

# The Ideal, the Neighborhood, and the Status Quo: Gaus on the Uses of Justice\*

*David Estlund*

## I. INTRODUCTION

Many political philosophers and theorists have been reflecting recently on methodological questions about realistic-ness, “realism,” idealization, and idealism in normative theoretical work about social justice.<sup>1</sup> A book called *The Tyranny of the Ideal* seems clearly to signal sympathy with opponents of “ideal theory” (which notoriously means multiple things, but still).<sup>2</sup> While there is some truth to that, one can also say that, like much in the book, it’s complicated. It is a nuanced engagement with the role of the ideal in political philosophy. A central problem that Gaus sets himself (call this the first question) is this: can there be a conception of normative theorizing about justice that (a) is practically relevant in making social decisions and (b) makes ineliminable use of the idea of full or ideal justice, (c) even when that ideal is unattainable? The problem is posed by Amartya Sen, who argues that this is harder than it might seem. Social decisions depend on rankings of available alternatives, and that’s all. What practical use, Sen asks, would an ideal of justice have beyond those rank-

\* This is a critical study of Gerald Gaus, *The Tyranny of the Ideal: Justice in a Diverse Society* (Princeton, NJ: Princeton University Press, 2016). Parenthetical citations in the text are to this book.

1. For a critical overview, see Zofia Stemplowska and Adam Swift, “Ideal and Nonideal Theory,” in *The Oxford Handbook of Political Philosophy*, ed. David Estlund (Oxford: Oxford University Press, 2012).

2. Laura Valentini, “Ideal vs. Non-Ideal Theory: A Conceptual Map,” *Philosophy Compass* 7 (2012): 654–64.

ings of available social worlds indicating which of them are more just?<sup>3</sup> Some have argued that it is important to stay on track in moving toward full justice,<sup>4</sup> and it is clear that steps that improve justice in the short run might take a society off that track. That suggests an important role for understanding full justice, and—contra Sen—not only what is more just than what. Gaus calls this central dilemma “The Choice”: “We often must choose between relatively certain (perhaps large) local improvements in justice and pursuit of a considerably less certain ideal, which would yield optimal justice” (81).<sup>5</sup> Gaus develops a version of that idea, adding to the dimension on which one scenario is more just than another (I will call that comparative justice) a second dimension, that of being farther from or “closer . . . to the ideally just condition” (let’s call that “ideal similarity”). This “multidimensional analysis allows us to see why the ideal would be necessary in orienting our judgments about justice” (11). It is a very robust and remote conception of the social ideal whose practical value Gaus is defending against Sen’s point: even unattainable perfect social justice. That makes the thesis—that it indeed can have a legitimate place in a practically useful theory of social choice oriented to justice—a bold and especially interesting one, and it is, as far as that one point goes, a limited defense of ideal theory.

Pushing back in the other direction, however, the book also develops, among other things, two central convictions. In what I will call the second question, pressed against one kind of idealistic theory associated with the Rawlsian idea of a just, “well-ordered society,” Gaus questions the value of consensus: roughly, he asks, isn’t it an undesirably homogeneous conception of the good society that contains within such a society a broad consensus on a conception of justice? Then, in a third question, against another strand of idealistic political philosophy, he argues against seeking to realize the ideal: roughly, shouldn’t political practice be conservative in a certain way—cautiously averse to structural change that is in the pursuit of social justice?<sup>6</sup> While there is much more going on than this, a central thread of argument goes roughly like this: while

3. Amartya Sen, “What Do We Want from a Theory of Justice?,” *Journal of Philosophy* 103 (2006): 215–38.

4. A. John Simmons, “Ideal and Nonideal Theory,” *Philosophy and Public Affairs* 38 (2010): 5–36. Simmons detects this line of argument in John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press of Harvard University Press, 1971), and is sympathetic.

5. Simmons states the tension between those criteria in “Ideal and Nonideal Theory,” 21.

6. This is the straightforward meaning of “conservative” operative in this review: some degree of resistance to structural social change as such, largely irrespective of what is true in the status quo. This kind of conservatism has no connection to the adherence to certain values sometimes called “conservative,” such as the importance of local community, nuclear family, relatively unregulated markets, homogeneity, or any such thing.

- i) we can make sense of a kind of ideal theory in which even an unattainable social ideal orients political practice (by counseling similarity to the ideal),
- ii) social moves toward the ideal or to worlds more similar to it tend to be uncertain and ill-advised, especially if there is a more certain, even if lesser, improvement available, and,
- iii) anyway, since the idea of full social justice typically includes a degree of social consensus that would never be advisable, the very idea of a social ideal “border[s] on incoherence” (xvii).

The conclusion is that a theory of ideal or full social justice, even if it had potentially orienting content, would tend often to be an inappropriate ingredient in deliberation about social choices. Indeed, there is a further claim, which we will look at in passing, that such a theory, not to mention even less practical ideal theory, is dangerous. I will raise challenges to each of these claims, which correspond to the questions I have called first, second, and third above.

## II. JUSTICE AND PRACTICALITY

Readers will naturally wonder where Gaus’s project is positioned with respect to the recent widespread interest in so-called ideal and nonideal theory. To partly address that, before considering his treatment of the three issues described, we should note a fourth strand of idealistic theory which Gaus briefly mentions and puts aside. Gaus counts himself “as immensely fortunate” that Marxism and its threat to the (liberal, democratic) Open Society are “no longer a practical worry,” but there is a qualification: “But within the academy . . . the allure of the ideal is as powerful as ever.” This takes us close to the core aim of the book, and to understand it better, it is important to draw out a couple of distinctions. Some recent writers (in this strand, Gaus attends mainly to work of G. A. Cohen and of the present author) have argued that sound standards of social justice are not guaranteed to lie within the bounds of what is likely, or feasible, or even orienting to realistic social choice under realistic conditions. That is, (it is argued that) the social worlds that lie within those boundaries of (let’s loosely call it) practicality might all be unjust to some degree. For that reason, there are philosophical questions about the nature of social justice that are not so constrained. It is important to notice that unlike certain critics of “ideal theory,” Gaus allows, along with some such idealistic theorists, that humans might indeed be too corrupt for justice (144).

Other critics of the idea that there is justice beyond those constraints of practicality—and here Gaus does count himself among the critics—believe that ideal justice seems bound to be too far from our experience

to be solidly within our intellectual ken. For that reason, thinking about it might or might not be available to help orient practice in some way (more on that shortly), but practice ought not aim at realizing it. Here, notably, Gaus agrees again with one of the main points of defenders of idealistic theory: ideal justice might well be such that society ought not, under the circumstances, aim to realize it.

There remains the question whether justice might lie beyond even the merely orienting kind of practical value. Some, as he says, hold that “a theory of justice may be the true theory, even if it has no practical value. Others have denied this—it is certainly a live issue. . . . For the most part, I set it aside” (15). That question, then, is independent of the several concerns of Gaus’s that I have described. The debate Gaus bypasses here would concern whether or not justice is bound to lie within the space this leaves. With that question put aside (which is fair enough), we can turn to the three questions that will be my focus: First, is there a role for a theory of ideal social justice in rational social choice, rather than merely rankings of comparative justice? Second, is the ideal of a consensus around a conception of social justice a bad idea? Third, do uncertainty and unfamiliarity about dramatically more just worlds warrant a strong conservative presumption against precipitous social change?

### III. ORIENTATION AND SIMILARITY TO THE IDEAL

Simmons argues, following Rawls, that we should care not only about which steps improve justice but also about staying on track toward much greater justice. It is clear that these two goals will sometimes conflict, but the point is enough to show that Sen is not obviously correct that all we want to know is which available condition is more just than the others—comparative justice. Simmons also points out, “Some may believe that we simply cannot confidently make the [needed] social-scientific (or moral?) judgments. . . . Given such uncertainty, why should we not simply seize whatever local and comparative gains we can get?”<sup>7</sup> And, indeed, Gaus argues that the pursuit of great justice must be discounted to some (unspecified) extent, insofar as it involves more uncertainty about both the evaluations and the likely effects of the changes. Those are less weighty difficulties, though hardly absent, in the decision to make a small change for the sake of a modest improvement. The challenge is from Sen, and the general strategy of reply is from Simmons, and it is at this point that Gaus enters. He points out that the reply needs further specification, since it is not clear what getting “closer” to ideal justice might mean if it is distinct from getting more just. One possibility is staying on a path where full justice might be achieved by steps that preserve the feasibility of the

7. Simmons, “Ideal and Nonideal Theory,” 24.

next necessary step. For reasons I will not pursue, Gaus rejects that version. A second possibility is that the path is one of increasing structural similarity to the case of full justice, and this is the one that Gaus explores.

Consider, then, Gaus's argument that the practical "orienting" value of theoretical investigation of even unattainable social ideals can be saved from the clutches of Sen. It depends, as he says, on whether there is a dimension of value in realistic social choice that both (a) requires reference to full or ideal justice and (b) is not already entailed by the method of balancing the degree of a social world's justness with the probability (or more broadly "feasibility") of achieving it. He proposes for that purpose what he throughout calls a world's structural "proximity" or "closeness" to the ideal. What he means is proximity in similarity space, or, more simply, structural similarity.<sup>8</sup> I will switch to that vocabulary, lest the idea of a scenario's "being closer" to the ideal connote the wrong thing, such as the better accessibility of the ideal from that world, a distinct dimension that will also concern us.

The question I will focus on is whether a world's structural similarity to the ideal is indeed a dimension of value for purposes of social decision making. But before considering that question, first focus on the point about how that would do the trick, which I agree with Gaus that it would: if there were reason to care not only about how much more just some alternatives are than others (along with how likely, or accessible, or feasible the various alternatives are) but also, in its own right, about how similar the worlds are in structural terms to an ideally or fully just world, then we would have an obvious practical importance for understanding at least some structural features of full or ideal justice even apart from whether it could ever be achieved. That would be an answer to Sen's skepticism about the value of understanding full justice rather than merely the comparative justice relation. (Gaus will still give reasons for thinking that orientation in terms of the ideal will have very limited application, as I mentioned, but they are importantly different from the present point of Sen's, which the value of similarity to the ideal is meant to answer.)

It may be helpful to have a concrete example to illustrate the difference between increased comparative justice and similarity to the ideal. Suppose that an ideally just society has three branches of government: executive, legislative, and judicial. Now consider a social world that has only the executive—or in effect, collapsing all three into one. As long as it is not too disruptive in other ways, that society would be made structur-

8. He writes, "When I say that social world a is 'very close' to social world b, I shall mean that it is very similar to it. . . . We move toward the ideal by making our world more like it, by changing our institutions and background facts so that they better align with the ideal world" (52). In other parts of the argument it is important that the similarity space is modeled as relative to various "perspectives," which disagree about certain aspects of justice and its measurement, though that does not affect the present point.

ally more like the ideal society if it added an independent legislative branch containing two of the three in the ideal. Still, it is a wholly separate question whether this would also make the society more just. So these are distinct dimensions of comparison, as Gaus makes very clear.

It seems to me correct that ideal similarity does not collapse into comparative justice plus achievability. But the question then becomes what kind of value this is supposed to be—in what way it should matter to us, in the status quo, not only how much more just some of our alternatives are than others, along with how accessible they are, but also how structurally similar a world is (or how much more so than alternatives) to the social ideal. It is natural to think that this structural similarity of some world to the ideal will be relevant to how accessible the ideal would be from that world, but that appeal to “feasibility” is rejected by Gaus for reasons we cannot go into. So putting that consideration aside, why else might we care how similar some world is to the ideal? We might suppose that a world’s being more similar to the ideally just world will entail that it is more just. But then the dimension in play would not be similarity but only the comparative justice relation. No essential reference to the ideal is needed for that.

At this point we might pause and insist that there could still be a practical value in understanding similarity to the ideal if similarity does tend to bring along greater comparative justice, and so long as a world’s relative similarity to the ideal is more easily ascertained than directly asking which worlds are more just than which. That is, suppose that Sen says that there is no need to posit an ideal because there is no need to ascertain how similar any world is to it, and that is because all we need is a ranking of which worlds are more just than alternatives. The reply we might imagine is this: so long as we have reason to believe that ideal similarity correlates with the comparative justice relation—more similar to the ideal entails more just—ideal similarity is a tractable heuristic or proxy for the other relation. And in that way the practical value of positing the ideal is established, even if it is resting on the shoulders of dimensions that need no reference to the ideal if only they were directly tractable.

In many contexts certain kinds of similarity to something good can be relied on to correlate with betterness itself. If my chair is very comfortable, then as between two other chairs, it may be that the one that is structurally most like my chair will also be the more comfortable. There are questions, of course, about what counts as similarity, and which similarities are being measured: color similarity will not affect comfort, and the chair that is in some ways more structurally similar might be less comfortable if it is, in addition, much smaller than the less similar alternative. Similar questions arise for comparing social worlds with respect to structural similarity. But waive those, and grant that similarity to a good thing often correlates with betterness. Still, as is well known, including to Gaus,

this is not always true. As he says, “Increase in inherent justice and proximity to the ideal are distinct dimensions” (140). The problem of “second-best” reasoning is precisely that when there is some possible good arrangement whose goodness consists in its having several features, say, five of them, if not all five of them will be present, it is invalid reasoning to suppose that having as many of those five as possible will be the best option. That is one salient kind of structural similarity of one world to another: having more of that world’s structural features than other worlds do. That kind of similarity does not, as a general matter, reliably correlate with betterness, in this case superior justice. So the question remains: why should we think that similarity of a world to the ideally just world is a valuable dimension in comparing alternative worlds with an eye to justice?

In his definition of “neighborhood,” Gaus writes that “a neighborhood delimits a set of nearby social worlds characterized by relatively similar justice-relevant social structures” (80). Much of Gaus’s discussion is conducted under the assumption that the space of alternative worlds has a certain structure, a “moderately rugged landscape,” namely where in the similarity “neighborhood” of the status quo, one world A’s being more similar to a more just world C than another world B is will correlate with A’s also being more just than B (141). This assumes away the structure that we know real alternatives can have, and which prevents the general validity of the kind of second-best reasoning refuted above. Certainly it is true that in some environments similarity correlates with betterness. By analogy (putting it roughly), in some environments we could infer from “*x* is a mammal” to “*x* is a dog,” namely, a world in which the only mammals were dogs. But unless we are, in reality, in that special class of worlds, both inferences are invalid, the dog one and the justness one. The restriction to moderately rugged landscapes—ones in which neighborhoods are such that similarity to the ideal correlates with improved justice—is not meant to be anchored in any observation or argument about the structure of our real situation. Rather, such a structure is stipulated as part of the model, and the stipulation is a convenient mathematical simplification (mathematical because the spaces and dimensions in question are defined quantitatively), but it also imports just the structure that allows Gaus’s solution to Sen’s challenge to the practical relevance of the ideal. Within the “neighborhood” of the status quo, then, ruled out by stipulation are spaces in which moving to a world more similar to the ideal will not always be a move to a better world. Sen could apparently reply that he never denied such a mathematical or conceptual possibility, but that everything hangs on whether we have reason to believe that our social alternatives really have this structure. Gaus does not argue that we do, though he seems to think it is likely.

There is an important caution about Gaus’s modeling idea of a moderately rugged similarity space, though it is not an objection. That space



of similarity relations can be presented graphically as resembling a terrestrial landscape, with hills and mountains, the obligatory analogy in recent thinking about ideal theory. This maplike space, on which we imagine locating our society at some time, might also seem to suggest that we traverse some continuous path through this space, as on a terrestrial map. However, while there are indeed possible continuous paths through it, many actual and possible paths taken by societies cannot be represented as paths in this similarity space at all. This is because among our options at any given time may be states that are not very similar to the status quo—literal or figurative revolutions, to mention especially dramatic examples. In Gaus’s model of similarity space, these are represented as leaps to points at some distance in the modeled space. There is, in reality, no necessary correlation between an option’s availability (by which I mean its being a possible next state of society) and its “proximity” in similarity space—that is, similarity. Gaus, as we will see, will strongly recommend moving in (almost?) a continuous path, rather than making such leaps, but it is just that—a recommendation. This spatial model, while useful for certain purposes, can easily mislead a reader into seeing continuous paths as the only possibilities to be considered. Similarity landscapes are fundamentally different from topographical terrestrial maps on which one’s path must be continuous, beginning with a point immediately adjacent to the status quo. That terrestrial model works fairly well for a kind of space with which Gaus’s similarity space might be easily confused, which we might call “availability space”: any given point is a social state surrounded by immediately adjacent possible next states, with height of the terrain representing better justice. That looks much like similarity space, but it is fundamentally different since movement through it (unlike similarity space) must describe a path. We return to Gaus’s reasons for recommending relatively continuous paths below.

#### IV. AGAINST REALIZING THE IDEAL

Excellent or ideal justice will (let’s grant for purposes of argument) be a social condition highly dissimilar from our society’s status quo (pick a society). For that reason, Gaus argues, we will know much less about “the way such a society would actually work out” (21) than we know about our own, and less than we know about societies that are more similar to the status quo. For this reason, Gaus says, it is generally too risky to knowingly set a course that would probably make things less just than the status quo or than available alternatives, in the uncertain hope of great steps (of increasing similarity) toward the ideal that will supposedly make it worth it. Any efforts to improve similarity to ideal justice are defensible only if they are small changes toward conditions so similar to the status quo that we can be relatively sure we understand them nearly as well as we under-



stand the status quo. So, while there is a possible role for movement that is guided not by increasing justice but by moving toward the ideal (recall the answer to Sen), it is highly limited by great uncertainty about worlds dissimilar from the status quo.

Here is a weaker claim than Gaus is apparently defending, one that is hard to disagree with:

**Weak Conservatism:** Some conceptions of an ideally just society would be (if not wholly implausible) too conjectural to defensibly serve as a practical goal or orienting direction. If unnecessary injustice is risked or accepted as part of the plan, then setting out on that path is reckless.

This weak claim is beyond dispute, and so we might assume that Gaus means something stronger. On the other hand, the book is peppered with reminders of how fascist and communist conceptions of an ideal society were recklessly and catastrophically pursued in the twentieth century. Perhaps Gaus does, after all, believe that the weak conservative principle is in some danger of being forgotten or neglected in recent political philosophy—that some significant trends advocate practical political programs that would knowingly risk increasing injustice, or forgoing easy gains, for the sake of conjectural steps toward full justice that make it all worthwhile. For example, Gaus writes, “A century witnessing Stalin, Mao, and Pol Pot disastrously confirmed Mill’s judgment [DE: that revolutionaries such as Robespierre and St. Just are reckless]; Robespierre is insignificant compared to this utopian trinity. The worry that certain judgments of the unchanging ideal will give rise to recommendations for immediate implementation is by no means a ‘utopophobia’ of the liberal fallibilist” (88).<sup>9</sup> His argument here is not, so far as I can see, challenging any extant theory or approach, but seems (not just here but throughout the book) to be rehearsing Popper’s arguments for the Open Society against totalitarian-

9. Stalin also comes up elsewhere in pursuit of a minor recurring theme that should not escape mention. Gaus writes that (to quote at length) while it may seem that “ideal theorists in the academy today . . . would never think of taking political power in their own hands to pursue their visions of the ideal . . . (Note 67:) . . . perhaps they would *think* of it; recall Cohen’s remark that his ‘own inclinations are more liberal’ (*Rescuing Justice*, p. 186), and so he rejects the Stalinist approach to assigning occupations; he wrestles with freedom v. Stalinism in *Rescuing Justice and Equality*, chap. 5” (149). Gaus’s point here is Cohen’s supposed “wrestling,” a possible sign that he “would *think* of [taking political power].” This is one of several cases where Gaus obliquely or directly, but unconvincingly, aligns views he is challenging with the horrors of twentieth-century communism, in the personages of Lenin (143), Stalin, Mao, and Pol Pot. Cohen’s ostensible giveaway is apparently his so much as discussing Stalinist infringement of freedom of occupation, since Cohen shows not the slightest sympathy for it.

ism.<sup>10</sup> If we are meant to find a challenge to some currently contending approaches to political philosophy, I do not see what it is, though perhaps that is not the intent. Or maybe examples that fit the criticized pattern could be cited.

In any case, however much energy Gaus means to put behind that weak conservatism, he also proposes a stronger conservatism, and in this case he is surely inviting a genuine debate. Suggesting a stronger view, Gaus writes, “We must wonder why, when we face The Choice, we should turn our back on relatively clear local optimization to pursue what may well be a wandering search for the ideal—perhaps in the end the global optimum lies in the opposite direction we initially supposed, and so toward, not away from, our local optimum. Making The Choice to pursue the ideal looks irresponsible” (87). He adds, “We must seriously question an approach to political life that inherently encourages its adherents to neglect what, on their own view, are clear improvements in justice for the sake of pursuing an ideal, the pursuit of which gives meaning to their political lives or fulfills their dreams” (143). Are we merely to “wonder” about and “seriously question” the riskier approach in all cases? Or by saying that it “looks irresponsible,” is Gaus claiming that it is irresponsible? To fix ideas, here is a conservatism that is presumably too strong, and more than Gaus proposes:

**Strong Conservatism:** Stipulating for simplicity that worlds that are very similar to the status quo will not be dramatically less just, the uncertainty about how much less just dissimilar worlds might be morally disqualifies them (renders them “irresponsible”) as appropriate directions of change (to be refined in (d) below).

Such a strong conservatism holds that in a stipulated environment in which

- a) worlds similar to the status quo will not be much less just,
- b) worlds much more just than the status quo will be very dissimilar to the status quo, and
- c) worlds very dissimilar to the status quo will be much more difficult for us to understand in relevant detail than the status quo (76, 81),

then

10. Despite the fact that I coined the term ‘utopophobia’, perhaps it is not my (or similar) work being challenged here, but rather some broader sanguinity about utopian thought. As against my own position, the argument from this parade of horrors would misfire. Indeed, a central point of mine is that even when, as we might normally expect, it is not the case that we should set out for a highly just condition, this fails to show that it is not what full justice requires. Gaus does not confront that position, as we have seen, having chosen (for fine reasons) to set it aside at the very beginning (as I point out above).

- d) there is no prospective improvement in justice over the status quo, and no sufficiently small decrease in justice, such that accepting the decrease for the prospect of the improvement is ever “responsible.”

We really have two questions here. First, even if (a)–(c) are true, do they prove or substantially support that strong conservatism? They patently do not. They leave entirely open whether for some small decrease (or opportunity cost in justness) and some prospect of great improvement it is “responsible” to go for the improvement. Plausibly, then, Gaus intends a more moderate conservatism than this. The idea of a “neighborhood” around the status quo is an important element of his position. Building that in, here is one form a more moderate conservatism might take: perhaps movements in the direction of apparently improved justice are sometimes responsible even if they reduce justice or forgo smaller but more certain gains, but only when the prospective improvement is itself within a neighborhood of structural similarity to the status quo and hence reasonably well understood.

So far, this isn’t much help in deciding whether Gaus’s view is the weak conservatism, the strong conservatism, or somewhere in between. None of those is yet ruled out, since it depends on the size of the neighborhood (supposing it has sharpish boundaries; or, if we model a neighborhood more realistically as a graded trade-off schedule between ostensible gains and degree of confidence, it all depends on what the trade-off schedule is). If available social worlds (possible next social states) must be extremely similar to the status quo in order to have the requisite familiarity or comprehensibility, then the view reverts to the implausibly strong conservatism: only very small structural changes will ever be responsible. If they need not be very similar, or if nothing more is settled about this, then the view may be no more than the weak version, but that position is obvious from the beginning.

A second question is this: suppose that the conclusion (d) is correct in many or most cases where (a)–(c) are true, perhaps owing to some additional premises (since it clearly doesn’t follow otherwise). What help is this in assessing whether real-world decisions are “responsible” or not, as Gaus appears to be doing (in, e.g., shepherding us away from the siren song of Lenin)? Those three conditions are never argued by Gaus to be true, or even probably true. They are stipulations of a certain analytical model that Gaus constructs and explores. So, even if (d) could be well supported in that model scenario, this would not bear on real-world choice unless we are given some reason to believe that (a)–(c) are true. This point is not an objection to Gaus’s reasoning, as far as it goes. The analytical model, at least supposing that with some supplementation it could show something like (d) to follow, does serve to raise the stakes of the dis-

putable claims (a)–(c). Any analytical model starts somewhere and so is limited. This one is limited in that it does not offer much support for those premises as applied to real social life. Nor, it's true, do I have the space here to consider their plausibility.

## V. AGAINST CONSENSUS

Turning to a different issue, it is one of the book's central aims to critique the practical aim of any thoroughgoing consensus about justice. Insofar as such consensus is built into the ideal of full social justice, Gaus takes the epistemic disvalue of consensus to count against the value of achieving any putatively ideal social condition. The example that draws Gaus's main attention is John Rawls's idea of a just and well-ordered society.<sup>11</sup> This imagined social consensus on the content (at least within some small family of variants) of social justice is obviously the opposite of viewpoint diversity on those matters.

It is important to note that a well-ordered society is not in any logical tension with protections of basic freedoms as broad as Gaus might wish, but he is correct, and it has not been much scrutinized, that it would involve (even entail) a lack of diversity of viewpoint on those particular matters. In arguing for what, following Popper, Gaus calls "the Open Society," Gaus is not discussing protected freedom of expression, but rather viewpoint diversity itself—and neither entails the other. (The terminology of "open society" might be misleading here, since consensus does not imply any lack of social, much less legal, "openness" to all points of view, including strongly protected liberties of thought, conscience, and expression.)

The narrowness—the consensus about justice—is troubling to Gaus not on liberty grounds but on epistemic grounds: a society in which there is little diversity of viewpoint on some matter is, he plausibly insists, at a disadvantage in the effort to arrive at more accurate understandings of that matter—in this case, the content of social justice (225). As found in Rawls, however, the idea of a well-ordered society is not to be celebrated independently of whether the shared conception of justice is correct or sound. The Rawlsian position, as I understand it, is that so long as the conception of justice in question is sound (or correct, or the best, or [Rawls's preferred formulation] the most reasonable, etc.),<sup>12</sup> it is a feature of a per-

11. Gaus also quotes Cohen as building into the ideal that the principles are both met and "affirmed" (19).

12. Rawls writes, e.g., "We think of a well-ordered society as a scheme of cooperation for reciprocal advantage regulated by principles which persons would choose in an initial situation that is fair." Here and elsewhere, the importance of well-orderedness is understood in the context of the ordering principles of justice being those Rawls thinks correct or "most reasonable." John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), 29. For the characterization of the two principles

fectly just society not only that the correct principles are met by the basic social structure but also that this conception would be shared and complied with by all and this would be common knowledge. A fully just society is not in accordance with principles of justice coincidentally, but intentionally in this sense. But then it turns out that there is no danger of the kind Gaus seems worried about: that such a consensus might freeze in place a flawed public conception of justice—that is not the scenario under discussion. So, while Gaus normally speaks of how consensus would interfere with any closer approach to true justice, that point would only count against the desirability of consensus around what might be an erroneous conception of justice. The epistemic value of viewpoint diversity would be an important point against any such alleged independent value of consensus. While Gaus does not cite anyone who takes that view, some might indeed do so, and Gaus's point against such a view is well taken.

There remain reasons why one might be concerned on epistemic grounds about a consensus even around the true or best conception of justice, and it may be that Gaus is partly moved by them. I have in mind the Millian argument that where opposing points of view are no longer heard and weighed, the ability of the society to defend the correct conception if it should, in the future, come under intellectual attack may diminish. Even mistaken views ought to be heard, according to this view, in order to keep the correct conception, supposing it is the established view, in fighting shape. Indeed, Mill argues,<sup>13</sup> it is epistemically especially effective if the opposing views are offered by people who believe them. In a Rawlsian well-ordered society, however, there would, by definition, be no such true believers in erroneous conceptions of justice, and so any such preemptive epistemic value would be missing. There is no reason to suppose that there would not still be “devil’s advocate” explorations and debates about other conceptions, and that is an important point. That might be enough for maintaining fighting shape, even if Mill is not so sure. Still, suppose it is not enough. Now suppose that a sound public conception of justice could achieve such a robust place in the course of a society’s history that even if it became a matter of broad consensus and no genuine debate, it might securely remain in place into the future. I see nothing in Gaus’s discussion to raise doubt about whether that would be a highly desirable achievement, if it somehow happened, since the value of viewpoint diversity for improving the conception is not applicable. Such a permanent achievement is hardly inconceivable, however unrealistic it might be (the question to be considered next). Keep in mind that we have stipulated full and robust basic liberties of debate and discussion.

---

as the “most reasonable,” see *ibid.*, 52, 309; and John Rawls, *Political Liberalism* (New York: Columbia University Press, 2005), xlv, and *passim*.

13. John Stuart Mill, *On Liberty* (Peterborough, ON: Broadview, 1999), chap. 2.

And, partly owing to those liberties, the stably agreed public conception could, and certainly would, come in for study, discussion, and exploratory and academic debate, even if no one genuinely held a conflicting view. That, as I understand it, is the Rawlsian idea of just, well-ordered society.

Of course, it sounds like an unrealistic scenario even if a highly desirable one. Gaus, indeed, holds that “an Open Society, in which each is free to pursue his or her own inquiry into justice . . . will be characterized by continued, deep diversity, with no shared ideal” (149). Much of his broader argument depends on that claim, and many would be inclined to agree.<sup>14</sup> Further reflection on Mill’s argument makes this less than obvious, however. Mill argued that there is reason to hope that, for any given issue (or perhaps only certain issues), a time will come when there is no longer serious disagreement, and also where there is little chance of ever falling back into the previous error.<sup>15</sup>

Mill unfortunately offers no clear historical or epistemological account to support that idea, but there are cases that may seem to support it. Begin with factual and scientific views. For example, consider the view, now a matter of consensus, that many diseases are caused by germs. The Millian point, applied to this case, would be that now, after a historical interval of ignorance and then debate about it, there is little or no chance of that consensus disintegrating, certainly not without massive oppression and infringement of free debate and inquiry. The same can be said for countless cases of scientific progress that have by now moved beyond the chance of significant controversy under free conditions.

The case that matters most for Gaus’s purposes is different, of course: consensus about the best conception of social justice, so consider that case. In the United States, it was once a matter of intense dispute whether people of African descent had a moral nature that was inferior to people of European descent, such that only the latter possessed the inalienable rights spoken of in the Declaration of Independence. As slow and uneven as progress has been on issues of racial equality, there is no longer debate about that. And while many momentous things are vulnerable to historical backsliding, there is very little chance that that will come back into

14. Gaus cites what he calls “Rawls’s core insight: the exercise of human reason under free institutions leads to disagreement” (149). But (as I pursue presently in the body of the text) that Rawlsian insight does not posit disagreement about everything under free conditions; that is not plausible, and Rawls (as I think Gaus acknowledges) does not himself accept it.

15. Mill famously writes, “The real advantage which truth has, consists in this, that when an opinion is true, it may be extinguished once, twice, or many times, but in the course of ages there will generally be found persons to rediscover it, until some one of its reappearances falls on a time when from favourable circumstances it escapes persecution until it has made such head as to withstand all subsequent attempts to suppress it” (*On Liberty*, 73).

significant dispute. Racial injustice, so long as it persists, will be conducted on other pretexts and with other causes. Something similar might characterize views about women's moral and political status. And many hope and expect that something similar will be achieved around issues of sexual identity and orientation. Consider some additional issues and their course in history. In parts of Europe it was, in the time of feudalism, believed and then eventually disputed that social and economic inequality was justifiable on grounds, roughly, of family lineage. There are vestiges still, and so perhaps there is no consensus that this is wholly false. Even so, the plausibility of a robust feudal conception of social rights and status has been permanently discarded. These all seem to represent consensus on partial principles of social justice that is historically secure, so long as appropriate freedoms are not too badly restricted. It is, at least, not ridiculous to think that there can be that kind of permanent achievement of consensus on some matters, even these matters of social justice. The question is not whether there is some strong tendency to moral progress. The Millian claim I am invoking here is not that this is bound or even likely to happen. The question is whether conceptions of at least parts of social justice can, in ways we can roughly understand, achieve a permanent consensus. That part of the Millian view seems to find some support in history.

On similar grounds, the achievement of a consensus around a conception of social justice that is more complete and refined than the few steps I have alluded to is comprehensible in terms of dynamics and mechanisms we know from real life and history, whether or not it is at all likely. Surely the Rawlsian theory of a just, well-ordered society makes no claim that such a thing is likely, much less guaranteed. Rather, it is held to be a comprehensible historical possibility, and a strong candidate for the title of full, ideal, perfect social justice. Gaus's arguments, then, may not strongly support his claim, upon which he rests much, that "absent institutions to manage diversity, what we should expect in a free society is that individuals will understand justice in deeply different ways" (145). (Talk of a shared ideal, by the way, might connote a shared "blueprint" including fine details about institutional structure, practices, laws, and mores, but that's not the relevant target, and Rawls's conception of well-ordered justice, Gaus's central example, is not like that.)

Certainly room is left to doubt whether achieving such a society (which we have described only in this one feature—consensus on a sound conception of justice) is an appropriate practical goal for historically situated agents facing obstacles, constraints, and likelihoods of various kinds. Understanding which conception of justice is the best, at least any reasonably complete one, is extremely difficult. On many matters—no longer on the question of the basic moral equality of the races, or on many other matters either, but still on many remaining matters—only a fool would have great confidence that the correct or best conception of justice has been arrived



at. For now. Given the reasons for uncertainty, it would be foolhardy to hope that the debates about justice should come to an end in some consensus any time soon. Gaus not only would agree but insists on this point. It is helpful, then, to pull the two questions apart: First, is an ideal of social justice that includes consensus on a sound conception of justice plausible and desirable (whether or not it is likely to be met)? I have argued that it is. Second, could we ever (even if not any time soon) be in a position to judge that we have collectively gathered around a sound conception, so that the loss of genuine controversy is not to be lamented? Gaus, I think, doubts it, although I think that the historical examples described leave this open.

## VI. CONCLUSION

Readers expecting a certain familiar ideal/nonideal theory frame may be surprised to learn that Gaus accepts all of the following, which in several ways count him as a friend of some signal aspects of “ideal theory”:

1. Ideal justice is not (necessarily) an appropriate practical goal under realistic circumstances.
2. Ideal justice might, indeed, be unattainable.
3. If ideal justice is neither an appropriate aim of realization nor a source of practical orientation, then understanding it is “useless,” though not necessarily any less real or true or valuable for that (questions he puts aside).<sup>16</sup>
4. Even unattainable justice might have a practical “orienting” function even if its realization ought not to be an aim.

In what way, then, does “the ideal” threaten to tyrannize, as the book’s title warns? Gaus writes, “A tyrant rules in a manifestly unjust way; for us to be under the sway of an ideal theory is for us to ignore relatively clear improvements in justice for the sake of a grander vision for the future” (142). The language of tyranny is not aimed at the kind of “dreaming” theory—theory of ideal or perfect justice in which it might not provide practical guidance under the circumstances—which he puts aside without arguing against. Perhaps it is not aimed at ideal theory at all, but at what Gaus sees as a reckless practical stance. Is the reckless stance that threatens tyranny the one that rejects the weak conservatism described above? If so, we might be comforted to observe that few if any writers now advocate such a stance. Due attention must certainly be given to the difficulties involved in knowing whether the social realization of any alleged ideal (or even any alleged vast improvement) is really desir-

16. Gaus writes, “I am not claiming that such dreaming has no value” (39). On the other hand, he pretty clearly doubts it.

able. I do not want to say that this point is not worth emphasizing, but it is far from supporting the stronger conservatism that Gaus might seem to draw from it, a strong if not ironclad objection to social change except of a minimal and hesitant kind. The fact, which no one could deny, that there is such a thing as moving too fast toward what appears to be a more just society is no support for the claim that there is no such thing as moving too slowly, too cautiously. What is difficult to determine is where between those polar positions—a region in which every plausible view resides—he understands his argument to land. As we have seen, and as Gaus makes clear, the challenge was from Sen, and the broad direction of reply was from Simmons, who explicitly noted both that improving justice and moving toward full justice are desiderata that can conflict and that some will take the uncertainty inherent in the latter to favor the incremental climbing model of Sen, making no essential use of the social ideal. It is not clear whether Gaus goes that far. In any case, however far he does go, once the added uncertainty that attaches to bigger changes is acknowledged by all sides, it is unclear what basis is offered for his preferred degree of conservatism.

While I have questioned several central arguments and stipulations that lie behind the model that Gaus builds and explores in analytical, and often formal, detail, this is by no means to deny the interest and importance of the arguments internal to the model, or to the choice between various ways of refining a model of that broad kind. I have not explored those here. Insofar as the model is found to be internally sound (or this is granted for the sake of argument), the burden shifts to the premises and presuppositions. Likewise, insofar as they are found wanting, the question becomes whether and how to devise a good model that would allow deeper understanding of the implications of more defensible premises. I expect questions of both kinds to attract broad interest from Gaus's readers, though I leave the latter questions to others. And I expect that even if some of Gaus's premises must be adjusted or replaced, Gaus's method of trying to sharpen them in a framework that is more formally rigorous than most political philosophy will serve as a worthy model in its own right.