideal theory—is the most just institutional structure that can be achieved within the constraints set by the more or less intractable aspects of human nature and the nature of the world. Fleshing out and arguing for such an ideal requires the use of counterfactual assumptions, since the current state of our institutions—and the resulting social and moral

35. Ibid., p. 855.
36. Ibid., p. 853.
38. Ibid., p. 168.
39. Ibid., p. 181.
condition of the persons living under them (not to mention the ways in which the world is being used)—is so distant from the best we can realistically hope to achieve.\footnote{What needs to be avoided is not all “falsehood” in ideal theory, but rather only those idealizations that “build into a normative theory a false . . . account of the social phenomena the theory itself aims to put under moral scrutiny, in this way severely undermining its potential for guiding action in the real world” (Laura Valentini, “On the Apparent Paradox of Ideal Theory,” \textit{The Journal of Political Philosophy} 17 [2009]: 332–55, at p. 352).} It is precisely the job of nonideal theory to measure that distance and to prescribe courses of action that are sensitive to real (i.e., nonideal) world facts like scarcity and noncompliance.

Ideal theorizing says rather: suppose that we could structure our basic institutions so that the persons living under them would feel valued and respected as individuals and would in consequence be committed to the principles that guide their shared activities, so that the world would be used responsibly to support both the current and future generations, and so that the different nations of the world could interact as equals, as well-ordered societies in an international society of peaceful peoples. If such changes are realistic, then many of the facts about the nonideal world we currently inhabit could cease to \textit{be} facts. That we are asked in such ideal theorizing thus to consider counterfactual claims about persons and world, that we are forced to consider “significant falsehoods” in our arguments, seems not objectionable, but rather to be precisely what a philosopher trying to identify our ultimate moral objectives ought to be considering.\footnote{41. For further discussion of the problem of false assumptions in ideal theory, see Stemplowska, “What’s Ideal about Ideal Theory?” pp. 326–29.} It is, of course, both straightforwardly true and perfectly consistent with Rawls’s claims that ideal theory will still leave an enormous amount of work to be done by nonideal theory—work both normative and social-scientific, as we have seen. So while having identified our ultimate objectives is hardly “of little use” in addressing real societal problems of injustice, it is perfectly fair (as we have also seen) to observe that practical political prescriptions cannot be expected to be easily extracted from ideal theory.

Sometimes, of course, these complaints about the unrealistic character of Rawlsian ideal theory turn out to be not really complaints about ideal theorizing at all, but rather only to be complaints about the specific ideal theory for which Rawls argues. Rawls’s ideal theory, it is said, pays no attention to the long histories of injustice to people of color, to
women, to various groups that constitute minorities in their religious convictions or sexual orientations, and so on; and it pays no attention to the modern destruction of community and social and family networks, to the neglected pursuit of genuinely common ends, to the threat posed by liberal society to the values of culture and ethnicity. But the kinds of “oversights” with which such criticisms charge Rawls are not in any way essential to the Rawlsian characterization of ideal theory or its relationship to nonideal theory; they are rather simply a function of the specific approach that Rawls employs to derive the content of ideal theory, to argue for his particular, favored conception of “justice as fairness.” Historical injustice goes unaddressed in Rawls’s theory because the derived principles of justice are purely “forward-looking,” because the choice problem given to Rawls’s original position contractors requires their choice of forward-looking principles. And the nature of that choice problem is determined not by Rawls’s version of the distinction and the relation between ideal and nonideal theory, but instead by Rawls’s general view of what social justice is, of the particular kind of political fairness that is given expression in the original position model (see, e.g., JAF, pp. 64–66). A quite different content to ideal theory (for instance, one that was more sensitive to the need to redress historical injustice) could be defended while still subscribing wholeheartedly to the Rawlsian version of the ideal-nonideal distinction. Similarly, that version of the distinction can be embraced while advocating an ideal theory that is more attuned to the need for community or for the protection of rights to “cultural membership.” If Rawls’s ideal theory is insensitive to such concerns, it is only because his insistence that the correct conception of justice must be a “political” conception—one that is “stable for the right reason,” the possible object of an “overlapping consensus” of competing comprehensive conceptions—implies that there can be no true political community in a modern, pluralistic society without tyranny and injustice (PL, pp. 37, 146; JAF, pp. 3, 21). But Rawls’s position on that subject is hardly the only possible position.

To the actual priority complaint—i.e., to the complaint that ideal theory need not, and perhaps should not, be completed before theorizing about what is necessary for actual progress in the cause of justice in

this nonideal world—Rawls always responded as follows: “until the ideal is identified . . . nonideal theory lacks an objective, an aim, by reference to which its queries can be answered” (LOP, p. 90). And that response seems to me importantly correct. We can hardly claim to know whether we are on the path to the ideal of justice until we can specify in what that ideal consists. While perhaps some of the finer details of ideal theory will make no practical difference to the policies we pursue now (that is, they may make no difference until our actual institutions begin to approach the ideal more closely), the shape of our political ideal must be reasonably precisely specified before nonideal policies can be endorsed by a theory of justice.

The requirement that nonideal policies be “likely to be successful” requires that we know how to measure success; and that measure makes essential reference to the ultimate target, the ideal of perfect justice. As we saw above (in Section III), the requirement that nonideal policies be morally permissible refers to standards of permissibility already set by the principles of ideal theory, just as the priority to be given to grievous (over less grievous) injustices is to be understood in terms of the lexical ordering of the principles of ideal theory violated by the injustices at issue. Making sense of all of these requirements, then, necessitates that ideal theory have priority over nonideal theory. We, as theorists of justice, simply should not care which policies are politically possible (etcetera) unless those policies are also on an acceptable path to a just institutional structure. To dive into nonideal theory without an ideal theory in hand is simply to dive blind, to allow irrational free rein to the mere conviction of injustice and to eagerness for change of any sort.43

These points seem to me not to be adequately acknowledged in the recent (and much-discussed) effort by Amartya Sen to answer the question, “What Do We Want from a Theory of Justice?” Sen suggests that we don’t need to know all that ideal justice requires in order to compare (as to their effects on injustice) our policy options here and now, any more

43. “A conception of justice must specify the requisite structural principles and point to the overall direction of political action. In the absence of such an ideal form for background institutions, there is no rational basis for continually adjusting the social process so as to preserve background justice, nor for eliminating existing injustice. Thus ideal theory, which defines a perfectly just basic structure, is a necessary complement to nonideal theory without which the desire for change lacks an aim” (PL, p. 285).
than we need to know that Everest is the tallest mountain in the world before we can compare the heights of lesser peaks. Ideal (or “transcendental”) theory is, according to Sen, neither necessary nor sufficient as a foundation for nonideal theoretical conclusions about what will improve societal justice. While Sen’s point about Everest and determining the heights of smaller mountains is obviously true, its use in Sen’s analogy is, I think, potentially misleading. As we have seen, on Rawls’s conception of the relation of ideal to nonideal theory, nonideal theory’s function is not primarily to judge the relative injustices of possible unjust arrangements (except insofar as more or less grievously unjust arrangements can be addressed with equal impacts on the achievement of the overall goal of perfect justice). In other words, which of two smaller “peaks” of justice is the higher (or more just) is a judgment that matters conclusively only if they are both on equally feasible paths to the highest peak of perfect justice. And in order to endorse a route to that highest peak, we certainly do need to know which one that highest peak is. Perhaps for a while we can just aim ourselves in the general direction of the Himalayas, adjusting our paths more finely—between Everest and K2, say—only when we arrive in India. But we need to know a great deal about where to find the serious candidates for the highest peak before we can endorse any path to them from here.

It is, of course, impossible not to be sympathetic to the concerns expressed in these criticisms of Rawls’s ideal-nonideal distinction. As people who care about injustice, we want to do something about it, and we want our philosophical theories of morality and politics to inform our actions. What we don’t want is to wait and wait for the conclusions of philosophers’ endless ideal theorizing, conclusions that at best will offer us no more than a vision of a world that seems impossibly distant from the deeply unjust world we actually inhabit. As activists in the cause of justice, ideal theory may come to seem to us simply irrelevant. But it is

45. Ibid., p. 216.
important to remember that even most nonphilosophers who are active in the cause of justice do in fact have in mind, however vaguely, an ideal of justice toward which they take their campaigns to be ultimately directed. While some of us may become preoccupied with particular targeted injustices that seem to us especially grievous, none of us in the end forgets that justice is an integrated goal and that activism in one domain has the potential to affect adversely the achievement of justice in another. This is all, really, on which Rawls’s model of the ideal-nonideal distinction insists that we focus. The political philosopher’s first job, on this model, is to refine and argue for an ideal of justice, to say as clearly as possible what goal(s) we must attend to and how we must weigh various factors in our efforts to eventually reach that goal.

Perhaps we already know enough about the broad outlines of that ideal (or about essential overlaps between conflicting ideals) and about possible paths to whatever more precise version of the ideal philosophers might convincingly defend that we can proceed to responsibly address particularly grievous injustices now, without waiting for further refinement in ideal theory. But to allow that is not to deny the priority of ideal theory.47 It is rather to affirm that enough of ideal theory is settled that we can already begin to derive from it (conjoined with our social-scientific knowledge) the rudiments of nonideal theory. To insist only that ideal theory have this kind of priority and this kind of role in informing our political activism in a nonideal world is surely not to undermine the cause of justice, but only to insist that we be careful with our political sledgehammers and seek justice thoughtfully. That, I take it, is the simple and convincing heart of the Rawlsian model of the ideal-nonideal theory distinction.

47. Gopal Sreenivasan has defended a conception of “nonideal theory as anticipatory theory,” for which “the priority [of ideal theory] assumption fails” (“Health and Justice in Our Non-ideal World,” Politics, Philosophy & Economics 6 [2007]: 218–36, at p. 221). Non-ideal theory, so conceived, “can define targets for practical action before a complete ideal has been worked out” (ibid.). What Sreenivasan means by this, though, is, I think, actually consistent with my claims here: such nonideal theory, in his view, “has to make assumptions about the minimum requirements that any plausible and complete ideal theory of justice will include” in order to “be confident that steps towards these targets are steps in the right direction” (ibid.).