DEVIATING FROM THE IDEAL

Jacob Barrett

University of Oxford

Ideal theorists aim to describe the ideally just society. Problem solvers aim to identify concrete changes to actual societies that would make them more just. The relation between these two sorts of theorizing is highly contested. According to the benchmark view, ideal theory is prior to problem solving, because a conception of the ideally just society serves as an indispensable benchmark for evaluating societies in terms of how far they deviate from it. In this paper, I clarify the benchmark view, argue that existing criticisms of it are unsuccessful, and develop a novel redundancy objection to the benchmark view and the claim of priority it allegedly entails. I then consider the extent to which ideal theory might facilitate problem solving without being prior to it and argue that it can only play a modest role in this regard. The upshot is that ideal theory is neither required for nor especially relevant to problem solving—but it is not completely irrelevant either. It facilitates problem solving to some limited degree, but no more, say, than theorizing about dystopia.

1. Introduction

Ideal theorists aim to describe the ideally just society. They make judgments about the principles that would be realized in that society, or about the institutions and other concrete features it would contain. On their face, such judgments provide us with little practical guidance. Except in the mythical case where we can realize the ideal in a single stroke, judgments about the ideally just society seem irrelevant to which changes we should implement here and now. To figure out how to reform our current society, we must instead rely on comparative evaluations of justice. We must ask which changes would produce a *more* just society, not what the *most* just society would be like.

On the basis of this sort of reasoning, Amartya Sen (2009, especially ch. 4) has influentially argued that ideal theory is practically irrelevant. Although Sen doesn't put things in quite these terms, we may reconstruct his argument as proceeding in two steps. First, Sen defends a *problem solving* approach to doing practically relevant political philosophy, on which our aim is to identify changes to our society that would make it more just. Our goal, in other words, is to identify problems or "remediable injustices" in our society—features that render it less just than it would otherwise be—as well as solutions to these problems, or concrete changes that would eliminate or mitigate them, thereby making our society more just (Sen 2009, vii; compare, e.g., Anderson 2010, ch. 1, Schmidtz 2011, Wiens 2012). Second, Sen argues that a conception of the ideally just society doesn't help with problem solving: it is not necessary, sufficient, or useful for this purpose. So ideal theory doesn't tell us which changes would make our society more just, and, if we endorse the problem solving approach, it is therefore practically irrelevant. At best, it amounts to "an engaging intellectual exercise in itself" (Sen 2009, 101).

Sen's argument is controversial, but this reconstruction provides a fruitful starting point because each step highlights a different ground on which ideal theorists resist Sen's critique and affirm the practical relevance of ideal theory. Some challenge the first step, calling into question the adequacy of problem solving. To responsibly engage in practically relevant political philosophy, they argue, we cannot only solve particular problems of injustice, but must also take into account the extent to which "solutions" to these problems set back or further our long-term target of eventually achieving an ideally just society; and for that, we need ideal theory (e.g., Rawls 1999b, 128, Simmons 2010). Others object to the second step, arguing that problem solving requires or at least is greatly facilitated by ideal theory. A conception of the ideally just society may even serve as an indispensable benchmark for evaluating societies in terms of how far they deviate from it. It may provide, as John Rawls famously puts it, "the only basis for the systematic grasp of these more pressing problems" (Rawls 1999a, 8; compare, e.g., Shelby 2013).

I have criticized the view that ideal theory provides us with a long-term target for reform elsewhere (Barrett 2020; see also Gaus 2016). So, in this paper, I confine my attention to the question of whether ideal theory is prior to problem solving or otherwise facilitates it. This topic should be of broad interest because problem solving is clearly one important sort of political philosophy regardless of what one thinks of other aspects of the ideal/nonideal theory debate. My conclusion will be largely, but not entirely, negative. Ideal theory is neither required for, nor especially relevant to, problem solving—but it is not completely irrelevant either. It facilitates problem solving to some limited degree, but no more, say, than theorizing about dystopia.

I begin with an explanation of the *benchmark view* (Wiens 2015) and of Rawls's claim that it implies the priority of ideal theory over problem solving, along the way showing why a common criticism of the view—namely, that we are often able to identify improvements in justice without any conception of the ideal in mind—rests on a misunderstanding. I then turn to Sen's own criticisms of the benchmark view, which I also show to be unsuccessful, but which I draw on to develop a more powerful *redundancy objection* to the benchmark view and the claim of priority it allegedly entails. The redundancy objection, in brief, is that benchmark theorists are committed to thinking that the truth of comparative judgments of justice are explained by *ideal deviation criteria* on which one society is less just than another when it deviates further from the ideal along some given dimension, and to thinking that such criteria make essential reference to the ideal. But it turns out that all plausible ideal deviation criteria make the ideal redundant: we can use the dimensions in which such criteria measure deviations to make comparative evaluations of justice directly, without ever referring to the ideal.

With my argument against the benchmark view complete, I turn to an examination of ideal theorists' standard reply to related arguments—that ideal theory greatly facilitates problem solving even if it is not prior to it—as well as a common rejoinder by critics of ideal theory—that ideal theory obstructs problem solving more than it facilitates. The conciliatory conclusion I reach is that both ideal theorists and their critics overplay their hands. Ideal theory can play some subsidiary role in problem solving: it may help us to develop abstract criteria of comparative justice, to work out novel solutions to the problems we face, and to better appreciate such problems. But there is nothing special or privileged about ideal theory's ability to play these roles, in comparison to theorizing about any number of actual or hypothetical nonideal societies. Ideal theory may provide us with some relevant data points, but these are a small set of data points among many others.

Two clarifications before I go on. First, my focus in this paper is on the extent to which ideal theory is practically relevant, and, specifically, on whether it yields practically relevant verdicts pertaining to justice. Nothing I say here bears on whether ideal theory may be valuable for some other reason (see, e.g., Estlund 2014, Estlund 2020, chs. 16-17). Second, throughout this paper, I use the term "ideal" as shorthand for the "ideally just society." There are many other perfectly acceptable usages of the term "ideal," but I never employ them here.

2. The Benchmark View

Like much else in contemporary political philosophy, recent interest in the practical relevance of ideal theory traces to Rawls. Rawls's own views on the matter are complicated. In later work, he suggests that the ideal serves as a "long-term goal of political endeavor" (Rawls 1999b, 138). But in *A Theory of Justice*, Rawls proclaims that our conception of the ideally just society—and, specifically, the principles that would govern its basic structure—is practically relevant because "[e]xisting institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason" (Rawls 1999a, 216). Tommie Shelby (2013, 153) explains:

> On the Rawlsian view, injustices are conceptualized as *deviations* from the ideal principles of justice… An injustice is a failure on the part of individuals or social arrangements to satisfy what the ideal principles of justice demand. Thus, charges of injustice *presuppose* ideals of justice, which particular individuals and institutions can and often do depart from. Such deviations can be small or great, minor or serious… depending on the size and nature of the gap between ideals and practice.

So, on this benchmark view, injustice is a deviation from ideal justice, and for one society to be less just than another is for it to deviate further from the ideally just society. Much as something is a worse replica of an original painting to the extent that it deviates from that painting, a society is less just to the extent that it deviates from the ideal.

Critics of ideal theory often assume that the benchmark view implies a strong claim of epistemic priority: that "we first need to know what an ideally just society would be, to identify the ways our current society falls short" or to identify ways of making it more just (Anderson 2010, 3). Their standard move is then to point to various blatant problems of injustice whose elimination would clearly improve justice—for example, slavery or caste and gender discrimination—to note that we can judge as much without referring to the ideal, and to declare the benchmark view defeated. Such examples show, after all, that we can identify problems of injustice without having an ideal in mind, and that we "can judge social option A as being better than social option B without starting with a view of the best society" (Appiah 2017, 168). Indeed, if we know anything at all about justice, it is that slavery is unjust and that abolishing it improves justice. At least when it comes to justice, then, "knowledge of the better does not require knowledge of the best" (Anderson 2010, 3).

These critics are right. We typically identify injustice long before we have any conception of ideal justice; we "recognize the existence of a problem before we have any idea of what would be best or most just" (Anderson 2010, 3). But the very obviousness of this claim should give us pause—do any benchmark theorists disagree? As far as I am aware, they don't. Rawls himself, for example, holds that theorizing about justice involves a process of moving "back and forth" between (among other things) our judgments about particular instances of injustice and our judgments about the principles that would govern the ideally just society—sometimes revising the former in light of the latter, and sometimes revising the latter in light of the former, until all our judgments hang together in "reflective equilibrium" (Rawls 1999a, 18). Our judgment that slavery is unjust, for example, serves as a "fixed point" in our theorizing; any plausible theory of justice must accommodate the datum that slavery is (very) unjust and so deviates (greatly) from the ideal (Rawls 2001, 29). So Rawls doesn't claim that we first need to complete our ideal theory before we can make any judgments about what would make our society more just. And neither does the benchmark view more generally.

Instead, the benchmark view claims that our conception of the ideal provides a benchmark of evaluation "comparison with which defines a *standard* for judging actual institutions" (Rawls 1999a, 199, emphasis added). In other words, the view is that *criteria* of comparative justice make essential reference to the ideal, since such criteria must be articulated in the following *ideal deviation* form: one society is less just than another if and only if it deviates further from the ideal (compare Sen 2009, 98). And from here, it is not hard to see why benchmark theorists insist that ideal theory enjoys a certain *theoretical priority* over problem solving, in the sense that the latter must be "worked out after an ideal conception of justice has been chosen" (Rawls 1999a, 8). The goal of problem solving, after all, is to identify concrete ways of making our society more just. And though, as we have seen, there are some easy cases where we can tell that a change would improve justice without resorting to any theory, problem solving is supposed to help us get beyond these cases. But to do this, it is commonly thought that we must invoke criteria of comparative justice—criteria of the form "society *x* is more just than society *y* if and only if…"—that explain which changes would improve justice. So if benchmark theorists are right that criteria of comparative justice make essential reference to the ideal, and if problem solving requires us to appeal to such criteria, it follows that we cannot engage in problem solving without referring to the ideal. Though we may make scattered pre-theoretic judgments about how to improve our society without any conception of the ideal in mind, we cannot engage in serious theorizing about this until we first complete our ideal theory (compare Valentini 2011, 306-307, Shelby 2013, 156).

One possible response to this claim of theoretical priority would be to deny that we need criteria of comparative justice to engage in problem solving. For example, one might defend a version of "methodological particularism" (Sinnott-Armstrong 1999) on which, regardless of whether criteria of comparative justice make essential reference to the ideal, such criteria are unnecessary or unhelpful when theorizing about how to improve our society: we can get by relying on our pre-theoretic convictions. Here I set aside this response, and grant that we need criteria of comparative justice to engage in problem solving. In what follows, then, I tackle head on the question of whether criteria of comparative justice make essential reference to the ideal, assuming that, if they do, the benchmark theorist's claim of theoretical priority follows. Sen's own criticisms of the benchmark view, though ultimately unsuccessful, provide a useful starting point.

3. Sen's Critique

The benchmark view claims that criteria of comparative justice make essential reference to the ideal. This makes such criteria atypical, since, as Sen notes, most criteria of comparative evaluation don't work this way. To take Sen's most famous example, we don't need to refer to the highest mountain to make a principled comparison of the heights of Mount Kilimanjaro and Mount McKinley (Sen 2009, 102). We can instead rely, say, on the criterion of which mountain is longer from its base to its summit.

Sen takes examples like this to show that ideals play no necessary role in comparative evaluation. If criteria of comparative height make no reference to the highest point, and if—to take another of his examples—criteria of comparative beauty make no reference to the most beautiful object, then there is no reason to think that criteria of comparative justice refer to the most just society (Sen 2009, 16, 101). But this argument is, by itself, too quick. Though Sen's examples do show that "[t]here would be something deeply odd in a general belief that a comparison of any two alternatives cannot be sensibly made without a prior identification of a supreme alternative," benchmark theorists are not committed to this *general* belief (Sen 2009, 102). Instead, they need claim only that *some* comparative criteria make essential reference to a supreme alternative, and that the correct criteria of comparative justice belong in this class. If the benchmark view holds, then criteria of comparative justice are more like criteria of how far a point deviates from a particular mountain, or how far a replica deviates from an original painting, than criteria of how high or beautiful an object is. The correct criteria of comparative justice must be formulated in ideal deviation form, and therefore make essential reference to the ideal, even if other candidate criteria of comparative justice need not.

The real upshot of Sen's argument, then, is that it is a substantive question whether criteria of comparative justice make essential reference to the ideal—not one we can settle by reflecting on the logic of comparative evaluation in general. So we have two positions in front of us. On the first, which Sen assumes, criteria of comparative justice are *essentially comparative*: they can yield verdicts about which of two societies is more just than the other simply by referring to features of those two societies. This is compatible with a monistic theory on which only one essentially comparative criterion is relevant to justice—say, a utilitarian criterion on which a society is more just when it contains more well-being, or a perfectionist criterion on which societies are more just when they exhibit more excellence. But it is also compatible with a pluralistic theory, on which a society is more just when it better satisfies a plurality of essentially comparative criteria—concerning, say, equality, freedom, and efficiency—given some way of balancing these criteria when they conflict. On the second, which benchmark theorists endorse, comparative evaluations of justice are instead explained by one or more *ideal deviation* criteria that make essential reference to the ideal, in the following sense. They can yield verdicts about which of two societies is more just than the other only by also referring to the ideally just society, and then asking which departs further from this ideal.

With this in mind, let us turn to Sen's second criticism, namely, that such ideal deviation criteria are implausible, because "descriptive closeness is not necessarily a guide to valuational proximity"—for example, "a person who prefers red wine to white may prefer either to a mixture of the two, even though the mixture is, in an obvious descriptive sense, closer to the preferred red wine than pure white wine would be" (Sen 2009, 16). In other words, Sen's worry is the familiar one that, of three options, the option that deviates less from the best option is not necessarily "second-best" of the three (Lipsey and Lancaster 1956;

compare, e.g., Goodin 1995). To see this in the case of justice, suppose we understand deviations in terms of institutional similarity: a society deviates from the ideal to the extent that its institutions differ from those realized in the ideal. Now suppose further that the ideal would contain no oppression, and so no institutions in place to ameliorate oppression or compensate its victims. Abolishing the ameliorative or compensatory institutions that we have in place would therefore make our society deviate less from the ideal—but surely this wouldn't make things more just (Estlund 2014, 120-121). Or, to take a case more analogous to Sen's wine example, consider healthcare. Although the view is empirically contentious, there is nothing incoherent in holding that the ideally just society would contain a purely market-based healthcare system—say, because it is most efficient—or a purely public system—say, because it is most equal—while also insisting that both options are more just than a "mixed" or two-tier system—say, because it is both less efficient and less equal. So, again, deviating further from ideal institutions doesn't necessarily make a society less just.

An obvious response to this worry is that I have employed the wrong deviation measure. There are all sorts of ways to measure deviations from the ideal, and these examples merely show that institutional similarity is not what is relevant. Indeed, as Pablo Gilabert (2012, 46) points out, something similar is true of Sen's wine example:

> [A] mixture of red and white wine will be descriptively closer to pure red wine, in one respect, than pure white wine would be. But this descriptive similarity may not be important, or decisive, for the agent making the comparisons, because the important principles rendering red wine best would be better instantiated in white wine than in the mixture. Perhaps color is not evaluatively important, while consistency of taste is.

So we can avoid second-best worries in the case of wine by measuring deviations with respect to those features of my favorite wine that fundamentally explain why I like it best, rather than its more incidental features. If, for example, I like a particular red wine best because of its taste, and I care not at all about color, then if a deviation measure is to track how much I like wine, it should measure departures with respect to taste, not color. And returning to the case of justice, Gilabert argues that we can avoid second-best worries in much the same way. Though features like taste are clearly irrelevant to justice, there are certain basic values or principles whose implementation by ideal institutions explains why those institutions would be contained in the ideally just society (Gilabert 2012, 48). Nonideal societies are those that fall short of realizing these basic values or principles, and an adequate ideal deviation criterion should judge that a society is less just to the extent that it falls further short in this regard.

Gilabert provides a compelling response to Sen's second challenge. As Sen himself stresses, the justice of institutions is explained (to a significant extent) by their "social realizations": by their effects on further features of society, for example, the distribution of resources, that are fundamentally relevant to justice (Sen 2009, especially ch. 10). So to pick out a particular institutional arrangement as ideal, we must invoke some basic values or principles that explain what makes that institutional arrangement ideal—for example, the fact that it produces an equal distribution. And because institutions interact with one another to produce their justice-relevant effects, a society that is closer to the ideal with respect to its institutions may fall further from it with respect to the realization of these basic values or principles—for example, by distributing resources less equally (see Barrett 2020, Gaus 2016, especially ch. 2). This is what generates second-best problems when we attempt to measure deviations with respect to institutional similarity, but there is no room for such problems to arise if we measure deviations with respect to basic values or principles themselves. We can thus avoid second-best worries, and so Sen's second objection, by ensuring that our ideal deviation criteria track deviations in this latter way. They must count one society as deviating further from the ideal when it falls further short of satisfying those basic values or principles, whatever they are, whose optimal satisfaction by an institutional arrangement is what fundamentally explains why that arrangement is ideal (compare Swift 2008, 372-378).[1]

This brings us to Sen's final worry, which targets the sufficiency rather than the necessity or plausibility of appealing to the ideal in our criteria of comparative justice:

> The difficulty lies in the fact that there are different features involved in identifying distance [from the ideal], related, among other distinctions, to different fields of departure, varying dimensionalities of transgressions, and diverse ways of weighing

[1] An anonymous reviewer notes that Wiens (2020) might seem to challenge this line of thought, since it shows that Lipsey and Lancaster's (1956) theorem can be interpreted in a way that generates second-best problems even when measuring deviations with respect to basic values or principles. However, as Wiens (2020, 14) explains, the theorem demonstrates only that second-best problems arise if we measure deviations from "ideal ratio principles," which specify the *ratios* between the extents to which different fundamental values would be realized in the ideal. Specifically, Wiens (2020, 16) shows that if we can't satisfy one ideal ratio principle (say, "the ideal freedom-to security principle"), it doesn't always improve justice to satisfy others (say, "the ideal freedom-to-equality ratio principle"). But this is consistent with my argument, since I don't propose measuring deviations from ideal *ratio* principles—nor am I aware of anyone else who does. Anticipating this concern, Wiens (2020, 13) suggests that ideal ratio principles correspond to familiar mid-level principles, from which some benchmark theorists do seem to think we should measure deviations. Granting as much, his argument seems best understood as showing that second-best problems arise if we measure deviations with respect to such *mid-level* principles—which is again compatible with my point above.

> separate infractions. The identification of [the ideal] does not yield any means of addressing these problems to arrive at a relational ranking of departures.… The characterization of spotless justice, even if such a characterization were to emerge clearly, would not entail any delineation whatever of how diverse departures from spotlessness would be compared and ranked (Sen 2009, 98-99).

As an illustration, Sen (2009, 99) notes that "in a Rawlsian analysis… departures may occur in many different areas, including the breaching of liberty, which, furthermore, can involve diverse violations of distinctive liberties" and that "[t]here can also be violations—again, possibly in disparate forms—of the demands of equity in the distribution of primary goods." We will return to Rawls shortly, but for now, consider another example. Suppose we identify the ideal as a society in which no one is oppressed. One way for a society to deviate from this ideal is for it to contain more oppression. But simply saying that the ideal contains no oppression doesn't tell us how to measure or aggregate different sorts of oppression to determine which of two societies deviates further from the ideal in this regard. Further, if no one is oppressed, then oppression is equal, so another way for a society to deviate from this ideal is for it to have a less equal distribution of oppression. But is this relevant, and, if so, how are we to measure such deviations and balance them against deviations with respect to total oppression? And these are only two of the many sorts of deviations we might consider. For example, perhaps we should instead measure deviations with respect to the level of oppression experienced by the most oppressed member of society, who would experience no oppression in the ideal.

Sen's worry, to be clear, is not that it is impossible to construct deviation measures that let us evaluate how far a society deviates from the ideal: we can, after all, stipulate whatever measure we like. It is rather that we can opt for many different deviation measures, and simply specifying an ideal doesn't tell us which to pick. With only a conception of the ideal in mind, we can judge that all nonideal societies are less just than the ideal, but we cannot make any further judgments about the relative justice of nonideal societies. For that, we must specify which deviations from the ideal are relevant, and in which dimensions we should measure them. If only one dimension of deviation is relevant, this will yield a single ideal deviation criterion that claims that one society is more just than another when it deviates less from the ideal in this dimension. But if more than one dimension is relevant, we will have a set of ideal deviation criteria, and we will need some way of balancing them when they conflict.

Once again, Sen's criticism yields an important lesson: a conception of the ideal cannot, by itself, generate comparative evaluations of nonideal options. But his conclusion that we must reject the benchmark view doesn't follow. Though benchmark theorists claim that criteria of comparative justice make essential reference to the ideal, they need not claim further that specifying a conception of the ideal is all we need to rank nonideal options. Instead, they can acknowledge that any ideal deviation criterion must include two elements: a conception of the ideal, and a deviation measure that specifies some dimension along which deviations from this ideal are fundamentally relevant to justice. Without employing a deviation measure we cannot determine the extent to which a society deviates from ideal justice, and so cannot make comparative evaluations of nonideal alternatives. But all this remains compatible with benchmark theorists' insistence that criteria of comparative justice also require us to specify some conception of the ideal from which we measure how far societies deviate in the relevant dimension, and that such criteria therefore make essential reference to the ideal.

4. The Redundancy Objection

So much for Sen's critique. There remains, however, a deeper worry in the vicinity, which I will now argue ultimately derails the benchmark view. The worry is that once we fully specify a plausible ideal deviation criterion, we make the ideal redundant: we end up with a criterion that is equivalent to an essentially comparative criterion, and that therefore makes no essential reference to the ideal. For notice where our replies to Sen have left us. If the benchmark view holds, then comparative evaluations of justice are explained by ideal deviation criteria (along, perhaps, with further criteria for balancing them) that each contain two components: a conception of the ideal, and a deviation measure that specifies some fundamentally justice-relevant dimension along which societies can come closer to or further from this ideal. But if this is so, then once we have a plausible deviation measure in place, it appears to do all the work. We can use the dimension picked out by our deviation measure to make comparative evaluations directly, without ever referring to the ideal. Or so I will argue—first, by applying this redundancy objection to a few existing views, and then by showing why it generalizes.

Before going on, it is worth distinguishing my redundancy objection from a related argument due to David Wiens. Wiens (2015, 437-438) proposes a "general model of normative political theory" on which fundamental criteria of comparative justice are essentially comparative, and the ideal is defined as whatever society these criteria deem more just than all others in the set of societies meeting relevant constraints. On the basis of this model, he infers that the ideal is redundant to comparative evaluations of justice: we can make such judgments by employing the same essentially comparative criteria we use to pick out the ideal (Wiens 2015, 442). Though illuminating, a key weakness of this argument is that it runs into the same difficulty as Sen's first objection: it simply assumes that fundamental criteria are essentially comparative, and so begs the question against benchmark theorists who insist that such criteria instead have an ideal deviation structure. The strategy of my redundancy objection, by contrast, will be to grant that fundamental criteria of comparative justice may be ideal deviation criteria, yet show that all plausible criteria of this sort are equivalent to essentially comparative criteria, and so make the ideal redundant.

To begin to get a feel for this redundancy objection, consider a toy example. Suppose you believe some institutional arrangement—say, property-owning democracy—is ideal, and, since you are a benchmark theorist, you hold that criteria of comparative justice must be formulated in ideal deviation form. But you then face the question of how to measure deviations from the ideal, so you turn to the prior question of what basic values or principles explain why this institutional arrangement is ideal. Upon reflection, you conclude that property-owning democracy is ideal because it promotes more total well-being than all other feasible arrangements. Further, you decide that since, on your view, total well-being is all that justice is fundamentally concerned with, you should measure deviations from the ideal in the dimension that arranges societies from less to more total well-being. So you come to the following ideal deviation criterion: a feasible society is more just than another when it differs less in its total well-being from the ideally just society. But you now realize that since the ideally just society has the most well-being of any feasible society, your principle need not be formulated in ideal deviation form after all, because it is equivalent to the essentially comparative utilitarian criterion that one feasible society is more just than another when it has a greater sum of well-being in it. Though you began as a benchmark theorist, you ended up with a criterion that makes no essential reference to the ideal. It makes the ideal redundant.

I call this a toy example because no one would really expect a utilitarian to accept the benchmark view. Utilitarianism is a paradigmatic essentially comparative theory because it claims that one society is more just when it produces more total well-being, and simply picks out the ideal as whichever society is more just than all other feasible societies given this prior essentially comparative criterion; in this sense, it perfectly fits Wiens's "general model" mentioned above. But, as we will now see, many actual benchmark theorists fall prey to the same objection as our hypothetical utilitarian. They begin with a conception of the ideally just society, specify some dimension *d* in which we can arrange societies from less to more *d*, and defend an ideal deviation criterion on which one society is less just to the extent that it deviates further from the ideal in dimension *d.* But this criterion turns out to be equivalent to the essentially comparative criterion that a society is more just whenever it is more (or whenever it is less) *d,* thus rendering the ideal redundant.

Consider the most explicit attempt to spell out an ideal deviation criterion in the literature: Thomas Christiano and Will Braynen's (2008) "common good" principle of equality, on which a society's justice depends on its deviation from the distribution of well-being that would be realized in the ideally just society. On their view, we pick out this ideal distribution by determining the greatest total sum of well-being that is feasible and then imagining that this well-being is distributed equally. We then measure deviations from this ideal according to a deviation measure that meets several intuitive desiderata, and judge that one society is less just than another if it deviates further from the ideal according to this measure. Specifically, Christiano and Braynen (2008, 415) propose that the deviation between society x and the ideally just society is given by the following formula, where i_1 through i_n represent the (equal) well-being levels of individuals 1 through n in the ideally just society, and x_1 through x_{n} represent the (perhaps unequal) well-being levels of individuals 1 through n in society x:

$$(1/x_1 + 1/x_2 + \ldots + 1/x_n) - (1/i_1 + 1/i_2 + \ldots + 1/i_n)$$

In other words, this deviation measure says that a society deviates from the ideal to the extent that the sum of individuals' well-being reciprocals differs from the sum of well-being reciprocals at the ideal (where if my well-being level is w, my "well-being reciprocal" is $1/w$).

Christiano and Braynen should be applauded for being unusually precise in specifying their ideal deviation criterion. But their precision comes at a cost: it allows us to identify exactly how their view makes the ideal redundant. For note that if we plug their deviation measure into the criterion that *x* is more just than *y* when *x* deviates less from the ideal than *y*, we get:

Society *x* is more just than society *y* if and only if:

$(1/x_1 + 1/x_2 + \ldots + 1/x_n) - (1/i_1 + 1/i_2 + \ldots + 1/i_n) <$

$(1/y_1 + 1/y_2 + \ldots + 1/y_n) - (1/\mathrm{i}_1 + 1/\mathrm{i}_2 + \ldots + 1/\mathrm{i}_n)$

And since the part of this criterion referring to the well-being reciprocals of individuals in the ideal—namely, $(1/i_1 + 1/i_2 + \ldots + 1/i_n)$—cancels out from each side of the above inequality, this is mathematically equivalent to the following essentially comparative criterion:

Society *x* is more just than society *y* if and only if:

$(1/x_1 + 1/x_2 + \ldots + 1/x_n) < (1/y_1 + 1/y_2 + \ldots + 1/y_n)$

So Christiano and Braynen's criterion need not be formulated in ideal deviation form after all, since it is equivalent to the essentially comparative criterion that one society is more just than another when the total sum of well-being reciprocals in the former society is less than the total sum of well-being reciprocals in the latter. Regardless of whether this criterion is plausible, it clearly makes no essential reference to the ideal. Just like the utilitarian ideal deviation criterion we considered, it makes the ideal redundant.

Let us turn, next, to Rawls's more complicated theory of justice. For Rawls, the task of ideal theory is to come up with a set of principles that would regulate the ideally just society's "basic structure." Specifically, Rawls (1999a, 266) argues that—given several further assumptions, including, for example, that the principles are public and (nearly) universally complied with — the ideally just society would realize the following principles:

> Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all (*Equal Liberty*).

> Social and economic inequalities are to be (a) attached to offices and positions open to all under conditions of fair equality of opportunity (*Fair Equality of Opportunity*), and (b) to the greatest benefit of the least advantaged (*Difference Principle*).[2]

Rawls adds that Equal Liberty has lexical priority over Fair Equality of Opportunity, which in turn has priority over the Difference Principle, where a lexically posterior "principle does not come into play until those previous to it are either fully met or do not apply" (1999a, 38).

So, in Rawls's ideally just society, everyone shares in the most extensive possible equal scheme of basic liberties and enjoys fair equality of opportunity, while remaining inequalities in social and economic benefits redound to the greatest benefit of the worst off. How, though, are we to measure deviations from this society? Rawls (1999a, 38) proposes that we appeal to the above lexical ranking and to some additional "priority rules." The relevance of the lexical ranking is that when comparing societies that fail to satisfy Rawls's principles, a deviation from Equal Liberty is always greater than a deviation from Fair Equality of Opportunity, and that each is always greater than a deviation from the Difference Principle. And to measure the extent of each type of deviation, Rawls (1999a, 266) suggests the following "priority rules":

> Equal Liberty: (a) a less extensive liberty must strengthen the total system of liberties shared by all; (b) a less than equal liberty must be acceptable to those with the lesser liberty.
>
> Fair Equality of Opportunity: an inequality of opportunity must enhance the opportunities of those with the lesser opportunity.[3]

[2] I paraphrase these principles slightly for simplicity, including by ignoring a reference to Rawls's "just savings principle."

[3] Rawls also includes a priority rule asserting the priority of justice over efficiency and another

So, as Rawls's exposition makes clear, a society that fails to fully realize Equal Liberty deviates less from the ideal than another if it either (a) provides more basic liberties to all or (b) provides more basic liberties to those with less in cases where some have more basic liberties than others. Since one way to provide more basic liberties to all is to provide more basic liberties to the worst off, this is equivalent to the idea that a society deviates less from Equal Liberty to the extent that it provides more basic liberties to the worst off (or to the second worst off in cases of ties for worst off, and so on). Likewise, a society deviates less from Fair Equality of Opportunity if it provides more opportunities to the worst off (or to the second worst off in cases of ties for worst off, and so on). And, though Rawls never explicitly states a priority rule for the difference principle, he is also clear that a society deviates less from the difference principle to the extent that it provides more social and economic benefits to the worst off (or to the second worst off in cases of ties for worst off, and so on) (Rawls 1999a, 279-280).

Putting Rawls's lexical ranking and priority rules together, then—and ignoring cases of ties for simplicity—we come to the following theory of comparative justice. A society is more just if it provides more basic liberties to the worst off. When basic liberties are not at stake, a society is more just if it provides more opportunities to the worst off. And when neither basic liberties nor opportunities are at stake, a society is more just if it provides greater social and economic benefits to the worst off. Though this is a more complicated theory of comparative justice than the last two we considered, it is again one that makes no essential reference to the ideal: to determine which of two societies is more just, we need only consider which provides more basic liberties, opportunities, or social and economic benefits to the worst off, and so need not refer to Rawls's ideal principles or the ideally just society that would

rule intended to cover his "just savings principle" which I set aside here.

realize them. So even Rawls falls prey to the redundancy objection. His ideal deviation criteria are equivalent to essentially comparative criteria, rendering the ideal redundant.[4]

In all three cases, then, we find the same pattern mentioned earlier. A benchmark theorist begins with the view that a society is unjust to the extent that it deviates from the ideal, but in the process of working out how to measure deviations from the ideal ends up defending an essentially comparative criterion (or set of criteria) that makes the ideal redundant. And other attempts in the literature, even by those sensitive to Sen's criticisms, meet a similar fate. For example, Philip Pettit defends a principle of "equal freedom as non-domination" on which ideal justice is achieved when everyone is nondominated—in roughly the sense that they can live as they please rather than by the leave of others with arbitrary power over them—and holds that societies are unjust to the extent that they deviate from this ideal (Pettit 2012, 124-125). Citing Sen, Pettit notes that such deviations can occur along various dimensions. But when he spells out the ideal deviation criteria that result from fleshing out these dimensions, they are all equivalent to essentially comparative ones: people may enjoy freedom as nondomination over more or less important choices, these freedoms may be better or worse entrenched, and so on (Pettit 2012, 125). Indeed, despite Pettit's adoption of the benchmark view, he appears to admit that, when evaluating nonideal societies, no reference to the ideal is needed. As he and his co-author José Luis Marti put it: "In any context the ideal

[4] A qualification: Rawls's theory of comparative justice—like his theory of ideal justice—only applies in "reasonably favorable conditions" (that is, to societies that avoid extreme economic, technological, or cultural deprivation). Although questions of justice arise outside of these conditions, "[t]he principles and their lexical order were not acknowledged with these situations in mind and so it is possible that they no longer hold" (Rawls 1999a, 216).

will argue for that improvement, assuming there is one available, that makes for a higher degree of nondomination overall" (Martí and Pettit 2010, 158). And verdicts about a society's degree of nondomination overall make no reference to the ideal.

To be clear, these examples don't demonstrate that no plausible criteria of comparative justice make essential reference to the ideal. But they do provide the beginnings of an inductive argument suggesting it would be surprising if the benchmark view held. Not only are many well-known theories of justice essentially comparative (say, utilitarianism, perfectionism, and various forms of pluralism), but the most explicit attempt (Christiano and Braynen 2008) and the best-known attempt (Rawls 1999a) to defend ideal deviation criteria that make essential reference to the ideal fall prey to the redundancy objection, as does an attempt by a theorist aiming to address Sen's worries (Pettit 2012). And, try as I might, I have been unable to find a single example of a benchmark theorist who articulates an ideal deviation criterion that doesn't similarly turn out to be equivalent to an essentially comparative one, making the ideal redundant. The upshot is that, at least among the benchmark theorists I am aware of, all fall prey to the redundancy objection, so none vindicate the theoretical priority of ideal theory over problem solving.

Still, it is worth considering whether all ideal deviation criteria are doomed to redundancy, or whether there is some principled reason we should expect this of any plausible ideal deviation criterion. Are the examples I have chosen representative, or might some ideal deviation criteria genuinely vindicate the benchmark view?

5. Generalizing the Redundancy Objection

It turns out that this question has a less straightforward answer than we might hope. On the one hand, there are indeed some logically possible ideal deviation criteria that are not equivalent to essentially comparative criteria. Endorsing such a criterion would, in principle, allow a benchmark theorist to avoid the redundancy objection. But on the other hand, criteria of this sort have a peculiar structure that renders them explanatorily deficient. In particular, any ideal deviation criterion that is not equivalent to an essentially comparative criterion implies the existence of a brute "sweet spot" at which continuing to change a society in a way that, until that point, makes things more just, suddenly and inexplicably ceases to improve justice. Such criteria are therefore implausible.

To begin with the claim about logical possibility, take any dimension *d* along which we can arrange societies from less to more *d*. There are two possible essentially comparative criteria that employ this dimension: one claiming that a society is more just whenever it is more *d*, and another claiming that a society is more just whenever it is less *d*. The only way for a benchmark theorist to defend an ideal deviation criterion that uses this dimension but is not equivalent to an essentially comparative criteria is for them to deny both claims. This requires them to endorse a *sweet spot criterion* on which making a society more *d* (or less *d*) only makes things more just up until a certain sweet spot, after which continuing to make society more *d* (or less *d*) either begins to make things less just or stops making any difference to justice at all.

For example, if—just to take a simple and familiar case—the relevant dimension is one arranging societies from less to more total well-being, then a benchmark theorist may say that the total level of well-being in the ideally just society serves as a sweet spot, such that, in measuring deviations from the ideal, a society can deviate by either having too much or too little well-being: if the ideal has *n* well-being, then a society with *n*-1 units of well-being deviates as much from the ideal as one with *n*+1 units. Or, alternatively, a benchmark theorist may claim that increasing well-being past the ideal, sweet spot value of *n* ceases to have any impact on justice, such that a society with *n*+1 units of well-being is no more just than one with *n*.

Unlike the utilitarian criterion, which yields comparative evaluations by referring to the total well-being in each of two societies and asking which is greater, such sweet spot criteria really do make essential reference to the ideal: without referring to the level of well-being realized at the ideal, they cannot yield verdicts of which of two nonideal societies is more just simply by looking at each society's total well-being. After all, if we were to continue increasing a society's total well-being, we would not continue to improve justice but would eventually overshoot the ideal value. And where, precisely, this overshooting occurs cannot be determined without referring to the level of well-being in the ideally just society—that is, without referring to the ideal or sweet spot value of *n*.

I have defined sweet spot criteria in such a way that they represent the only logically possible sort of ideal deviation criteria that are not equivalent to essentially comparative criteria, and so the only logically possible ideal deviation criteria that can avoid the redundancy objection and vindicate the benchmark view. But this vindication is pyrrhic: sweet spot criteria are highly implausible. The problem is not that they posit the existence of a sweet spot, but that they posit the existence of a *brute* sweet spot for which no explanation seems available. If increasing *d* only improves justice up until a sweet spot *n* after which further increases in *d* set back justice or at least no longer improve it, then there must be some explanation of this. But, as I will now argue, benchmark theorists can provide such an explanation only by reopening themselves to the redundancy objection, since the sort of explanations on offer for the existence of sweet spots presuppose that we are operating with an essentially comparative view. It follows that benchmark theorists can avoid the charge of redundancy only by running into the charge of explanatory deficiency.

There are, as far as I am aware, three ways to explain the existence of a sweet spot in some justice-relevant dimension—say, why *n* is the ideal or sweet spot value of well-being.

The first is that, thanks to certain feasibility constraints (say, certain facts about human nature) it is impossible to have more than *n* well-being. Or, if we take another dimension such as one referring to the level of oppression in society, we might claim that zero oppression is ideal because we here run into a constraint of logical possibility: it is logically impossible, and so infeasible, to have less than zero oppression. This explanation is straightforward and is readily available to an essentially comparative theorist who wishes to explain why they are committed to the existence of a sweet spot. But it is not available to the benchmark theorist, because if a sweet spot criterion is to avoid making the ideal redundant, it cannot merely claim that continuing to increase *d* past some point fails to increase justice because doing so is infeasible, but must rather claim that continuing to increase *d* past some point ceases to make things more just even when this is feasible. Otherwise, such a criterion ends up equivalent to an essentially comparative criterion: it simply claims that, whenever it is feasible, increasing *d* always makes things more just, such that there is no need to refer to the ideal when making comparative evaluations of justice with respect to dimension *d*.

A second explanation of why a sweet spot might exist in some justice-relevant dimension is available to those who adopt a pluralistic theory of justice on which multiple criteria are relevant to our all-things-considered evaluations of justice. For example, such theorists might claim that *n* is the ideal value of well-being because, taking into account other criteria of justice—say, one again concerned with oppression—the society that best balances these criteria (given certain feasibility constraints) would have *n* well-being. But while this is a perfectly fine explanation of why a sweet spot might exist, it again presupposes an essentially comparative view and so renders a benchmark theorist vulnerable to the redundancy objection. For even if a benchmark theorist adopts a pluralistic view on which other criteria are relevant to justice besides the sweet spot criterion in question, they are still committed to thinking that, holding all other criteria fixed, deviating further from the ideal in the dimension picked out by their sweet spot criterion makes things less just. A sweet spot criterion that doesn't make the ideal redundant must therefore claim that even when it is feasible to increase *d* without this increase in *d* conflicting with any other criterion, increasing *d* past some sweet spot level ceases to improve justice. Otherwise, such a criterion again proves equivalent to an essentially comparative criterion: it simply claims that, holding all other criteria fixed, increasing *d* always makes things more just when such increases are feasible. So such a criterion can again yield comparative evaluations of justice without referring to the ideal.

A third explanation of why a sweet spot might exist in one dimension is that this is explained by some more fundamental criterion (or criteria) of justice concerning some more fundamentally justice-relevant dimension (or dimensions). If, for example, there is an ideal tax rate such that either raising or lowering this tax rate would make things less just, then this may be explained by other criteria: perhaps the sweet spot in our tax rate is explained by its maximization of well-being, or by its ability to optimally balance efficiency against equality. But this explanation won't help the benchmark theorist, in the first place, because our earlier discussion of Sen has revealed that ideal deviation criteria must refer to dimensions that are fundamentally relevant to justice, on pain of running into second-best problems. And, even setting this aside, explaining the existence of a sweet spot in one dimension by reference to other more fundamental criteria provides no solace to the benchmark theorist, since it implies that we don't really need to invoke their sweet spot criterion to make comparative evaluations of justice: we can appeal to the more fundamental criteria that explain the existence of the sweet spot in our less fundamental criterion. While such an explanation can satisfy essentially comparative theorists who are happy thinking that fundamental criteria are essentially comparative, it is therefore not available to benchmark theorists who deny this. It does nothing to explain the existence of a sweet spot in a fundamental criterion of justice, and so nothing to rebut the charge that fundamental sweet spot criteria are explanatorily deficient.

We have now examined three ways how, for some given dimension *d*, increasing *d* might make things more just up until some sweet spot *n*, after which further increases no longer generate improvements in justice. But we have also seen that to endorse any of these explanations is to abandon a sweet spot criterion that makes essential reference to the ideal in favor of one that is equivalent to an essentially comparative criterion on which, for some fundamentally justice-relevant dimension *d*, increasing *d* always makes things more just when it is feasible to do so without setting back other criteria. So, in order for an ideal deviation criterion to avoid making the ideal redundant, it must claim that, for some justice-relevant dimension *d*, making a society more *d* makes it more just until it hits some sweet spot where making a society more *d* suddenly ceases to make it more just—where the existence and position of this sweet spot is not explained by any feasibility constraint, by any conflict with other criteria, or by any more fundamental criteria. Unless a benchmark theorist can provide some fourth explanation for the existence and position of a sweet spot, they must therefore claim that it is simply a brute, inexplicable fact that ranking higher in some dimension that is fundamentally relevant to justice makes things more just until, suddenly, it doesn't. But it is difficult to imagine what this fourth explanation could be. And though I cannot rule out its possibility, it suffices to say that no benchmark theorist has ever even tried to provide one, and that until a benchmark theorist proves otherwise, it therefore seems reasonable to proceed on the assumption that no such explanation is available.

Our earlier observation that all ideal deviation criteria articulated by actual benchmark theorists make the ideal redundant should therefore come as no surprise. None commit themselves to a criterion that contains a brute, inexplicable sweet spot. Indeed, to head off a possible objection, it is worth noting that even theorists that defend sweet spots in other normative domains insist that these sweet spots are not brute facts but are rather explained in one of the ways we have considered. For example, though Aristotle famously claims that virtue involves a mean between excess and deficiency—such that courage, for example, involves being neither too fearful nor not fearful enough—even he explains such sweet spots by reference to the more fundamental criterion of practical wisdom. In the case of courage, practical wisdom determines that there are certain things one should fear (to a particular degree), and others one shouldn't fear, such that failing to hit the mean of courage involves fearing things one shouldn't fear, or not fearing things one should fear (to the right degree) (Aristotle 2000, 1006b). So it is this prior standard of what one should and shouldn't fear (to what degree) that explains what counts as either fearing too much or too little, rather than a brute fact about fear being good up until a sweet spot, and then suddenly bad. So even Aristotle doesn't posit a *brute* sweet spot, and this is what allows his view to retain whatever plausibility it has. As he puts it: though virtue is a "mean" in one dimension, it is an "extreme" in the dimension of practical wisdom, and its extreme position in the latter dimension is what explains the sweet spot in the former (Aristotle 2000, 1107a; compare Annas 1993, 59-61).

Crucially, my argument for this conclusion hasn't *assumed* that fundamental criteria of comparative justice are essentially comparative. It has rather taken seriously the benchmark theorist's claim that criteria of comparative justice have an ideal deviation structure, and asked what would have to be true of such criteria for them to avoid rendering the ideal redundant. And what it has found is not very reassuring for the benchmark theorist. If an ideal deviation criterion is to avoid the redundancy objection, it must contain a sweet spot, since, otherwise, it is equivalent to an essentially comparative criterion. There are three ways to explain the existence of sweet spots, but, upon examination, we find that these explanations are only available on the presupposition that fundamental criteria are essentially comparative. So benchmark theorists are committed to positing the existence of ideal deviation criteria that contain not only sweet spots, but brute, inexplicable sweet spots—rendering such criteria explanatorily deficient. In other words, benchmark theorists face a dilemma. If ideal deviation criteria don't contain brute sweet spots, then they are equivalent to essentially comparative criteria, and so make the ideal redundant. But if ideal deviation criteria do contain brute sweet spots, then they are implausible on grounds of explanatory deficiency, and so must be rejected. Either way, the benchmark view is defeated: *plausible* criteria of comparative justice make the ideal redundant. Even though we may articulate them in ideal deviation form, all plausible criteria turn out to be equivalent to essentially comparative criteria, capable of yielding verdicts about when one society is more just than another simply by referring to features of those two societies. The redundancy objection generalizes.

6. Facilitation and Obstruction

To this point, I have been concerned with rebutting the benchmark view, according to which criteria of comparative justice make essential reference to the ideal. This was important to do because, otherwise, ideal theory would have a strong claim to theoretical priority over problem solving, in the sense that we would first need to work out the former before engaging in the latter. For, as noted earlier, if problem solving requires us to invoke criteria of comparative justice—as I have granted—and if criteria of comparative justice make essential reference to the ideal—as the benchmark view claims—it follows that problem solving requires us to refer to the ideal. Now that we have seen that plausible criteria of comparative justice don't make essential reference to the ideal, this claim of priority falls by the wayside. But there remain the further possibilities that ideal theory might facilitate problem solving without being prior to it, as ideal theorists argue, or that it might instead obstruct problem solving, as critics retort.

To begin to investigate these possibilities, consider David Estlund's recent criticisms of two versions of a view he attributes to Sen, "categorical comparativism" and "methodological comparativism," which respectively hold that only comparative judgments of justice are true or meaningful, and that we should only ever appeal to comparative judgments about justice in political philosophy (Estlund 2016, 11-12). Against categorical comparativism, Estlund notes that many common judgments imply a partition between "just" and "unjust"—for example, that slavery is unjust or that perfect equality is (ideally) just—and that it is implausible to deny the meaningfulness or truth of all such non-comparative claims (Estlund 2016, 12). Against methodological comparativism, Estlund argues that even if our goal is to arrive at criteria of comparative justice, it would be a mistake to throw out our pre-theoretic judgments and intuitions about non-comparative justice en route to this goal, since such judgments are often clearer and more reliable than our comparative judgments. For example, we can make more reliable "eyeball" judgments about perfect equality than about relative equality, and, as a matter of the "epistemology of theory-building," judgments about perfect equality or ideal justice more generally often seem to inform our theorizing about comparative criteria (Estlund 2016, 26; see also, e.g., Estlund 2020, ch. 15, Gilabert 2012, 46-47, Stemplowska 2008, 336-338, Swift 2008, 372-378).

Estlund's criticisms are persuasive and provide a welcome warning against those who might otherwise take Sen's proposed turn toward comparative theorizing too far. Indeed, his rejection of categorical and methodological comparativism is plausible and consistent with everything I have said here. I have argued that ideal theory enjoys no relevant priority over problem solving, but this is a far cry from claiming that non-comparative judgments about justice are uniformly false, meaningless, or irrelevant to theorizing, such that we should swear off them altogether (in fact, I earlier suggested that our theorizing is often properly informed by non-comparative judgments like "slavery is unjust").[5] Still, it is one thing to note that non-comparative judgments can be true and relevant to comparative theorizing, and another to specify exactly what role ideal theory plays here. So it is worth disentangling different versions of the view that judgments about the ideal inform our criteria of comparative justice.

Consider first a strong version of this view. In response to my argument against the benchmark view, one might object that even if criteria of comparative justice make no essential reference to the ideal, a conception of the ideal might nevertheless *justify* these comparative criteria.[6] For example, even though Rawls's criteria of comparative justice don't refer to his ideal, perhaps his ideal principles justify his comparative ones, since they imply that his comparative principles must pick out a society realizing his ideal principles as most just. Similarly, on theories where injustice is a "defect" relative to ideal justice—such as a theory on which ideal justice involves the absence of oppression, domination, or inequality—ideal justice may be of special conceptual relevance since it implies that criteria of comparative justice must

[5] Nor does my argument against the existence of brute sweet spots commit me to thinking that can be no partition between "just" and "unjust": all I claim is that there aren't points at which continuing to change a society in some way that, until that point, improves justice, inexplicably stops having this effect. For example, if making a society more *d* makes the society more just until we hit a partition where the society switches from "unjust" to "just," this won't be a brute sweet spot if the partition is explained by the society meeting the upper limit of comparative justice. This corresponds to what Estlund (2016, 20) calls a "ceiling partition." Thanks to an anonymous reviewer for suggesting this clarification.

[6] Thanks to two anonymous reviewers for this objection and for the examples that follow.

pick out this ideal as their conceptual upper limit.

Now, if we interpret this objection as claiming that a conception of the ideal can *fully* justify our comparative criteria, then this can't be right. As we saw in our earlier discussion of Sen, all sorts of criteria of comparative justice are compatible with any given ideal, so it is impossible to invoke a conception of the ideal to justify the choice of any particular criterion of comparative justice. For example, there are various ways to measure deviations from Rawls's ideal that conflict with his proposed "priority rules," and I earlier flagged several ways we might measure deviations from an ideal of no oppression—showing that simply specifying that this is the ideal doesn't yield a justification for any given way of comparing the justice of societies that contain some oppression. The same is true for all other conceptions of the ideal, including those that understand ideal justice in terms of the absence of defects. To call back to Sen (2009, 99): "The characterization of spotless justice…[does] not entail any delineation whatever of how diverse departures from spotlessness would be compared and ranked."

That said, a weaker version of the view in question survives. While we cannot use Rawls's ideal to justify any particular criteria of comparative justice, his ideal principles do imply a constraint on any comparative criteria: whatever else is true of them, they must pick out a society realizing his ideal principles as most just. Similarly, if the ideal society conceptually involves no oppression, this implies a conceptual constraint on our comparative criteria: they must pick out a society involving no oppression as their conceptual upper limit. Or, to take a more concrete example, suppose we develop a conception of an ideally just society in which race has no influence on one's treatment or life prospects and laws are "color-blind," making no reference to race at all. This is clearly an important data point when working out our criteria of comparative justice since it places one constraint on them: they must be able to explain why a color-blind society is ideal.

The upshot is that while a conception of the ideal can't fully justify any particular criteria of comparative justice, it can nevertheless inform our criteria of comparative justice by *constraining* them in this way. Since ideal theory is in the business of delivering such a conception, this suggests one way ideal theory might facilitate problem solving. In a moment, we will see that this is not as significant a concession to ideal theorists as it might seem. But first, let us briefly consider how else ideal theory might facilitate problem solving.

Although I have been focusing primarily on abstract criteria of comparative justice throughout this paper, recall that the ultimate goal of problem solving is not merely to develop such criteria, but to apply them by working out solutions to problems or concrete ways of making our society more just. This suggests two further ways that ideal theory might facilitate problem solving. First, ideal theory might help to expand our understanding of the space of social possibilities, thereby helping us to come up with new and creative solutions (e.g., Rawls 2001, 4, Gilabert 2012, 47). For example, in thinking through a market socialist ideal in which a free market sets individuals' pre-tax income, redistributive taxation equalizes post-tax income, yet individuals remain motivated to work by an egalitarian "ethos" (Carens 1981, Cohen 2008, 189-196), we may come to recognize the possibility of using similar ethos-based solutions to solve more immediate problems that we have historically attempted to solve via material incentives. Second, since the concrete problems we face—the ways our society is less just than it would otherwise be—are themselves often hidden from view or difficult to understand, ideal theory might help us to identify or appreciate such problems, thereby guarding against a complacent acceptance of the status quo (e.g., Gilabert 2012, 47, Stemplowska 2008, 332-333, Estlund 2020, ch. 13). If, for example, we spell out a conception of an ideally just society free of gender oppression, then comparing this society to our actual society may help us notice the ways our current gender relations fall short. Perhaps in the ideally just society individuals of different genders wouldn’t only enjoy formal equality under the law, but equality with respect, say, to patterns of economic dependence, and realizing this might help open our eyes to such patterns in our own world.

These three facilitating roles of ideal theory may seem uncontroversial enough: ideal theory can help us to develop our criteria of comparative justice, to come up with new solutions, or to notice existing problems. But against each role, critics have warned that ideal theory may obstruct problem solving more than it facilitates.

Against the first role, critics argue that focusing on the ideal may distort our theorizing about criteria of comparative justice, in much the same way that attempting to fit a scientific theory to a small set of data points may lead to a distorted theory (compare e.g., Anderson 2010, 4-5, Goodin 1995, 45-55, Schmidtz 2011, 775-778). For example, though a color-blind ideal may put a constraint on our criteria of comparative justice—they must be able to explain why this is ideal—there are many criteria that can meet this constraint, and focusing too much on the ideal may lead us to endorse one that most simply explains why this is ideal rather than one that best fits our convictions about a variety of actual and hypothetical nonideal societies. In the case at hand, we may jump to the conclusion that we should accept a criterion of comparative justice on which color-blind policies are always more just than color-conscious policies. But it might turn out that had we thought carefully about a variety of nonideal cases, we would have come to realize that the actual criterion that explains why a color-blind society is ideal (if it is) also explains why color-conscious policies are more just than color-blind ones in some nonideal circumstances (if they are)—say, because this criterion concerns equality of opportunity (compare Anderson 2010, 4).

More generally, while judgments about the ideal constrain our criteria of comparative justice, this doesn't suggest any privileged role for ideal theory, since our judgments about all sorts of other actual and non-hypothetical societies play an identical constraining role: if we judge that one nonideal society is more just than another, then this similarly constrains us to picking criteria of comparative justice that match this judgment. Nor does ideal theory enjoy a privileged role because we must in any sense start with our judgments about the ideal, at least if we adopt anything like Rawls's own method of reflective equilibrium: much as we need to reject criteria of comparative justice that have implausible implications about the ideal, so too must we reject conceptions of the ideal that imply implausible constraints on our handling of nonideal cases. And again, even if we grant Estlund's claim that pre-theoretic "eyeball" judgments about the ideal are often especially clear and reliable, this doesn't imply a special role for ideal *theory*, whose verdicts are anything but pre-theoretic.

Turning to the second role, though theorizing about the ideal is one way to expand our understanding of the space of social possibilities, critics argue that ideal theory is liable to focus us on less relevant regions of this space: on unrealistic rather than viable solutions (e.g., Anderson 2010, 3-4, Schmidtz 2011, 775-778, Wiens 2012, 48-53). For example, though the above variant of market socialism is interesting, the history of attempts to structure an economy without relying on material incentives should certainly give us pause, and make us turn to the various ways that actual societies have dealt with their problems. More generally, we often develop more promising solutions by beginning with a careful analysis of the causal workings of the problems we face and by brainstorming various ways we might fix them—or by engaging in cross-historical and -societal surveys of how other past and present societies have tried to solve similar problems—than by imagining an ideally just society free of such problems altogether. And in this regard, ideal theory may distract rather than facilitate. We may explore more relevant regions of the space of social possibilities by focusing directly on our problems and by exploring the regions of this space that correspond to social actualities than by attempting to figure out the ideal point in this space.

Third, though it is again possible that comparing an ideally just society to our own helps us notice or better understand ongoing problems, critics claim that, in practice, focusing on the ideal is more likely to blind us to such problems than to reveal them. In the first place, we often develop our conception of the ideal in response to problems we have already identified. Indeed, some argue that this it not only typically but necessarily true of our moral and political thinking (see Anderson 2010, 3, who follows Dewey 1922, especially 260-261). But even if a conception of the ideal sometimes helps us to appreciate problems in our current society, it strains plausibility to claim that turning away from our actual, problem-ridden society and focusing instead on trying to identify an ideally just, problem-free society is generally a better way of identifying and understanding our problems. For example, if the ideal would contain no oppression, then ideal theory will involve no theorizing about the workings of oppression, and so will leave us ill-equipped to identify or understand the subtle and insidious forms of oppression that pervade actual societies.

The most famous proponent of this idea, Charles Mills, goes further, arguing that the disciplinary dominance of ideal theory serves as an *ideology* "in the pejorative sense of a set of group ideas that reflect, and contribute to perpetuating, illicit group privilege" (Mills 2005, 166). Writes Mills (2005, 172):

> Can it possibly serve the interests of *women* to ignore female subordination…? Obviously not. Can it possibly serve the interests of *people of color* to ignore the centuries of white supremacy…? Obviously not. Can it possibly serve the interests of the *poor and the working class* to ignore the ways in which an increasingly inequitable class society imposes economic constraints that limit their nominal freedoms…? Obviously not. If we ask the simple, classic question of *cui bono*? then it is obvious that ideal theory can

> only serve the interests of the privileged.

So, Mills argues, when we engage in ideal theory and imagine a society free from oppression and other injustice, 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… will never be achieved" (Mills 2005, 170).

Mills overstates his case. It is not plausible that the dominance of ideal theory in political philosophy really *guarantees* that we will never reach the ideal. But his weaker claim that the widespread practice of ideal theory hinders the pursuit of greater justice, or at least our theorizing about it, has more bite. And the same is true of the other two criticisms. In each case, critics go too far if they deny that ideal theory ever helps us to develop our criteria of comparative justice, to come up with new solutions, or to appreciate ongoing problems. Ideal theory can certainly play these roles. Yet at the same time, critics identify genuine limits of ideal theory that warn against an overreliance on it. Our judgments about the ideal serve as relevant data points that inform our criteria of comparative justice, to be sure, but we risk overfitting our theory to this data if we fail to investigate actual or hypothetical nonideal societies that provide equally important data. Ideal theory can help us to explore the space of social possibilities by probing its outer limits, but this is no substitute for engaging in empirical analysis of the problems we face and of actual attempts to solve similar problems. And though a conception of the ideal can sometimes help us to guard against complacency and to notice instances of injustice, the risk of complacency also arises for those who spend their time gazing at the ideal rather than examining and gathering evidence about problems we currently face.

Now, it is difficult to articulate precisely to what extent ideal theory is useful, or at what point relying on it becomes problematic. But the limited relevance of ideal theory to problem solving nevertheless comes out clearly when we note that it has no more claim to relevance than theorizing about dystopia. Much like a conception of the ideal, a conception of dystopia may help us to develop our criteria of comparative justice by providing us with new data points, or to explore the space of social possibilities by alerting us to ways that "solutions" might work out horribly—just think of how adeptly Kurt Vonnegut's "Harrison Bergeron" illustrates the folly of criteria that license attempts to resolve inequality through leveling down, or how George Orwell's *1984* warn us of the dangers of totalitarianism. And theorizing about dystopia can similarly help us to notice problems in our current society through alerting us to resemblances rather than gaps—here, Margaret Atwood's *Handmaids Tale* stands out for its ability to alert us to existing gender inequalities through its depiction of a dystopian society in which such inequalities are exaggerated yet still alarmingly recognizable. But despite its ability to sometimes play these helpful functions, no one would claim that dystopian theorizing plays, or should play, a central role in problem solving. And ideal theorists provide us with no grounds for viewing ideal theory any differently.

7. Conclusion

If my arguments are successful, then ideal theory plays a modest, ancillary role in problem solving. Ideal theory enjoys neither epistemic nor theoretical priority over problem solving, and working out a conception of the ideal is not centrally important to problem solving, but is rather something we might sometimes find helpful along the way—just like it may sometimes be helpful to think about dystopia, or, for that matter, about history, or science fiction scenarios, or anything else that helps us focus our minds on relevant social phenomena.

Still, for all I have said here, ideal theory might derive additional value by some other route. Although I have argued elsewhere that ideal theory cannot provide us with a long-term goal for reform (Barrett 2020), this leaves open the possibility that ideal theory—like many other forms of philosophy and abstract theorizing—might be valuable for some non-practical reason (Estlund 2014, Estlund 2020, chs. 16-17).[7] Perhaps it is intrinsically important, or of intrinsic interest, to understand what ideal justice would be. So my point is not that political philosophers shouldn't do ideal theory. It is just that they shouldn't do so under the misconception that it is a necessary precondition of, or of central relevance to, working out how to solve the pressing problems of injustice we face.

Acknowledgements

For helpful comments and discussion, I would like to thank Matthew Adams, Samuel Bagg, Allen Buchanan, Tom Christiano, Jerry Gaus, Christopher Howard, Sarah Raskoff, Connie Rosati, Santiago Sanchez Borboa, David Schmidtz, two anonymous reviewers, and participants in the Philosophy Work in Progress group at the University of Arizona.

[7] Though see Barrett (2022) for some worries about this strategy of defending ideal theory. An alternative strategy is to argue that ideal theory is indeed practically relevant, not because it provides a long-term goal or facilitates problem solving, but for some less commonly discussed reason. For example, an anonymous reviewer suggests that ideal theory may be practically relevant because the extent to which a society deviates from the ideal determines how severely unjust it is, which bears in turn on when individuals are permitted to resist unjust institutions, which reforms are permissible, and the cost agents must shoulder in pursuing reforms. At first pass, I find it hard to see why the ideal is relevant here: insofar as the severity of injustice is practically relevant, why isn't it enough to appeal, say, to how much less just certain institutions are than their absence, or to how much less just the status quo is than feasible reforms? But regardless, my argument doesn't rule out such alternative avenues to practical relevance.

References

Anderson, E. (2010). *The Imperative of Integration*. Princeton: Princeton University Press.

Appiah, K. A.. (2017). *As if: Idealizations and Ideals*. Cambridge: Harvard University Press.

Annas, A. (1993). *The Morality of Happiness*. New York: Oxford University Press.

Aristotle. (2000). *Nicomachean Ethics*. (R. Crisp, Trans.). Cambridge: Cambridge University Press.

Barrett, J. (2020.) Social Reform in a Complex World. *Journal of Ethics and Social Philosophy,* 17, 103-132. doi: 10.26556/jesp.v17i2.900

Barrett, J. (2022). *Utopophobia: On the Limits (If Any) of Political Philosophy*, by David Estlund. *Mind*, 131, 691-700. doi: 10.1093/mind/fzaa087

Carens, J H. 1981. *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*. Chicago: University of Chicago Press.

Christiano, T & Braynen, W. (2008). Inequality, Injustice and Leveling Down. *Ratio*, 21, 392-420. doi: 10.1111/j.1467-9329.2008.00410.x

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge: Harvard University Press

Dewey, J. (1922). *Human Nature and Conduct*. New York: Henry Holt and Company.

Estlund, D. (2014). Utopophobia. *Philosophy & Public Affairs*, 42, 207-235. doi: 10.1111/papa.12031

Estlund, D. (2020). *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton: Princeton University Press.

Estlund, D. (2016). Just and Juster. *Oxford Studies in Political Philosophy*, 2, 9-32. doi: 10.1093/acprof:oso/9780198759621.003.0002

Gaus, G. (2016). *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press.

Gilabert, P. (2012). Comparative Assessments of Justice, Political Feasibility, and Ideal Theory. *Ethical Theory and Moral Practice*, 15, 39-56. doi: 10. 1007/sl0677-01 1-9279-6

Goodin, R. E. (1995). Political Ideals and Political Practice. *British Journal of Political Science*, 25, 37-56. doi: 10.1017/S0007123400007055

Lipsey, R. G. & Lancaster, K. J. (1956). The General Theory of the Second Best. *The Review of Economic Studies*, 24, 11-32. doi: 10.2307/2296233

Martí, J. L., & Pettit, P. (2010). *A Political Philosophy in Public Life: Civic Republicanism in Zapatero's Spain.* Princeton: Princeton University Press.

Mills, C. W. (2005). 'Ideal Theory' as Ideology. *Hypatia*, 20, 165-184. doi: 10.1111/j.1527-2001.2005.tb00493.x

Pettit, P. (2012). *On The People's Terms: A Republican Theory and Model of Democracy.* New York: Cambridge University Press.

Rawls, J. (1999a). *A Theory of Justice* (Rev. ed.) Cambridge: Belknap Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement.* Cambridge: Belknap Press.

Rawls, J. (1999b). *The Law of Peoples.* Cambridge: Harvard University Press.

Schmidtz, D. (2011). Nonideal Theory: What it is and What it Needs to Be. *Ethics,* 121, 772-796. doi: 10.1086/660816

Sen, A. (2009). *The Idea of Justice.* Cambridge: Belknap Press.

Shelby, T. (2013). Racial Realities and Corrective Justice: A Reply to Charles Mills. *Critical Philosophy of Race*, 1, 145-162. doi: 10.5325/critphilrace.1.2.0145

Simmons, A. J. (2010). Ideal and Nonideal Theory. *Philosophy & Public Affairs*, 38, 5-36. doi: 10.1111/j.1088-4963.2009.01172.x

Sinnott-Armstrong, W. (1999). Some Varieties of Particularism. *Metaphilosophy*, 30, 1-12. doi: 10.1111/1467-9973.00108

Swift, A. (2008). The Value of Philosophy in Nonideal Circumstances. *Social Theory and Practice,* 34, 363-387. doi: 10.5840/soctheorpract200834322

Valentini, L. (2011). A Paradigm Shift in Theorizing about Justice? A Critique of Sen. *Economics & Philosophy*, 27, 297-315. doi: 10.1017/S0266267111000228

Stemplowska, Z. (2008). What's Ideal about Ideal Theory? *Social Theory and Practice*, 34*,* 319 340. doi: 10.5840/soctheorpract200834320

Wiens, D. (2015). Against Ideal Guidance. *The Journal of Politics*, 77, 433-445. doi: 10.1086/679495

Wiens, D. (2012). Prescribing Institutions Without Ideal Theory. *The Journal of Political Philosophy*, 20, 45-70. doi: 10.1111/j.1467-9760.2010.00387.x

Wiens, D. (2020). The General Theory of Second Best is More General Than You Think. *Philosophers' Imprint*, 20(5), 1-26.